MASTER THESIS · MSC STRATEGIC PRODUCT DESIGN

Empowering Consultants as Innovators

A Framework for Strengthening Innovation Capabilities for AI Adoption in Strategic Consultancy

THE EMPOWERED CONSULTANT · FOUR LEVELS OF WORKING WITH AI AGENTS



AUTHOR

Ferhat Dursun
November 2026

SUPERVISORY TEAM

Bart Bluemink, chair
Pan Wang, mentor
Khalid Maliki, company mentor

GRADUATION COMPANY

Supply Value
Zeist

TU Delft Delft University of Technology

Supply Value
We make strategy work

Empowering Consultants as Innovators

A Framework for Strengthening Innovation Capabilities for AI Adoption in Strategic Consultancy

Master thesis

November 2026

Ferhat Dursun

MSc Strategic Product Design

Faculty of Industrial Design Engineering, Delft University of Technology

Supervisory team

Bart Bluemink, chair, Faculty of Industrial Design Engineering

Pan Wang, mentor, Faculty of Industrial Design Engineering

Khalid Maliki, company mentor, Supply Value

Graduation company: Supply Value, Zeist

Cover illustration: drawn by the author, rendered with Magnific. Typeset in Inter and Calibri.

Preface

This report is the result of my graduation project for the master Strategic Product Design at the Faculty of Industrial Design Engineering of Delft University of Technology, carried out at Supply Value in Zeist.

The arrival of AI has changed many sectors and the work of many people. The question I wanted to answer for myself was not whether AI replaces creative people, but whether it can help people, myself included, to work more innovatively. As people we should use the technology and the methods available to us to add value and skills to ourselves, not to replace ourselves with technology. That became the subject of this thesis. It looks at consultants, but I believe the question behind it is a general one, for people who have never heard of design thinking as much as for people who work with it every day. In a world in which AI may well make us lazy, we need it to make us more innovative, not less. If this work helps a few people to do that, it has done what I hoped for.

I owe thanks to many people. To my chair, Bart Bluemink, and my mentor, Pan Wang, who guided me patiently through a process in which I sometimes worked very much on my own and too slowly, and who, at the moments it mattered, said so plainly enough that I picked up speed. To my company mentor, Khalid Maliki, who made me feel at home at Supply Value, was always open to help and made time for me even during his holidays. To the consultants of Supply Value, who found time for me next to full agendas whenever I asked, in the interviews, in the co-creation session and in the evaluation; this report is built on what they said, and their names are deliberately absent from it.

To my wife, Hilal, who looked after our children far more often than her share in these months and still supported me, every time, in finishing my degree. To my son Mustafa Musa, who turns four in December and then starts school, and who told me that papa works a lot and is less at home: I am sorry, and thank you. I am glad to finish this study before his school begins. To my daughter Meryem, who has just turned one and whose sweetness gave me the energy to keep going. To my parents, who gave up a great deal to bring me this far and never stopped pushing me to finish my studies. To my sisters, my sisters-in-law and my mother-in-law, who took care of the children whenever a deadline came close. And to the many friends and family members who helped along the way: if I named everyone, this preface would be longer than the thesis itself, so I stop here, with gratitude.

Ferhat Dursun

Amsterdam, November 2026

Executive summary

IN NUMBERS **8** interviews · **681** fragments, 559 coded · **14** requirements · **88** methods, 8 in depth · **10** formative test runs · **3** consultant sessions

Generative AI has made analysis and drafting fast and easy and is moving towards AI agents that carry out multi-step tasks; for a consultancy the question shifts to what its consultants must be able to do with them. This thesis asks how Supply Value, a mid-sized consultancy, can strengthen the innovation capability of its consultants, so that they can direct AI agents at client problems, create more value for clients and develop their own skills.

Eight interviews and two co-worker journey maps (maps of a consultant's working week) indicate two differences between consultants who create value with AI and those who do not, neither technical: whether a consultant has a use case and time to try it, and, among those who use AI, whether they frame the question and judge the result. In the way stand client rules not at hand at the moment of use, knowledge lost between assignments, and learning time that competes with billable work. Framing the problem is the craft clients keep paying for, and it is learned by watching, without a shared approach.



The Empowered Consultant
Four dimensions, four levels, five mechanisms, three guardrails (figure 5.2).



The Methode-coach
A web application in which an AI coach makes the consultant frame the problem first (figure 5.5).



The implementation roadmap
A limited pilot and a decision on wider use, drawn up with the coach itself (figure 6.1).

In a co-creation session two teams both designed a sparring partner that asks before it produces and leaves every decision with the consultant. The design has three connected parts: The Empowered Consultant, a framework of four dimensions and four levels, with framing and judgement as the entry condition of every level, five learning mechanisms and three guardrails; the Methode-coach, a web application in which an AI coach makes the consultant frame the problem before anything is produced, coaches through 88 methods, eight in depth, applies the rules agreed with the client and gives less help as the consultant progresses, built as a prototype with parts designed only and tested in ten formative runs on fictional cases, nine by an AI playing a junior and one by the researcher; and a roadmap that proposes a limited pilot at Supply Value, followed by a decision on wider use.

Three consultants tested the prototype: two worked through part of their session, and one completed the journey-mapping method while also using a second AI assistant; none completed the final comparison task. The sessions show a coach that asks what a senior would ask, problem statements in the consultants' own words, the client's rules applied by the coach when a name was typed, the client's assumption lost between two steps, and a generated report that claimed more than its evidence. Whether thinking first costs or saves time is not established. Management did not formally evaluate the framework; that is left to the decision after the pilot.

The capability can be strengthened by describing it in terms of client work, by support that hands the thinking back, and by a limited pilot followed by a management decision on wider use (section 6.4). No cost or return was measured. Whether the coach builds capability remains to be shown; the pilot starts there.

Good to know before reading this report

Terminology

Four terms about the technology are used in a fixed sense in this report; section 2.1.1 defines them from the literature, with their sources. In short: **AI** is the field as a whole and is used only when a statement holds for the whole field; **GenAI** is prompt-and-response use of a generative model, a person asks, the model answers, the person decides; **an AI agent** is a system on top of a language model with a role, a goal, tools and the autonomy to carry out a sequence of steps, with a person checking at defined points; **an AI tool** is any concrete product a consultant uses. In quotations the term the participant used is kept. The **frontier** of a model is the edge of what it can do reliably, which the user cannot see because the model answers with the same confidence on both sides of it (section 2.1.2).

Three names refer to the design. In chapters 3 and 4, before its form was chosen, the design is called **the support**. From chapter 5 onwards **The Empowered Consultant** is the framework that describes the capability (section 5.2), and the Methode-coach (Dutch methode, method), also called the tool, is the application, built as a prototype, with which consultants build and apply that capability on their assignments (sections 5.3 and 5.4). Within that application, **the coach** is the AI agent the consultant talks to; the word always refers to the AI and never to a person. People who guide a consultant are called a senior, a colleague or a manager.

Innovation capability is used in the sense defined in section 2.3.5: the ability to recognise where a client problem can be approached differently, to frame it as a workable problem, to choose and apply a structured method to it, and to bring AI into that process without surrendering judgement.

Participants and anonymisation

Eight people at Supply Value were interviewed for this research; five consultants took part in the co-creation session (C1 to C5) and three in the evaluation (C-A, C-B and C-C). Interviewees gave consent on the TU Delft form and orally (section 2.5.1), the session participants orally (Appendix G.2), and the evaluating consultants were informed orally at the invitation that their material would be used (section 6.1); the ethics paragraph of section 7.3 states what this does and does not cover. The interviewees are referred to as P1 to P8 throughout the report and in the appendices. Participants are described as a group and referred to by their code only: roles, seniority and other details are not linked to a code, so that no individual within Supply Value can be identified. All participants are referred to as "they". I worked at Supply Value during the project and knew the participants as colleagues, and a supervisory relationship within the sample may have influenced what participants felt comfortable sharing; this is discussed in the limitations of chapter 7. A client is an organisation Supply Value works for; clients are described by sector and type, never by name. Quotations are translated from Dutch by the researcher; the translation follows the Dutch wording, and an omission within a quotation is marked with […]. The fragment identifier after each quotation (for example P4-041) refers to the coded fragment in Appendix D; the transcripts are deleted after the defence (section 2.5.1). Appendices C, D, E, F and J hold the codebook, the coded fragments and the interview material in full; they are confidential and not included in this public version, and are part of the version assessed by the graduation committee.

Supply Value reads the report after it has been submitted; at submission no written confirmation of the company on publication had been received.

Use of AI in this research

AI was used in this research in the ways listed here, and in each case I kept the decision. The language model in every application was Claude (Anthropic; the models Claude Opus 5.5 and Claude Fable 5.1 on a Max subscription, with the data not used for training AI models), operated through Claude Code. First, the interviews were transcribed automatically with the built-in transcription of Microsoft Teams and Microsoft Word and I corrected the transcripts by hand; how the transcripts were processed is stated in section 2.5.1. Second, in the coding of the interviews the language model proposed codes per fragment on the basis of my codebook, as a first reader; I reviewed every proposal, changed or rejected it where needed, and the final coding is mine (section 2.5.1 and Appendix C). Third, the prototype of chapter 5 was built with the same model as an AI coding assistant that I directed with written specifications per building block, and the first layer of the method database was written with it (section 5.4.4). Fourth, nine of the ten formative tests (section 5.7) were carried out by an AI agent, the same model, playing the role of a junior consultant on fictional cases that I designed; the tenth was a dry run by me. In those tests the AI tester also simulated the case data and wrote the test reports, as it did in the second dry run of the evaluation (section 6.1); I reviewed every test report and decided what was changed. Fifth, the implementation roadmap of section 6.4 was drawn up with the Methode-coach, which runs on the same model. Sixth, the illustration on the cover was rendered with Magnific. Seventh, AI was used in writing this report: I supplied the material, the analyses, the decisions and the argument, drafts of sections were produced and revised in dialogue with the language model, and I checked the sentences against the sources before they stayed. The quotations were translated from Dutch by me with the help of Claude and Google Translate. That check caught errors such as a mechanical replacement of "the firm" by "Supply Value" that produced a false statement about the company's headcount, a figure that belongs to McKinsey (Merritt, 2025; section 2.1.2); section 7.5 reflects on what such errors taught me about working with a language model. The research questions, the interpretation of the interviews, the design decisions and the conclusions are mine. No interview finding or quotation was generated: every quotation is taken from a corrected transcript and carries a fragment identifier. The texts of the 88 methods in the coach were drafted with the same model, from the Delft Design Guide as index and the literature of each method; eight were then deepened and checked by me, and 80 are at catalogue depth (section 5.4.4).

Reading conventions

The report follows the Double Diamond and is organised in seven chapters; figure 1.2 shows how the chapters relate to its phases. In chapter 2, the literature sections end with a short paragraph headed "What this means for the research question". Chapter 1 is written from the point at which the research started, after the orientation at Supply Value: it states the problem as it was framed then and what was still to be established, with the sources updated to the time of writing, and it does not use what the later chapters found. Parenthetical citations use "and" rather than an ampersand, and page ranges in the reference list are written with "to"; both are deliberate deviations from APA 7, in line with the rule of this report to use no dashes.

Abbreviations and codes

Abbreviation	Meaning
AI, GenAI, LLM	Artificial intelligence, generative AI, large language model (section 2.1.1)
API	Application programming interface: the route by which the prototype calls a language model
EU AI Act	Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence
GDPR	General Data Protection Regulation (EU) 2016/679

Abbreviation	Meaning
RACI	Matrix that records per task who is responsible, accountable, consulted and informed
RAG	Retrieval-augmented generation: a language model that retrieves passages from a private knowledge base before answering
RQ, SQ1 to SQ5	Research question; the five sub-questions stated in section 2.6

Table 0.1 Abbreviations used in this report.

Code	Meaning
P1 to P8	The eight interview participants (table 2.1)
P4-056	A coded interview fragment: participant and fragment number (Appendix D)
Theme 1 to 11	The eleven interview themes (table 2.2; Appendix J)
A1 to A8	The eight assumptions of chapter 1, tested in section 2.6 (table 2.3)
Cluster 1 to 5	The five insight clusters (section 3.1)
Map A, map B; A/02, B/06	The two co-worker journey maps (section 3.2) and their phases; MT-A1 to MT-A3 and MT-B1 to MT-B3 in table 3.3 are their moments of truth
DR1 to DR14	The design requirements: DR1 to DR13 from the design brief (table 3.5), DR14 added by the co-creation session (section 4.1.7); the letters of DR4 and DR13 mark parts that are judged separately
Indicator 1 to 3	The success indicators of the design brief (section 3.3.5)
C1 to C5	The participants of the co-creation session (table 4.1)
Question 1 to 12; 1a, 3e	The how-might-we questions (table 4.2) and the directions per question (Appendix H.3)
Concept A, B, C	The three concepts of section 4.3
D1 to D4, L1 to L4, M1 to M5, G1 to G3	The dimensions, levels, learning mechanisms and guardrails of the framework (section 5.2)
Test 1 to 11	The ten formative test runs (section 5.7; table I.1) and the second dry run of the evaluation, test 11 (section 6.2)
C-A, C-B, C-C	The three consultants who returned material in the evaluation (section 6.2)
Task A, task Z	The two written tasks of the evaluation, before and after the session with the coach (section 6.1)

Table 0.2 Codes used in this report. M1 and G2 are framework codes; M.1 and G.2, with a full stop, are appendix sections.

Contents

Preface iii

Executive summary iv

Good to know before reading this report v

Contents viii

1 Introduction 1

1.1 Problem context 1

1.2 Problem definition 2

1.3 Project assignment and what needs to be known 3

1.4 Initial research question 3

1.5 Project approach 4

2 Discover 6

2.1 Artificial intelligence: what it is and what is changing for consultants 6

2.2 Consultancy as a service: where the value for the client lies 11

2.3 Innovation capability: what it is and why this thesis studies it 13

2.4 The client environment: regulation, data governance and organisational readiness 16

2.5 Interviews: the situation at Supply Value 18

2.6 Sharpened research question and sub-questions 24

3 Define 27

3.1 Insights: from codes to clusters 27

3.2 How a consultant at Supply Value works: the assignment process and two journey maps 32

3.3 Design brief 38

4 Develop: directions 43

4.1 Co-creation session with consultants 43

4.2 How-might-we questions and concept directions 48

4.3 Choice of direction 51

5 Develop: the concept 56

5.1 From the chosen direction to a concept 57

5.2 The framework: The Empowered Consultant 59

5.3 The Methode-coach: what it is and what it does 65

5.4 How the Methode-coach is built, and why 69

5.5 How the framework and the coach fit together 78

5.6 From requirements to design decisions 80

5.7 Formative tests of the prototype 81

6 Deliver 83

6.1 Evaluation set-up 83

6.2 What the sessions showed 85

6.3 What changes in the design 91

6.4 Implementation roadmap for Supply Value 92

7 Conclusion, discussion and reflection 97

7.1 Answer to the research question 97

7.2 Discussion 99

7.3 Limitations 101

7.4 Recommendations and further research 103

7.5 Personal reflection 104

References 105

Appendices 110

A Project brief 110

B Interview guides 116

C Codebook 119

D Coded fragments 120

E From codes to sub-themes to clusters 121

F Evidence behind the journey maps 122

G Co-creation session: materials, board and outputs 123

H Ideation: boards, idea lists, directions and evaluation 137

I Screens of the prototype 156

J The eleven interview themes in full 164

K Traceability from requirements to design decisions 165

L Recommendations beyond the scope of this thesis 167

M Materials of the evaluation 169

1

CHAPTER 1

Introduction

Why a strategic consultancy must ask what its consultants can do with AI, and how this thesis approaches that question.



IN THIS CHAPTER 1.1 Problem context · 1.2 Problem definition · 1.3 Project assignment and what needs to be known · 1.4 Initial research question · 1.5 Project approach

1.1 Problem context

Generative artificial intelligence (GenAI) is changing what clients expect from strategic consultancy. Large language models are strongly exposed to writing and information-processing tasks, and higher-wage occupations are more exposed than lower-wage ones (Eloundou et al., 2024); many of these tasks are part of consulting work, and clients can increasingly carry them out themselves. If generative AI and related technologies can technically automate work activities that take up 60 to 70 percent of employees' time, as McKinsey Global Institute (2023) estimates, this thesis expects the knowledge asymmetry on which consultancy has traditionally rested to narrow. At the same time, organisations report gains at the individual level that do not yet show in their results, which this thesis reads as a difficulty of implementation rather than of strategy: in McKinsey's global survey of 1,719 respondents, eighty percent reported that AI had improved their individual productivity, while these gains had not yet translated into broad financial impact for organisations (Tinkoff et al., 2026). Supply Value observes that client demand is shifting from advice about AI towards concrete guidance on applying AI in daily operations, responsibly and within the applicable rules; this thesis treats that observation as part of the assignment (section 1.3). Surveys by consultancies that sell AI services, McKinsey's and PwC's among them, are used in this thesis for the trends they report and are read with their commercial interest in mind. Recent sources describe AI agents as systems that carry out multi-step tasks in a defined role (PwC, 2026; L. Wang et al., 2024); this thesis reads that as a further shift beyond prompt-based use of GenAI, and section 2.1.3 examines whether it reaches consulting practice.

This research takes place at Supply Value, a mid-sized strategic consultancy of about fifty people in Zeist, the Netherlands, with three business units: Procurement, Supply Chain and Operations, and Digital. Supply Value combines advice with implementation. Its consultants, the second party in this context, are embedded in client organisations, many of them in the public sector, and carry out the change they advise on. The management of Supply Value regards AI as an opportunity rather than a threat. An internal AI programme, initiated by management, sets the boundaries for AI use, provides tooling and training, and encourages experimentation.

The clients of Supply Value form the third element of the context. Public organisations such as municipalities, government agencies and healthcare providers operate under strict rules on data protection, procurement and safety. AI use within these organisations may be restricted, bound to the client's own systems, or not permitted at all, and established workflows and employee behaviour in such environments change slowly. A consultant who wants to use AI in advising and implementing for these clients cannot simply bring in a consumer tool; the value has to be created within the client's legal and organisational limits. The EU AI Act (Regulation (EU) 2024/1689) adds to this by placing obligations on organisations that deploy AI, including a duty to take measures that support AI literacy among their staff (Article 4, applicable since 2 February 2025).

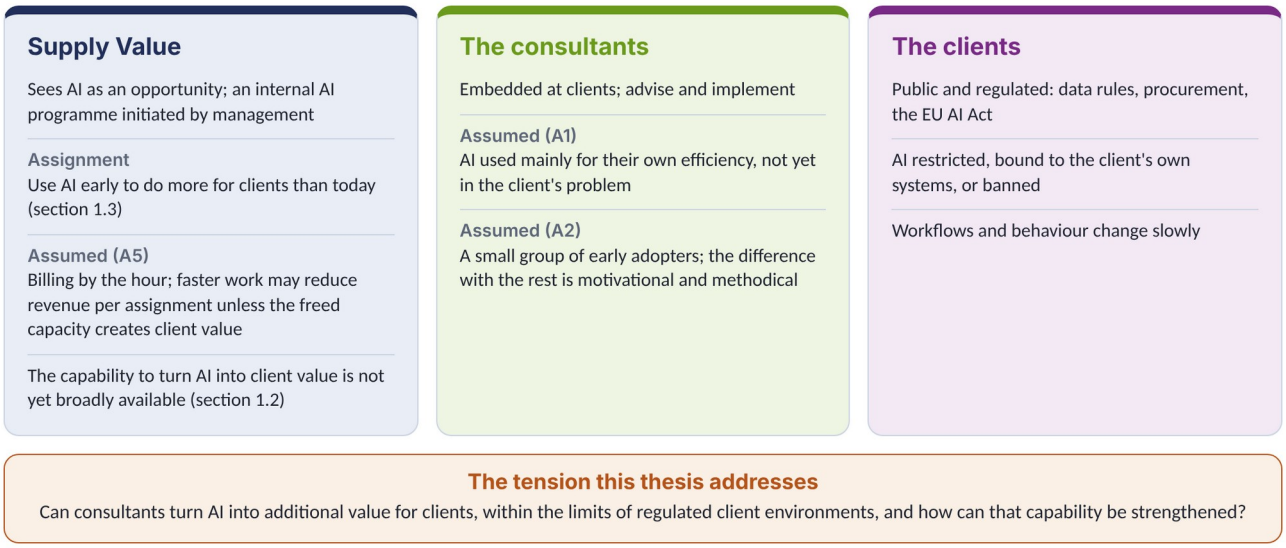


Figure 1.1 The problem context: the three parties and the tension between them.

Together these three parties, against the backdrop of the shift in the technology, define the tension this thesis addresses (figure 1.1). The opportunity is present and the intention of Supply Value's management is clear. What is not self-evident is whether, and how, the consultants themselves are able to turn AI into additional value for clients in regulated environments. Being able to operate an AI tool is not the same as being able to identify where AI can help a client, frame that problem well enough for AI to contribute, and implement the result responsibly. Using AI does not build that capability by itself, and AI does not replace it; if anything, use exposes whether a consultant has it. That is assumption A4 below, and whether and how AI can help build the capability, rather than only speed up the work, is part of what this research has to find out. The capability itself, which section 2.3 defines from the literature as innovation capability, is the subject of this research.

1.2 Problem definition

The problem this thesis addresses is formulated as follows.

> **Supply Value sees AI as an opportunity to create more value for its clients, but the capability to translate AI into client value in advisory and implementation work is not yet broadly available among its consultants. The organisation has no structured way to develop this capability across all consultants within the limits of regulated client environments.**

This problem statement is based on the assignment as formulated by Supply Value (Appendix A) and on the literature discussed in chapter 2. It rests on a number of assumptions that have not yet been established. They are stated explicitly here so that the literature and the interviews in chapter 2 can confirm, refine or reject them (section 2.6).

A1. Consultants' current use of AI is directed mainly at their own efficiency, such as routine tasks, summarising and drafting, rather than at creating value in the client's problem.

A2. The capability to use AI for client value is concentrated in a small group of early adopters, and the difference between this group and the other consultants is motivational and methodical rather than technical.

A3. Uncertainty about what is permitted, both at Supply Value and at clients, leads in practice to non-use of AI rather than to responsible use.

A4. A consultant who cannot frame a client's problem cannot direct AI towards it. Problem framing and a structured way of working are prerequisites for AI to add value in consultancy work.

A5. If GenAI shortens the time consultants need for their current work while Supply Value continues to bill by the hour, revenue per assignment may decline unless the freed capacity is used to create additional value for clients. This assumption is formulated at the start of the research from Supply Value's hourly billing model; it is a logical consequence of that model, not an observed effect, and how participants see it is examined in section 2.5.5. It is the reason the problem described in A1 to A4 matters to Supply Value.

Assumptions A1 to A5 were formulated after the orientation at Supply Value, which included the first four interviews of section 2.5.1; A6 to A8 in section 1.3 come from the assignment. The problem is formulated as a question about a capability, and what the appropriate response to it is, is a question for this research.

1.3 Project assignment and what needs to be known

Supply Value hosted this study and provided access to consultants to explore how AI could support their work with clients. The assignment as given by Supply Value is the starting point of this thesis, not its conclusion; the project brief approved at the start of the project is in Appendix A. It contains three further assumptions that are not yet established and that the research tests alongside A1 to A5.

A6. Automating tasks to make consultants faster is by itself not sufficient to remain competitive, because competitors gain the same efficiency; what distinguishes a firm is what its consultants do with the time and the capability that AI frees up.

A7. Adopting AI tools without a shared view of what they are for leads to scattered use with little effect on client value. Whether Supply Value needs an explicit vision on AI use, and what such a vision should contain, is a question for this research, not a given.

A8. Strengthening the innovation capability of consultants is an appropriate response to the problem in section 1.2. What form that response should take follows from the research. What would count against A8 is evidence that the barrier lies in access to tools or in client permission alone; table 2.3 reports A8 against that criterion.

To test these assumptions, the research first needs to establish what is actually happening. Four things need to be known before any design decision can be made: how consultants at Supply Value currently work and use AI, and why some use it more than others; how they currently approach and solve client problems, and where in that process AI could contribute value; what the client environment allows, in terms of regulation, data governance and organisational readiness; and what the literature says about building a capability of this kind in a consultancy. These four questions guide the Discover phase in chapter 2. The formal sub-questions that follow from them are stated in section 2.6, once the findings allow them to be formulated precisely. In scope are the consultants of Supply Value and their work for clients, most of them in the public sector; the business model of Supply Value is an assumption (A5) and not a matter for the design; the technical development of AI and the training of people who are not consultants are outside this research. The constraints and the open points of the design brief are in sections 3.3.6 and 3.3.7.

1.4 Initial research question

The research question below is the question with which this research starts. It is formulated at the level of AI in general, because it is not yet clear which type of AI is relevant to the problem, and it is formulated around innovation capability rather than around any particular intervention. Chapter 2 distinguishes between AI, generative AI and AI agents (section 2.1.1) and reports the literature and the interviews. On that basis the research question is sharpened and subdivided into sub-questions at the end of chapter 2 (section 2.6).

How can the innovation capability of consultants at Supply Value be strengthened, so that they can use AI to create more value for clients in their advisory and implementation work within regulated client environments, while developing rather than replacing their own professional skills?

The last clause is part of the question from the start because Supply Value's interest and the consultant's interest are not automatically the same. The argument at the start was that a consultant who hands work to AI may deliver faster today and be worth less to the client tomorrow; the evidence on skills is in section 2.1.3. Innovation capability is used in the sense that section 2.3.5 defines from the literature (see also the terminology on p. v).

1.5 Project approach

The research follows the structure of the Double Diamond, developed by the Design Council in 2005 (Design Council, 2007): two rounds of diverging and converging, first on the problem and then on the solution. The Double Diamond is used as a way of organising the research, not as a method in itself; the methods used within each phase are described where they are applied. The choice for a broad, iterative structure follows from the problem: AI in consultancy is recent, in the literature reviewed for this thesis (chapter 2) few studies address mid-sized consultancies, and a capability-building intervention cannot be designed correctly in one pass. It has to be tested and adjusted with the people who are going to use it. The researcher was present at Supply Value during the graduation period as a graduate student; that presence informed the orientation and the choice of participants and is reflected on in section 7.3, but no observation protocol was kept and no field notes beyond those of the interviews were used as data. Within this structure the research is a qualitative design study in a single firm: the knowledge it produces is grounded in what eight people at Supply Value said and in how a designed intervention behaved, not in a sample from which to generalise. That choice fits a question about one firm's consultants and limits what the conclusions can claim beyond Supply Value, which section 7.3 returns to. Figure 1.2 shows the phases and the chapters that report them. The research iterated within the second diamond: the formative tests of section 5.7 changed the design before the evaluation, and the evaluation of chapter 6 changed it again (section 6.3).

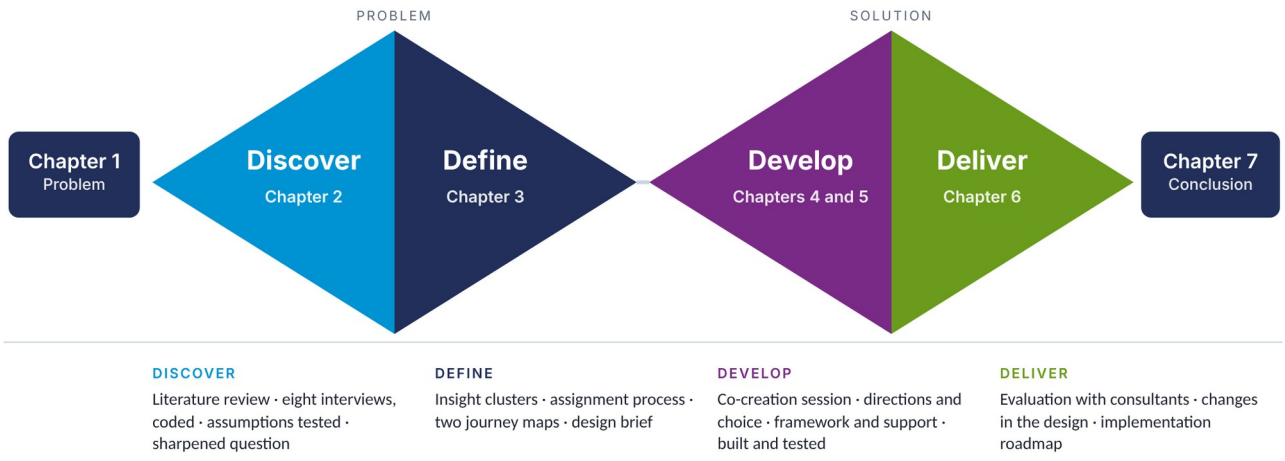


Figure 1.2 The Double Diamond as the structure of this thesis, with the chapters that report each phase. Adapted from Design Council (2007).

Because this thesis argues that consultants should work in a structured, human-centred way, the research itself was set up to do the same: every phase starts from what consultants said and did, and consultants took part as interviewees in Discover and as co-designers and testers in Develop and Deliver. Table 1.1 shows per phase of the

Double Diamond what was done with whom and where it is reported; the codes used in this report (P1 to P8, DR1 to DR14 and the others) are listed in table 0.2.

Phase	What was done, with whom	What it yielded	Where
Discover	Literature review on AI, consultancy and innovation capability. Eight semi-structured interviews with management and consultants of Supply Value, coded thematically; test of assumptions A1 to A8; sharpened research question and sub-questions.	How consultants work, learn, seek help and use AI; what the client environment allows; what the literature says about building the capability; sharpened research question and sub-questions.	2.1 to 2.6 (2.5.1 for the counts), App. B to D
Define	Grouping of the codes into sub-themes and clusters; the assignment process of a consultant; two interviews worked into co-worker journey maps; design brief with requirements.	Insight clusters, journey maps, design brief.	3.1 to 3.3, App. E and F
Develop	Co-creation session with five consultants; how-might-we questions and ideation; a framework and a support developed and tested on fictional cases, nine of the ten test runs by an AI playing a junior consultant and one by the researcher.	The chosen direction; the framework and the support, worked out and tested; what the tests changed.	4.1 to 4.3, 5.1 to 5.7, App. G to I
Deliver	Evaluation of the support: a second dry run by an AI tester (test 11); three consultants invited, each for one session on a fictional case; two returned partial sessions and one a complete session; the changes in the design that follow from it; implementation roadmap, drawn up by the researcher with the support.	What the two sessions showed; the design after the evaluation; the roadmap; conclusions.	6.1 to 6.4, 7

Table 1.1 What was done in each phase of the Double Diamond, with whom, and where it is reported.

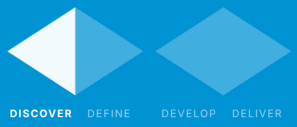
2

CHAPTER 2 · DISCOVER PHASE

Discover

Literature on AI, consultancy and capability, and eight interviews coded into 681 fragments: what the problem looks like from inside Supply Value.

WHERE YOU ARE



IN THIS CHAPTER 2.1 Artificial intelligence: what it is and what is changing for consultants · 2.2 Consultancy as a service: where the value for the client lies · 2.3 Innovation capability: what it is and why this thesis studies it · 2.4 The client environment: regulation, data governance and organisational readiness · 2.5 Interviews: the situation at Supply Value · 2.6 Sharpened research question and sub-questions

Section 1.3 identified four things that must be known before the research question can be answered with any precision. Sections 2.1 to 2.4 address them from the literature; section 2.5 addresses the same questions empirically, through eight interviews with management and consultants of Supply Value; section 2.6 tests the assumptions of chapter 1 against both and states the sharpened research question and the sub-questions that guide the rest of the thesis.

2.1 Artificial intelligence: what it is and what is changing for consultants

This section answers the first of the four questions of section 1.3 from the literature: which type of AI is relevant to the problem, and what does it change in the work of consultants.

2.1.1 Terminology: AI, generative AI and AI agents

The terms AI, generative AI and AI agents are often used interchangeably in practice and in the trade literature. In this thesis they refer to three different things, and the distinction matters for the design: the capability a consultant needs to work with a chatbot is not the same as the capability needed to direct an agent that carries out a multi-step task on the consultant's behalf. Figure 2.1 shows how the terms nest; the four terms below are used consistently throughout the report.

Artificial intelligence (AI) is the umbrella term. It refers to computational systems that perform tasks normally associated with human cognition, such as reasoning, learning, perception and problem solving (McCarthy et al., 1955/2006; Russell and Norvig, 2020). The term covers everything from rule-based systems and classical machine learning to the generative models discussed below. In this thesis, AI is used only when the statement holds for the field as a whole, for example when discussing regulation (the EU AI Act applies to AI systems in general), AI literacy as a capability, or the AI programme of Supply Value, which covers more than one type of technology.

Generative AI (GenAI) is the branch of AI that produces seemingly new content, such as text, images or audio, from a prompt, using models trained on very large datasets (Feuerriegel et al., 2024); the same models also write code. The best known form is the large language model (LLM), a model trained on text that can answer questions, summarise, draft and reason in natural language. In this thesis, GenAI is used for the prompt-and-response use of these models: a person asks, the model answers, and the person decides what to do with the answer. This is how most consultants at Supply Value currently use AI, and it is what interviewees mean when they speak of "using AI" or "using ChatGPT".

An AI agent is a system built on top of an LLM that is given a role, a goal, a set of tools (such as search, file access or the ability to call other systems) and the autonomy to plan and carry out a sequence of steps towards that goal (L. Wang et al., 2024), with defined points at which a person checks the result (Schluntz and Zhang, 2024); in this thesis an agent always has such points, and a system without them is outside the design. The difference with GenAI is not the underlying model but the degree of autonomy: a GenAI application produces one answer per prompt, an agent decides which steps to take, executes them and returns an outcome. Between the two lies what Schluntz and Zhang (2024) call a workflow: a system in which the steps are predefined and the model directs its own process only within a step; the support of chapter 5 is of that kind (section 5.3.1). When several agents with different roles cooperate on a task, this is referred to as a multi-agent system. **Agentic AI** is used in the literature for the broader class of orchestrated, often multi-agent systems (Sapkota et al., 2025); it is used sparingly in this thesis, and only to name the trend, never as a synonym for a single agent.

An AI tool is the word used for a concrete product that a consultant uses, whatever the technology behind it. Where the type of technology matters, the tool is qualified as a GenAI application or as an AI agent.

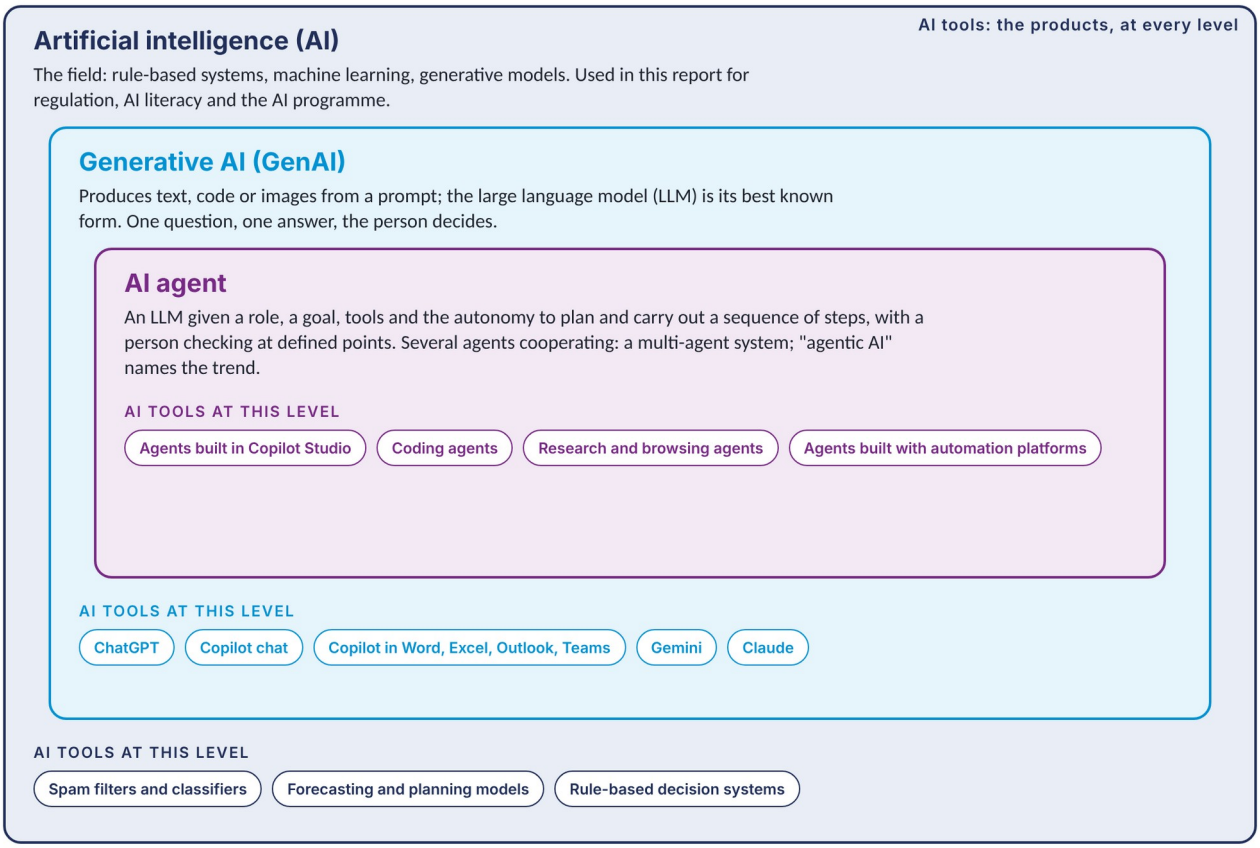


Figure 2.1 How the terms nest, and the AI tools as products at each level.

2.1.2 How consultancies use generative AI and AI agents

Consultancy has traditionally depended on knowledge asymmetry; this thesis infers that the asymmetry narrows now that clients can produce analyses with GenAI themselves, given the share of knowledge work that McKinsey Global Institute (2023) estimates generative AI can automate. The analytical and writing tasks of knowledge workers are among the most exposed to LLMs (Eloundou et al., 2024), which shifts the consultant's contribution from information processing to judgement, contextualisation and stakeholder engagement. In McKinsey's global survey of early 2024, 65 percent of respondents reported that their organisation regularly used GenAI, nearly double the share ten months earlier, and the largest increase in adoption was found in professional services

(McKinsey, 2024). Taken together, the sources describe AI as more than a productivity tool for professional services: it changes what clients can do themselves and so where the consultant's contribution lies.

The large firms have moved from experimentation to structural commitment. McKinsey launched its internal platform Lilli in 2023, a GenAI application that searches McKinsey's own knowledge sources and cites them, is used by about 72 percent of the firm and handles more than 500,000 prompts per month (McKinsey & Company, n.d.). About 40 percent of McKinsey's projects are AI-related, according to a statement of the firm cited in a blog post (Endo, 2025). In the same period McKinsey's headcount fell from about 45,100 at the end of 2023 to about 40,000 by mid-2025, which the firm attributes to attrition and performance reviews (Merritt, 2025).

The most cited evidence on the effect of GenAI on consultants themselves comes from a field experiment with 758 consultants of Boston Consulting Group (Dell'Acqua et al., 2026). The study introduces a term that this thesis uses throughout: the **frontier** of an AI model. The frontier is the edge of what a model can do reliably: on one side lie the tasks it handles well, on the other the tasks it handles badly. The difficulty is that the edge cannot be seen by the user, because the model answers with the same fluency and confidence on both sides. Three conclusions from the experiment are carried forward. First, for tasks inside the frontier, consultants with AI worked faster and delivered work of higher quality, and the consultants with the weakest scores gained the most. Second, for a task outside the frontier, consultants with AI were more often wrong than consultants without it, because they trusted an answer that looked as good as the others. Third, since the edge is invisible from the inside, effective use of AI in consulting requires a human check by someone whose own judgement is good enough to notice when the model has crossed it. This thesis takes that third conclusion as a design requirement (section 2.2.4) and treats recognising the frontier as a skill to be learned, not only as a rule.

AI use also produces failures. Deloitte agreed to a partial refund for a report prepared for the Australian government that contained AI-generated hallucinations (Paoli, 2025), and a Deloitte report for a Canadian provincial government was found to contain non-existent citations, reported as likely AI-generated (Dobson, 2025). For a firm that works with public sector clients, such an error is not only embarrassing but a liability issue, a point that returns in the interviews (section 2.5.4).

For Supply Value the comparison with the large firms cuts both ways. Clients often expect the same output from a mid-sized firm as from a firm with an internal platform like Lilli. Developing an enterprise-wide platform was outside the scope of this project. The difference is a capability gap and, this thesis infers, an opportunity: a smaller firm can embed AI in its work without the overhead of an enterprise programme, and can adapt faster to a specific client. Whether that inference holds is one of the things the interviews examine.

2.1.3 Trends: what is changing and what it means for Supply Value

How the trends were identified

The trends in this section and the sources in sections 2.2 and 2.4 were identified through desk research during the graduation period, through Google and Google Scholar with search terms around generative AI, AI agents, consulting, professional services and knowledge work, and by following the references of the sources found. Peer-reviewed studies are used for the effects on individuals, because they are the only sources with a controlled design; reports from research firms, consultancies and research institutes are used for the scale and recency of adoption, because academic evidence lags the technology by one to two years, and they are grey literature, marked as such where a claim rests on them alone; regulatory texts are used for obligations. A development counts as a trend when at least two independent sources report it, at least one of them from 2025 or 2026, when the sources show a change over time rather than a state, and when it has a plausible consequence for the work of consultants at a mid-sized consultancy; developments that met the first three conditions but not the fourth, such as the growth of sustainability advisory, are not discussed. The statements below describe the situation as

of 2026, and what this thesis infers from a trend for Supply Value is an assumption that the interviews of section 2.5 test. The innovation capability and design literature of section 2.3 was found from the Delft Design Guide, the course literature of the programme and backward citation from the sources found.

Trend 1: The shift from generative AI to AI agents

Since 2024 the use of AI in organisations has moved from prompt-based generative AI towards AI agents, systems that plan and execute multi-step tasks in a defined role (section 2.1.1). In McKinsey's global survey of 1,719 respondents (Tinkoff et al., 2026), 40 percent of respondents from organisations with more than one billion dollars in revenue report scaling AI agents, up from 27 percent a year earlier, while the share among smaller organisations stayed flat at 22 percent; the Stanford AI Index, which draws its organisational adoption figures from the previous wave of the same McKinsey survey, reports that agent deployment "was in the single digits across nearly all business functions" in 2025 (Stanford HAI, 2026, p. 173). Gartner (2025b) predicts that over 40 percent of agentic AI projects will be cancelled by the end of 2027 because of escalating costs, unclear business value or inadequate risk controls, notes that many products marketed as agents are rebranded assistants and chatbots, which it calls "agent washing", and at the same time expects 33 percent of enterprise software applications to include agentic AI by 2028, up from less than 1 percent in 2024; these are forecasts, not observations. In a field study comparing a conversational assistant with an autonomous agent, Yang et al. (2026), employees of the vendor whose product they studied, found that the agent not only reduced task time but expanded the scope of work people attempted, shifting the user's role from operator to supervisor; PwC (2026) reports from its own practice that early-career workers equipped with AI agents contribute faster than before. Taken together, the sources show a shift that is real in the large organisations and early everywhere else: the figures describe large organisations and evidence on mid-sized professional service firms is scarce, and the cancellations Gartner expects show how immature the shift is rather than a reversal, because the three causes it names are the causes of failed implementations.

For Supply Value this thesis takes two positions, which are design choices rather than findings of the sources. First, the relevant level for building capability is the agent, not the chatbot: the shift that changes how a consultant works is the delegation of multi-step tasks with supervision, and that requires a different capability than prompting (assumption A4). The sources carry this for large organisations only, the figures for smaller organisations are flat, and the interviews of section 2.5 contain no consultant who directs an agent; the position is chosen because the support has to hold the context of an assignment across steps, and because it is where the capability gap will open if the trend reaches mid-sized firms. Second, the gap between large and mid-sized organisations in scaling agents suggests that a firm of Supply Value's size cannot rely on tooling alone to keep up; the difference will have to come from how its consultants direct agents at client problems (assumption A6). Any support this thesis designs carries the three risks Gartner names, cost, unclear value and risk controls; what that means for the design is taken up in chapter 3. The idea that consultants will coordinate and supervise several agents at once, found in McKinsey (2025) and in the trade press, is treated here as a scenario rather than a finding.

Trend 2: Adoption is broad, value is not; the bottleneck is implementation

Nearly nine in ten organisations report regular use of AI in at least one business function, and 80 percent of respondents report that AI has improved their individual productivity, yet 37 percent attribute any effect on earnings to AI and the share of "AI high performers" has remained flat at about 6 percent (Tinkoff et al., 2026; grey literature). The same survey finds that nearly three quarters of the high performers report fundamentally redesigning workflows because of their AI use, against one quarter of the other respondents. A study by the MIT NANDA initiative of more than 300 enterprise AI initiatives, 52 interviews and 153 surveys, preliminary, based on a convenience sample of conference attendees and without a public sector analysis, concluded that "just 5% of integrated AI pilots are extracting millions in value" (Challapally et al., 2025, p. 3) and that only 5 percent of

custom enterprise AI tools reach production; the core cause it names is that "most GenAI systems do not retain feedback, adapt to context, or improve over time" (Challapally et al., 2025, p. 3). The same study reports a "shadow AI economy" (p. 8): only 40 percent of companies bought an official LLM subscription while workers at over 90 percent use personal AI tools for work. The peer-reviewed evidence (Dell'Acqua et al., 2026; Cui et al., 2026) concerns individual productivity, not firm-level value, so the step from "individuals are faster" to "Supply Value earns more" is not established by any of the sources.

The literature supports assumptions A1 and A6: individual efficiency gains are common and do not by themselves distinguish a firm; what distinguishes the organisations that see value is that they redesign how work is done around the technology. For a consultancy this places the question at the level of the consultant and the assignment: not whether tools are available, but whether consultants can redesign how they approach a client problem with AI in it. It also supports Supply Value's implementation model (section 2.2): what clients struggle with is making AI work in their operations, which is the kind of embedded work Supply Value already does.

Trend 3: AI changes how professional skills develop, in both directions

AI raises the output of knowledge workers and at the same time creates a risk for the skills that produce judgement. Across three randomised trials with 4,867 software developers, access to an AI coding assistant increased completed tasks by 26.08 percent, and "less experienced developers had higher adoption rates and greater productivity gains" (Cui et al., 2026, abstract). The Stanford AI Index notes that productivity gains are largest in structured, measurable work and that "recent evidence raises concerns that heavy AI reliance may carry long-term learning penalties that slow skill development over time" (Stanford HAI, 2026, p. 174); that evidence is recent and largely correlational, so it cannot rule out that people who were learning less anyway rely more on AI. Kellogg et al. (2025) find that junior professionals cannot be expected to teach seniors to use GenAI, because novices lack a deep understanding of the technology and of the risks it carries for the work. Anthropic's analysis of one million conversations, a usage study by the vendor, finds that experienced users "are more likely to use Claude to iterate on their work, and much less likely to delegate greater responsibility through directive use patterns", and have a higher success rate; the authors offer learning by doing as one explanation, next to cohort and survivorship effects, and state themselves that the study cannot separate the two (Massenkoff et al., 2026, p. 15). A field study with 523 early-career professionals at one firm, reported as a practitioner article without full methodological detail, found three profiles, which the authors call AI apprentices, AI delegators and AI amplifiers, and reports that traditional measures of capability did not explain which profile someone fell into, while the delegators scored lowest on foundational skills (Agarwal et al., 2026). The World Economic Forum (2025) expects roughly 39 percent of current skill sets to be overhauled by 2030, and PwC (2026) warns from its own practice that the skills most valued in the AI era, including critical judgement and the ability to direct AI systems, are exactly the skills that used to develop through junior analyst experience. None of these studies concern consultants; the developer studies are the closest, because programming, like consulting, combines a structured craft with judgement about what the client needs; and what the sources show is a growing evidence base since 2023 rather than a measured shift in how skills develop. Trend 3 therefore meets the condition of several independent sources but not that of a measured change over time; it is kept as a trend for the direction the sources agree on and treated in the design as a working assumption (conclusion 4).

The sources suggest, without establishing it for consultants, that iterating with AI on a self-framed problem builds capability and that delegating the whole task does not. The design treats this as a working assumption that chapter 6 tests; it is the main literature support for assumption A4 and for the last clause of the research question. If it holds, support that only teaches tool use risks accelerating the decline it is meant to prevent, and the support has to make the consultant do the thinking first.

The three trends are carried forward into the conclusions of section 2.2.4, trend 1 as the distinction between prompting and directing agents, trend 2 as the shift from adopting tools to redesigning how consultants work on client problems, trend 3 as the condition that support makes the consultant think first and fades as capability grows and as input for the learning mechanisms of section 2.3, and into the interviews of section 2.5, which test whether consultants already use anything beyond prompt-based tools and whether the inferences above hold at Supply Value.

2.1.4 What this means for the research question

The literature confirms that GenAI has become common in professional services and that AI agents are the next step, but it also shows that adoption and value are two different things: most organisations report individual productivity gains, few report an effect at firm level, and the difference lies in whether work is redesigned around the technology rather than the technology inserted into existing work (Tinkoff et al., 2026; Challapally et al., 2025). For a consultancy this places the question at the level of the consultant: not whether the tools are available, but whether consultants can direct them at client problems. Assumptions A1 and A6 are supported by the literature; whether they hold at Supply Value is tested in section 2.5. The type of AI that is relevant to the problem is the AI agent, with prompt-based GenAI as the entry level: a chatbot that answers questions does not change how a consultant works, while an agent that carries part of a method, holds the context of the assignment and asks the consultant the right question at the right moment can. The gap this thesis addresses is therefore stated as follows, as of 2026: organisations adopt AI widely but few turn it into value, the sources locate the difference in how work is redesigned rather than in access to tools, and no study yet shows how consultants direct AI agents at client problems; it is a capability gap before it is a technology gap.

2.2 Consultancy as a service: where the value for the client lies

This section answers the second question of section 1.3 from the literature: where does the value of a consultancy for its client lie, what remains human when AI takes over analytical work, and what does AI do to the business model.

2.2.1 Consultancy as a knowledge-intensive service, and Supply Value's model

In this thesis a management consultancy is understood as a professional service organisation that helps clients solve complex problems, improve performance and achieve strategic goals; the description is the author's. What makes consulting firms distinct from other service providers is that their core product is knowledge itself: expertise that internal staff typically cannot replicate on their own. In the academic literature these firms are classified as knowledge-intensive business services (Muller and Doloreux, 2009), where reputation and trust are the primary currency because clients cannot fully assess the quality of the advice before they receive it (Glückler and Armbrüster, 2003). Demand for external consultants grows as the environment organisations operate in becomes harder to navigate on their own: rapid change means that even large enterprises regularly face challenges their internal teams are not equipped to handle (P. Wood, 2002), and consultancies act as a bridge between specialised knowledge and the organisations that need it (Muller and Doloreux, 2009).

Supply Value operates within this landscape with a deliberately different approach. Rather than delivering recommendations and stepping back, it embeds its consultants in client organisations to guide and execute change from the inside. As it describes itself in the orientation interviews and on its website, the firm specialises in procurement, supply chain and digital transformation, and works with public and private sector clients across the Netherlands. Clients do not just receive a plan but get the people to carry it out. This thesis infers, and explores in the interviews, that in a market where AI is commoditising the advice itself, this implementation focus is Supply Value's most defensible position; section 2.6 states that it remains an inference.

2.2.2 The pressure on the billable hour (assumption A5)

The traditional billable-hour model is under pressure: when AI can complete in hours what previously took weeks, charging by the hour penalises a firm for being efficient; this is a logical consequence of the billing model, not an empirical finding. This thesis infers that fees may move towards models that tie them to outcomes rather than time, such as a fixed price for solving a defined problem, a share of the savings generated or a subscription for ongoing access to analytical tools and insights; no source measuring such a shift was found, and no peer-reviewed source from 2025 or 2026 on the consulting business model specifically. In the McKinsey survey, 39 percent of respondents expect AI-related declines in their organisation's total employment in the coming year, but only 14 percent report that AI contributed to an overall decline in the past year (Tinkoff et al., 2026), which suggests that the pressure is anticipated more than experienced. Against the conditions of section 2.1.3 this development does not qualify as a trend: the claim about the business model rests on trade commentary, and no measured shift in how consultancies bill has been found; it is therefore carried as an assumption.

Assumption A5 remains what it was in chapter 1: a conditional expectation rather than an observed effect. Its importance is not that revenue is declining, which the sources do not show, but that it explains why the opportunity in A1 to A4 matters to Supply Value: freed capacity has to be turned into something a client pays for. Participants at Supply Value discussed how AI might change the value of consulting work (section 2.5.5); the study did not measure the effect on revenue or establish a preferred pricing model, and A5 is not input for the design of the capability itself.

2.2.3 Advice, implementation and the human element

This thesis infers that the difference between firms that only give advice and firms that help implement it will determine which firms remain relevant as AI takes over more of the analytical work that traditional consulting was built on: the advice-only model is losing its value not because consultants are less capable but because the product they were selling, information and analysis, is no longer scarce. What clients struggle with is less knowing what to do than making it happen (Trend 2, section 2.1.3). This is where Supply Value's model of embedding consultants inside client organisations becomes distinctive: the consultant who recommends a change is also the one responsible for implementing it, which creates an accountability that arm's-length advisory firms cannot offer, and a proximity that builds an understanding of the client's culture, politics and real capabilities.

Even as AI takes on more of the analytical and drafting work, the judgement required to navigate organisational reality remains human. The World Economic Forum (2025) reports, from a GPT-4o self-assessment, that skills such as empathy, active listening and nuanced judgement show little potential for substitution by AI; in consulting this translates, in this thesis's reading, into knowing when to trust instinct over data, building the trust that makes a client follow through, and recognising when a technically correct recommendation is organisationally wrong. This thesis infers that consultants who combine this judgement with the ability to direct AI will be the ones clients seek out, and that Supply Value's challenge is to ensure that the consultants delivering its model are equipped to lead conversations about AI at client organisations. That inference is tested in section 2.5.5.

2.2.4 Conclusions from the literature

Sections 2.1 and 2.2 set out to answer two of the questions of section 1.3: which type of AI is relevant to the problem and what it changes in consulting work, and where the value of a consultancy for its client lies. This section draws the conclusions; each is linked to the trend it is derived from and to the assumption of chapter 1 it bears on.

Conclusion 1. Adoption is broad, value is not (Trends 1 and 2). Nearly nine in ten respondents report regular use of AI in their organisation and most individuals report productivity gains, but few organisations report an effect

on results, and the difference lies in redesigning work around the technology rather than inserting it into existing work (Tinkoff et al., 2026; Challapally et al., 2025). Assumptions A1 and A6 are supported by the literature.

Conclusion 2. The relevant unit for this thesis is the AI agent, and directing an agent is a different capability from prompting (Trend 1); this is an inference and a positioning of the thesis, not a finding of the sources, which show the shift in large organisations only. Assumption A4 is supported in principle: the shift that changes how a consultant works is the delegation of multi-step tasks with supervision, and that asks for a different capability than prompting.

Conclusion 3. Generative AI raises the quality of work inside its frontier and lowers it outside, and consultants cannot see the frontier from the inside (Dell'Acqua et al., 2026). It is the one finding with a controlled design that concerns consultants themselves.

Conclusion 4. The sources suggest, without establishing it for consultants, that working with AI can slow the development of professional skills when it replaces practice, and that professionals who iterate with it on a problem they framed themselves do better than those who delegate to it (Trend 3). The evidence is recent and largely correlational; the design treats this as a working assumption.

Conclusion 5. The value of a consultancy for its client lies increasingly in implementation and in the human reading of the organisation, not in analysis that clients can now produce themselves (sections 2.2.1 and 2.2.3). Supply Value's embedded model fits this shift. Assumption A5 is plausible but rests on grey literature and is carried forward as an assumption.

Two further conclusions concern the client environment and are drawn in section 2.4: the limits set by regulation and data governance (assumption A3), and the obligation under Article 4 of the EU AI Act to support AI literacy among staff. Together these seven conclusions are the literature-based input to the design brief; section 2.5 tests them against the situation at Supply Value, and section 3.3 derives the requirements from them and from the insight clusters. Figure 2.2, at the end of section 2.4, summarises the seven conclusions and the evidence behind them.

2.3 Innovation capability: what it is and why this thesis studies it

The research question of section 1.4 asks how the innovation capability of consultants can be strengthened. This section explains why the question is framed in terms of innovation, defines what innovation capability means in this thesis, and sets out what the literature says about how such a capability develops. It answers the fourth question of section 1.3: how is a capability of this kind built in a consultancy.

Why innovation. The problem of section 1.2 is not that consultants lack access to AI but that they are not yet able to turn it into additional value for clients. Sections 2.1 and 2.2 showed where that value lies: not in analysis, which clients can increasingly produce themselves, but in recognising where a client's situation can be approached differently, framing that as a workable problem, and implementing a solution that the client can carry on. That is, in the terms of the innovation literature, the capacity to carry out new combinations (Schumpeter, 1934), understood in the later literature as new combinations of knowledge (Kogut and Zander, 1992). For a consultancy this capacity works in two directions, and the research on it is uneven (Lemus-Aguilar et al., 2015): consultants help clients innovate, while the firm itself must innovate its own ways of working. AI touches both at once. It changes what consultants can do for clients and it changes how consultants work. The thesis therefore treats innovation, and not AI adoption or productivity, as the goal, and AI as one of the means.

What innovation capability means here. Innovation capability is described in the literature as the ability of an organisation "to continuously transform knowledge and ideas into new products, processes and systems for the benefit of the firm and its stakeholders" (Lawson and Samson, 2001, p. 384). This thesis makes its own

operationalisation of that organisational concept at the level of the individual consultant, because in a consultancy the service is produced by individual consultants; it is a working definition of this thesis, not a construct taken from the literature on individual innovation capability or innovative work behaviour, and it narrows the concept to four connected abilities: the ability to recognise where a client problem can be approached differently, to frame it as a workable problem, to choose and apply a structured method to it, and to bring AI into that process without surrendering judgement. The first three come from the innovation and design literature discussed in sections 2.3.1 to 2.3.3; the fourth follows from sections 2.1 and 2.2. This working definition is used throughout the thesis and is revisited in section 2.6 after the interviews.

2.3.1 Problem solving: the problems consultants work on

At the heart of every consulting engagement is a problem that needs to be solved. But not all problems are the same. Simon (1973) made a distinction that is still relevant today: well-structured problems have a definite test for a solution and a problem space that can be represented, ill-structured problems lack one or more of these, and the boundary between them is fluid. Rittel and Webber (1973) add that problems in organisations and planning are typically "wicked": ambiguous, involving competing interests, and resisting a single right answer. This thesis assumes, and the interviews in section 2.5.3 examine, that the problems clients bring to consulting firms almost always fall into the second category, and that figuring out what the real problem is, before attempting to solve it, is the most important and most difficult part of the work.

This has direct consequences for how AI fits into consulting practice. AI is powerful when the problem is already well-defined: it can synthesise information quickly, apply frameworks and generate structured outputs at a speed no human can match. The evidence of section 2.1 suggests that it is weaker at the messy, ambiguous front end of an engagement: recognising when a problem is being framed in the wrong way, picking up on resistance that a client never explicitly states, or judging which recommendation will work given the culture and politics of an organisation; this is the thesis's reading of the exposure estimates of Eloundou et al. (2024) and of the frontier that Dell'Acqua et al. (2026) describe, not a finding of either. This thesis infers that problem structuring, the ability to translate a vague and complex client situation into a clear and workable problem statement, is one of the most important skills a consultant can develop, and that it is the point where the innovation agenda and the AI agenda connect: a consultant who cannot articulate what the client's problem is will not be able to direct AI to help solve it (assumption A4). Whether problem structuring is learnable through practice with a structured method is the question of the next two sections.

2.3.2 Design thinking as one structured approach

Section 2.3.1 concluded that the problems clients bring to consultants are ill-structured and that framing them is the most demanding part of the work. The question that follows is which structured approaches exist for such problems and whether consultants can learn them. Several candidates are documented in the literature, among them systems thinking, lean and agile methods, and design thinking. This thesis examines design thinking in most detail for three reasons, stated here as the reasoning of the thesis rather than as a finding: it is explicitly built around ill-structured, human-centred problems, which matches the problem type of section 2.3.1; it has a documented track record of use by managers and professionals without a design background (Liedtka, 2018; Micheli et al., 2019); and its methods are catalogued with a consistent structure, the same elements per method, which makes them a usable index for the AI-supported approach that sections 2.1 and 2.2 point towards. Whether design thinking is in fact the approach that fits how consultants at Supply Value solve problems is examined in the co-creation session (chapter 4); the interviews (section 2.5) ask how consultants currently approach problems without imposing any method on them.

Design thinking is an iterative, human-centred approach to problem solving that has moved well beyond its origins in the design world. Its influence reaches beyond design: design thinking has become part of mainstream

management practice (Micheli et al., 2019). The process is usually described in five phases, empathise, define, ideate, prototype and test (Micheli et al., 2019). What makes it different from conventional analytical problem solving is not just the steps but the mindset: Liedtka (2018) highlights that it creates space for experimentation without fear of failure, grounds decisions in direct observation rather than assumption, aligns teams around shared design criteria, and makes thinking visible through prototyping so that ideas can be tested rather than debated in the abstract. In ten case studies in the public and social sector, Liedtka et al. (2017) describe how design thinking produced solutions that the people involved accepted and carried forward, because they had taken part in developing them; this thesis infers that this point matters particularly for a firm whose consultants are responsible for implementation.

For this thesis the most important point is that the reviews treat the core attributes of design thinking, among them user empathy, problem framing and iteration, as practices that can be learned rather than as a designer's talent (Liedtka, 2015; Micheli et al., 2019): design thinking has been taken up by managers and professionals without a design background, and whether training produces these attributes is still an open question in that literature. The practical challenge is that design thinking requires repetition to use well under pressure; knowing the rules of a method and performing it fluently are different stages of skill (Dreyfus and Dreyfus, 1980). Sreenivasan and Suresh (2024) report that AI can support ideation, prototyping and collaboration within the design process, which this thesis reads as support for practitioners who lack formal training, and Hoßbach and Isaksen (2025) argue that AI in this context can act as a collaborator, playing roles such as brainstorming partner, potential customer or coach. Verganti et al. (2020) add an important limit: AI is strongest at generating variations within a known space and weaker at reframing problems or identifying new directions, which remains a human responsibility. This thesis assumes (A4, A8) that AI can provide part of the structure that consultants without formal training need, in the framing and definition steps by prompting the right questions and in ideation and prototyping by generating options the consultant evaluates; the assumption is tested in chapters 4 and 5.

2.3.3 Design methods and approaches

Design thinking is only as useful as the methods that support it in practice. Methods like stakeholder mapping, persona development, journey mapping and how-might-we questions each serve a specific purpose in the innovation process. Stakeholder mapping forces consultants to think explicitly about who is affected by a problem and how. Personas embed empathy into the work. Journey mapping surfaces the experiences and frustrations that clients actually have rather than the ones a consultant assumes they have. These are not complicated methods in principle, but using them well under the pressure of a real client engagement requires practice and structure. The Delft Design Guide (Van Boeijen et al., 2020) catalogues such methods per phase of the design process, each with the same elements: what the method is for, when to use it, how to carry it out in outline, and where to read further. The catalogue is an index rather than a manual: the descriptions are short, and turning one into guidance that a consultant can follow step by step requires the method's own literature and the practice of the firm, which is what chapter 5 does for the methods it deepens (section 5.4.4). Supply Value has its own methods that are used in the same way, such as the Kraljic matrix and spend analysis in procurement; these are part of the same repertoire and are treated as such in chapter 5.

2.3.4 How innovation capability develops in consulting firms

There is a documented tension in the consulting industry: firms help clients innovate while their own internal innovation is both necessary for survival and little studied (Lemus-Aguilar et al., 2015); this thesis calls that the innovation consultant dilemma. When consultants work alongside clients on complex challenges they can act as co-producers of innovation rather than only as carriers of existing knowledge (Den Hertog, 2000); in this thesis's reading a digital transformation at a municipality is not solved by applying a fixed template, and the quality of

that co-creation depends on how capable the consultant is of facilitating an innovative process rather than an analytical one.

The internal side is harder. Taminiau et al. (2009) found, in interviews at Dutch consulting firms, that innovation within consulting firms tends to happen through informal knowledge sharing among practitioners rather than through formal knowledge-sharing channels, and Anand et al. (2007) found that new practice areas emerge from motivated consultants with their own expertise and networks, but only take hold with organisational support and a defensible turf of their own. Learning and knowledge theory point in the same direction: capability develops through experience and reflection on real problems (Kolb, 1984), through scaffolding (D. Wood et al., 1976), rooted in the zone of proximal development (Vygotsky, 1978, p. 86), that is gradually withdrawn as competence grows (Collins et al., 1989; Van de Pol et al., 2015), and through the conversion of tacit into explicit knowledge within a community (Nonaka and Takeuchi, 1995). Acceptance, finally, depends on whether people expect a personal benefit (Davis, 1989), and motivation that is autonomous produces different results from motivation that is controlled (Ryan and Deci, 2000): a programme that relies on mandates produces compliance, not capability. The 70-20-10 model that is often cited in this context is deliberately not used here, because its empirical basis is weak (Clardy, 2018).

2.3.5 What this means for the research question

Innovation capability, as used in this thesis, is the ability to recognise where a client problem can be approached differently, to frame it as a workable problem, to choose and apply a structured method to it, and to bring AI into that process without surrendering judgement. The literature suggests that this capability is learnable, that it develops through practice on real problems rather than through training alone, and that AI can support the method steps where variation and structure help while the framing and the reading of the client remain human work (Liedtka, 2018; Verganti et al., 2020). Assumption A4 (a consultant who cannot frame the problem cannot direct AI) is supported in principle. Three mechanisms carry forward into the design: learning by doing on live assignments, scaffolding that fades as capability grows, and peer exchange within Supply Value.

2.4 The client environment: regulation, data governance and organisational readiness

This section answers the third question of section 1.3 from the literature: what does the client environment allow, and what does that mean for responsible AI use by consultants. It is short, because the empirical answer comes from the interviews (section 2.5.4); its purpose is to set out the three structural conditions that any design has to respect.

Data protection and the location of data

Geopolitical tensions, national security concerns and increasingly stringent data protection frameworks have driven a structural shift in AI infrastructure towards sovereign AI: the migration of data, workloads and AI capabilities from global cloud platforms to regionally or nationally controlled environments; Gartner (2025a) expects that by 2029 more than half of multinational organisations will have a digital sovereignty strategy, up from fewer than one in ten today. The use of commercially available LLMs, where submitted data may be processed and potentially used for training, poses data security risks; the MIT NANDA study reports that workers at over 90 percent of organisations use personal AI tools for work, often without the knowledge or approval of their IT department (Challapally et al., 2025).

A mid-sized consultancy cannot build sovereign infrastructure. The workable path is a combination of enterprise agreements that contractually exclude training on submitted data, retrieval on private data stores, and explicit rules on what may be submitted to which system. Because the rules differ per client, and within a client per

project, those rules cannot be a single firm-wide policy: they have to be agreed at the level of the client and the project and be present at the moment a consultant submits something. Supply Value has an internal AI policy for client-related work. The interviews test whether rules per client exist, whether they are clear enough to be followed, and whether consultants know them at the moment of use (assumption A3).

Governance, liability and the EU AI Act

As AI systems take on more autonomous roles, the role of the human in the loop becomes more important. Gartner (2025c) predicts that by the end of 2026 more than 2,000 legal claims will have been filed over harm caused by AI, which it calls "death by AI" claims; this is a forecast, not an observation. The most significant regulatory response so far is the EU AI Act (Regulation EU 2024/1689), the first general law on AI. It works with risk classes: a small number of uses are prohibited, such as social scoring; a list of high-risk uses carries heavy obligations on documentation, human oversight and quality, among them systems in critical infrastructure, education, recruitment, credit scoring and systems that evaluate emergency calls or dispatch emergency services (Annex III); everything else is permitted with light transparency duties. The Act distinguishes providers, who build or market an AI system, from deployers, who use one, and it enters into force in phases: the prohibitions and Article 4 on AI literacy apply since 2 February 2025, the rules for general-purpose models since August 2025, most high-risk obligations from August 2026. Article 4 obliges providers and deployers to take measures that support the development of AI literacy among their staff and other persons operating AI systems on their behalf; the original text of 2024 required them "to ensure, to their best extent, a sufficient level of AI literacy", a formulation relaxed by the Digital Omnibus on AI of July 2026 (Regulation (EU) 2024/1689, Article 4; European Commission, 2025).

Three things. Supply Value is a deployer of the AI its consultants use, so Article 4 applies to it directly, and a description of the capability consultants need for working with AI is, in this context, also a compliance instrument: it documents the measures with which Supply Value supports AI literacy. Its clients are deployers of the AI the consultants help implement, and at public clients some uses fall in the high-risk class, which is why a consultant at a public client speaks of certification that takes years (section 2.5.4). And the Act turns responsible use from an intention into a legal duty, which is why rules on what may be submitted to which system have to be specific to the client and the project rather than a general policy.

Organisational readiness in the public sector

Public organisations operate under strict rules on data protection, procurement and, in healthcare, on medical devices, and are known to change workflows and behaviour slowly. The most recent systematic picture of the Dutch public sector is the audit of the Netherlands Court of Audit of October 2024 (Algemene Rekenkamer, 2024): 70 central government organisations reported 433 AI systems, 167 of which were still experiments or pilots; 88 percent of the organisations used no more than three systems; for more than half of the reported systems the organisation had not weighed opportunities against risks; and for 35 percent of the systems it was unknown whether the system met expectations. The audit covers central government, not municipalities or care institutions, and it predates the AI Act obligations of 2025; the literature therefore gives no current picture of the readiness of the clients Supply Value works for, and this condition rests for the rest on the interviews, which is noted as a limitation in chapter 7. What the audit does show is that AI in the Dutch public sector is in an experimental phase in which governance lags use, which is the environment in which a consultant has to agree rules per client (cluster 3). The interviews (section 2.5.4) provide the empirical picture: rules that differ per client, that depend on leadership, and that range from encouragement to outright bans.

What this means for the research question

Assumption A3 (uncertainty leads to non-use) is plausible in the light of the shadow AI finding but cannot be settled from the literature. Two conclusions are carried forward. Conclusion 6: the rules on what may be

submitted to which system differ per client and per project, and clients in this phase rarely have them ready, so they have to be agreed at the start of an assignment and be present at the moment a consultant submits something, rather than left to the consultant's memory. Conclusion 7: AI literacy in the sense of Article 4 is a legal duty for Supply Value as a deployer, so a description of the capability also documents compliance. Section 3.3 derives the requirements from both.

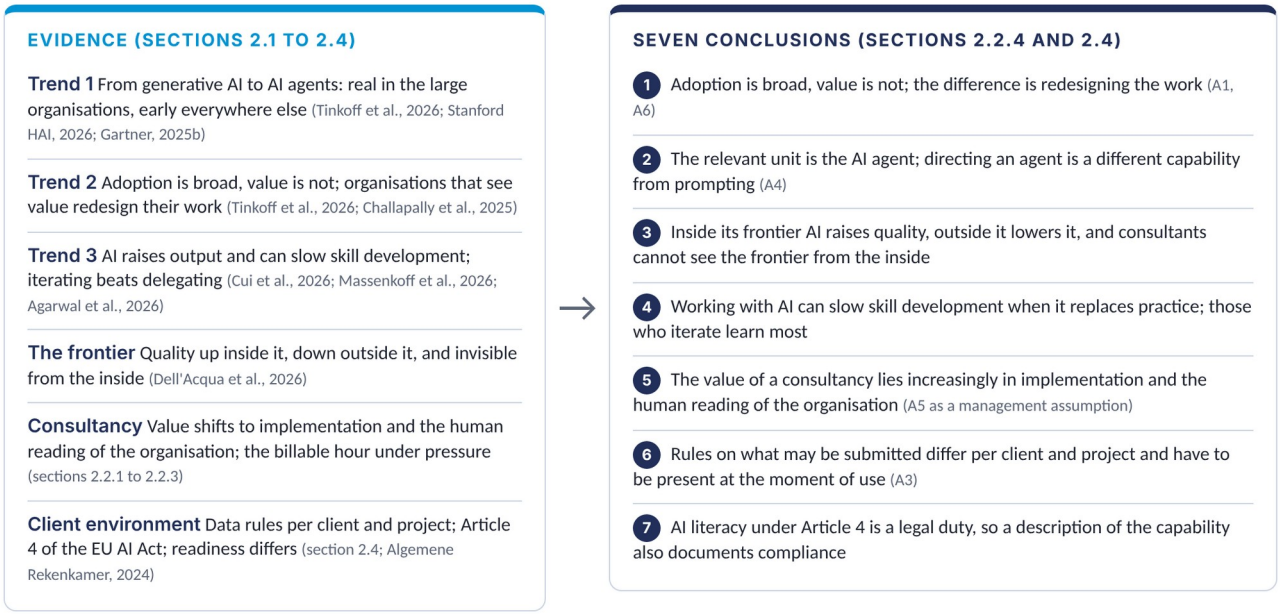


Figure 2.2 The seven conclusions of the literature and the evidence behind them. The requirements that follow from them are formulated in the design brief (section 3.3).

2.5 Interviews: the situation at Supply Value

This section reports eight semi-structured interviews with management and consultants of Supply Value, coded thematically, as eleven themes built from codes; each theme is summarised here with its evidence, and Table 2.2 gives the overview. The full account per theme, with what participants said in their own words, the counter-voices, the researcher's inferences and what is carried forward, is in Appendix J. The codebook is in Appendix C, the coded fragments in Appendix D.

2.5.1 Interview design and analysis

Purpose

The interviews serve the empirical side of three of the four questions of section 1.3: how consultants at Supply Value currently work and use AI and why some use it more than others; how they approach client problems and where AI could add value; and what the client environment allows and how they deal with it. They also ask what clients continue to need from a human consultant, as seen from Supply Value. The literature question of section 1.3 is addressed in sections 2.1 to 2.4. The interviews also test the assumptions A1 to A8 of chapter 1.

Sequence

Four interviews (P1 to P4) were held before the problem was formulated in its present form, with a topic list on Supply Value's AI ambitions, the tools in development and the daily work of consultants who use AI. Four (P5 to P8) were held afterwards, with guides derived from the four questions of section 1.3, and reached people the first four had not. The first four interviews were coded afterwards with the same codebook as the others, so that all eight contribute to the same analysis. Ten interviews were planned; eight were held, because of the time

available and the availability of the consultants in that period. The eight cover management and consultants of different seniority; they do not cover clients, external experts or consultants at medior level.

Sampling and participants

Participants were selected in two rounds: the first four for the orientation on Supply Value's AI programme, the second four purposively, to fill the gaps on three dimensions, role (management, consultant), seniority (junior, senior) and current AI use (from daily use to little or no use), so that the sample contains contrasts on each dimension rather than a representative average. What the frontrunners in the sample do differently may partly come with their role. The participants come from all three business units and from across Supply Value, with one business unit over-represented. The roles overlap, as they do at Supply Value: managers can also work on client assignments, as senior consultants. No consultant at medior level was interviewed: those available had no time in the period of the interviews; the sample therefore contrasts seniors and management with juniors, which is noted as a limitation in chapter 7. Table 2.1 describes the participants as a group: eight people with experience in consulting or management, who differed in seniority and frequency of AI use. A supervisory relationship within the sample may have influenced what participants felt comfortable sharing. Names, roles and other details per person are not given, and all participants are referred to by their code and as "they".

Characteristic	The eight participants
Background	Experience in consulting or management
Seniority	Junior to senior; no consultant at medior level
Current AI use	From daily use to little or no use
Interviews	Four before and four after the problem was formulated (section 2.5.1); 27 to 62 minutes each

Table 2.1 The interview participants as a group. Roles, seniority and AI use are not given per person, so that no participant can be identified; in the text participants are referred to by their code only.

Interview guides

Five instruments were used; they are included in Appendix B. The guide for P4 (Appendix B.2) is organised along the working day in five phases, from the start of the day through client contact, analysis and reporting to the end of the day, and asks per phase what the consultant does, which tools are used, and where AI helps or falls short; it was written to collect the material for a co-worker journey map. The interviews with P1, P2 and P3 followed a topic list on Supply Value's AI ambitions, the tools in development, data security and daily use (Appendix B.1). One guide of round 2 (Appendix B.4), used in two interviews, has four parts: onboarding at Supply Value, the current client assignment including a recent problem the consultant had to solve step by step, AI use and what holds it back, and a closing part. A second guide (Appendix B.5) has seven parts: the daily work, the organisation, the AI tools and who decides about them, how Supply Value decides on training and the development of consultants, the participant's own use of AI, and the future of Supply Value. The third guide (Appendix B.3) covered the same ground as the guide of B.4 in six phases, with more emphasis on judging the work of others and on what the participant expects of Supply Value. The guides of round 2 follow the same rules: ask for the last concrete occasion rather than for an opinion, one question at a time, no enthusiasm or disappointment about AI shown by the interviewer, and the questions about AI placed after the questions about the work so that they do not colour the earlier answers. The round 1 instruments, written for the AI programme and the journey map, put AI earlier, which is a reason to read the round 1 interviews partly as accounts of the programme rather than of everyday practice.

Procedure and consent

Interviews were held in Dutch, in person or via Teams, and lasted between 27 and 62 minutes. The TU Delft consent form was sent beforehand and consent was also recorded orally at the start of the recording, after the

same opening statement in every interview (Appendix B, Procedure). Each interview was recorded, transcribed automatically and corrected by the researcher, and directly after it the researcher made five minutes of field notes. Interviews were transcribed with the built-in transcription of Microsoft Teams and Microsoft Word; full interview transcripts containing names were submitted to an external AI service for coding assistance (Claude Opus 5.5 and Claude Fable 5.1, Anthropic, Max subscription), under account settings in which the data is not used for training AI models, and the service was instructed to omit names from its output; no application for prior human research ethics review was submitted (section 7.3); recordings and transcripts are stored on the SharePoint of Supply Value and locally on the researcher's laptop and are deleted after the defence.

Coding

The transcripts were coded thematically in a process that combined inductive and deductive coding; the approach is closest to a codebook form of thematic analysis, with one coder, a growing codebook and a language model as first reader. Each transcript was split into fragments of roughly 25 to 60 words, 681 in total, of which 559 received one or more codes; the 122 fragments without a code contain no content that bears on the questions, such as the opening of the interview or practical talk. Every coded fragment received up to three codes, a marker for whether its content was stated by the participant, inferred by the researcher, or both, a memo where needed, and in the final round a sub-theme. The codebook grew cumulatively: six starting codes derived from the four questions of section 1.3 (for example "Human verification of AI output" and "Building blocks and reuse"), and 101 codes that emerged from the interviews, each with a definition, a note on when the code applies and when not, and the interview in which it first appeared. After each interview the earlier transcripts were re-read for the new codes. The result is 107 codes over 559 coded fragments; 528 of these are marked as stated, 30 as both stated and inferred, and 1 as inferred only. Of the 101 inductive codes, 79 are recorded with the interview in which they first appeared: 8 in P1, 7 in P2, 10 in P3, 16 in P4, 15 in P6, 16 in P7 and 7 in P8; none is recorded as first appearing in P5, and 22 carry no first interview. New codes were still appearing in the last interviews, so saturation was not reached, and the themes are a first map of Supply Value rather than a complete one. The codebook is in Appendix C with the Dutch code names, an English translation and the definitions; the coded fragments are in Appendix D.

A large language model was used in the coding as a first reader. For every interview it proposed codes per fragment on the basis of the codebook; the researcher then went through every fragment in the total workbook, accepted, changed or rejected the proposed codes, added codes where the model had missed content, and recorded the evidence status and a memo where needed; the share of proposals kept unchanged, changed or rejected was not recorded. The final coding is the researcher's; no separate independent coding was done, so no inter-coder agreement is reported. The eleven themes of sections 2.5.2 to 2.5.5 were built by the researcher, without the model, by grouping the codes that answer the same question of section 1.3 and writing each theme from its fragments, with the counter-voice recorded where the fragments disagree (Appendix J); the sixteen sub-themes and five clusters of chapter 3 are a second grouping of the same codes, made for the sub-questions of section 2.6 and described in section 3.1.1 and Appendix E. The model version and the prompt used for the coding are given in Appendix C.

Presentation

The findings in sections 2.5.2 to 2.5.5 are organised by the four questions of section 1.3, not by participant. Each theme is summarised with its main finding, the fragments it rests on, the participants in which it occurs and where it is carried forward; Table 2.2 gives the overview with the assumption each theme bears on. The full account per theme, with what participants said in their own words, a counter-voice where one exists, the researcher's inference and what is carried forward, is in Appendix J, so that the clusters of section 3.1 and the quotations there can be checked against it. Fragment identifiers refer to Appendix D. Themes that rest on one participant are presented as that participant's view, not as a finding about Supply Value.

Theme	Finding	Occurs in	Bears on
1 Most use is generative AI for routine tasks and as a reference, in the consultant's own work	Copilot and ChatGPT for mail, text, summaries and look-ups; frontrunners also for approach and analysis; client-directed AI exists, but not yet in how most consultants work on client problems	All eight	A1; cluster 2
2 After the first wave, adoption has split into a small group that leads and a majority that waits for a use case	The difference is a use case, time and a first experience, not the tools and not motivation; one failed experiment sets the attitude for months	Seven (not P1)	A2, A7; cluster 2
3 What frontrunners do differently: they frame the question, they iterate, and they keep some work by hand	AI skill is tacit and learned by doing; the question determines the output; deliberate manual work is a boundary on the design	All eight	A4; cluster 5
4 Framing the problem is the demanding part, and it is done from experience	A craft of experienced people; no shared approach a junior can follow; consultants carry practice methods they do not call methods	Five (P2, P5, P6, P7, P8)	A4; cluster 4
5 Knowledge sits with people and is lost between assignments	Onboarding alone and under time pressure; building blocks exist but are hard to find; tacit knowledge is the obstacle, retrieval the proposed mechanism	All eight	Cluster 4
6 Consultants want a sparring partner and an overview, not the automation of their thinking	Structure and dialogue, not output; what a consultant keeps by hand is part of staying in control	Seven (not P3)	A8; cluster 5
7 Compliance at the moment of use costs time and is uncertain for some	Anonymising and GDPR-proofing cost time; the governance baseline exists on paper but is not at hand at the moment of use; uncertainty is named by one participant	Six (P1, P3, P4, P5, P6, P8)	A3; cluster 3
8 Client environments differ, and what is allowed depends on leadership	From encouragement to a ban, and a client's own model as the only one permitted; the variation is the finding	Seven (not P8)	A3; cluster 3
9 Verification of AI output is non-negotiable and is a question of trust and liability	Shared across roles; a skill to recognise the frontier, not only a rule	Six (not P2, P8)	Clusters 1 and 3
10 Listening, translating and connecting remain human	Framing the question and reading the organisation are what clients keep paying for	Six (P1, P2, P4, P5, P7, P8)	Cluster 1
11 How AI might change the value of consulting work	Discussed by participants, not measured; a consultant's critique of AI as a goal in itself; two tracks, personal efficiency and firm-wide knowledge	Five (P1, P2, P3, P7, P8)	A5, A6; not clustered (strategy)

Table 2.2 The eleven themes of the interviews: finding, occurrence and what each bears on. The full account per theme is in Appendix J. "Occurs in" counts any coded occurrence of the theme in an interview, supporting or contradicting; the interview instruments differed between the rounds (section 2.5.1). "Bears on" names the assumption of section 1.3 and the insight cluster of section 3.1 to which the theme is carried.

2.5.2 How consultants currently work with AI, and why some more than others

Theme 1: Most use is generative AI for routine tasks and as a reference, in the consultant's own work

Consultants describe using Copilot and ChatGPT to draft mail, rewrite text, summarise documents, prepare for meetings and look things up (P2-028, P3-004, P4-023); P8 calls it "my extended Google reference, but with interaction in it" (P8-059). The frontrunners go further, with daily use for analysis and drafting and an agent for presentations, and with GenAI to shape an approach at the start of an assignment (P5-025). Client-directed AI exists at Supply Value (P1-013, P1-034, P8-029, P8-033), but not yet as part of how most consultants work on client problems. Assumption A1 therefore holds for the consultant's own work and is sharpened: use is

individual, in the consultant's own work rather than in the client's problem. Occurs in all eight interviews; carried into cluster 2.

Theme 2: After the first wave, adoption has split into a small group that leads and a majority that waits for a use case

P8 describes a first half year that went well and then a valley, "a kind of threshold that we still have to clear" (P8-041, P8-042), and a split between a few who are "extremely far" and a group in which they place themselves (P8-043). The consultants outside the first group give concrete reasons: no use case yet (P6-069, P7-054), an AI hour that competes with the work they are judged on (P7-072), a cost-benefit consideration rather than disinterest (P7-073), one failed experiment that settled it (P7-045), and for some a fear of failing (P7-084). The frontrunners do not describe a valley; P3 frames the split as a question of transferring knowledge (P3-023). The difference is not technical, because the tools are the same: it is a use case, time and a first experience, and for some a fear of failing, which supports A2 with a correction. Occurs in seven of eight interviews (not in P1); carried into cluster 2.

Theme 3: What frontrunners do differently: they frame the question, they iterate, and they keep some work by hand

P1 puts the difference in one sentence: "It is no longer about having the right answers. It is about asking the right questions" (P1-024); P2 adds that what is not made explicit in the question the agent does not know (P2-031); P3 compares AI skill to driving, something learned at the wheel (P3-019, P3-020). Frontrunners use AI as a starting point and go back and forth with it rather than accepting the first answer (P4-056, P4-084, P5-025, P5-057). Both P4 and P8 keep some work by hand because it is faster, more precise or more personal: P4 writes actions straight into the notebook (P4-067, P4-068), and P8 would like to have mail written with AI but finds that the personal tone does not come through (P8-053). The difference between the groups is methodical, the empirical counterpart of Trend 3, and the deliberate manual work is a boundary on the design; part of the difference between frontrunners and others may come with their role. Occurs in all eight interviews; carried into cluster 5.

2.5.3 How consultants approach client problems, and where AI could add value

Theme 4: Framing the problem is the demanding part, and it is done from experience

P8 describes problem solving as peeling the problem before solving it, in "a small system of cause and effect" (P8-049 to P8-051). P5 describes work agreements and expectation management as what gives an assignment structure, for example a 30-60-90 day plan, a practice method that appears in no design catalogue (P5-026, P5-027), notes that clients do not want to hear method terms (P5-091), and observes that juniors arrive with theory and without the practice examples (P5-089, P5-090). P6 describes the opposite experience: not knowing what the first step should be (P6-016, P6-036). Framing at Supply Value is a craft carried by experienced people and learned by doing, without a shared approach a junior can follow, and consultants carry methods of their own that they do not call methods. This supports A4 and makes framing the entry condition of the capability. Occurs in five of eight interviews (P2, P5, P6, P7, P8); carried into cluster 4.

Theme 5: Knowledge sits with people and is lost between assignments

Every assignment starts with getting into the content and the organisation, largely alone and under time pressure (P5-024). P6 searched the SharePoint for building blocks and put the risk precisely: "there really is still good information somewhere that I have not yet got to" (P6-013), and looked for help at the client and the predecessor rather than at Supply Value (P6-043, P6-044); P7 describes ideas and agreements that get lost because there is no place to keep them (P7-018). Participants described two practical needs: finding reusable knowledge from earlier assignments and deciding which tools they could use for a particular client task, with tacit knowledge as the main obstacle to reuse (P3-016, P3-017, P3-029). P5 does have working channels for help, the daily start and the business unit meeting (P5-031); whether a consultant finds help depends on whether they

look for it there. This is the most concrete opportunity in the interviews: it is where time is lost, where AI is not used, and where context could be held across steps and assignments. Occurs in all eight interviews; carried into cluster 4.

Theme 6: Consultants want a sparring partner and an overview, not the automation of their thinking

P8 uses GenAI as a reference tool but wants "a sparring partner" (P8-056); P2 wonders whether agents could take part in a brainstorm (P2-046), and P1 shows an agent that prepares one (P1-025, P1-046, P1-047). P6 wants an overview of when to do what and whom to ask (P6-036); P7 wants a system that signals proactively (P7-028); P6 makes the condition explicit: applying a method with AI is a way to learn, letting AI do the work is not (P6-068). P4 deliberately keeps actions out of AI, because correcting a generated text is slower than writing the line (P4-067, P4-068). Consultants ask for structure and dialogue, not output, which matches the scaffolding literature and contradicts the assumption in the original assignment that automation is the main value. Occurs in seven of eight interviews (not in P3); carried mainly into clusters 1 and 5.

2.5.4 What the client environment allows, and how consultants deal with it

Theme 7: Compliance at the moment of use costs time and is uncertain for some

P3, a frontrunner, names information security as what holds them back (P3-042), and anonymising first as not efficient (P3-047); P6 spends more time making things GDPR-proof than the time saving is worth (P6-050). Participants reported that permitted AI use varied by client, task and environment (P1-065, P1-066, P5-069, P5-081). The frontrunners work within the rules, in the approved environment, on tasks without client data, or on the client's own model (P3-046, P5-038, P7-040). Assumption A3, that uncertainty leads to non-use, is indicated by one participant (P3-042); the others describe the cost of compliance and clear rules. The governance baseline exists on paper; what it permits for a given client and task is not at hand at the moment of use. Occurs in six of eight interviews (P1, P3, P4, P5, P6, P8); carried into cluster 3.

Theme 8: Client environments differ, and what is allowed depends on leadership

P2 states that use for the client "is bound to limits" and that one public client has its own language model (P2-042); P7's client permits only its own AI assistant (P7-040, P7-049). At one public client an AI error has serious consequences, and advice on its operational decisions would need certification (P4-005, P4-048, P4-049). Whether a public organisation encourages, ignores or bans AI often depends on who is on the board (P4-053, P4-045), and the AI literacy gap between Supply Value and public clients is a cliff (P4-043, P4-089); at P7's client the willingness is there but developer capacity is scarce (P7-032). Not every client restricts AI; some encourage it and provide training (P4-107). The client environment is a set of profiles that the consultant has to read, a client that prescribes its own model being one of them; this supports keeping client and sector rules as a separate layer and makes the model the support uses a rule per project. Occurs in seven of eight interviews (not in P8); carried into cluster 3.

Theme 9: Verification of AI output is non-negotiable and is a question of trust and liability

P5 treats verification as the same responsibility one has for a colleague's work (P5-063) and names hallucination and repetition in long chats as reasons (P5-055, P5-056); P4 describes a model that "just makes up a number" (P4-058); P3 wants to know themselves what is in a document rather than rely on a summary (P3-005, P3-027); P1 overrules the model where its output is wrong (P1-034, P1-042). The theme is shared across roles, without a counter-voice. Verification is therefore a fixed part of the capability and of the support, and a skill, recognising when the model is outside its frontier, rather than only a rule; the design question this leaves is how the check becomes a skill a consultant learns, and not a gate that costs time and is skipped. Occurs in six of eight interviews (not in P2 and P8); carried into clusters 1 and 3.

2.5.5 What clients continue to need from a human consultant, and what Supply Value expects

Theme 10: Listening, translating and connecting remain human

P1 assigns the roles: asking good questions, making creative connections and social interaction are for people (P1-021); P5 states that soft skills cannot be replaced by AI (P5-083); P7 is the translator between the operation and the automation team and the outsider who may be critical of the client's figures (P7-016, P7-009, P7-010); P4 is the connecting party between the client and the other parties involved (P4-015, P4-016); P2 names "the real listening and looking at the client" (P2-029). One participant wondered whether AI would make exceptional work harder to recognise (P1-112, P1-113). What clients need when AI is available is, in the view of management and consultants at Supply Value, the reading of the organisation and the framing of the question, the same capability this thesis has to build; AI does not compete with it but exposes whether a consultant has it. No client was interviewed, so this is the provider's view of what clients need. Occurs in six of eight interviews (P1, P2, P4, P5, P7, P8); carried into cluster 1.

Theme 11: How AI might change the value of consulting work (assumption A5)

Participants discussed the potential to reuse work across assignments and questioned what clients would value (P1-075, P1-076, P8-029, P8-033). The study did not measure the effect on revenue or establish a preferred pricing model. One participant described a tension between the growing effort of manual work and the cost of automating it (P8-017), with work descriptions as the precondition for agents (P8-024, P8-025). P3 describes AI at Supply Value as two tracks: catching up on personal efficiency in daily tasks, and gathering knowledge firm-wide to leverage it in AI use (P3-010 to P3-013). One consultant warns against making AI a goal rather than a way: start where the value is and then see whether AI can do it. A5 therefore remains an assumption, and the mechanism it describes is the organisational side of the cancellations Gartner counts (Trend 1). Occurs in five of eight interviews (P1, P2, P3, P7, P8); not carried into a cluster, because the codes concern Supply Value's business model.

2.6 Sharpened research question and sub-questions

Table 2.3 states what the literature and the interviews together say about the assumptions of chapter 1. The clusters that the interviews produce are built in section 3.1; the table anticipates them where an assumption depends on them.

	Assumption (chapter 1)	Origin	Literature (2.1 to 2.4)	Interviews (2.5)	Verdict and consequence
A1	Current AI use is directed at own efficiency, not client value	Orientation	Indirect (grey literature: Tinkoff et al., 2026)	Supported and sharpened: for most consultants use is individual or absent; frontrunners use it for approach and analysis; client-directed AI exists, but not yet in how consultants work on client problems (themes 1, 3)	Holds for the consultant's own work, with two groups; the client-directed products are the exception. Entry via a concrete own task.

	Assumption (chapter 1)	Origin	Literature (2.1 to 2.4)	Interviews (2.5)	Verdict and consequence
A2	Capability sits with early adopters; the difference is motivational and methodical	Orientation	Partly (Dell'Acqua et al., 2026: largest gains for lower-skilled users when they do use AI)	Supported with a correction, also in the four interviews held after the problem was formulated: same tools; the difference is a use case, time and a first experience, not motivation in general (theme 2)	Per claim: concentration in a small group supported; a motivational difference not supported; a methodical difference supported. The capability targets framing and method, not tool training; the implementation makes time.
A3	Uncertainty about rules leads to non-use	Orientation	Plausible (shadow AI, Challapally et al., 2025)	Indicated by one participant (P3-042); the others describe the cost of compliance and clear rules, and at some clients a real ban (themes 7, 8)	Weakly supported. Rules per client and project, checked at the moment of use, follow from theme 8 and section 2.4 rather than from A3.
A4	Who cannot frame the problem cannot direct AI	Orientation	Plausible, the thesis's inference from sections 2.3.1 and 2.3.2 (Simon, 1973; Liedtka, 2018) and from Dell'Acqua et al. (2026)	Supported (P1, P2, P3, P8: the question determines the output; theme 4: framing is the craft)	Holds as design principle; its effect is tested in chapter 6.
A5	Faster work plus hourly billing may reduce revenue per assignment	Orientation	Weak (grey literature)	Discussed by participants, not measured (theme 11)	Remains an assumption; not a design requirement.
A6	Automation alone does not distinguish a firm	Assignment	Supported (workflow redesign distinguishes high performers)	Supported (P7: value first; P8: work descriptions before agents)	Holds. The design starts from the work process, not from the tool.
A7	Without a shared vision, scattered use	Assignment	Partly (Gartner, 2025b: unclear value as reason for cancellation)	Nuanced: there is a vision, the AI programme, but it does not reach the second group (theme 2)	Adjusted: the missing element is not the vision but its translation into the consultant's work. Roadmap.
A8	Capability building is an appropriate response; form open	Assignment	Supported: learning by iterating works (Trend 3)	Consultants themselves ask for an overview, a checklist, a sparring partner (themes 4, 6)	Holds against the criterion of section 1.3: the barriers found are not access to tools or client permission alone (themes 2, 4, 7, 8); the form is decided in chapter 4.

Table 2.3 Test of the assumptions of chapter 1 against literature and interviews. Origin: formulated after the orientation at Supply Value (A1 to A5) or taken from the assignment (A6 to A8). Because the first four interviews fed the formulation of A1 to A5, the test is partly circular for those: A1, A2 and A4 recur in the four interviews held after they were formulated, A5 recurs in both rounds, and A3 rests on one interview of the first round (P3-042).

Two things changed between the initial research question and this point. First, the type of AI is now specific: the relevant unit is the AI agent, with prompt-based GenAI as the entry level (section 2.1.1, Trend 1). Second, the capability is now specific: the interviews indicate two differences, whether a consultant has a use case and time to try it, and, among those who use AI, whether they frame the question and judge the result; neither lies in access to tools. The sharpened research question is therefore:

> **How can Supply Value strengthen the innovation capability of its consultants, so that they can direct AI agents at client problems and create more value for clients in advisory and**

implementation work within regulated client environments, while developing rather than replacing their own professional skills?

The main question is subdivided into five sub-questions. Each covers something that must be known to answer the main question; the findings it comes from are given in brackets, and table 1.1 names the chapter that answers it. The working definition of section 2.3 stands after the interviews: they locate the difference between consultants in its second and fourth ability, framing and judgement, and add that the context of the assignment must be held across steps, which is a condition for the support rather than an ability of the consultant.

SQ1. How do consultants at Supply Value currently use AI in their work, and what distinguishes the consultants who use it extensively from those who use it little or not at all? (themes 1 to 3; conclusion 1)

SQ2. How do consultants at Supply Value currently approach and solve client problems, and where in that process could an AI agent add value for the client? (themes 4 to 6; conclusion 2)

SQ3. Which limits set by regulation, client policy and data governance determine what AI use is possible in the client environments of Supply Value, and how do consultants deal with those limits today? (themes 7 to 9; conclusions 6 and 7)

SQ4. How is innovation capability built in a consultancy, and what does the literature say about how professionals learn to work with AI and AI agents without losing their own skills? (section 2.3; conclusion 4)

SQ5. What do clients continue to need from a human consultant when AI agents can take over part of the analytical work, and what does that mean for the capability consultants need? (theme 10; conclusion 5) SQ5 has no counterpart among the four questions of section 1.3; it follows from the premise of section 1.1 and from theme 10 of the interviews.

This chapter answers SQ1, SQ3 and SQ5 from the interviews and the literature; the clusters of section 3.1 consolidate those answers. In plain words: most consultants use AI for their own efficiency and a small group uses it to shape approach and analysis, and what separates them is a use case, time and a methodical way of working, not the tools (SQ1; themes 1 to 3). What the client environment allows differs per client and per project, and the uncertainty about it, more than the rules themselves, holds consultants back (SQ3; themes 7 to 9). Clients keep paying for framing the question, reading the organisation and judgement, which AI does not replace (SQ5; theme 10). SQ5 is answered from the provider side: no client was interviewed, so what clients need is what management and consultants of Supply Value say clients need. SQ2 is answered in part: where in the consultant's process an agent adds value is the subject of chapters 3 and 4 and is tested in chapters 5 and 6. SQ4 is answered from the literature; whether those mechanisms work for consultants at Supply Value is the subject of chapter 6. Section 7.1 gives the final answers. The question of how a firm keeps up with a rapidly changing AI environment, which was raised at the start of the research, is not carried forward as a sub-question; it is not a question the Discover phase can answer and is left to Supply Value.

3

CHAPTER 3 · DEFINE PHASE

Define

From 681 coded fragments to five insight clusters, two journey maps from inside the work, and a design brief with fourteen requirements.

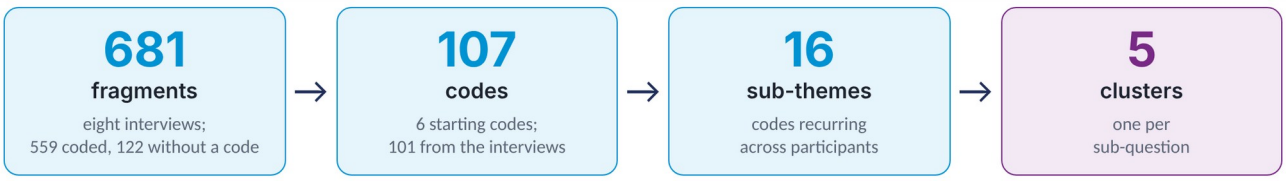


IN THIS CHAPTER 3.1 Insights: from codes to clusters · 3.2 How a consultant at Supply Value works: the assignment process and two journey maps · 3.3 Design brief

3.1 Insights: from codes to clusters

3.1.1 How the insights were derived

The clusters in this chapter are built from the coded interviews of section 2.5 in three steps, shown in figure 3.1. First, following the coding approach of section 2.5.1, the 107 codes were grouped into sub-themes: a sub-theme is a claim about the situation at Supply Value that several codes support together. The retention rules are this study's own: a sub-theme was retained when it was supported by at least two participants or, for a single participant, when it concerned that participant's specific role, for example a view that only one participant's role made possible. This produced sixteen sub-themes. Second, the sub-themes were grouped into clusters on the basis of the question they answer for the research question, so that each cluster bears on one or two of the sub-questions of section 2.6. This produced five clusters; themes and clusters thus group the same codes along different axes, the themes by the questions of section 1.3 and the clusters by the sub-questions of section 2.6. Third, for each cluster the evidence base was recorded: the number of codes and fragments, the number of participants in which it occurs, and the counter-voices. Every coded fragment in the workbook is also marked as stated by the participant, inferred by the researcher, or both; in every cluster more than 90 percent of the fragments are marked as stated and only one fragment in the whole workbook is marked as inferred only, so the clusters rest on what participants said; the claims that group them are the researcher's and are marked as inference where they go beyond the fragments. This is not repeated per cluster. A cluster supported by fewer than half of the participants would be presented as tentative; none of the five is. The same rule applies to the sub-themes: table 3.1 gives the number of participants per sub-theme, and a sub-theme with fewer than four participants is read as tentative. The full assignment of codes to sub-themes and clusters is in Appendix E. The literature and the desk research of sections 2.1 to 2.4 are not an input to this clustering; their conclusions (section 2.2.4) feed the design brief next to the clusters (table 3.5), and sections 2.2 to 2.4 hold the relation of each cluster to the literature. Two clusters keep the name of the theme that dominates them.



Not carried into clusters: 12 codes on strategy and business model (65 fragments, used in 2.2.2, 2.5.5 and chapters 6 and 7) and 4 codes on role context (73 fragments, used in 2.5.1 and 3.2). A sub-theme needs at least two participants, or one participant in a unique role.

Figure 3.1 From 681 fragments to five clusters. Fragments and codes are counted per cluster; a fragment can carry up to three codes and can therefore count in more than one cluster.

Sixteen of the 107 codes were deliberately not carried into a cluster. Twelve codes concern the value of consulting work and Supply Value's business model; they answer a question about Supply Value, not about the capability of consultants, and are used in section 2.2.2 and in the discussion. Four codes describe the role context of the participants and are used to describe the sample (section 2.5.1) and the assignment process (section 3.2).

An insight is a statement about the situation at Supply Value that is grounded in fragments that participants stated rather than in the researcher's inference, that recurs across participants with different roles or levels of AI use or is contradicted in a way that is itself informative, and that has a consequence for the research question. Where an insight relies on inference, this is stated. Eight interviews within one firm do not support claims about consultancy in general; the insights are about Supply Value and are used as such.

3.1.2 Overview of the clusters and their evidence base

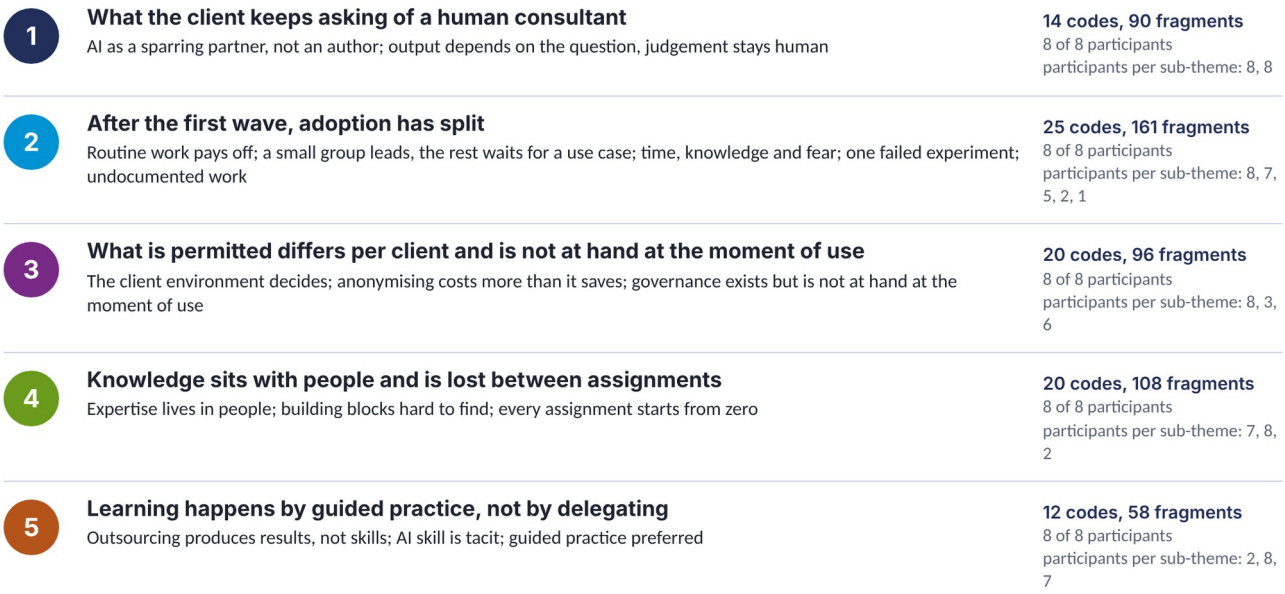


Figure 3.2 The five insight clusters with their sub-themes and evidence base.

Cluster	Sub-themes (participants)	Codes	Fragm.	Partic.
1. What the client keeps asking of a human consultant (SQ5)	AI as a sparring partner, not as an author (all eight); Output depends on the question, judgement stays human (all eight)	14	90	8 of 8
2. After the first wave, adoption has split (SQ1)	Routine work is where AI already pays off (all eight); A small group leads, the majority waits for a use case (P2, P3, P4, P5, P6, P7, P8); Not disinterest but cost and benefit: time, knowledge and fear (P2, P3, P4, P7, P8); One failed experiment can set the attitude (P4, P7); Undocumented manual work cannot be handed to AI (P8)	25	161	8 of 8
3. What is permitted differs per client and is not at hand at the moment of use (SQ3)	The client environment decides what is possible (all eight); Anonymising costs more than it saves (P3, P5, P6); Governance exists but is not at hand at the moment of use (P1, P3, P4, P5, P7, P8)	20	96	8 of 8
4. Knowledge sits with people and is lost between assignments (SQ2)	Expertise lives in people, not in documents (P1, P2, P3, P5, P6, P7, P8); Building blocks exist but are hard to find and use (all eight); Every assignment starts from zero (P5, P6)	20	108	8 of 8
5. Learning happens by guided practice, not by delegating (SQ1, SQ4)	Outsourcing to AI produces results, not skills (P5, P6); AI skill is itself tacit knowledge (all eight); Guided practice as the preferred route (all except P2)	12	58	8 of 8

Table 3.1 The five clusters and their evidence base. Fragments are counted once per cluster; a fragment with codes in two clusters counts in both.

3.1.3 Cluster 1: What the client keeps asking of a human consultant

EVIDENCE BASE **14** codes · **90** fragments · all eight participants · bears on SQ5

Sub-themes. AI as a sparring partner, not as an author: participants who use AI well use it to test their thinking, to get a first version or to brainstorm, and keep authorship of the result. Output depends on the question, judgement stays human: the quality of what AI produces is set by the quality of the question, and verification, soft skills and the reading of the client remain with the consultant.

What participants said. Across roles, participants describe the human part of the work in the same terms, asking the good question, listening, translating between operations and technology and connecting parties, and they describe verification of AI output as a non-negotiable responsibility. Where AI is used well it is a starting point or a sparring partner: "they had something to build on, so that you do not really have to start completely from zero" (P4-056). Fragments: P1-021, P1-024, P2-029, P5-083, P5-063, P4-058 (Appendix D); the codes per sub-theme are in Appendix E.

Counter-voice. None; the theme recurs across management, seniors and juniors. The cluster answers SQ5 from the provider side: no client was interviewed (section 7.3).

Researcher's inference. The capability Supply Value has to build is the same as the value the client continues to pay for: framing the question and reading the organisation. AI does not compete with that capability; it exposes whether a consultant has it, and one participant wondered whether AI would make exceptional work harder to recognise (P1-113).

For the design brief. The capability is defined in terms of value delivered at the client, and human validation of AI output is part of it as a skill, not only as a rule. Confirms conclusions 3 and 5 of section 2.2.4.

3.1.4 Cluster 2: After the first wave, adoption has split

EVIDENCE BASE **25** codes · **161** fragments · all eight participants · bears on SQ1 · The largest cluster.

Sub-themes. Routine work is where AI already pays off: mail, summaries, presentations, a macro. A small group leads, the majority waits for a use case: the valley after the first wave. Not disinterest but cost and benefit: the time, the knowledge and the fear that keep the majority where it is. One failed experiment can set the attitude: P7's comparison of records across systems, P4's Copilot link that has not worked for weeks. Undocumented manual work cannot be handed to AI: P8's view.

What participants said. P8 describes a first half year in which the AI programme went well, followed by "a valley" with a growing distance between a small group that sets up agents and the rest (P8-042, P8-043). The consultants outside the first group give concrete reasons, no use case that is obviously worth the effort, no time next to the work they are judged on, a failed experiment, and a cost-benefit consideration rather than disinterest; P8 adds that the firm is too small to automate its operations and too large to run them by hand, and that work descriptions have to exist before agents can take work over. Fragments: P6-069, P7-054, P7-072, P7-045, P7-084, P7-073, P8-017, P8-022, P8-024, P8-025 (Appendix D).

Counter-voice. The frontrunners do not describe a valley (P3, P4, P5), and P3 sees the same split as a problem of transferring knowledge rather than of motivation (P3-023).

Researcher's inference. The split is not technical. What differs is whether a consultant has a concrete problem in which AI has proven useful, whether there is time and permission to find one, and whether a first attempt failed; the early adopters and the majority differ in how they weigh a new practice, as the adopter categories of diffusion research describe (Rogers, 2003). Support for the majority therefore has to start from the consultant's

own assignment and pay back within the first session, and the implementation has to make time that does not compete with billable work.

For the design brief. The entry point is a concrete task on the consultant's own assignment; the capability describes work processes before it introduces agents; the implementation makes time for learning. Confirms conclusion 1 of section 2.2.4.

3.1.5 Cluster 3: What is permitted differs per client and is not at hand at the moment of use

EVIDENCE BASE **20** codes · **96** fragments · all eight participants; per sub-theme eight, three and six · bears on SQ3

Sub-themes. The client environment decides what is possible: the client's own model, a ban, medical device regulation, an AI literacy gap, scarce developers. Anonymising costs more than it saves: the workaround that does not work. Governance exists but is not at hand at the moment of use: the firm has rules, guardrails and a compliance check; one participant names uncertainty as what stops them.

What participants said. One participant, a frontrunner, names uncertainty about what is permitted as what holds them back: "information security is what is now holding me back" (P3-042). The others describe the cost of complying at public clients and the range of client environments, from a client with its own language model to one where an AI error has serious consequences and one that wants nothing to do with AI, with the board deciding which; inside Supply Value, participants reported that permitted AI use varied by client, task and environment. Fragments: P3-047, P6-050, P2-042, P7-049, P4-005, P4-049, P4-045, P4-053, P5-069, P1-065, P1-066 (Appendix D).

Counter-voice. Frontrunners, and P7 at a client that prescribes its own model, work within the rules rather than avoiding AI, in the approved environment, on the client's own model or on tasks without client data (P3-046, P5-038, P7-040), and some clients encourage AI and provide training (P4-107).

Researcher's inference. The accent lies on the moment of use rather than on prohibition: the rules exist, but what they permit for this client and this task is not at hand when a consultant works, and the cost of complying falls on the consultant. A rule that lives in a policy document does not change behaviour; the rule has to be agreed per client and project and be present at the moment of use. Helping the client use AI responsibly is itself part of the consultant's work (P4).

For the design brief. Rules on what may be submitted to which system are agreed per client and project and checked by the support at the moment of use; client and sector context is a configurable layer, not baked into methods. Confirms conclusions 6 and 7 of section 2.4.

3.1.6 Cluster 4: Knowledge sits with people and is lost between assignments

EVIDENCE BASE **20** codes · **108** fragments · all eight participants · bears on SQ2

Sub-themes. Expertise lives in people, not in documents: tacit knowledge, peeling the problem, method knowledge through colleagues, help that stays within the assignment. Building blocks exist but are hard to find and use: Supply Value's reusable components and the wish for an overview, a checklist or a signal. Every assignment starts from zero: onboarding pressure, role ambiguity at the client, vague assignments.

What participants said. Every new assignment starts with building context, largely alone and under time pressure (P5-024). P6 found onboarding incomplete and building blocks hard to find, and did not know what they had not seen: "there really is still good information somewhere that I have not yet got to" (P6-013); for help they went to the predecessor and to colleagues at the client rather than to Supply Value; ideas and agreements at the

client get lost; P8 describes framing as peeling the problem before solving it, P5 describes expectation management as the structure of an assignment and juniors as arriving with theory and without the practice examples; reuse of knowledge meets tacit knowledge as its main obstacle. Fragments: P6-004, P6-043, P6-044, P7-018, P8-049, P8-050, P5-026, P5-089, P3-017, P3-029 (Appendix D).

Counter-voice. P5 has working channels for help, the daily start and the business unit meeting (P5-031), which P6 did not use for the assignment: whether a consultant finds help at Supply Value depends on whether they look for it there.

Researcher's inference. The fragmented-context problem is the most concrete opportunity in the data: it is where time is lost, where AI is not used, and where an agent that holds context across steps and assignments can add value without replacing judgement. Problem framing has no shared structure at Supply Value; it is a craft learned by watching.

For the design brief. The support is an agent that holds the context of an assignment across steps and time, helps with the framing step explicitly, and captures knowledge during an assignment in a form that flows back to Supply Value. Confirms conclusion 2 of section 2.2.4.

3.1.7 Cluster 5: Learning happens by guided practice, not by delegating

EVIDENCE BASE **12** codes · **58** fragments · eight of eight participants · bears on SQ1 and SQ4 · The smallest cluster in fragments.

Sub-themes. Outsourcing to AI produces results, not skills: P6's observation about the macro, P5's about theory and practice. AI skill is itself tacit knowledge: the driving analogy, the AI frontrunner as the go-to person, the question how to transfer AI knowledge to the average consultant. Guided practice as the preferred route: learning with AI as a practice partner, a feedback loop, a change approach that gives time and removes fear.

What participants said. P6 states the mechanism: "if you let it write the code and check the output, you do not learn how to write that code. But if you have a method..." (P6-068). P3 describes AI skill as tacit knowledge learned by doing, which does not transfer by explanation and makes the frontrunner the person colleagues bring their questions to; P1 describes AI as a junior colleague that has to be instructed at the same level of detail as a person and expects investment in training, guardrails and explanation; P4 and P5 see AI as a practice partner; P7 describes the change approach the majority needs: give people time, do not impose. Fragments: P3-019, P3-020, P3-057, P3-022, P3-026, P1-095, P1-096, P1-098, P1-071, P4-098, P4-100, P5-088, P7-083, P7-084, P7-089 (Appendix D).

Counter-voice. None as a counter-voice; P4's deliberate choice to keep some work by hand (P4-068) is a boundary on the cluster: not everything should be handed over, and what a consultant keeps by hand is part of how they stay in control.

Researcher's inference. This cluster is the empirical basis for the pedagogical core of the design: the support makes the consultant think first (conclusion 4 of section 2.2.4), and it fades as the consultant's capability grows. It is also the thinnest cluster in evidence for its central claim: that delegating does not build skill rests on two participants (P6, P3) and on literature that is suggestive rather than settled, which is why success indicator 1 tests exactly this claim in chapter 6. It also reframes assumption A7: the vision exists, but it reaches consultants as tooling, not as a way of working with AI that builds capability.

For the design brief. The support helps with framing and method choice before it produces content, and support fades per level; the implementation gives time and removes the fear of failing. Confirms conclusion 4 of section 2.2.4.

Set aside: Supply Value's strategy and business model

Twelve of the sixteen codes that were not clustered describe Supply Value's strategy and business model, among them codes on the value of consulting work, on AI products, on two tracks, individual and firm-wide, and on AI as a goal instead of a means. Together they carry 65 fragments from six of the eight participants (not P5 and P6), more than cluster 5, and they are set aside deliberately: a cluster answers a sub-question about the capability of consultants, and these codes answer a question about Supply Value. They are not lost. They are the evidence behind section 2.2.2 and theme 11, and they return in the discussion of chapter 7. Appendix E lists them with their counts.

3.1.8 What the clusters say together

Read together, the five clusters tell one story. The value a client pays a consultant for, in the view of management and consultants, is the framing of the question and the reading of the organisation (cluster 1), and that is also what the consultants who use AI describe as what they do differently: they frame the question, hold the context and judge the result (clusters 1, 4 and 5). That capability is learned by guided practice on real problems, not by delegating to AI or by training about tools (cluster 5). Cluster 2 adds what separates the consultants who use AI from those who do not before that capability comes into play: a concrete use case, time to find one and a first experience that did not fail. What stands in the way is therefore not a lack of tools but three things: what is permitted differs per client and is not at hand at the moment of use (cluster 3), knowledge that sits with people and is lost between assignments (cluster 4), and the absence of a use case and of time for the majority (cluster 2). These are the conditions the design brief in section 3.3 has to meet.

3.2 How a consultant at Supply Value works: the assignment process and two journey maps

3.2.1 The assignment process

The interviews describe not only what consultants think about AI but how their work is structured. Figure 3.3 summarises the assignment process as it emerges from the eight interviews, with per phase where AI is and is not used today; the evidence per phase, with the fragments, is in Appendix F. The phases were derived from the coded fragments on role context and daily work; they are the researcher's inference from what the participants said about their work and have not been validated with consultants as a description of the process. The figure locates the opportunity, which lies in the phases where no AI is used today and where the problem of the thesis sits (context building, framing, hand-over), and it is the basis for the structure of the support in chapter 5, which models phases 2 to 4.

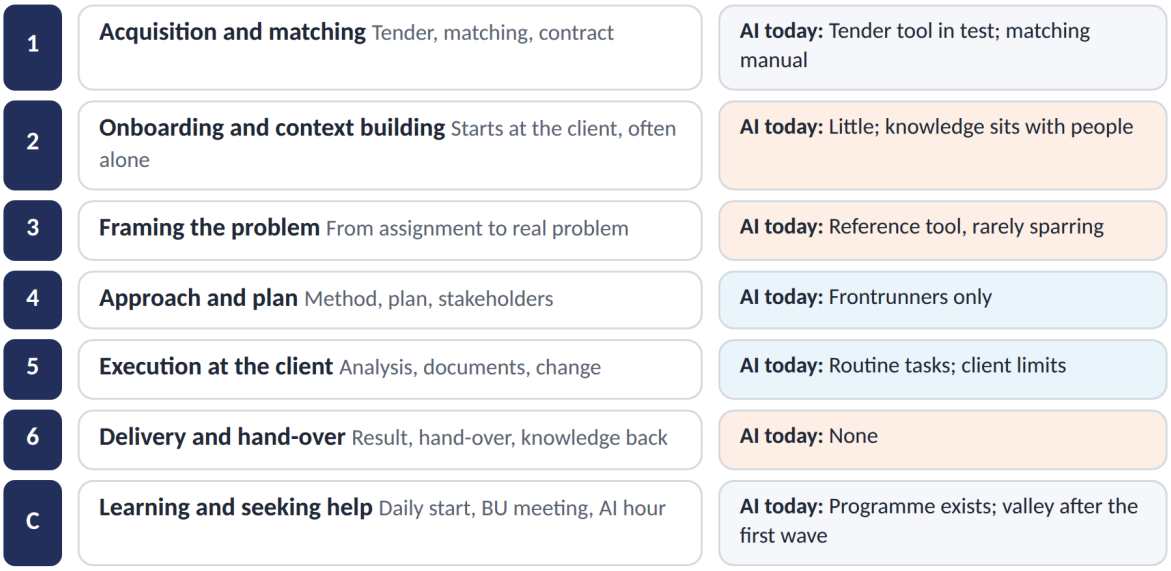


Figure 3.3 The assignment process with, per phase, where AI is used today. Colour indicates the extent of current AI use as described in the interviews.

3.2.2 Co-worker journey maps

Where the interviews in section 2.5 show what consultants say about AI, they do not show where AI sits in a working week, or what has to happen before it gets there. To make that visible, two co-worker journey maps were built, applying the customer journey mapping method (Van Boeijen et al., 2020; Stickdorn et al., 2018) to an employee rather than a customer (Gibbons, 2018; Kalbach, 2020, ch. 3). A journey map is built for one actor in one scenario (Kaplan, 2016a), so two maps were built rather than one composite, chosen to span the range that theme 2 identified as the core of the adoption problem: map A is the week of a consultant who uses AI daily in text work (P4), map B the week of a consultant whose one experiment with AI failed and who has not used it since (P7). Both are consultants at a similar level on long secondments in regulated public or semi-public settings, so the difference is neither seniority nor the kind of client; it is the use of AI itself, so map A shows what a consultant who uses AI does with it, not what any consultant reaches by starting to use it.

In one assignment, a consultant tracked the same action across several channels. In another, a consultant tried to compare records across systems and had to check the results manually. Each map draws on one semi-structured interview, coded as described in section 2.5.1, and was not validated with the participant; the phases are taken from the participants' own descriptions of their week, with one working week as the common time scale. Per phase the map shows what the participant does, with the touchpoints and channels, what they expected, one verbatim quotation and how the phase felt, scored on a five-point scale with the status of each score printed on the map; below the grid run the emotional curve, the use of AI with its friction, the moments of truth and the opportunities. Every claim carries an evidence status, researched, inferred or assumption, recorded per cell in Appendix F, which also describes the method in full. The maps are shown as figures 3.4 and 3.5.

Co-worker journey map A

One interview · evidence per cell in Appendix F

A working week of a consultant who uses AI

P4, consultant
Consultant at Supply Value, seconded to a public client.

The week that is mapped
From the overview built on Monday to what is still on the "other" list on Friday. Six phases in their own words, with a two-week sprint running underneath.

What this map asks
Where does support deserve priority in the week of a consultant who already uses AI, and where should we stay away?
Researched · one participant · not yet validated

	01 Building the week overview *Monday morning, every week*	02 Mailboxes and actions *daily, several times*	03 Meetings and calls *daily; supplier once a week*	04 Finding out and analysing *ad hoc, when something breaks*	05 Producing output *through the week, two-week sprint*	06 Ticking off and what stays behind *end of day and week*
Does	• Lists tasks and meetings per day • Sets the sprint board beside it • Notebook • Sprint board • Calendar	• Works several mailboxes • Copies actions into the notebook • Tried an AI prompt to pull in meetings • Mailboxes • AI assistant • Notebook	• Calls the supplier • Trains partner organisations • Checks in with colleagues at the client • Supplier • Partner organisations • Client staff • AI assistant	• Summarises long documents • Hunts for reliable sources • Probes instead of taking the first answer • AI assistant • Search • Supplier	• Presentations, training material, documents • Builds reusable components • Uses a code-based AI tool for slides • Code AI tool • Presentation software	• Ticks off the list • Moves the rest to "other", for when there is time • Own list • Sprint board
Expected beforehand	*not asked*	that the prompt would pull the meetings into the overview	*not asked*	*not asked*	*not asked*	*not asked*
Thinks	*"I make it every Monday so that I know whether I still have time, and so that it is immediately clear."*	*"So then I did it by hand again. That was a bit disappointing."*	*"So much of this is people work, so calling, you just do that one to one."*	*"You do of course still have to look very critically at the output."*	*"The perfect scenario is of course that you no longer have to do administrative work yourself, but that you can just think creatively yourself and really have a good sparring partner."*	*"But this one I am not going to manage. And that does not really matter."*
Feels	**satisfied or confident** +1 ◌ estimate	**irritated or uncertain** -1 ● expressed	**neutral** 0 ◌ estimate	**satisfied or confident** +1 ◌ estimate	**satisfied or confident** +1 ◌ estimate	**neutral** 0 ○ from the participant's words
AI and friction	No AI Rebuilt by hand every Monday	Tried, failed One action, recorded in several places	Drafting mail, preparing questions Scepticism and a knowledge threshold at the client	Summarising and sourcing Weak sources, then a phone call anyway	In routine use here Many separate subscriptions	No AI Work without a deadline rolls on



Expressed by the participant: phase 02 and the peak from the earlier assignment. From the participant's words: phase 06. Estimate: the other four. The curve is the researcher's reading and was not confirmed by the participant.

The three moments that decide this week

Moment of truth 1 phase 02
The only place where AI touches coordination work, and it has been failing for weeks.

Moment of truth 2 phase 04
AI as a real sparring partner, on a dossier carried alone. The reference point for what AI is worth, but from a previous assignment.

Moment of truth 3 phase 03
Where the AI work either lands with the client or does not.

Where to act, and where to leave things alone

Opportunity phase 02
One place for actions from mail, meetings and board, without retyping
Internal AI work, with IT
Places where one action is recorded

Opportunity phase 03
Raising AI with a low-literacy client without lecturing
Internal AI work and the consultant
Client staff asking their own first question

Protect phase 04
Keep sparring available when someone carries a dossier alone
The consultant with their manager
Solo dossiers with a sparring partner

Opportunity phase 05
Bring the scattered tools into one permitted environment
Information security
Number of separate subscriptions

Open question phase 06
Does anything survive the week? Not researched, so validate before designing
To be decided
Still to be defined

Evidence status of a score: ● expressed by the participant ○ read from the participant's words ◌ researcher's estimate ◆ moment of truth **Touchpoints:** ● own tools ● client ● partners and suppliers ● external ● personal

Quotations are translated from Dutch; omissions are marked with [...]. Evidence status per cell is recorded in Appendix F. This figure is an A3 landscape page; print it at A3 or folded.

Supply Value · We make strategy work

Figure 3.4 Co-worker journey map A: a working week of a consultant who uses AI (P4). Current state; individual journey, n = 1; not validated with the participant. Evidence per cell in Appendix F. A3 landscape page: print at A3 or folded.

Co-worker journey map B

One interview · evidence per cell in Appendix F

A working week of a consultant who barely uses AI

P7, consultant
Consultant at Supply Value on a long secondment at a large client. Part of the week is reserved for the firm. Places themselves "a bit more towards the back" of the adoption curve.

The week that is mapped
From the morning stand-up to the call that arrives at the end of the day and is pushed to tomorrow. Five phases run at the client; the hours for the firm run alongside them, not after them.

What this map asks
What would have to be true before AI adds anything to this week, and where does the value already sit without it?
Researched · one participant · not yet validated

	01 Morning stand-up with the department *daily, every day*	02 Meetings about the automation *through the week, runs on for weeks*	03 Own analysis of supplier reliability *weekly, when the planners' lists arrive*	04 Mail and running issues *daily, between the meetings*	05 End of the day *daily, pushes on to tomorrow*	06 The hours for the firm *weekly, alongside the other phases*
Does	● Joins the department stand-up ● Picks up the day's logistics issues ● Departmental stand-up ● Issue lists	● Collects requirements from planners ● Translates them for the automation team ● Follows the supplier data links ● Planners ● Automation team ● Client systems and spreadsheets ● Suppliers	● Analyses the weekly reason-code lists ● Builds their own alternative calculation ● Spreadsheet ● Planners' lists ● Management team	● Traces records from months ago ● Retypes supplier documents by hand ● Ran one AI experiment on them ● Mail ● Client system ● Supplier documents ● AI assistant	● Takes a late call and moves it to tomorrow ● Leaves ideas from conversations unrecorded ● Phone ● Meeting software ● Colleagues	● Joins the firm's stand-up and BU meeting ● Drops the AI session when it lands on top ● Firm stand-up ● BU meeting ● AI session
Expected beforehand	*not asked*	*not asked*	*not asked*	that an AI assistant could compare records across two systems	*not asked*	*not asked*
Thinks	*"The average day will just start with a stand-up with my supply chain department."*	*"I actually make the translation between the operation, what they need and the requirements […]"*	*"I tinker with it a bit myself, to draw better conclusions than what is being done with it now, in my opinion."*	*"It started bluffing, it started making mistakes in the quantities, it did not recognise some of the lines. […] And then I thought: do not rely on it."*	*"I should actually have written that down just now, but I have not. So I have to go and ask about it again."*	*"[…] I find it interesting, but as long as I am held accountable for other things that I am then not doing, I am not going to take part."*
Feels	*not asked*	**satisfied or confident** +1 ○ estimate	**satisfied or confident** +1 ○ from the participant's words	**irritated or uncertain** -1 ○ from the participant's words	**irritated or uncertain** -1 ○ from the participant's words	**irritated or uncertain** -1 ○ from the participant's words



No score was expressed; phases 03 to 06 follow from the participant's own words, phase 02 is an estimate, phase 01 was not asked. The curve is the researcher's reading and was not confirmed by the participant.

	01	02	03	04	05	06
AI and friction	No AI *Not discussed in the interview*	No AI, it runs through developers Waits on a handful of developers	No AI, spreadsheets at a basic level The system does not give the picture; lists are kept beside it	One experiment, failed, not repeated A different format per supplier; all retyped by hand	None, the automatic notes go unused Ideas evaporate; reminders are kept by hand	Barely, waiting for a use case Learning competes with billable work

The three moments that decide this week

Moment of truth 1 phase 04
Their only first-hand experience of AI at work: a careful prompt failed, and that settled it.

Moment of truth 2 phase 03
Their analysis accepted by the client's management team: the high point, and no AI touches it.

Moment of truth 3 phase 06
Where the AI programme loses them, not through unwillingness, but because the bill is settled elsewhere.

Where to act, and where to leave things alone

Opportunity phase 04
A second attempt at the comparison, with someone alongside them
Internal AI work and the consultant
Items compared correctly in a test

Opportunity phase 06
Time for learning AI that is genuinely free, and safe to fail in
Firm leadership with the consultant's manager
Participation by fully booked consultants

Opportunity phase 02
Prototype before the developer queue, where the client environment allows it
The consultant with the automation team
Requirements shown as a prototype first

Protect phase 03
Repeat what made the management-team moment work
The consultant with their manager
Own analyses that lead to a decision

Open question phase 05
Capture ideas without extra work. The idea came from the researcher, not from them
To be decided
Still to be defined

Evidence status of a score: ● expressed by the participant ○ read from the participant's words ◌ researcher's estimate ◆ moment of truth **Touchpoints:** ● own tools ● client ● partners and suppliers ● external ● personal

Quotations are translated from Dutch; omissions are marked with […]. Evidence status per cell is recorded in Appendix F. This figure is an A3 landscape page; print it at A3 or folded.

Supply Value · We make strategy work

Figure 3.5 Co-worker journey map B: a working week of a consultant who barely uses AI (P7). Current state; individual journey, n = 1; not validated with the participant. Evidence per cell in Appendix F. A3 landscape page: print at A3 or folded.

Map A: a week in which AI is already in use

The map of P4 (figure 3.4) runs from the overview built on Monday morning to the tasks that are still on the "other" list at the end of the week. AI appears in three of the six phases, where the task is text-based and individual, drafting emails, summarising documents, preparing questions, producing slides, and it is absent where the work is coordination: the two phases that hold the week together, building the overview and working through the mail, contain no AI at all. That absence is not for want of trying, and it produces the first moment of truth. A prompt that was meant to pull meetings from the calendar into the weekly overview failed for weeks, and nobody owned its repair:

> *"The last few weeks it did not work at all, because it says it cannot see any meeting in my calendar. So then I did it by hand again. That was a bit disappointing."* (P4-066, P4-067)

This is the only point in the map where AI is pointed at coordination work. The highest point of the map sits outside the mapped week: on a previous assignment the participant was the only person working on a complex dossier and used AI to test whether their reasoning held (P4-037 to P4-041), and the map shows this as a separate, marked cell. In the current assignment that kind of sparring has no place yet. The friction in this week is mostly about switching between tools that hold the same action in more than one place, with cost as a second, smaller threshold (P4-031 to P4-034).

Map B: a week in which AI is absent

The map of P7 (figure 3.5) covers the same span, from the morning stand-up to the call that arrives at the end of the day and is pushed to tomorrow; five phases run at the client, and the hours reserved for Supply Value run alongside them, which the map shows as a parallel band. AI appears once in the whole week, and it appears as a failure. The participant tried to have an AI assistant compare records across two systems, a task that is otherwise done by retyping figures by hand:

> *"It started bluffing, it started making mistakes in the quantities, it did not recognise some of the lines. That was about half a year ago; it was not a success. And then I thought: do not rely on it."* (P7-045)

The attempt was not careless: the prompt contained definitions and an example taken from the work instructions, and the participant explains why the task is hard, because every supplier uses a different format (P7-046). Six to seven months later it had not been tried again (P7-048). This single experience carries more weight in their attitude than any policy or training session, which is why it is marked as a moment of truth rather than as an ordinary pain point. The highest point of this map also lies outside AI: the participant built an alternative calculation for the planners' weekly reason codes and presented it to the client's management team, where it was accepted (P7-009 to P7-011). The third moment of truth is internal. Of the hours reserved for Supply Value each week, the AI session is the item that gets dropped:

> *"I find it interesting, but as long as I am held accountable for other things that I am then not doing, I am not going to take part."* (P7-072)

The participant is explicit that this is not disinterest but a cost-benefit judgement (P7-073), and adds a condition that Supply Value can act on: an hour being freed up does not mean it is free for them (P7-089). Read together with what they say holds them back (P7-084, P7-085), this locates the barrier in how learning time is accounted for, not in the individual.

What the two weeks show together

Placed side by side, the two maps make three things visible that neither shows alone (table 3.2).

	Map A (uses AI)	Map B (barely uses AI)
Where AI sits in the week	Text work: drafting, summarising, preparing questions, slides	Nowhere, apart from one failed experiment
Where AI is absent	The two phases that hold the week together: building the overview and processing mail	All phases; the work is meetings, analysis and coordination
The high point	Sparring on a dossier carried alone, from a previous assignment	An own analysis accepted by the client's management team
The threshold	Fragmentation and time: the same action recorded in several places	A cost-benefit judgement: learning competes with billable work

Table 3.2 The two mapped weeks compared.

First, AI has not entered coordination work in either week: in both it was tried on exactly the kind of task the literature promises AI will absorb, collating, comparing, keeping track, and in both the participant returned to doing it by hand. The difference between the two consultants is that one has found value in text work and the other has not yet found an entry point at all. Second, the high point in both maps is human, thinking a problem through where no colleague has the knowledge and an analysis that changes a decision, and both are the part of the work the participants describe with the most energy; the target of the design is not to automate these moments but to make them available more often. Third, the barriers are of a different kind: map A shows a fragmentation problem that better-integrated support could address, map B an accounting problem that no support can address, because as long as learning competes with work that is measured, the hour will keep losing. The design brief separates what the intervention can solve from what Supply Value has to solve around it.

Moments of truth

Carlzon (1987) calls every contact that shapes how the whole experience is judged a moment of truth; following later journey-mapping practice, each map marks at most three that are taken to weigh most, and the limit of three is a choice of this study. Six moments result (table 3.3), and they are the points at which the design brief is aimed.

	Moment	Why it is decisive	Evidence
MT-A1	The AI link that keeps failing (phase 02)	The only point where AI touches coordination work; weeks of failure set the picture of what AI can do in this week	Stated (P4-066, P4-067)
MT-A2	AI as a real sparring partner (phase 04)	The reference point for what AI is worth, but it comes from a previous assignment	Stated (P4-037 to P4-041)
MT-A3	Raising AI with the client (phase 03)	Where the AI work either lands with the client or does not; the participant describes an AI literacy cliff at public clients	Stated (P4-089)
MT-B1	The failed experiment (phase 04)	The only first-hand experience of AI at work; a careful prompt failed and that settled it	Stated (P7-045 to P7-048)
MT-B2	Own analysis accepted by the client's management team (phase 03)	The highest point of the map, and no AI is involved	Stated (P7-009 to P7-011)
MT-B3	Dropping the AI session (phase 06)	Where Supply Value's AI programme loses the participant, on a cost-benefit judgement	Stated (P7-072, P7-089)

Table 3.3 The six moments of truth across both maps, with the evidence status of each.

Opportunities

In a Define-phase artefact an opportunity is a direction or a question, not a solution. Each opportunity in table 3.4 is attached to a low point or a moment of truth and carries an owner and an indicator, because maps without a

clear purpose, and maps that are not acted on, are among the most common reasons journey maps fail in practice (Kaplan, 2016b). Provenance per opportunity is recorded in Appendix F.

Opportunity (direction)	Owner	Indicator
A/02: bring actions from mail, meetings and the board together in one place, without retyping	Internal AI work, with IT and information security	Places where one action is recorded
A/03 and B/06: make Supply Value's AI support reach consultants who are fully booked at a client, and make it safe to fail	Management	Participation by fully booked consultants; failed attempts shared rather than hidden
A/04 and B/03: protect what already works without AI: sparring on a solo dossier, an own analysis that leads to a decision	The consultants themselves, with their manager	Solo dossiers with a sparring partner; own analyses that lead to a decision
B/04: retry the failed comparison task with support, on a current model	Internal AI work with the participant and the owner of the AI assistant	Items compared correctly in a test; time the manual work now costs
B/02: let consultants prototype before the developer queue, where the client environment allows it	The participant with the client's automation team	Requirements shown as a working prototype first

Table 3.4 Opportunities, each attached to a phase in map A or map B.

Limitations of the maps

Each map rests on a single interview, which makes it an individual actual journey rather than a typical one (Halvorsrud et al., 2016); neither is supported by observation or validated with the participant it describes, and the emotional layer is the weakest: it is the researcher's estimate, with most scores read from the participant's words or estimated by the researcher, as the figures show, and it was not confirmed by the participants. The maps show how two consultants describe their week and where AI does and does not appear in it; they do not establish how common either pattern is, and section 7.3 states what follows for the claims of this thesis.

Input for the design brief

Three conclusions carry into section 3.3, and each feeds a requirement there. First, AI currently earns its place in individual text work and not in coordination, so an intervention that promises to organise a consultant's week starts from the weakest evidence in this research; this bounds the direction and feeds DR10, the entry through a concrete task on the consultant's own assignment, and it is why the direction of organising the week is set aside in section 4.3. Second, the moments that both participants value most are moments of thinking, which suggests supporting the reasoning rather than replacing the output; this feeds DR8 and DR4. Third, the barrier in map B is structural: unless learning time is accounted for differently, support aimed at consultants who are fully booked at a client will meet the same fate as the AI session that gets dropped; this feeds DR12. Table 3.5 records section 3.2 as an origin of these four requirements.

3.3 Design brief

3.3.1 Design goal

To design a way for Supply Value to strengthen the innovation capability of its consultants, so that they can direct AI agents at client problems and create more value for clients in advisory and implementation work within regulated client environments, while developing rather than replacing their own professional skills. The research of chapters 2 and 3 points to three connected parts: a description of the capability itself, what it consists of and how it develops; support that helps consultants build and apply it in their daily work; and an implementation at Supply Value that gives the first two a place in how Supply Value already works. The direction each part takes is chosen in chapter 4 and worked out in chapter 5.

3.3.2 Target group

The primary users are everyone at Supply Value who works on client assignments, across seniority levels and business units; at Supply Value that includes managers who work on assignments themselves next to their management role (section 2.5.1). The design needs to work for consultants who are already actively using AI and want to go further, and for consultants who have not yet started and need a clear and personally relevant reason to begin; the second group is described as the larger one (P8-043), and its entry point has to be a concrete task on its own assignment. The management role is a second use of the same design rather than a second group of users: a manager uses the description of the capability to see where a consultant stands and to support development in their conversations with consultants, without needing to be an AI expert, and a manager who has a level of their own can speak in those conversations from their own practice. Clients are not users, but the design has to respect their rules and, where the consultant helps the client use AI responsibly, support that.

3.3.3 What is to be designed

The deliverable consists of three connected parts. The first is a description of the capability for working with AI agents: what it consists of, how a consultant progresses from prompt-based use with guidance to directing and building agents, which learning mechanisms carry that progression and which guardrails bound it, and what managers can use it for. The second is support that consultants use on their assignments: it holds the context of an assignment, helps the consultant frame the problem and choose a structured method, produces content only after the consultant has done the thinking, and applies the rules of the client context. The literature points to an AI agent as the form of this support, as a direction to be tested in chapter 4 (conclusion 2 of section 2.2.4, a positioning of this thesis; cluster 4 asks for context held across steps); the interviews contain no consultant who directs agents, and chapter 4 chooses the form within that positioning, informed by the co-creation session. The third is an implementation at Supply Value that ties the first two to Supply Value's existing AI programme, its daily start and business unit meetings, and the time consultants have for learning.

3.3.4 Design requirements

The requirements below merge the conclusions of the literature (sections 2.2.4 and 2.4) with the requirements derived from the insight clusters (section 3.1). Each requirement is one sentence and states what it applies to, the capability, the support or the implementation, and where it comes from; table 3.5 lists them. The contrasts that explain a requirement, for example that the support is not a chatbot that answers single questions (DR7) and that the capability is not the adoption of tools (DR1), are given in the sections the origin column points to and are not repeated in the requirement. Where a requirement has parts that are judged separately, the parts carry a letter. The traceability from requirement to the component of the design that realises it is completed in chapter 5.

The list distinguishes hard requirements from criteria, the demands and wishes of the Delft Design Guide and of Pahl and Beitz (Van Boeijen et al., 2020; Pahl and Beitz, 1996). A hard requirement is a go or no-go that is not scored; a criterion is what concepts are compared on, and a concept may score low on one and still be chosen. Only the rule in DR4 and the agreements and data in DR13 are hard; the last column of table 3.5 records this and section 4.3.2 applies it.

	Requirement	Origin	Applies to	Hard or criterion
DR1	The capability is described in terms of how consultants work on client problems; it describes the work process before it introduces agents.	Conclusion 1; cluster 2; A6	Capability	Criterion

	Requirement	Origin	Applies to	Hard or criterion
DR2	The capability distinguishes levels, from prompt-based use with guidance to directing and building agents, each with framing the problem as its entry condition.	Conclusion 2; clusters 4 and 5; A4	Capability	Criterion
DR3	The capability is defined and assessed in terms of value delivered at the client: a problem framed, a change implemented, a client able to continue.	Conclusion 5; cluster 1	Capability	Criterion
DR4a	Every AI output is validated by a consultant before it moves on.	Conclusion 3; clusters 1 and 3; theme 9	Capability, support	Hard
DR4b	Validating AI output is treated as a skill to be learned: recognising when the model is outside its frontier.	Conclusion 3; clusters 1 and 3; section 3.2	Capability, support	Criterion
DR5	Support fades as the consultant's capability grows: what the support does at first, the consultant later does with less support.	Conclusion 4; cluster 5	Capability, support	Criterion
DR6	AI literacy in the sense of Article 4 of the EU AI Act is the base layer of the capability.	Section 2.4	Capability	Criterion
DR7	The support holds the context of an assignment across steps, methods and time.	Conclusion 2; cluster 4	Support	Criterion
DR8	The support helps the consultant frame the problem and choose a method before it produces content, and writes nothing into the consultant's work without confirmation.	Conclusion 4; cluster 5; theme 6; section 3.2	Support	Criterion
DR9	Rules on what may be submitted to which system are agreed per client and project and checked by the support at the moment of use, and a submission that breaks them is stopped; client and sector context is a separate, configurable layer, not built into the methods.	Section 2.4; cluster 3 (the client environment); theme 8	Support	Criterion
DR10	The entry point is a concrete task on the consultant's own assignment that pays back within the first session, without configuration or prior AI knowledge.	Cluster 2; A1, A2; section 3.2	Support, implementation	Criterion
DR11	Knowledge captured during an assignment flows back in reusable form to the consultant and to Supply Value.	Cluster 4	Support, implementation	Criterion
DR12	The implementation redesigns at least one recurring workflow per business unit (a quantity chosen by the researcher so that every unit has one worked example), connects to the existing AI programme, the daily start, the business unit meetings and the AI hour, and gives consultants time for learning that does not compete with the work they are judged on.	Conclusion 1; clusters 2 and 5; A7; section 3.2	Implementation	Criterion
DR13a	The support works within Supply Value's existing agreements.	Constraints (3.3.6)	Support	Hard
DR13b	The support submits no sensitive client information to external systems without approval.	Constraints (3.3.6)	Support	Hard
DR13c	The support is deployable without significant IT investment.	Constraints (3.3.6)	Support	Criterion
DR13d	The support is usable at client sites where tools are restricted.	Constraints (3.3.6)	Support	Criterion

Table 3.5 Design requirements with their origin, what they apply to and whether they are hard or a criterion. "Conclusion" refers to section 2.2.4; "cluster" to section 3.1; "theme" to section 2.5 and Appendix J; "section 3.2" to the journey maps; A1 to A8 to the assumptions of chapter 1. Letters mark the parts of a requirement that are judged separately.

3.3.5 Success indicators

Whether the design succeeds is assessed on three success indicators within the time of this thesis and one aim beyond it. Indicator 1 is behavioural and concerns the consultant: when consultants use the support on a realistic case, do they frame the problem and choose a method with less help at the end of the session than they needed at the start of it? This is observed in the evaluation with consultants in chapter 6. A drop within one session is at most a first signal, because it can come from learning the tool, the case or the sequence of the method; the indicator proper compares the help a consultant needs on a new case after use, across at least two cases or sessions. It is the most important indicator, because it is the direct test of the claim that the support builds capability rather than replacing it. Indicator 2 concerns the quality of the work: are the outputs that consultants produce with the support judged usable by a senior consultant who does not know who made them, and do they remain the consultant's own work? If the support made consultants more independent but their work worse, or produced good work that the consultant cannot explain, the design would have failed. Indicator 3 concerns acceptance: do consultants and managers judge the description of the capability as understandable, usable and applicable to their own situation, and do they see themselves and their colleagues in its levels? The consultants' acceptance of the support is asked after their test session; the managers' judgement of the description, including placing consultants of their own team in its levels, is left to the decision after the pilot (section 6.4). The long-term aim lies beyond the thesis: do consultants progress through the levels over time, and does the capacity that AI frees up show up as value at the client? This cannot be measured within the graduation period and is stated as a limitation and as a recommendation for a pilot in chapter 7.

3.3.6 Constraints

The support must work within Supply Value's existing commercial AI agreements and cannot require the submission of sensitive client information to external systems without explicit approval under Supply Value's governance; eventually its model calls have to run through the environment Supply Value has approved for client-related work (section 2.4). It must be deployable without significant IT infrastructure investment, because developing an enterprise-wide platform was outside the scope of this project (section 2.1.2). It must be usable by consultants working at client sites where access to certain tools may be restricted. It must actively support the development of junior consultants rather than remove the experiences through which they build their professional foundation (Trend 3, cluster 5). And it must be honest about what is and is not autonomous in it, so that the design is not undermined by the "agent washing" that Gartner (2025b) warns against. The fifth constraint is carried by DR5 and DR8; the sixth is a reporting rule for chapter 5, which separates what is built from what is designed.

3.3.7 What is fixed and what remains open

At the end of the Define phase three things are fixed. The problem is a capability problem before it is a technology problem, and the capability is the one defined in section 2.3, and the interviews put its weight on framing the question and judging the result, which management and consultants say clients keep paying for; holding the context across steps is what the support must do so that the consultant can exercise them. Among the consultants who use AI this is what the interviews indicate separates them, while a use case and time decide whether a consultant uses AI at all (section 3.1.8). The design has to work for the majority that waits for a use case, which means it starts from the consultant's own assignment, makes the consultant think first and respects the rules of the client at the moment of use. And the design has three parts, a description of the capability, support for building and applying it, and an implementation at Supply Value, with thirteen requirements (DR1 to DR13) that trace back to the literature and the interviews; DR14 follows from the co-creation session. Two things remain open for the next chapter, where consultants inform the choice and the researcher makes it: where exactly in the assignment process an AI agent adds value for the client (sub-question SQ2), and whether a

structured design method is the right backbone for the support, given that consultants at Supply Value solve problems by experience and that clients do not want to hear method terms.

4

CHAPTER 4 · DEVELOP PHASE

Develop: directions

A co-creation session with consultants, twelve how-might-we questions and three concept directions, one of which is chosen.

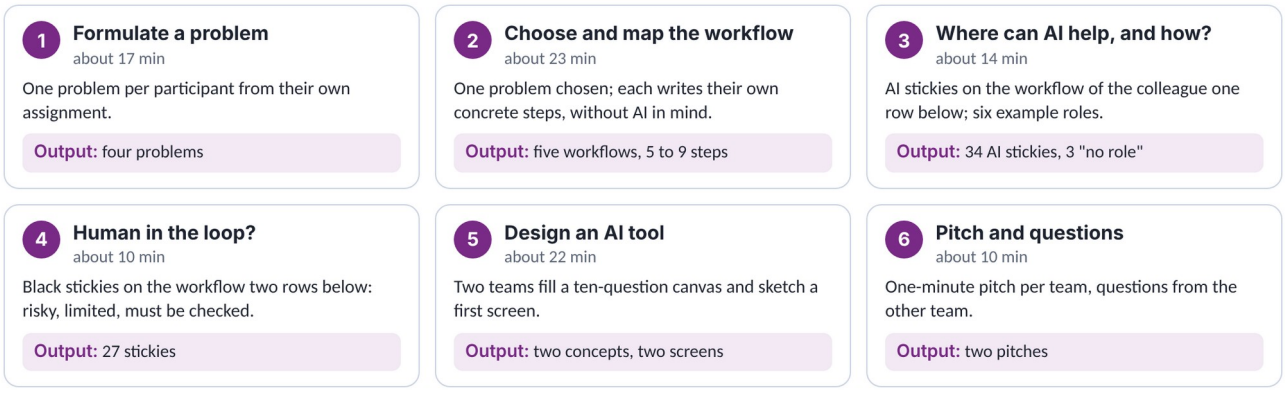


IN THIS CHAPTER 4.1 Co-creation session with consultants · 4.2 How-might-we questions and concept directions · 4.3 Choice of direction

4.1 Co-creation session with consultants

4.1.1 Purpose and set-up

The Discover phase answered SQ2 only in part: the interviews showed how consultants approach client problems and where AI is absent today, but not where in that process an AI agent could add value, and the design brief left open whether a structured method is the right backbone for the support (section 3.3.7). The co-creation session, a generative session in the sense of Sanders and Stappers (2008) in which consultants contributed problems, workflows and tool concepts within a frame set by the researcher, with the ideation and the concept choice done by the researcher afterwards, was set up to watch consultants work on a problem rather than ask them about it: to see how they break a client problem down into steps, where they place AI and where they keep the human in the loop, and to let them design, in teams, the AI support they would want. The session took place at the Supply Value office, lasted about an hour and a half, partly online, and followed the six steps of figure 4.1; the materials, the board, the transcribed outputs and the procedure are in Appendix G; consent for the session was oral (Appendix G.2), and the recordings and transcripts are handled as described in section 2.5.1.



1 h 35 min, Miro board; three participants in the office with the researcher, two online. Durations from the recordings.

Figure 4.1 The six steps of the co-creation session, with what each step asked and produced.

Step 1 collected a real problem from each participant's own assignment; step 2 chose one of them and asked every participant to write, alone and without thinking about AI, the concrete steps they would take to solve it. Steps 3 and 4 were done on a colleague's workflow: each participant added where AI could help and in which role to the workflow of the person one row below, and where AI is risky, cannot do the step or must be checked to the workflow two rows below, so that every workflow was seen by three people. Step 5 asked two teams to design an AI tool along a canvas of ten questions and to sketch its first screen; step 6 was a one-minute pitch per team with questions from the other team.

4.1.2 Participants

Five consultants worked in two teams to develop early concepts (table 4.1). Four had also taken part in the interviews, so the two sources partly reflect the experiences of the same people; which participant is which is not stated, so that the C-codes cannot be linked to the P-codes. The sample therefore also shares the limits of the interview sample: one business unit is over-represented, and there is no consultant at medior level. In step 5 team A had two participants and team B three.

Participant	Team in step 5
C1	B
C2	B
C3	B
C4	A
C5	A

Table 4.1 The participants of the co-creation session and their team in step 5. Function, seniority and location are not given per person.

4.1.3 Step 1: the problems consultants brought

Three of the four problems written on the board describe the same situation from different positions, and it is the situation of cluster 4: changing environment, background and stakeholders at every new assignment, finding out how the client's processes run while the people involved keep their information in different places, and a predecessor's OneNotes, logbooks and action lists from which a monthly programme report has to be produced. C4 brought a problem of a different kind: the client works with a third-party logistics partner (3PL, an external party that stores and moves goods for the client) whose stock system is not linked to the client's own; movements are shared as an export and entered by hand, bookings run late or go wrong, and stock in the system differs from reality. The group chose C4's problem for the rest of the session, on the researcher's proposal, because two of the other problems overlapped and because it was concrete and could be broken into many steps; the choice was therefore steered, and the problem was a live case for which its owner already had a solution direction in mind (section 7.3). C5 joined after this step.

4.1.4 Step 2: how five consultants would approach the same problem

The five workflows are shown in figures G.12 and G.13 in Appendix G, aligned on eight phases that the researcher derived afterwards from the steps the participants wrote, together with the stickies of steps 3 and 4. Three things stand out. First, all five start by talking to the people involved, and none starts with an analysis or a document; for four of them the second step is to define the problem and its context with those people. This is the assignment process of section 3.2.1 seen from the inside: framing is done by asking people, not by applying a method, which is what theme 4 of the interviews described. Second, every workflow that goes beyond the problem ends with a decision maker, not with the consultant deciding: the outcomes of a brainstorm go to "those with decision authority", a value case goes to the management team for a decision, the options are discussed with "whoever has to choose"; only C1 stays at the level of data, systems, requirements and cost. Third, the workflows differ in where they stop: C4 and C5 are the only ones who reach a pilot and an implementation, and both name the concrete organisational steps that belong to it. In the discussion that clarified the case the group reframed the problem within a quarter of an hour, by asking questions, from "two systems that cannot talk" into "an internal manual process with a check that only partly needs a person" (transcript part 1, 32:06 to 38:04). This is collective problem framing by experienced questioning, and it is exactly what the tools of step 5 were then designed to do for one consultant alone.

How to read the two figures: one consultant per row, the phases A to H as columns; the green tags are the AI roles proposed in step 3 and the black tags the human-in-the-loop remarks of step 4; the texts of the stickies are in Appendix G.3.1.

4.1.5 Steps 3 and 4: where AI helps and where the human stays

The green tags in figures G.12 and G.13 are the 34 AI stickies of step 3, the black tags the 27 human-in-the-loop stickies of step 4, each written by a colleague of the workflow's owner, so the figure shows how the group as a whole reads each workflow. Sparring partner is the most chosen role, on every workflow and at thirteen steps, followed by knowledge bank, assistant, draft maker, analyst and risk checker; coach appears only in combination with sparring partner. The AI stickies sit almost entirely in the first six phases, from talking to people to drafting the case, and hardly at all in planning, piloting and implementing, and the steps explicitly marked "no role for AI" are establishing the available budget and hours and discussing the options with the person who has to choose. None of the black stickies says that AI should not be used at a step. They say that the output must be checked, because AI can hallucinate "logical" outcomes, white spots and formulas, that privacy and data security apply throughout, that AI "can miss organisational sensitivity" and context, that stakeholders must recognise their own words, and that judgement stays with the consultant, "keep human in the loop" at every step where something is presented or decided. The two teams took exactly these elements into their tools.

4.1.6 Step 5 and 6: the two tools the teams designed

Figure 4.2 shows the two tool concepts side by side, transcribed from the canvases and the pitches. The two teams designed their tools separately, after working on the same workflow together, and arrived at the same kind of tool. The frame they worked in was the researcher's: the role list of step 3 offered the sparring partner first, the canvas of step 5 required a human-check moment and asked what AI may not do, with examples, and an example name was given, so the convergence is agreement with the offered frame rather than independent invention.

Tool A · a sparring agent for the whole implementation process

Team A: C4 and C5 (online), from C4's workflow

Moment	No single step: sparring across the whole process, from the interview questions to the implementation plan
For whom	Everyone, "but the less you know, the more you get out of it"
Input	Situation sketch; interview transcripts or recordings; the consultant's reviews of its output
What AI does	Asks critical questions to fill the gaps rather than assuming it has everything; structures what it has been given; only then makes a draft
Output	Drafts, "never the final version": interview questions, questions for the verification round, a value case, an implementation plan
Consultant checks	"Actually everything": sensitivities, feasibility, client context, stakeholders, quality, the final decision
Limits	No final advice by AI alone; no confidential data without permission and the right systems; nothing sent to the client; responsibility stays with the consultant
Why use it	Time; better quality was proposed and withdrawn inside the team
Beyond ChatGPT?	*"The confidential data it can use, and not much more"*

Tool B · a sparring partner that sharpens the problem into figures

Team B: C1, C2 and C3 (office), from C2's step

Moment	One step: sharpen the impact of the problem in specific figures, so that a management team can decide
For whom	Everyone with basic knowledge of the domain
Input	Data; the problem context per role; the organisational context; the stakeholders; the client's solution priorities
What AI does	"Draws it out of you": asks the right questions, makes logical connections, works the answers out in a model, inspires creative thinking
Output	Stakeholder overview; models; the right questions; the reasoning behind its output; a presentation for the management team
Consultant checks	Whether it makes sense and the figures are logical; whether the stakeholders can decide on the proposed solutions
Limits	Responsibility for the process and the decisions stays with the consultant; no end product goes to the client
Why use it	Overview of large data; time; sharper framing; possibly more creative insights
Beyond ChatGPT?	*"The back end of the model is set up" on best practices for sparring and asking good questions, with guardrails; the user's own words*

WHAT BOTH TEAMS DESIGNED, WITHOUT HEARING EACH OTHER

Support that asks before it produces, delivers drafts and never end products, and leaves every check and decision with the consultant. What separates them is what sits behind the conversation: confidentiality alone (A) or methods, best practices and guardrails (B).

Figure 4.2 The two tool concepts, transcribed from the canvases of step 5 and completed from the pitches where the canvas was left brief.

Team A's design: a sparring agent for the whole implementation process

Team A (C4 and C5), who described themselves as inexperienced with AI, designed one sparring partner for the whole implementation process rather than for one step, because "sparring partner" had been written at several steps of C4's workflow. The tool takes a situation sketch, the transcripts or recordings of interviews and the consultant's reviews of its output; it asks critical questions to fill the gaps in the information rather than assuming it has everything, structures what it has been given and only then makes a draft, "never the final version": interview questions, a value case, an implementation plan. The consultant checks "actually everything", and the limits are hard: the final advice is never determined by AI alone, confidential data is never processed without permission and the right systems, it sends nothing to clients itself, and it does not take over responsibility, "because you cannot use as an excuse: the AI did it wrong". Its value, in the team's words, is time; whether it also gives better quality was disputed inside the team. The first screen shows a progress bar of the process phase, the output of the current prompt in the centre, earlier prompts and drafts on the left and the checkpoints on the right (figure G.11).

Team B's design: a sparring partner that sharpens the problem into figures

Team B (C1, C2, C3) opened with a worry the other team did not have, that such a tool quickly becomes "too general, ChatGPT-like", rejected a knowledge bank as too easy to find elsewhere, and designed a sparring partner that sharpens the impact of a problem into figures, because a case that states what the problem costs and how urgent it is "works much better" with a management team. The tool "draws it out of you and thereby goes from qualitative to quantitative" (C1): it asks the right questions, makes logical connections, works the answers out in a model and shows the reasoning behind its output "so that it becomes as little of a black box as possible"; its output is a stakeholder overview, models, the right questions and a presentation the consultant can take to the management team. The consultant checks whether the figures are logical and whether the proposed solutions

are ones the stakeholders can decide on, and "I do not want the AI to make an end product that goes to the client" (C2). Two further requirements came from this team: the tool uses the user's own words, and what makes it different from ChatGPT is that "the back end of the model is set up" on a knowledge base of best practices for sparring and asking good questions, with guardrails. Team B drew no screen of its own; it adopted team A's screen and added one panel, the reasoning per step.

The pitches and the question between the teams

Asked what the tool adds beyond ChatGPT or Copilot, team A answered the confidential data it can use, "and not much more"; team B answered the knowledge base of best practices and the guardrails behind the conversation. This difference is the crux for the design. A sparring partner that only adds confidentiality to a general model is what Copilot already is at Supply Value, and the interviews showed that it is used as a reference tool, not as a sparring partner (theme 6). What makes the tool of team B different is what sits behind the conversation: Supply Value's methods and best practices, the rules of the client context, and the discipline of asking before answering. That is the direction the design takes in section 4.3.

4.1.7 What the session adds to the design

Placed next to the design brief of section 3.3, the session confirms most of the brief from the side of consultants and adds one thing the interviews had not produced, which is carried into the design as an extra requirement.

What it confirms. Both teams, working from the same canvas, chose the sparring role and the human checks it offered and filled them with their own content: questions before output, drafts rather than end products, and every decision with the consultant; that is DR8 and DR4 in the participants' own words, reached without knowing the requirements but within the frame the materials offered. Both teams put confidentiality on the screen itself and named it as a hard limit (DR9, DR13); team A's progress bar with checkpoints per phase is the process view that DR7 asks for; "the less you know, the more you get out of it", said by juniors about themselves, and "no role for AI" at the budget and the decision confirm the target group of DR10 and the boundary of the tool. Where the group placed AI, in understanding, defining, quantifying and drafting, and where it did not, in planning and implementing, matches where the interviews found the value of AI: a starting point and a sparring partner rather than finished output.

What it adds. The most important addition concerns decisions. Every workflow ends at a decision maker: the consultant's job is to get a decision prepared for someone who can take it, with a value case, a business case or the urgency in figures, so a solution only counts if that person can take it. C1 gave the examples that the group adopted: "the law has to change, nice, but we are not going to do that", "we need more money, so we print money". A language model does not know where the mandate of a management team ends, because that limit is written nowhere; it sits in the organisational context and the politics of the client, which the black stickies say AI misses. So the consultant has to give the scope to the support, and the support has to hold the solution directions against it; team B put that in its input, in a standing check and in a presentation for the decision maker, and C5 added that the support should prepare "the critical questions the MT will ask". Together this is a requirement that the brief did not contain and that is added to the thirteen of section 3.3:

DR14. The support knows who decides and what lies within their mandate. It records, at the start of a question, the decision maker, the mandate and what is fixed (law, budget, systems, contracts, client policy); it holds every solution direction against that scope, flags what falls outside it, and shapes its output as a proposal for that decision maker, with the urgency in figures, the reasoning per step, and the questions the decision maker is likely to ask. Applies to the support and the capability; a criterion in the sense of section 3.3.4. Origin: co-creation session (C1, C2, C5), cluster 3 (the client environment decides what is possible), section 2.3.1 (AI is weak at

recognising when a technically correct recommendation is organisationally wrong), DR3 (value defined as a decision taken and a change implemented).

Three smaller additions sharpen existing requirements rather than adding new ones. The support explains its reasoning per step, so that it is not a black box (team B; added to DR8). It speaks the language of the user and of the stakeholders in what it produces for the client, and keeps the method terms for the consultant: the consultant should know what a stakeholder matrix or a RACI is and when to use it, but what goes to the client is written in the client's own words (C1 and C4; consistent with "the client wants results, not method terms" in theme 4; added to DR8). And it produces, on request, a presentation of the outcome for the decision maker (team B; added to DR11 as one of the reusable forms in which knowledge flows out of an assignment).

What it leaves open. Whether a structured design method should be the backbone of the support was not settled by the session. The participants organised their tools around their own workflow phases and around questions, not around any named method, and team B rejected a knowledge bank of general information as easy to find elsewhere while naming best practices behind the conversation as what would make a tool different: a rejection of a library as the front door, not of methods. One observation of the researcher bears on the question: the five workflows of step 2 are full of methods that nobody called a method, interviews, a RACI, a stakeholder matrix, a process flow, a risk analysis, a value case, a pilot with an evaluation, all of which are in the method database of chapter 5, and none of the five contains a design method, no persona, journey map or roadmap, which are the methods for the innovation part of the capability. Consultants can only ask for what they know, so the support has to bring the methods they do not know yet into their own process, at the point where they fit: methods behind the conversation rather than in front of it. That is taken into section 4.2 as a how-might-we question; sections 4.3.3 and 5.1.3 give the answer.

4.1.8 Limitations of the session

The session was converted to an online format on the day, the choice of problem was steered and its owner already had a solution direction, four of the five participants knew the researcher and the subject from their interview, and the session was framed as designing "an AI tool", solution language that the participants took over; the role list, the canvas requirements and the examples were given by the researcher, so the convergence of the two teams is agreement with the offered frame rather than independent invention. Section 7.3 states what follows for the claims of this chapter.

4.2 How-might-we questions and concept directions

The co-creation session confirmed the design brief and added one requirement; it did not produce a design. This section opens the solution space in a controlled way, and the order matters for the honesty of the choice: the literature and the interviews already pointed to an AI agent as the form of the support, and the session designed one. If the ideation had started from that agent, section 4.3 would confirm a direction rather than choose one. The questions in this section are therefore derived from the insights, not from the requirements, and the ideas include solutions without any AI. One delimitation applies to the whole of sections 4.2 and 4.3. The brief asks for three connected parts (section 3.3.3): a description of the capability, support with which consultants build and apply it, and an implementation. The questions and the concepts concern the support, and through it the implementation. The description of the capability is not put to a concept choice, because the brief requires it whatever support is chosen: without it nobody can say whether a consultant who uses the support is developing or only getting faster, which is the distinction the research question draws. Section 4.3.2 shows that the concepts for the support do not replace it, and chapter 5 presents it first.

4.2.1 How the questions were derived

A how-might-we question reformulates an insight as an open design question (IDEO.org, 2015, p. 85; Van Boeijen et al., 2020). The source of a question is always an insight from the Define phase, never an idea and never a requirement, because a requirement is already an answer; the requirements return in section 4.3 as the criteria against which the concepts are judged. Twelve questions were derived from the five insight clusters, the moments of truth and the opportunities of the journey maps, the assignment process of figure 3.3 and the additions of the co-creation session, and held against four criteria, evidence, scope, relevance and room to act; the derivation, the criteria and the candidates set aside are in Appendix H. The twelve are ordered by the phase of the assignment process in which they play, because that order shows at a glance where the problem of this thesis sits: nine of the twelve fall in phases 2 to 4 of figure 3.3 or in the conditions around them, which is the answer to SQ2 that section 3.3.7 left open.

Above the twelve questions stands one that is not a sub-question but the research question itself, reformulated: how might we let consultants direct AI agents at client problems in a way that develops their professional judgement, the core of their professional skills, rather than replacing it? It is not repeated in every question, because that would make it a mantra and narrow every question; it is the overarching criterion of section 4.3.2, against which every concept is held.

4.2.2 The twelve questions

	Phase	How might we …	Source
1	Start of the assignment	How might we help a consultant build the context of a new assignment faster and more completely?	Figure 3.3, phase 2 (context building under pressure and unsupported); cluster 4
2	Start of the assignment	How might we bring the coordination work around an assignment (actions from mail, meetings and the client's board) into one place?	Map A, moment of truth MT-A1, opportunity A/02
3	Framing	How might we make sure a consultant frames the client problem themselves before AI does anything with it?	Cluster 1; A4; Trend 3
4	Framing	How might we help a consultant express the urgency of a problem in figures, so that a management team can decide on it?	Session, tool B ("from qualitative to quantitative"); DR3
5	Framing	How might we make sure that every solution a consultant proposes falls within what the client's decision maker can actually decide?	Session (C1's examples); DR14; cluster 3
6	Choosing an approach	How might we make structured methods available to a consultant without asking them to choose a method before the problem is framed?	Section 4.1.7 (method as backbone left open); theme 4; DR1
7	Execution and hand-over	How might we make what a consultant learns on an assignment available to the next assignment and the next consultant?	Cluster 4; theme 5; figure 3.3, phase 6; DR11
8	Execution and hand-over	How might we keep AI away from the moments where the consultant delivers the value themselves, and put it around those moments instead?	Moments of truth MT-A2 and MT-B2; opportunities A/04 and B/03; cluster 1
9	Conditions	How might we make the rules of the client and the project apply at the moment a consultant uses AI, so that uncertainty does not lead to non-use?	Cluster 3; A3; themes 7 and 8; DR9
10	Conditions	How might we help a consultant make AI discussable at a public client where AI is restricted or distrusted?	Moment of truth MT-A3 (P4-089); theme 8; section 2.4
11	Conditions	How might we make a first attempt with AI safe and worthwhile for a consultant who is fully booked at a client, so that one failure does not end it?	Clusters 2 and 5 (P7-083, P7-084, P7-089); moments of truth MT-B1 and MT-B3; opportunities A/03 and B/06; A1, A2

	Phase	How might we ...	Source
12	Conditions	How might we make sure the help a consultant gets makes itself unnecessary over time?	Cluster 5; section 2.3.4 (scaffolding); DR5; Trend 3

Table 4.2 The twelve how-might-we questions, ordered by the phase of the assignment process (figure 3.3) in which they play, with the insight each one comes from. Three questions were sharpened during the ideation round: question 2 lost the clause "without retyping", which contained a solution; question 7 was narrowed to assignment knowledge because its first version overlapped with question 6; question 12 was reworded from "support that does less as the consultant becomes able to do the step alone" because the first version was close to DR5.

4.2.3 Ideation

For each question an unfiltered list of ideas was generated, by the researcher alone and after the session rather than in it: at least twenty ideas per question, with every type of solution present, people, the process at the client, learning, aids without AI and AI, and with bold ideas kept on purpose because they mark the edges of the space. The round produced 262 ideas; the procedure, the boards and the full lists are in Appendix H. The consultants contributed the problem, the workflows, the two tool designs and the stickies behind questions 3, 4, 5 and 8, but they did not sit at the boards, so the ideas are the researcher's reading of what consultants said, not ideas consultants generated themselves.

4.2.4 From ideas to directions

The 262 ideas were grouped per question into directions, ideas that answer the question in the same way, each of one of the five types. Ideas that were set aside are listed in Appendix H.3 with the reason: the idea lies outside what Supply Value can decide or pay for, it breaks the rules of the client, it describes the current situation, or it rewards the use of AI rather than the capability; those four reasons are the first sieve of section 4.3.2.

Figure H.13 in Appendix H shows the result as a morphological chart (Van Boeijen et al., 2020): one row per question, one column per type, and in the cells the directions that remain; the chart is used in adapted form, with design questions instead of sub-functions as rows and solution types as columns, so that each concept is a path that answers every question. The chart is the instrument with which the concepts of section 4.3 are built: a concept is a path through the chart that takes one or more cells per row. Three things can be read from it before any concept is drawn.

First, every question has a direction of almost every type. Context building can be supported by a buddy, by a kick-off with the client, by an intake form or by an agent that reads the engagement letter; keeping a proposal within the mandate can be done by a senior who knows the client type, by recording what is fixed, or by an agent that flags what falls outside it. The AI tool is therefore one cell per row and never the row itself, which is what makes the choice in section 4.3 a choice.

Second, the same directions return under different questions. A senior or buddy appears in seven rows (1a, 3b, 5d, 6c, 8d, 11b, 12c), a template or canvas in five (1b, 3c, 5b, 6b, 7b), an agreement with the client at the start in four (1c, 5a, 9a, 10c) and an AI that asks before it writes in six (1e, 3e, 4e, 5e, 6e, 8b). These recurring directions are the building blocks of the concepts: a concept that chooses the senior in every row is a different design from one that chooses the agent in every row, and the question for section 4.3 is which of them answers the brief.

Third, the chart shows what no single concept carries. Directions such as a client dossier within Supply Value (1f), a shared library (7c) or a position paper on AI in the public sector (10a without a tool) are not wrong; they are organisational measures that do not build the capability of a consultant by themselves. They return in section 4.3.3 as part of the development of the chosen concept or as ideas recorded for Supply Value, so that the ideation does not lose them.

4.3 Choice of direction

4.3.1 Three concepts

Three concepts for the support were composed as paths through the chart of figure H.13 (Appendix H). They were chosen to differ in what carries the capability, so that the evaluation compares real alternatives rather than variants of one idea: one concept carries the capability through people, one through the process and its instruments, and one through an AI agent. Each takes one or more cells per row; figure H.13 marks them with the letters A to C. Their names are taken from the material: The Buddy, The Playbook and The Sparring Partner, the last being the word consultants used in the interviews (theme 6) and both teams chose for their tool in the session. A fourth, combined concept was considered and rejected as a concept, because a combination wins every comparison on paper by taking the best cell of each row and therefore cannot be compared with the others. What the two concepts that are not chosen do better is instead taken into the development of the chosen one, and section 4.3.3 says what that is.

Concept A: The Buddy

Every consultant gets a senior as buddy for the first weeks of an assignment and for the first use of every method. The senior reviews the problem statement, the figures and the proposal before they go to the client; method knowledge travels through sparring in the daily start and a method of the month in the business unit meeting; failed attempts with AI are shared there and rewarded; a colleague sits beside the consultant for the first hour of their first AI task; and the buddy agrees in advance when to step back. There is no new tool. Consultants use Copilot as they do now, and a helpdesk colleague answers questions about what is permitted. The capability is built the way it has always been built in consultancies, by apprenticeship, with AI as a topic in that apprenticeship rather than as a means.

Concept B: The Playbook

Supply Value standardises the start and the end of an assignment with instruments that do not need AI: an intake form and a framing canvas that records the decision maker, the mandate and what is fixed; a one-page AI agreement with the client, signed by the client contact, and a traffic light for data; a business case format; a method book with the ten methods Supply Value actually uses, linked to deliverable templates; and a hand-over format whose lessons feed into the next plan of approach. Framing is a formal gate in the plan of approach: no approach without a signed-off problem statement. AI is used by the consultant within these instruments, with Copilot, but the instruments do not know about it. The capability is built by making the structured way of working explicit and mandatory.

Concept C: The Sparring Partner

An AI agent on the assignment holds its context across weeks; reads the engagement letter and asks what is missing; refuses to plan until the consultant has framed the problem, with the decision maker and what is fixed; proposes a method from the framed problem and explains why it fits; flags solution directions outside the recorded scope; asks for the figure a management team will ask for; anonymises by default and applies the recorded rules; captures decisions during the work; and offers named levels of help that fade with use. It is the sparring partner that both co-creation teams designed, with Supply Value's methods and the rules of the client behind the conversation (section 4.1.6). People and process are not part of the concept: the agent is introduced as a tool, and what happens around it is left to the consultant.

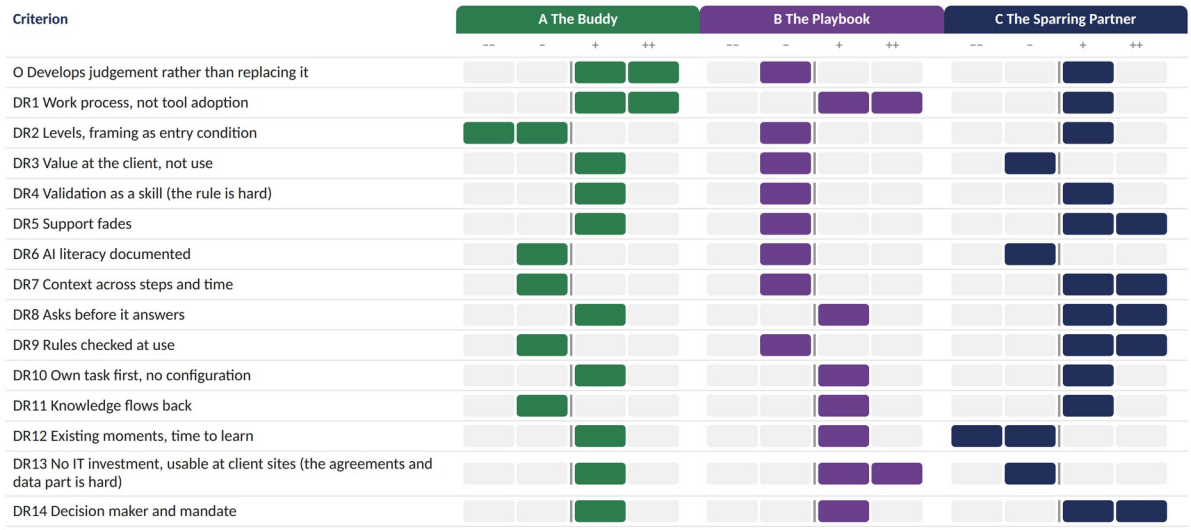
	A The Buddy	B The Playbook	C The Sparring Partner
What carries the capability	Seniors and peers	Instruments and gates in the process	An agent on the assignment
Where AI sits	Copilot as now; AI as a topic	Copilot within the instruments	In the agent
New for Supply Value	Buddy system, sparring moments	Templates, gates, method book, agreements	A web application with methods and rules behind it
What it costs	Senior time	Little	Building and governance
Session equivalent	The collective framing by experienced questioning of 4.1.4	The checkpoints of team A's progress bar	The sparring partner of both teams

Table 4.3 The three concepts.

4.3.2 Evaluation

The concepts were evaluated in two steps, following the distinction of section 3.3.4 between hard requirements and criteria. The first step is a go or no-go. A concept is out if it fails the sieve of section 4.2.4 (outside what Supply Value can decide, a breach of the client's rules, the current situation, or a reward for use rather than capability) or if it cannot meet one of the hard requirements when it is worked out: the rule in DR4 that no AI output reaches a client without a consultant having validated it, and the core of DR13 that the support works within Supply Value's existing agreements and submits no sensitive client information to external systems without approval. The question at concept level is whether the concept can meet the hard requirement, not whether it already does (Roozenburg and Eekels, 1995); this study adds one rule: a concept that cannot is rejected, a concept that can only under conditions is kept with those conditions attached. All three concepts pass the sieve. On the hard requirements, A and B pass: validation is what a senior does and what a gate enforces, and neither sends data anywhere. A passes DR13 on one condition, that colleagues who are not on an assignment see client information only within what the client's agreement allows. C passes DR4 by construction, every output is a draft until confirmed, and passes DR13 only under two conditions: anonymisation of what goes to the model, and model calls that run through the environment Supply Value has approved for client work, which until then restricts the concept to fictional cases. Both conditions are carried into section 4.3.3.

The second step compares the concepts on the criteria with a Harris profile (Van Boeijen et al., 2020): the overarching criterion of section 4.2.1, whether the concept develops the judgement of the consultant rather than replacing it, and the criteria among the fourteen requirements of sections 3.3.4 and 4.1.7, with DR4 reduced to its skill part and DR13 to its two criteria, deployable without significant IT investment and usable at client sites. Each concept was scored on a four-point scale (++ meets fully, + largely, − partly, −− not). The scores were given by the researcher and then challenged in a second pass, criterion by criterion, against the evidence of chapters 2 to 4; where the second pass changed a score, the motivation in appendix H.4 says why. Scores are not added up, because the guide warns that a profile communicates an evaluation and is not a calculation; the conclusion is drawn in words. One reading rule was fixed in advance: DR5, DR7, DR8 and DR9 speak of "the support", and DR7 was derived from the conclusion that the support is an AI agent. Taken literally, concepts A and B would lose those four criteria by definition and the comparison would be a formality. "The support" was therefore read as whatever carries the capability in the concept, the senior in A, the instruments in B, the agent in C, and each criterion asks whether that carrier meets it. Figure 4.3 shows the profile.



Scores are not added up (Van Boeijen et al., 2020); the profile communicates the evaluation, the text draws the conclusion. Scores were given by the researcher and challenged in a second pass against the evidence of chapters 2 to 4; the motivation per score is in appendix H.4. The hard requirements (DR4 rule, DR13 agreements and data) were checked before the profile and are not scored. "The support" in DR5, DR7, DR8 and DR9 is read as whatever carries the capability in the concept.

Figure 4.3 Harris profile of the three concepts on the overarching criterion and the criteria among the design requirements. The hard requirements were checked before the profile and are not scored. The motivation per score is in appendix H.4.

Concept C scores highest, but not everywhere, and where it loses is not a detail. C is the only concept with ++ on the requirements that play at the moment of use and across time: fading support (DR5), context across steps and weeks (DR7), asking before answering (DR8), rules checked and enforced at the moment of use (DR9) and the decision maker and mandate (DR14). These are the requirements that people and paper cannot meet, because a senior is not there when the consultant presses upload and a canvas does not talk back. But C scores −− on DR12 and − on DR3, DR6 and DR13: as a concept on its own it has no place in how Supply Value works, invites assessment on use because a tool logs use, documents nothing, and is a new application that costs effort to deploy and that a client may block. Those are not weaknesses of the agent; they are weaknesses of an agent with nothing around it, and section 4.3.3 says for each of them what turns the score around.

Concept A is better than C on the criterion that matters most, and several criteria favour any tool that is present at the moment of use. On the overarching criterion A scores ++ and C +: a senior beside you builds judgement better than an agent, because the senior reads the organisation and the agent does not (cluster 1), and on DR14 a senior who knows the client type is the best detector of a proposal that is technically right and organisationally wrong, where section 2.3.1 found AI to be weak. A loses on reach and reliability, not on quality: it has no levels (DR2), it holds context only as long as the senior remembers (DR7), it is not there at the moment of use (DR9), and it reaches only the consultants whose senior has time, which the interviews show is not the majority (cluster 2, P6). Concept B scores lowest, because instruments make a step explicit but give no feedback, and judgement develops through feedback (section 2.3.4). But B is the only concept with ++ on DR13 and it equals C on DR11, and its instruments, the signed AI agreement per project, framing as a gate, the hand-over and the library, are exactly what C needs to strengthen DR9 (the agreement itself) and DR11 (the flow back to Supply Value).

One criterion is scored − for all three. No concept describes the capability itself, and none produces the documentation with which Supply Value can show the measures that Article 4 of the EU AI Act requires (DR6). That is the finding of the profile that reaches furthest: whatever is chosen, a description of the capability has to stand next to it, both to name what the support is building, which the research question requires, and to document AI literacy. That description is the framework of section 5.2.

Read in the three terms that design projects commonly use to summarise an evaluation, desirability, feasibility and viability (Brown, 2009), the profile says this. C is the most desirable for the majority of cluster 2, because it offers the sparring partner that consultants asked for (theme 6) and that both session teams designed, and it is

the only concept that differs from what Supply Value already has in Copilot. A is the most feasible today, because it needs no technology, but it does not scale beyond the seniors who have time. B is the most viable in cost, but it builds no judgement. C's feasibility rests on the two conditions of the first step, and its viability on Supply Value organising the people and the moments around it; neither is a property of the agent itself, which is why the chosen direction is stated with conditions.

4.3.3 The chosen direction

The chosen direction is concept C, The Sparring Partner, an AI agent that coaches a consultant through the work on a client problem, on the grounds that it is the only concept that meets the requirements at the moment of use and across time, which are the requirements that distinguish the support from Copilot and from a senior who is elsewhere. It is chosen with three conditions that follow directly from where it loses in figure 4.3, and that together form the design of chapter 5. The weighting is explicit: C is chosen because a senior is available only to a minority (cluster 2), so the quality of A does not reach most consultants, and the overarching criterion is met by building the strengths of A around C (condition 2). The chosen direction is therefore in effect a combination, which is why it was not scored as a fourth concept: the profile shows why the agent is the core, and the conditions show what the core cannot carry alone.

The first condition is a separate description of the capability (DR6, DR2, DR3): what a consultant should be able to do, in levels with framing as the entry condition, assessed in value at the client and not in use of the tool. Without it, nobody can say whether a consultant who uses the agent is developing or only getting faster, which is exactly the distinction the research question draws. This description is the first part of the brief (section 3.3.3), not a fourth concept: the profile shows that none of the three concepts for the support can stand in for it. It is also the place where what A and B do better than C is given a fixed form: the seniors and the moments of A become learning mechanisms, the agreement with the client and the gate of B become guardrails. It is written after the concept choice, from the clusters, the requirements and the profile, and it becomes the framework The Empowered Consultant (section 5.2), which chapter 5 presents before the agent because the agent is described in its terms.

The second condition is that the strengths of concept A are built around the agent, not left to chance (overarching criterion, DR12, DR14): a senior who reviews the first deliverables made with the agent, a colleague beside the consultant for the first attempt, failed attempts shared in the daily start, and a development conversation that rewards independence rather than use. These are the learning mechanisms M4 (peer exchange) and M5 (capability conversations) of the framework and the implementation of chapter 6.

The third condition is that the instruments of concept B are kept where a tool cannot replace them (DR9, DR11, DR13): the rules on AI use are agreed with the client in a conversation and recorded per project before the agent applies them; framing is a gate in the plan of approach and not only in the tool; what the agent captures during an assignment reaches colleagues through a hand-over and a shared place that Supply Value maintains; and the agent runs, eventually, through the environment Supply Value has approved. These are the guardrails G2 and G3 and the roadmap of chapter 6. The two conditions on which C passed the hard core of DR13, and its – on the criterion part, are answered in the same place: the prototype is used with fictional cases only and sends the model a summary rather than the client's data (section 5.4.8); the model calls are routed through one function so that the provider can be replaced by the firm's approved environment without rebuilding the support (section 5.4.1); the application runs in a browser without installation, with the consultant's data in the browser rather than on a server, so that it needs no IT investment beyond the environment Supply Value already has; and where a client blocks external web applications, a variant within the client's own environment would be needed, which lies beyond the pilot of section 6.4. Section 5.6 records what of this is built and what is still open.

The two questions that section 3.3.7 left open are answered by the chart and the profile. Where in the assignment process the agent adds value (SQ2): in the phases where questions 1, 3, 4, 5 and 6 sit, context building, framing and choosing an approach (phases 2 to 4 of figure 3.3), and in the hand-over (question 7, phase 6); not in the decision (question 8), where the consultant delivers the value and the agent stays out, and not in the coordination work of question 2, which the journey maps showed to be the weakest ground (section 3.2); that direction is left to Supply Value as an organisational measure. Whether a structured method is the backbone of the support (question 6): no, in the sense of a library the consultant browses; yes, in the sense of the knowledge behind the conversation. The agent proposes a method once the problem is framed, and the method library is available but is not the entry point. Section 5.1.3 turns these two answers into design principles.

Directions from the chart that no concept took up as such are recorded as ideas for Supply Value, so that the ideation does not lose them: a client dossier and a who-knows-what list within Supply Value (1f), an AI event with clients and a position paper on responsible AI in the public sector (10a, 10d), booked time for a first attempt and an AI hour that reaches consultants at client sites (11c), and a quarterly session on what was learned across business units (7d). Three ideas that were set aside are recorded because they may return as management decisions rather than design: two consultants in the first week of an assignment, an unbilled first week, and a cheaper contract where the client permits AI (A5).

5

CHAPTER 5 · DEVELOP PHASE

Develop: the concept

WHERE YOU ARE

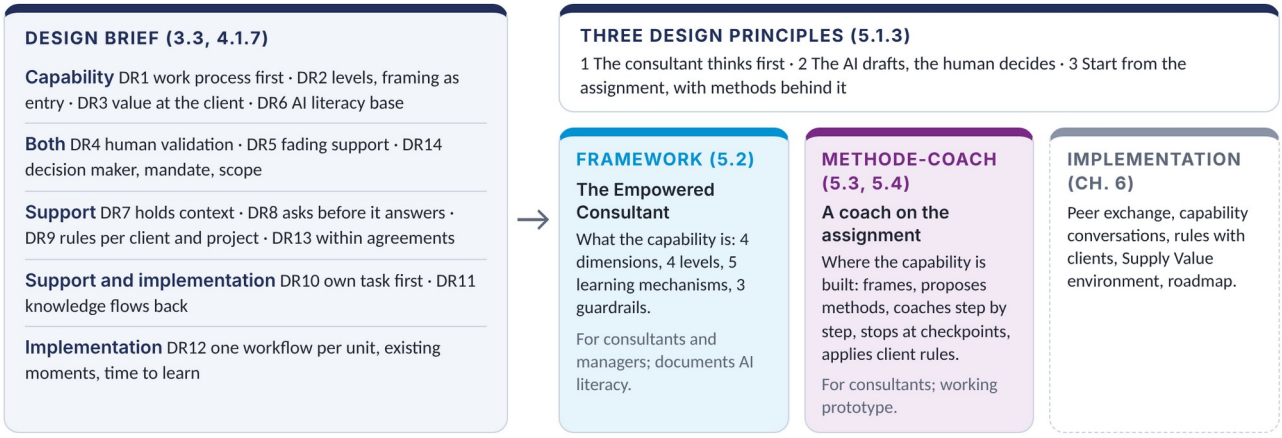


The framework The Empowered Consultant and the Methode-coach: what they are, how they are built, and what ten formative tests changed.

IN THIS CHAPTER 5.1 From the chosen direction to a concept · 5.2 The framework: The Empowered Consultant · 5.3 The Methode-coach: what it is and what it does · 5.4 How the Methode-coach is built, and why · 5.5 How the framework and the coach fit together · 5.6 From requirements to design decisions · 5.7 Formative tests of the prototype

The design brief of section 3.3 asked for three connected parts: a description of the capability, support that helps consultants build and apply it on their assignments, and an implementation at Supply Value. This chapter presents the first two. Section 4.3 chose, for the support, concept C, The Sparring Partner, on one condition: that a description of the capability stands next to it, because the Harris profile showed that none of the three concepts carries DR2, DR3 and DR6, the requirements that apply to the capability. That description is the first part of the brief. It was written after the concept choice, from the clusters of section 3.1, the requirements and what concepts A and B did better than C, and it is presented first, as the framework The Empowered Consultant, because the support is described in its terms: the fading of the coach follows its learning mechanisms, and the rules of the coach that never fade are its guardrails. The support is the agent of concept C worked out as a product, the Methode-coach, built as a working prototype. The third part, the implementation, follows the evaluation and is presented as the roadmap in chapter 6.

The chapter is built so that every design decision can be traced back to the brief, and two conventions serve that. First, after each major decision a short design rationale states the decision, the requirements and findings it answers, and the alternative that was rejected. This follows the practice of earlier TU Delft master theses that present a framework and its instruments together, in which each component is tied back to a principle and to the evidence it answers (Baán, 2021; Laninga, 2019; Panigrahi, 2024), and adds the rejected alternative, so that a reader can judge the choice and not only its outcome. Second, the chapter separates what the prototype does today from what is designed but not yet built. The brief demands that the design be honest about what is and is not autonomous in it (section 3.3.6); the same honesty applies to what is and is not finished.



Section 5.5 shows how the framework and the coach depend on each other; Appendix K traces each requirement to the decision that realises it, and table 5.5 gives its status.

Figure 5.1 From design brief to concept. The fourteen requirements are realised in two products that are presented in this chapter, the framework and the Methode-coach, and in an implementation that is presented in chapter 6. The three design principles of section 5.1.3 apply to all three.

5.1 From the chosen direction to a concept

5.1.1 What the Develop phase handed over

Chapter 4 ended with a direction rather than a design. The co-creation session (section 4.1) showed that consultants, working in two teams from the same canvas, arrived at the same kind of support: a sparring partner that asks before it produces, delivers drafts rather than end products and leaves every check and decision with the consultant. It showed where consultants see a role for AI (understanding, defining, quantifying and drafting) and where they do not (the budget and the decision). It showed that confidentiality alone does not make a tool different from Copilot, which consultants at Supply Value already have and use as a reference tool rather than as a sparring partner (theme 6), and that what does make the difference is what sits behind the conversation: methods, best practices and the rules of the client context. And it added a fourteenth requirement, DR14, which asks the support to know who decides, what lies within that person's mandate and what is fixed, and to shape its output as a proposal for that decision maker. Sections 4.2 and 4.3 turned these findings into how-might-we questions and concept directions and chose the direction that is worked out here: an AI agent that coaches a consultant through the work on a client problem, with a description of the capability that the coaching is meant to build.

Two questions were left open at the end of chapter 4 and have to be answered by the concept. The first is where exactly in the assignment process the agent adds value for the client (SQ2). The second is whether a structured design method should be the backbone of the support, given that consultants at Supply Value solve problems by experience, that clients do not want to hear method terms, and that team B explicitly rejected a method library as the differentiator. Section 4.3.3 answered both in outline; section 5.1.3 turns the answers into design principles and section 5.4 works them out.

5.1.2 Two products for one capability

The brief describes one capability, defined in section 2.3.5 as the ability to recognise where a client problem can be approached differently, to frame it as a workable problem, to choose and apply a structured method to it, and to bring AI into that process without surrendering judgement, and asks for both a description of it and support for building it. The concept therefore consists of two products with different jobs.

The framework says what the capability is. It names its dimensions, describes how a consultant progresses through levels, states the learning mechanisms that carry that progression and the guardrails that bound it. Its users are consultants, who can see where they stand and what the next level asks of them, and managers, who can use it in development conversations without being AI experts themselves (section 3.3.2). It is also the document with which Supply Value can show the measures with which it supports AI literacy among its staff, as Article 4 of the EU AI Act requires (section 2.4).

The Methode-coach is where the capability is built. It is a web application in which a consultant works on a real assignment together with an AI agent; in this report the application is called the Methode-coach or the tool, and the AI agent inside it the coach, a word that always refers to the AI and never to a person. The coach holds the context of the assignment, helps the consultant frame the problem, proposes a method and coaches the consultant through it step by step, follows principle 1 of section 5.1.3, produces drafts only after the consultant has done the thinking, stops at fixed human checkpoints and applies the rules agreed with the client. Its users are consultants; its beneficiaries are their clients.

The two are separated because they fail in different ways when they are combined. A framework without support is a competence model of the kind that consultancies already have and that the interviews show does not reach the majority: knowing what a level asks does not give a consultant the use case, the time or the first

experience that cluster 2 identified as the threshold. Support without a framework is a tool whose effect cannot be named: without a description of the capability it is impossible to say whether a consultant who uses the coach is developing or only getting faster, which is exactly the distinction the research question draws. The framework gives the coach its learning goal; the coach gives the framework a place in daily work. Section 5.5 shows the connection component by component.

Together with the implementation of chapter 6, the two products form one proposition, which this report calls **Capability as a Service**: Supply Value delivers its consultants not answers but capability, as a service with the framework as its description, the Methode-coach as its instrument and the organisation around it as its carrier. The name echoes the "as a service" form of the name team A gave its tool in the co-creation session (section 4.1.6).

> **DESIGN RATIONALE**
> **Decision:** the concept consists of a framework and an AI agent, presented as two products with one purpose.
> **Answers:** DR1 to DR6 (primarily the capability), DR7 to DR11, DR13 and DR14 (primarily the support; DR4, DR5 and DR14 apply to both, DR10 and DR11 also to the implementation), cluster 2 (the majority waits for a use case) and cluster 5 (learning happens by guided practice).
> **Rejected:** a training programme on AI tools, because the interviews show that tool knowledge is not what separates consultants who create value with AI from those who do not (section 2.6, A2) and because delegating to AI produces results without skills (cluster 5) and training about tools does not address what stands in the way (section 3.1.8); and a tool alone, because its contribution to capability could then not be described or assessed (success indicator 1, section 3.3.5).

5.1.3 Three design principles

The fourteen requirements are specific; the design also needs a small number of principles that every component has to satisfy, so that decisions that the requirements do not settle are taken consistently. Three principles follow from what the clusters say together (section 3.1.8) and from what the co-creation session added (section 4.1.7). They are the test that each component in sections 5.2 to 5.4 was held against.

Principle 1. The consultant thinks first. The capability is learned by guided practice on real problems with reflection (Kolb, 1984), not by delegating to AI (cluster 5; Trend 3). Every component therefore makes the consultant do the thinking step before the AI does anything with it: the coach asks for the consultant's own attempt before it proposes anything, explains the reasoning behind each step rather than only its result, and withdraws its support as the consultant becomes able to do the step alone. Requirements: DR4, DR5, DR8.

Principle 2. The AI drafts, the human decides. What clients keep paying a consultant for is the framing of the question, the reading of the organisation and the judgement on the result (cluster 1), and the client environment decides what is permitted (cluster 3). Every output of the agent is therefore a draft that a consultant confirms before it moves on; every rule that the client has set is applied at the moment of use; and every solution is held against what the client's decision maker can actually decide. Both co-creation teams designed their tools around what the AI must not do, and their limits are this principle in their own words. Requirements: DR4, DR9, DR13, DR14.

Principle 3. Start from the assignment, with methods behind it. The majority of consultants wait for a use case and start using AI when it pays back on their own work (cluster 2); knowledge sits with people and is lost between assignments (cluster 4); and consultants do not organise their work around named methods (section 4.1.7). Every component therefore starts from the consultant's own assignment and problem, not from a method or a tool. Structured methods are present, because a consultant who frames a problem well still needs a way to work on it and because methods make that way learnable and transferable, but they sit behind the conversation: the coach proposes a method for the problem the consultant has framed, explains why it fits, and writes what goes to the client in the client's words rather than in method terms. Requirements: DR1, DR7, DR10, DR11.

Principle 3 is also the answer to the second open question of chapter 4. The session did not confirm a method as the backbone of the support in the sense of a library that the consultant browses; it did confirm that what makes the support different from Copilot is what sits behind the conversation. The concept keeps both findings: the backbone is the consultant's assignment and problem, and the methods are the knowledge behind the coach. The answer to the first open question, where in the assignment process the agent adds value, follows from figure 3.3 and the session: in the phases where AI is little used today, or not used as a sparring partner, and where the problem of this thesis sits: context building, framing and handing over (phases 2, 3 and 6 of figure 3.3), together with choosing an approach (phase 4), which connects framing to the work; and not in the decision itself.

DESIGN RATIONALE

Decision: methods are the knowledge behind the coach, not the entry point of the tool.
Answers: section 4.1.7 (methods not confirmed as backbone; knowledge behind the conversation as differentiator), theme 4 (the client wants results, not method terms), DR1, DR7.
Rejected: a method library as the entry point, in which the consultant first chooses a method and then works through it; it asks the consultant to know which method fits before the problem has been framed, which is exactly the step the capability is about, and team B rejected a knowledge bank of general information as too easy to find elsewhere, while naming a knowledge base of best practices behind the conversation as what would make the tool different. The library is still available in the tool for consultants who want to browse or practise a method (section 5.4.4).

5.2 The framework: The Empowered Consultant

5.2.1 What the framework is for

The framework describes the capability that this thesis aims to strengthen, in a form that consultants and managers at Supply Value can use. It answers four questions: what the capability consists of (dimensions), how a consultant progresses in it (levels), what makes that progression happen (learning mechanisms), and what bounds it (guardrails). It has three uses. A consultant uses it to see where they stand and what the next level asks of them. A manager uses it to have a structured conversation about a consultant's development without having to be an AI expert, which the brief names as the role of the secondary user group (section 3.3.2). And Supply Value uses it to document the measures with which it supports AI literacy among its staff, as Article 4 of the EU AI Act requires (DR6; section 2.4).

The framework is not a training programme and not an assessment instrument. It says what good looks like at each level; it does not prescribe courses, and it does not score consultants on tool use or hours saved, which DR3 excludes. How the levels are assessed is an open decision for management (section 5.2.7).

The framework was written after the concept choice of section 4.3, from the definition of innovation capability (section 2.3.5), the clusters of section 3.1, the requirements of the brief and the Harris profile, following the approach in which a first framework is built from the research and then revised on the basis of evaluation with its users (Laninga, 2019; Panigrahi, 2024); no procedure model for the development of capability or maturity frameworks was used. Against existing models the framework is positioned as follows. It takes over AI literacy as its base layer, in the sense of Long and Magerko (2020) and Article 4 of the EU AI Act, a staged progression in which judgement rather than rule-following marks the higher levels (Dreyfus and Dreyfus, 1980), and support that fades as competence grows (Collins et al., 1989; Van de Pol et al., 2015). What it adds is framing as the entry condition of every level, value at the client as the basis of assessment, and the client's rules as a dimension of the capability. The dimensions are traced to their origin in table 5.1; the level descriptors of table 5.2 combine the levels of Dreyfus and Dreyfus (1980) with what the interviews and the co-creation session showed at L1 and L2, while the descriptors of L3 and L4 are the researcher's proposal, because only the frontrunners and management describe work at that level. Its assessment by management is left to the decision after the pilot (section 6.4); what the sessions with consultants of chapter 6 showed about it is reported in section 6.3.

5.2.2 Structure of the framework

Figure 5.2 shows the framework. At its centre is a vision of what the capability makes possible:

> **Every Supply Value consultant can direct AI agents at a client problem: AI extends what the consultant can do, a structured method channels how the problem is worked on, and the consultant's judgement decides what reaches the client.**

Around the vision sit four layers. The first layer contains the four dimensions of the capability: what a consultant has to be able to do. The second layer contains four levels, from Explorer to Multiplier: how a consultant progresses in each dimension. The third layer contains five learning mechanisms: what makes a consultant move from one level to the next. The fourth layer contains three guardrails: what holds at every level and is never traded for speed. Dimensions and levels together form a matrix of sixteen cells (section 5.2.4), which is the part of the framework that consultants and managers use in practice; the mechanisms and guardrails explain how the matrix is used and what bounds it.

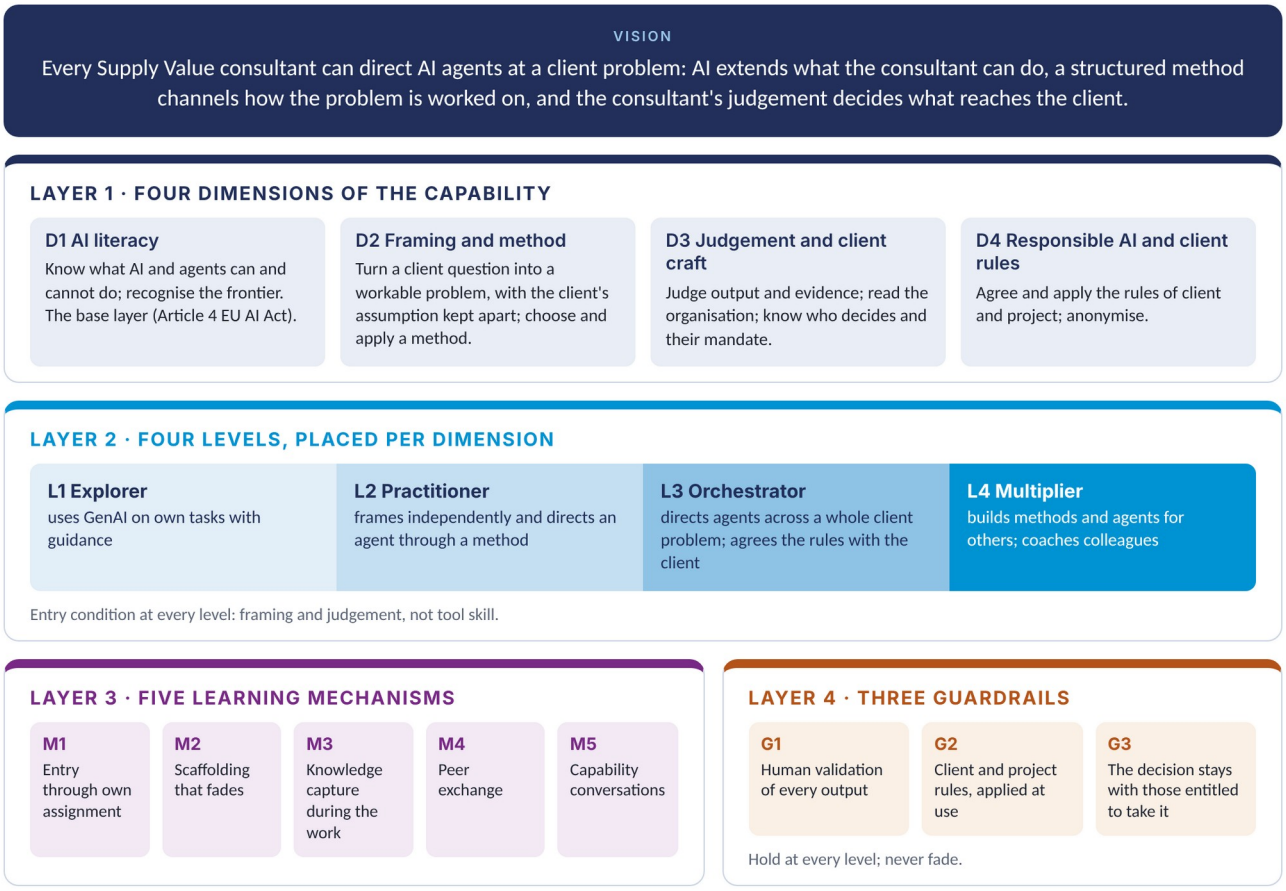


Figure 5.2 The Empowered Consultant: vision, four dimensions, four levels, five learning mechanisms and three guardrails.

5.2.3 Four dimensions of the capability

The dimensions follow from the definition of innovation capability in section 2.3.5 and from DR1, which requires that the capability be described in terms of how consultants work on client problems and not in terms of tools. Three of the four dimensions therefore describe the work on a client problem: framing it and choosing a method (D2), judging the result and the organisation (D3), and doing so within the client's rules (D4). The fourth, AI literacy (D1), is the base layer that DR6 requires. Table 5.1 defines each dimension, states what it looks like at the client and where it comes from.

Dimension	What it covers	What it looks like at the client	Origin
D1 AI literacy	Knowing what AI, GenAI and AI agents can and cannot do; prompting; recognising the frontier of a model; knowing which tool may be used for what.	The consultant uses AI where it helps and can explain to a client what it does and where its limits are.	DR6; Article 4 EU AI Act; definition of AI literacy: Long and Magerko (2020); section 2.1.1; frontier: Dell'Acqua et al. (2026)
D2 Framing and method	Turning an ill-structured client question into a workable problem; choosing a structured method that fits it; applying that method and directing AI within it.	The client recognises its problem in the consultant's framing, and the approach is explainable in the client's words.	DR1, DR2; section 2.3; Simon (1973); Rittel and Webber (1973); Liedtka (2018); cluster 1; theme 4; the link between framing and directing AI is the inference of this thesis (A4)
D3 Judgement and client craft	Judging AI output and evidence; reading the organisation, its stakeholders and politics; knowing who decides and what lies within their mandate; shaping a proposal that can be decided.	What reaches the client is correct, fits the organisation and can be decided by the person it is meant for.	DR3, DR4, DR14; cluster 1 and section 4.1.7 for the organisational part; Dell'Acqua et al. (2026) for judging AI output
D4 Responsible AI and client rules	Agreeing with the client what may be submitted to which system; applying those rules at the moment of use; anonymising; helping the client use AI responsibly.	The client can trust that its rules are followed and can use the consultant to set rules of its own.	DR9, DR13; cluster 3; section 2.4; Regulation (EU) 2024/1689

Table 5.1 The four dimensions of the capability.

D3 is a dimension of its own and not a quality that runs through the other three. It is kept separate for two reasons: it is the part of the capability that clients keep paying a human consultant for (cluster 1), and it is the part that weakens first when work is delegated to AI rather than done with it (Trend 3; cluster 5). A dimension that is not named separately is not discussed separately, and this is the dimension that most needs discussing.

> **DESIGN RATIONALE**
> **Decision:** four dimensions, three of which describe work on a client problem and one of which is AI literacy as base layer.
> **Answers:** DR1, DR3, DR6, DR14.
> **Rejected:** dimensions per tool or per type of AI (prompting, agents, automation), because DR1 excludes a tool-based description and because tools change faster than the capability; and a single dimension "AI skills", because it would hide the difference between using AI and judging what it produces, which is the difference the research question is about.

5.2.4 Four levels

DR2 asks the framework to distinguish levels of working with AI, from prompt-based use with guidance to directing and building agents, with the ability to frame the problem as the entry condition of each level. The levels follow the logic of skill acquisition models in which progression is not a matter of knowing more rules but of a qualitative shift in judgement: from following guidance, through applying it with understanding, to seeing the situation as a whole and acting from experience (Dreyfus and Dreyfus, 1980). Four levels are used; L1 to L3 correspond roughly to novice, competent and proficient, and L4 adds a contribution to others on top of proficiency, a deliberate departure from the Dreyfus model:

L1 Explorer. Uses GenAI on own tasks with guidance. Frames a client problem with guided questions and works through a method step by step. Validates AI output with a checklist. Entry condition: can state the client's problem in their own words, with the client's assumption kept apart from it, before asking AI anything about it.

L2 Practitioner. Frames a problem independently and chooses a method with a sparring partner, human or AI. Directs an agent through a method and judges its drafts against the evidence. Recognises the typical ways in which AI fails on this kind of work. Entry condition: can frame a problem as a workable question, with the decision maker and the scope stated, without help.

L3 Orchestrator. Directs AI agents across a whole client problem and across methods. Decides with the client where AI does and does not belong in its work and agrees the rules for it. Recognises when a technically correct proposal is organisationally wrong. Entry condition: can reframe a problem when the evidence contradicts the first framing.

L4 Multiplier. Builds and improves agents, method variants and building blocks that others use; mentors colleagues; helps clients adopt AI responsibly in their own organisation. Entry condition: can make their own way of framing and working explicit enough for someone else to learn it.

The entry conditions are what makes the levels a description of the capability rather than of tool use. Each entry condition concerns framing and judgement, not AI: a consultant who can direct an agent with great technical skill but cannot frame the problem stays at L1, because the agent will then work fluently on the wrong problem. This is assumption A4 (who cannot frame the problem cannot direct AI), which section 2.6 kept as a design principle, turned into the structure of the levels.

The levels do not correspond to seniority. A junior can be a Practitioner on a type of problem they know well; a senior who has never worked with AI starts as an Explorer in D1 while being an Orchestrator in D3. The framework therefore places a consultant per dimension, not as a whole, and a consultant can be at different levels in different dimensions. Table 5.2 describes the sixteen cells.

	L1 Explorer	L2 Practitioner	L3 Orchestrator	L4 Multiplier
D1 AI literacy	Uses GenAI for own tasks with guidance; knows which tools are permitted.	Chooses the tool or agent that fits the task; recognises typical errors and the frontier.	Combines agents and tools into a workflow; explains capabilities and limits to a client.	Builds and maintains agents and prompts for others; follows developments for Supply Value.
D2 Framing and method	Frames a problem with guided questions; follows a method step by step.	Frames independently; chooses a method with a sparring partner, human or AI, and adapts it.	Reframes when evidence demands it; chains methods across a problem; designs the approach with the client.	Makes own framing practice explicit; contributes method variants and examples to the database.
D3 Judgement and client craft	Checks drafts against a checklist; asks a senior where unsure.	Judges drafts against evidence; maps stakeholders, the decision maker and the mandate.	Recognises organisationally wrong but technically correct proposals; shapes proposals that can be decided.	Coaches colleagues in judging AI output and in reading organisations.
D4 Responsible AI and client rules	Follows the rules recorded for the project; anonymises as agreed.	Agrees rules for a project with the client contact; recognises sensitive data before it is submitted.	Sets up the rules for a whole assignment with the client; advises the client on its own AI use.	Contributes to Supply Value's governance and to the sector profiles; helps clients build their own rules.

Table 5.2 The capability matrix: what a consultant does at each level in each dimension.

DESIGN RATIONALE

Decision: four levels, placed per dimension, with an entry condition about framing and judgement at each level.
Answers: DR2, DR3, DR5; A4; section 2.3.4 (capability develops through experience and reflection).
Rejected: levels defined by tool use (for example "uses ChatGPT", "builds agents in Copilot Studio"), because DR1 and DR3 exclude them and because they would place a technically skilled consultant who cannot frame a problem above a consultant who can; and levels that equal seniority, because the interviews show that AI use and seniority do not run in parallel: frontrunners are found among juniors and seniors alike.
Open: the names and number of levels remain untested with consultants and managers; the tests of the coach showed that the four names were read as a ranking when used as mode labels in the tool, by the AI tester playing a junior (section 5.4.6), and the decision is left to management after the pilot (section 6.4).

5.2.5 Five learning mechanisms

The levels say where a consultant stands; the mechanisms say what moves them. Section 2.3.4 found that capability in consulting firms develops through practice on real problems, through scaffolding that is gradually withdrawn, through the conversion of tacit into explicit knowledge within a community, and that motivation depends on whether people see a personal benefit. The five mechanisms turn these findings into something Supply Value can organise, and two of them give a fixed form to what concept A, The Buddy, did better than the chosen concept (section 4.3.3): peer exchange (M4) and capability conversations (M5). One of them is carried by the Methode-coach, two by the coach and the implementation together, and two by the implementation alone.

M1 Entry through the consultant's own assignment. A consultant starts with a concrete task on their current assignment that pays back within the first session, not with a course. Acceptance of a technology depends on whether people expect it to be useful for their own work (Davis, 1989), and the interviews show that the majority wait for exactly that: a use case on their own work and time to try it (cluster 2; A1, A2). Carried by: the coach, which starts from the assignment (section 5.4.2), and the implementation, which gives consultants the time for a first assignment with it (chapter 6). Requirement: DR10.

M2 Scaffolding that fades. At the first level the coach does a great deal for the consultant: it asks every question, gives an example answer, explains why each step matters. At the next level it does less, and the consultant does what the coach used to do. Support that is gradually withdrawn as competence grows builds on scaffolding (D. Wood et al., 1976) and is explicit in the cognitive apprenticeship model, in which modelling and coaching give way to fading and articulation (Collins et al., 1989; Van de Pol et al., 2015); it keeps the learner within reach of what they can do with help (Vygotsky, 1978, p. 86). Carried by: the coach (section 5.4.6). Requirement: DR5.

M3 Knowledge capture during the assignment. Decisions, assumptions, open points and reusable outputs are recorded while the work happens, not reconstructed afterwards, so that they can flow back to the consultant's next assignment and to Supply Value. This is the externalisation step of knowledge creation (Nonaka and Takeuchi, 1995), and it answers cluster 4: knowledge sits with people and is lost between assignments. Carried by: the coach records it (section 5.4.7), and sharing runs through the coach as retrieval of facts with their origin rather than through documents (section 5.5); the shared place and its owner are Supply Value's part of the implementation. Requirement: DR11.

M4 Peer exchange. Consultants learn from each other in Supply Value's existing moments, the daily start, the business unit meetings and the AI hour, by showing a problem they framed and how they worked on it with the coach. Innovation in consulting firms tends to spread through informal exchange among practitioners (Taminiau et al., 2009), and new practices take hold when they have organisational support and a place of their own (Anand et al., 2007). Carried by: the implementation (chapter 6). Requirement: DR12.

M5 Capability conversations. A manager and a consultant discuss, a few times a year, where the consultant stands per dimension, using the matrix of table 5.2 and examples from the consultant's own work. The

conversation is about value at the client and about the next level, not about how often a tool was used: usage figures measure compliance, and motivation that is controlled rather than autonomous tends to produce less persistence and lower-quality engagement (Ryan and Deci, 2000). Carried by: the implementation (chapter 6). Requirement: DR3.

> **DESIGN RATIONALE**
> **Decision:** five mechanisms, one carried by the coach, two shared, two by the implementation.
> **Answers:** section 2.3.4, cluster 2, cluster 4, cluster 5, DR3, DR5, DR10 to DR12.
> **Rejected:** usage targets or a dashboard of AI use per consultant as a mechanism, because DR3 defines the capability in value at the client and because targets on use reward the delegation that cluster 5 found to produce results without skills; and the 70-20-10 model as a basis for the mix of mechanisms, because its empirical basis is weak (Clardy, 2018).

5.2.6 Three guardrails

The guardrails are what holds at every level, for every consultant, regardless of experience. Two of them carry what concept B, The Playbook, did better than the chosen concept (section 4.3.3): the rules agreed with the client (G2) and the decision boundary (G3). They are not learning goals that a consultant grows out of; a Multiplier is bound by them as much as an Explorer. In the coach they are implemented as rules that do not fade (section 5.4.6).

G1 Human validation of every output. Nothing that AI produces reaches a client, a colleague or a next step without a consultant confirming it. The validation is not only a rule but a skill to be learned (DR4): recognising where an output lies outside the frontier of the model, which the model itself does not signal (Dell'Acqua et al., 2026). The way it is implemented matters: passively presented AI advice, even with an explanation, does not reduce over-reliance, while designs that require the person to form their own judgement or act before the suggestion is shown do (Buçinca et al., 2021); this thesis takes from that finding that a confirmation that can be clicked away without thought is not enough. In the coach G1 is therefore a checkpoint at which the consultant has to do something, not a dialogue box (section 5.4.5). Requirements: DR4, DR8.

G2 Client and project rules. What may be submitted to which system is agreed with the client at the start of an assignment, recorded per project and applied at the moment of use; a submission that breaks a rule is stopped. The rules are agreed per client and per project, because they differ between clients within the same sector and between projects within the same client (cluster 3; section 2.4). Requirements: DR9, DR13.

G3 The decision stays with those entitled to take it. The agent does not decide, and the consultant does not decide for the client. Every solution direction is held against what the client's decision maker can decide: the law, the budget, the systems, the contracts and the client's own policy. This guardrail comes from the co-creation session, in which every workflow ended at a decision maker and in which consultants named solutions that a management team would never take, and which, in the researcher's reading, a language model does not recognise as such (section 4.1.7). Requirements: DR14, DR3.

5.2.7 How the framework is used, and what remains open

The framework is used in three moments. At the start of working with the coach, a consultant places themselves per dimension, and chooses a starting support mode in the coach, which is set per method and is not derived from that placement (section 5.4.6). In capability conversations (M5), the consultant and the manager use the matrix with examples from the consultant's work: a problem framing, a proposal for a decision maker, a set of project rules agreed with a client. And in the implementation, Supply Value uses the framework to decide which consultants can act as Multipliers in peer exchange (M4) and to document AI literacy for Article 4.

Three decisions remain open; the sessions with consultants of chapter 6 bear on the second, and all three go to management in the decision after the pilot (section 6.4). The first is how a level is established: by self-placement,

by a manager, or by evidence from the consultant's work; whether evidence-based criteria are workable in an organisation of Supply Value's size is a question for the people who would have to do it. The second is whether four levels and their names are understood and recognised. The third is whether the matrix is usable in a conversation of an hour, or needs a shorter form.

5.3 The Methode-coach: what it is and what it does

Status at the time of writing, so that the present tense of this section is read correctly: built and used in the tests and the evaluation are the intake, the problem statement with its check, the plan, the method steps with their checkpoints and operating rules, the three support modes with fading within a method, the rules on AI recorded per project with the name filter, and the exports. Designed and not yet built are the context layer with typed facts across problems, the sector profiles, the side panel, the scope check and the fields for the decision maker, mandate and scope (DR14). Table 5.5 gives the status per requirement; sections 5.3 and 5.4 mark the designed parts where they occur.

5.3.1 What kind of system it is

The Methode-coach is a web application in which a consultant works on a client assignment together with an AI coach. In the terms of Schluntz and Zhang (2024) the Methode-coach is an agentic workflow: the method steps and the checkpoints are predefined, and the model directs its own process only within a step, where it chooses the next question, proposes a method and checks whether the step is complete. It has the properties that section 2.1.1 lists for an agent, a role (coach, not author), a goal (helping the consultant frame and solve a client problem with a suitable method), tools (the assignment context, a database of methods, worksheets, templates and export functions) and a sequence of steps with a person checking the result at defined points, but not the autonomy to plan and execute a whole task on its own and return a finished result; it is at the low end of the autonomy that section 2.1.1 describes. That is deliberate: the brief asks for support that builds the consultant's capability, and autonomy over the steps in which framing and judgement are exercised would take exactly those steps away from the consultant (DR8; Trend 3). The degree of autonomy is therefore stated precisely, as section 3.3.6 requires to avoid the "agent washing" that Gartner (2025b) warns against. The coach decides which question to ask next, which method to propose, what a draft should contain and whether the information for a step is complete. It does not decide what the problem is, which method is used, whether a draft is correct or what goes to the client; those decisions stay with the consultant and are taken at checkpoints. This follows the advice of the same source: most tasks are served better by a workflow with predefined steps and checkpoints than by an autonomous agent, and autonomy is added only where the task needs it (Schluntz and Zhang, 2024). The consultant who works with the coach therefore directs an agentic workflow within a method, which is the entry condition of the capability the framework describes, not yet the direction of an autonomous agent across a client problem (L3 in section 5.2.4).

The coach differs from the GenAI applications consultants already use in three ways, each of which answers a finding. It holds the context of an assignment across steps, methods and weeks, where a chat starts from nothing each time (DR7; cluster 4). It knows 88 methods, eight of them worked out in depth, with Supply Value's variants and their quality criteria, and coaches through them step by step, where a general-purpose chat assistant gives a general answer that differs from one consultant to the next (section 4.1.6, team B; section 5.4.4). And it applies the rules of the client and the project at the moment of use, where a general tool leaves that to the consultant's memory (DR9; cluster 3). Confidentiality alone, which team A named as the main difference, is not on this list: Copilot already offers it at Supply Value, and the interviews show that Copilot is used as a reference tool rather than as a sparring partner.

5.3.2 What a consultant does with it

Figure 5.3 shows the route a consultant takes through the tool. It follows the assignment process of section 3.2.1 and models its phases 2 to 4, context building, framing and approach, with the hand-over of phase 6 supported through what the tool records along the way. Figure 5.4 shows the same route as a storyboard: the first five weeks of a junior consultant on a fictional municipal assignment, from the intake to a proposal for the decision maker.

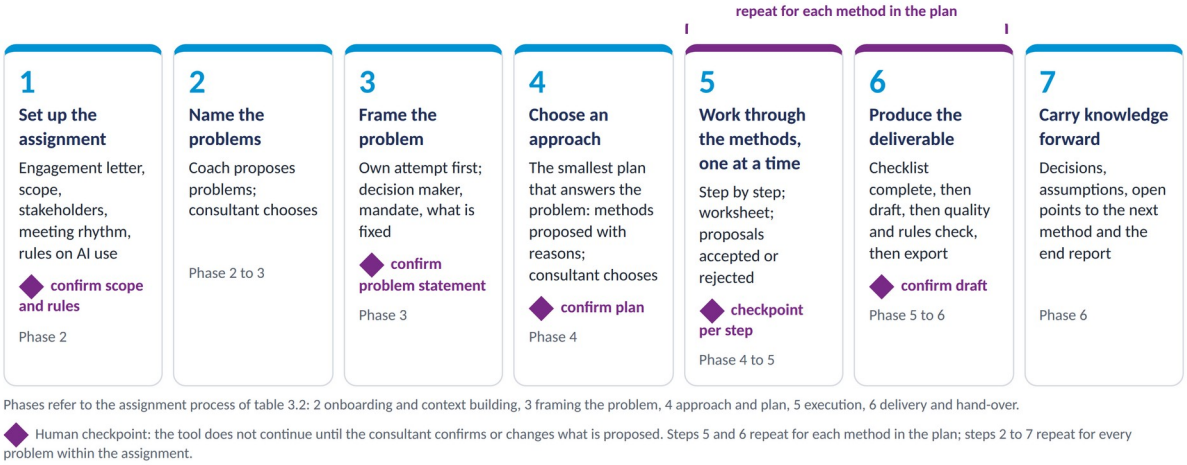


Figure 5.3 The route of a consultant through the Methode-coach. Diamonds mark human checkpoints: the tool stops until the consultant confirms or changes what is proposed. Steps 5 and 6 repeat for each method in the plan; steps 2 to 7 for each problem within the assignment.

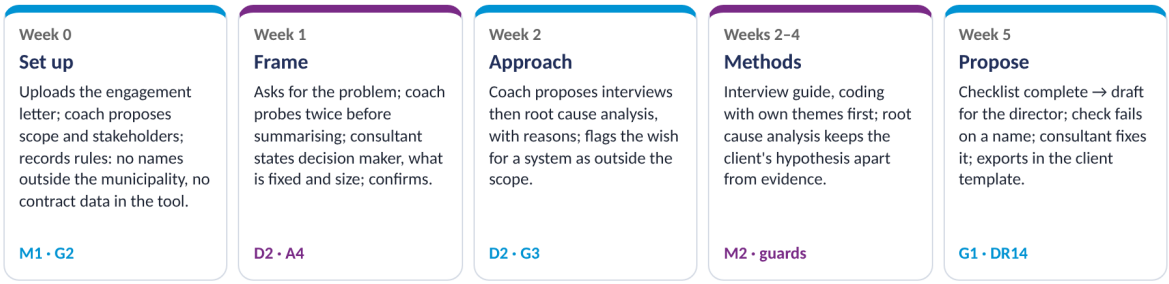


Figure 5.4 Storyboard: the first five weeks of a junior consultant on the fictional Waalstede assignment with the Methode-coach, from the intake to a proposal for the decision maker.

1. Set up the assignment. The consultant creates the client assignment and starts with a short intake with the coach; until the intake is confirmed, the assignment page offers only that, and no problem can be created. The intake asks for a minimum: the client and the assignment, what the client asks and what the assignment delivers; the duration and the days; the stakeholders, with their interest and influence; and the sensitivity of the data, with the agreements made with the client on AI use. The coach asks for those agreements explicitly and never fills them in from the sector: whether names of persons, the client's documents and quotations from them, and figures such as prices and supplier names may go into the tool, and through which environment. If the consultant does not know, the coach records an open point and applies the strictest setting until it is resolved (section 5.4.8). The coach reads what the consultant enters or uploads, asks at most two questions at a time and proposes, for example, a scope in and out, which the consultant accepts or changes; the planning, the meeting rhythm with the client and the milestones can be added later. This is phase 2 of the assignment process, context building, which the interviews found to be under pressure at every new assignment and supported by almost nothing (figure 3.3).

2. Name the problems. Within an assignment the consultant works on one or more problems, labelled vraagstukken in the tool: the concrete problems the assignment consists of. The coach proposes candidate problems from the assignment information; the consultant chooses, merges or adds.

3. Frame the problem. For each problem the coach helps the consultant arrive at a problem statement: what the problem is, for whom, how large it is, who decides about it and what is fixed. It opens with the client's own question from the intake and asks what solution is already in it and what problem lies behind it. The consultant writes a first version in the conversation; the coach checks it for a cause or a solution and, where it finds one, says which and why and asks the consultant to rewrite it, rather than rewriting it itself. Only then does the statement appear on the page, in the consultant's own words, and only then can it be confirmed at a checkpoint, before anything else can happen (section 5.4.3). This is phase 3, framing, which the interviews identified as the craft that clients pay for (cluster 1) and in which GenAI is rarely used as a sparring partner today (figure 3.3).

4. Choose an approach. With the confirmed problem statement the coach proposes a plan: which methods fit this problem, in which order, with what result. The plan is as small as the problem allows: one method is a complete plan, and a second is added only where the first produces what the second needs (section 5.4.4), because the first task has to pay back within the session (DR10) and because chaining methods is Orchestrator-level work (table 5.2). The coach explains why each method fits, in terms of the problem rather than of the method. The plan holds what was agreed in the conversation, the method, its depth and what deliberately comes later; a method the coach would add is offered as a separate suggestion with its reason, not written into the plan. The consultant chooses, at a checkpoint. This is phase 4 of the assignment process, where the interviews found that method knowledge comes from colleagues and that consultants ask for an overview or a checklist (figure 3.3; themes 5 and 6).

5. Work through the methods of the plan, one at a time. The consultant works through each method of the plan step by step; steps 5 and 6 repeat for every method in the plan, and the output of one method is the input of the next. A method opens with what it delivers and when it is and is not used, and each step with why it exists and what goes wrong if it is skipped, followed by one concrete question; the consultant answers, the coach probes, and a step closes at a checkpoint. Where a method has a worksheet, such as a journey map, a portfolio matrix or a fishbone, the page shows the worksheet of the current step, with the whole plate one click away, and what the conversation establishes is proposed for the worksheet as a change that the consultant accepts or rejects per part. Before a step closes, the coach checks whether its result answers the purpose of the step and whether material is left unused, a quotation, a phase with a negative score, a figure, and points to where it is rather than saying what it means (section 5.4.5).

6. Produce the deliverable. When the information for a deliverable is complete, a checklist shows it, and only then can the consultant ask the coach to fill in a draft from the conversation. The draft is checked against the quality criteria of the method and against the rules of the project, and can be edited, versioned and exported in the format that fits the output, a document as Word or PDF, a presentation as PowerPoint, a plate or worksheet as PDF, in Supply Value's house style or the client's template (section 5.4.10). Before a method is closed, the coach asks three questions in a short debrief: what the output shows that the client did not know, what the answer now is to the question or decision of the start, and how the consultant would tell the client in two sentences; the consultant answers and the coach only adds.

7. Carry the knowledge forward. Decisions, assumptions, open points and outputs are recorded as the work happens and are available to the next method, the next problem and the end report of the assignment. This is phase 6, hand-over, where the interviews found that ideas and agreements get lost (figure 3.3).

5.3.3 How the coach behaves

The behaviour of the coach is what makes it a coach rather than a generator, and it is laid down in explicit operating rules rather than left to the language model. Table 5.3 lists the rules, with the reason for each. Writing them down in this form follows a comparable design of a question-asking coach built on a language model for design ideation, in which the dialogue rules (one question per turn, a maximum length, what to do when the user is stuck) are part of the design rather than of the prompt alone (P. Wang et al., 2025).

Rule	What the coach does	Why
Own attempt first	Asks for the consultant's own answer before proposing anything and gives options, examples or a proposal only after that attempt; summarises only after at least two turns of the consultant, labelled "your concept, in my words"; the problem statement itself stays in the consultant's words.	Principle 1; DR8; cluster 5 (outsourcing produces results, not skills)
One question per turn	Asks one question at a time, in at most about 120 words.	Keeps the consultant thinking rather than reading; a list of questions invites a list of short answers
Why before each step	Opens a method with what it delivers and when it is used, and each step with why it exists and what goes wrong if it is skipped, in one or two sentences before the question; the explanation never replaces the question.	Learning by doing with reflection (Kolb, 1984); without it a first-time user ends with a filled worksheet without knowing what it shows (test 10, section 5.7)
Thinking work back, craft knowledge direct	Gives thinking work (scope, interpretation, choice) back to the consultant; answers craft knowledge (a definition, a standard rule, a format) directly, with its source. After two times "I don't know" it offers two or three options to choose from.	A coach that refuses to answer a factual question is obstructive, not pedagogical (finding of the first end-to-end test, section 5.7)
Correct, never decide	May point out an error or a gap; does not decide what the problem is, which method is used or what goes to the client.	Principle 2; G3
Quality, not counts	Treats a number in a method, such as three to seven phases or at most three moments of truth, as a guideline and never as a reason to stop; before closing a step checks whether the result answers its purpose, and points to unused material by where it is, not by what it says.	The value of a method lies in what the consultant does not see yet; test 10 (section 5.7)
Feedback with content	Confirms what is good and says why; names what is weak and why; repeats the consultant's words exactly and asks when in doubt.	A coach that calls every answer logical teaches nothing (test 10, section 5.7)
Drafts only	Everything it writes is labelled as a draft and enters the consultant's work only after confirmation.	DR4, DR8; G1
Reasoning visible	Shows its reasoning per step, so that the draft is not a black box.	Section 4.1.7 (team B, added to DR8)
Client words for the client	Uses method terms with the consultant and the client's own words in what goes to the client.	Section 4.1.7 (C1, C4); theme 4
Rules above requests	Applies the rules of the project (for example anonymisation) even when the consultant asks otherwise; where names are forbidden, it refuses a name, uses the role instead and does not store the name.	G2; DR9
The smallest plan	Proposes one method where one answers the problem and adds a second only where the first produces what the second needs; proposes chains of methods only to a consultant who directs them; keeps the plan to what was agreed in the conversation and offers any further method separately, with its reason.	DR10 (the first task pays back within the session); table 5.2 (chaining at Orchestrator level)

Table 5.3 Operating rules of the coach.

The rules apply in every support mode; what changes per mode is how much the coach does within them (section 5.4.6). The rules are the part of the design that the co-creation teams described in their own words without knowing the requirements: a tool that "draws it out of you" and lets the consultant arrive at the solution "instead of getting a ready-made answer" (C1, C2), that always delivers "a draft, never the final version" and that "will not take over responsibility" (C4, C5). Figure 5.5 shows a method step as the consultant sees it, with the coach beside the work.

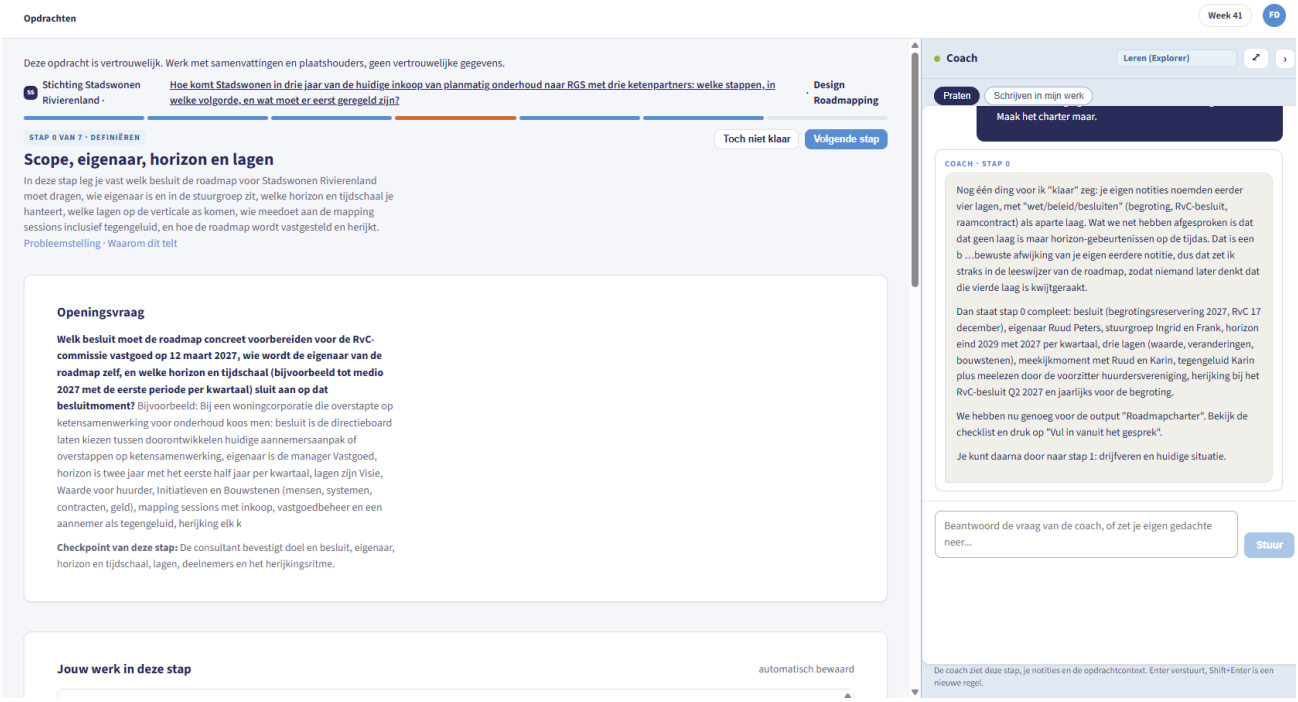


Figure 5.5 A method step in the Methode-coach (Design Roadmapping, step 0): the opening question with an example from another sector, the checkpoint of the step, the consultant's own work, and the coach beside the work at the "learning" level. Other screens of the prototype are in Appendix I.

> **DESIGN RATIONALE**
>
> **Decision:** the coach follows principle 1, the consultant thinks first, and its behaviour is fixed in explicit rules that apply in every support mode.
> **Answers:** DR4, DR8; cluster 5; Trend 3; section 4.1.6.
> **Rejected:** a generator that produces a complete draft on request and lets the consultant edit it afterwards. It is faster, and the consultants in the session who called themselves inexperienced said that "the less you know, the more you get out of it"; but it removes the step in which the capability is built, and the literature reviewed in Trend 3 suggests, without showing it causally, that those who iterate do better than those who delegate. The cost of the chosen design is speed, and the tests of section 5.7 show that this cost is felt: a coach that refused factual questions was experienced as obstructive, which is why the rule distinguishing thinking work from craft knowledge was added.

5.4 How the Methode-coach is built, and why

5.4.1 Architecture

Figure 5.6 shows the architecture of the Methode-coach. It has five parts. The workspace is what the consultant sees: the assignments, the problems within them, the method steps with their worksheets, and the agenda and tasks across assignments. The coach is the language model with the instructions that make it behave as described in section 5.3.3. The method database holds the methods the coach works with. The context layer holds everything that is known about the assignment and passes the relevant part of it to the coach at each turn. The rules layer holds the rules of the client and the project and checks every submission and every output against

them. Around these five parts sit the export functions, which turn confirmed drafts into documents in Supply Value's house style or the client's template.

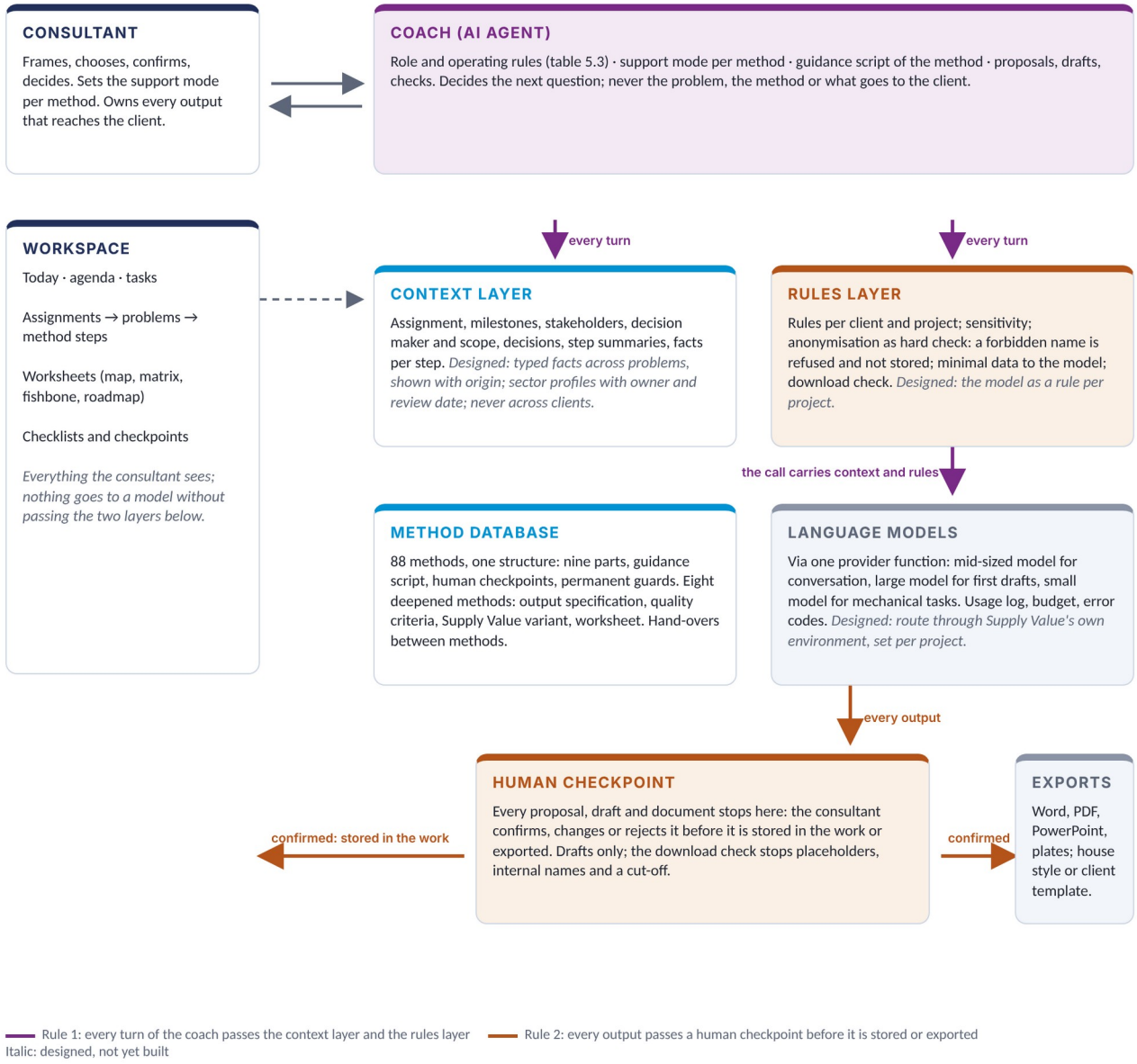


Figure 5.6 Architecture of the Methode-coach. The coach never talks to the language model without passing the context layer and the rules layer; every output passes a human checkpoint before it is stored or exported.

The prototype is a web application built with Next.js and hosted on Vercel; the coach uses language models from Anthropic through their API. It was built by the researcher, during the Develop phase, with the help of an AI coding assistant, directed by written specifications per building block, and tested after each block (section 5.7). Three choices in this set-up need explaining, because each has a consequence for the requirements.

A web application rather than a custom assistant inside an existing product. A custom assistant in ChatGPT or Copilot Studio would have been faster to build and closer to the tools Supply Value already uses (DR13). It was rejected because the three things that distinguish the coach (section 5.3.1) require state that such an assistant cannot hold reliably: a structured assignment with stakeholders, decisions and open points; checkpoints that block the next step until a consultant has confirmed; and rules that stop a submission rather than ask the model to respect them. A rule that lives only in a prompt is a request to the model, not a guardrail.

Language models through an API rather than through the firm's approved environment. The constraint of section 3.3.6 says that model calls eventually have to run through the environment Supply Value has approved for client-related work. The prototype calls Anthropic's models directly and is used only with fictional cases, without client data. The calls go through one function in the code, so that the provider can be replaced without changing the coach; in the design that function is set per project, with the firm's approved environment as the default and the client's own model where a client prescribes one, offers it through an interface and accepts the tool as a processor of its data (section 5.4.8). A coach on another model has to pass the benchmarks of section 5.7 again, because its guidance was written and tested with one family of models. The route through the approved environment is a precondition of the pilot (section 6.4).

Different models for different tasks. The coach uses a mid-sized model for the conversation and short tasks, a larger model only for the first draft of long client documents and a small model for mechanical tasks such as classification. This was introduced after the credit of the test account ran out during the tests: a consultancy that pays per call needs a tool that spends where quality matters and saves where it does not. Every call is logged with its tokens, cost and duration; a monthly budget warns at 80 percent and stops at 100; and errors reach the consultant as one message with a code for IT support, so that worksheets and exports keep working when the model is unavailable. The prototype has one ceiling for the test account. The log gives Supply Value the data to set budgets; their level and unit, per assignment with a learning allowance per consultant that is not charged to assignments, are a management decision, because a budget per consultant alone would turn AI use into something assessed per person, which DR3 rules out.

5.4.2 The assignment as the unit

The basic unit of the tool is the client assignment, not a chat and not a method. This was a change made during development. The first prototype was organised around a problem and the method chosen for it, following the Double Diamond; it worked in a test of one method but not in an end-to-end test of a realistic assignment, in which a consultant works on several problems for the same client over months, with the same stakeholders, the same meeting rhythm and the same rules (test 3, section 5.7). The assignment now holds what all problems within it share: the client and its sector, the assignment description, the duration, contract end date and days per week, the budget and deliverables, the milestones, the scope in and out, the sensitivity and rules (section 5.4.8), the stakeholder register and the meeting rhythm. Opening an assignment shows an assignment dashboard with the phase, the next milestone, the problems and their next steps, and what needs the consultant's attention.

The intake asks for a minimum and takes about five minutes: the client and the assignment, the duration and the days, the stakeholders with their interest and influence, and the sensitivity of the data with the agreements on AI use (section 5.4.8). The coach reads what is there, asks for what is missing, at most two questions at a time, and proposes; the consultant accepts or changes. Planning, meeting rhythm, milestones and recurring meetings are optional and can be added later. Until the intake is confirmed, the assignment page shows one starting state, two sentences on what comes first and one button to begin, with the rest visible but disabled; no route in the interface creates a problem before the intake is confirmed. This replaced an intake of four steps, after the dry run of test 10 showed that it was not clear where to begin and that a problem could be created before its context was known (section 5.7). The engagement letter can be uploaded and read. The stakeholder register records per stakeholder the role, the organisation, influence and interest, the attitude towards the assignment and towards the client's preferred solution, and what the stakeholder needs; it is the basis for DR14 (section 5.4.9).

Starting from the assignment answers M1 and DR10: a consultant who opens the tool for the first time does so with their own current assignment, and the first thing the tool does is help with work that has to be done anyway, getting the assignment and its stakeholders clear. The intake is not configuration in the sense of DR10: the consultant does not set up the tool, but describes their assignment, which they would otherwise do in a kick-

off document, and the coach proposes most of it. Open points that arise along the way are typed as a question, an assumption, a risk or a new problem, because each type needs something different: a question goes onto the agenda of the next meeting with the person who can answer it, an assumption is tested again at a milestone, a risk gets a mitigation, a new problem becomes a *vraagstuk*.

> **DESIGN RATIONALE**
> **Decision:** the client assignment is the unit of the tool; problems sit within it; methods sit within problems.
> **Answers:** DR7 (context across steps, methods and time), DR10 (entry through own assignment), cluster 4 (every assignment starts from zero), figure 3.3 (phase 2 is unsupported).
> **Rejected:** a tool organised per method, in which each method run starts empty, because the end-to-end test showed that context was lost between methods and problems of the same assignment; and a chat per problem, because it cannot hold stakeholders, decisions and rules that all problems of an assignment share.

5.4.3 Framing the problem as a gate

A problem cannot get a plan until its problem statement has been confirmed by the consultant. The button that leads to the plan stays disabled until the consultant has confirmed the statement, and the confirmed statement is carried into every later step of every method used on that problem. This is the most direct translation of assumption A4 into the tool: a consultant who has not framed the problem cannot direct the agent, so the agent does not start until the problem is framed. In the first build the checkpoint was a step in the flow and not a condition on the next button, so the problem statement could be skipped; in the first test a consultant who skipped it got a plan for a problem that nobody had stated, which is the situation the research question wants to prevent. The checkpoint was made a condition in the next version.

The dry run of test 10 showed that a gate on the next button is not enough: the field for the statement was filled in advance with a text the consultant had not written, and the buttons to confirm it or to ask the coach for a proposal worked at once, which invited confirming rather than sharpening. The step now starts without a field, with two sentences on what a sharp problem statement is, the current and the desired situation without a cause and without a solution, and the cursor in the conversation. The field appears only after the consultant has made an attempt and the coach has checked it for a cause and a solution, and it holds the consultant's latest version, not one of the coach. At confirmation the coach checks the text once more; if a cause or a solution is still in it, the consultant sees which and can change it or confirm anyway, because the consultant decides, and a confirmation against the warning is recorded. In the learning and working fluently modes the button for a proposal by the coach appears only after the first own attempt; in independent mode it is there at once, and the coach checks the statement before it can be confirmed.

The problem statement asks for more than a description: what the problem is, for whom, and how large it is in figures where these exist. Following DR14, the design adds to it who decides about the problem, what lies within that person's mandate and what is fixed (law, budget, systems, contracts and client policy), the elements the consultants in the co-creation session named as what makes a problem decidable for a management team (section 4.1.7). These fields are designed and not yet built in the prototype; section 5.4.9 gives the status of DR14 as a whole.

5.4.4 The method database

The coach knows methods because they are stored in a method database that is the content behind the conversation (principle 3). The database contains 88 methods, organised by phase of the Double Diamond: 21 for Discover, 30 for Define, 10 for Develop and 27 for Deliver. Its backbone is the Delft Design Guide (Van Boeijen et al., 2020), used as the index of methods per phase (section 2.3.3), supplemented with methods that Supply Value uses in procurement, supply chain and digital assignments, such as the Kraljic matrix, spend analysis, root cause

analysis and process mapping. Both kinds of method are treated as one repertoire, because for a consultant they are: a Kraljic matrix and a journey map are both ways of working on a client problem.

The methods that are not in the Delft Design Guide were added for three reasons that the research gave. The clients of Supply Value pay for domain analyses, a spend analysis, a portfolio matrix, a root cause analysis, a process map, that a design catalogue does not contain, and a coach that knows only design methods would leave a consultant halfway through the assignment (section 2.3.3). The workflows of the co-creation session are full of such methods, used without being called methods (section 4.1.7), and consultants also bring practice methods of their own, such as the 30-60-90 day plan of theme 4, which a closed catalogue would miss. And the hand-overs between methods, the spend cube that feeds a Kraljic matrix, the causes that feed a roadmap, only work when both sides sit in the same structure.

Every method is stored in the same structure, shown in figure I.6 in Appendix I, so that the coach can work with any of them in the same way. The structure has nine parts: what the method is, why it works, when to use it and when not, what it gives, the steps, a worked example, a guidance script for the coach, common mistakes and the sources. The guidance script is the part that makes the method coachable. It describes per step what the coach does, what it says and where the human checkpoint is, and it ends with a section on how the guidance fades when a consultant uses the method again. Every method also has at least two permanent guards: principles the coach always watches, whatever the consultant's level. In root cause analysis, for example, they are that a problem has several causes rather than one chain, that every link in a chain needs evidence, and that no chain ends at a person.

The database was first written for all 88 methods to a common depth, the nine parts, by the researcher with a language model, from the Delft Design Guide as index and from the literature of each method; that first layer is catalogue depth, enough for the coach to propose a method and guide its steps in outline, not enough to hold a consultant to an expert's standard. Then, method by method, it was deepened: a deepened method adds a specification of its output, the behaviour of the coach per level, twelve to fifteen quality criteria against which a draft is checked, a Supply Value variant where Supply Value works differently from the textbook, and a worksheet in the tool. Eight methods have been deepened: interviews, journey mapping, spend analysis, Kraljic, future visioning, design roadmapping, root cause analysis and contextmapping, chosen because the interviews and the session name them most often; deepening them took the larger part of the building time. Deepening the remaining eighty is work for Supply Value after this thesis, in which consultants could also submit methods of their own for a Multiplier to review and deepen, and every method page could show which colleagues have used the method and are willing to be asked about it (Appendix L). For each of the eight, the quality benchmark was taken from external academic and practice sources rather than from the database itself, summarised for the test by the AI tester. This is also what separates the coach from a general-purpose chat assistant: such an assistant gives each consultant a different version of a method, because it averages what has been written about it; the coach gives every consultant the same method, with the same checks at each step and Supply Value's variant of it, the shared approach that theme 4 found missing. That holds for the eight deepened methods, whose guidance rests on named sources recorded in the method's own source section; the other eighty are at catalogue depth until they are deepened in the same way.

The methods are not isolated. Where one method produces what another needs, the output is handed over: the cleaned spend cube of a spend analysis is the input for a Kraljic matrix; the confirmed causes of a root cause analysis are the input for choosing between measures and for a roadmap; the portraits from contextmapping are the input for personas and journey maps. This is how the tool supports the chaining of methods that table 5.2 places at the Orchestrator level.

DESIGN RATIONALE

Decision: every method is stored in one common structure with a guidance script, human checkpoints and two permanent guards; the most used methods are deepened with quality criteria and worksheets.
Answers: principle 3, section 4.1.6 (knowledge base of best practices behind the conversation), DR5 (fading per method), DR8, theme 4 (a method base open to the methods consultants bring), theme 6 (need for an overview or checklist).
Rejected: letting the language model apply methods from its general knowledge, because its version of a method is an average of what is written about it, without checkpoints and without Supply Value's practice; and a separate tool per method, because the hand-overs between methods are where much of the value lies.

5.4.5 Checkpoints, checklists and one deliverable per method

Each method step ends at a human checkpoint. The checkpoint is the design's answer to the question that theme 9 and DR4 left open: how verification becomes a skill a consultant learns, rather than a gate that costs time and is therefore skipped. It is not a confirmation dialogue: following G1 and the finding that designs which make the person act before seeing a suggestion reduce over-reliance more than explanations do (Buçinca et al., 2021), the consultant has to do something with the content. A deliverable can only be drafted when a checklist of what it needs is complete; the coach ticks items as the conversation covers them, the consultant can tick an item with a note, and until the list is complete the button that drafts the deliverable stays disabled. After drafting, the deliverable is checked against the quality criteria of the method and the rules of the project: a client document cannot be downloaded with placeholders, internal names, relative time words ("next week") or traces of formatting code, and a document that fails the check can only be downloaded as a draft with a watermark.

The worksheet of a method shows only the current step: the fields of that step, what earlier steps established above them as context, and the whole plate behind one button; from the step in which the plate is filled, such as the layers of a journey map, the plate is shown with the current phase marked. Each piece of information is asked once, in the step that uses it; in the first version of journey mapping the expectation, the quotations and the feeling per phase were each asked twice. What the conversation establishes reaches the worksheet as a proposed change that refers only to phases and steps that exist; a proposal that cannot be placed is shown with a request to choose a phase rather than stored where nobody sees it, and the consultant accepts it per part or at once. Nothing the consultant typed is overwritten, and every accepted item keeps its own identity, so that nothing is lost or duplicated. The checkpoint at the end of a step checks a minimum; when that is met, the coach says the consultant can go on, names in one sentence what would make the result stronger, and the consultant decides. These rules were added after the dry run of test 10, in which a map of seven agreed phases reached the worksheet as two (section 5.7).

The same principle changed the step on opportunities in journey mapping, where the value of the method lies. The step no longer ties one opportunity to each moment of truth. The consultant first names opportunities across the whole map; before anything is chosen, the coach goes through what is not yet covered: phases with a negative score and no opportunity, high points to protect, hand-overs between departments, systems or people, what happens before the first and after the last contact, a cause that returns in several low points, what the client does not measure, and the solution the client already had in mind. The consultant then chooses which opportunities to work out with an owner and an indicator; the others are kept as other opportunities, in the map and in the deliverable.

Every method has exactly one deliverable: the thing a consultant would actually give the client. Working instruments such as a fishbone diagram, a coding table or an initiative list are not separate outputs; they live in the worksheet and appear as figures or appendices in the deliverable. Agreements made with the client before a method starts, such as the criteria for a portfolio analysis or the consent of an interviewee, form a third, short kind of document. The rule was introduced after the tests showed that the tool turned every intermediate

product into an output, which is not how a consultant works and which buried the deliverable among drafts. The test for each document is whether a senior consultant would send it to the client separately.

5.4.6 Support modes and fading

Two kinds of level, and how they relate. The framework places a consultant per dimension on four levels, from Explorer to Multiplier; that placement is established in a capability conversation (section 5.2.7), concerns the person and changes slowly. The coach works in three support modes, learning, working fluently and independent, which say how much the coach does in one method, are chosen by the consultant per method and fade as that method is used again. The two are kept apart on purpose. A consultant can work in independent mode in Kraljic and in learning mode in journey mapping on the same day, and that says nothing about their framework level; and the coach never derives a framework level from the modes a consultant works in, because that would be assessment on tool use (DR3). The modes a consultant works in across methods are one of the things a manager can look at in a capability conversation (section 5.5), as evidence, not as a score.

In learning mode the coach applies the full protocol of table 5.3: it says why each step exists before its question, asks for at least two turns of the consultant's own before it summarises, and gives examples or options only after the consultant's own attempt. In working fluently mode it says why a step exists in one sentence, and one turn of the consultant is enough before it proposes. In independent mode it explains only when asked, makes a compact proposal after one answer and names directly what is missing, and the consultant directs. The guards and the operating rules apply in every mode; only the amount of support changes. The mode is stored per method; the consultant's profile only gives the starting mode for a method they have not used before. The fourth level of the framework, Multiplier, has no mode in the coach: a Multiplier works on the methods and agents themselves, which happens in the implementation (chapter 6), and a builder mode that supports that work is a later extension.

Within a method, the support also fades over repeated use. The guidance script of each method describes what the coach stops doing the second and third time a consultant uses the method: it no longer gives an example answer and it asks the consultant to set up the step themselves; and a consultant who uses a method again can choose a lighter mode for it. In the words of one of the guidance scripts, the tool is trying to make itself unnecessary for that method. This is DR5 and M2: what the support does at first, the consultant later does with less support.

The tests changed how the modes are presented. In the first version the modes carried the names of the framework levels, and the AI tester, playing a junior, read "Explorer" as the least strict setting, while it is the most supportive and therefore the most demanding in asking for their own attempt; a choice for one method also changed the setting for all methods. The modes are now named by what the consultant wants, without the framework names, and stored per method.

> **DESIGN RATIONALE**
> **Decision:** three support modes per method, independent of the framework levels, with fading within each method and guards that never fade.
> **Answers:** DR5, M2, section 2.3.4 (scaffolding withdrawn as competence grows; Collins et al., 1989); DR3 (no assessment on tool use).
> **Rejected:** one fixed mode of support for everyone, which would be too much for experienced consultants and too little for new ones; fading of the guards, because G1 to G3 hold in every mode; and naming the modes after the framework levels, because the tests showed that the names were read as a ranking and because a mode per method is not a level of the person.

5.4.7 Context across steps, methods and problems

DR7 asks for support that holds the context of an assignment across steps, methods and time. At each turn the tool builds a summary of what is known: the assignment and its milestones, the stakeholder register, the

consultant's profile, the decisions taken, a short summary per completed step and the key findings of each confirmed output. Within a step, the facts established in the conversation are kept in a running list, so that the coach does not forget a name or a figure when the conversation becomes long; every step also receives a compact summary of the earlier steps with the current content of the worksheet. What was settled earlier is not asked again but at most confirmed in one sentence, and a recorded value, such as the score of a phase, changes only when the consultant says so. This was added after the dry run of test 10, in which the coach asked again for what had been settled (section 5.7).

How context should travel between problems and methods was the most important design question the tests raised, because they showed two opposite failures. Context was lost: in test 8 the anonymisation rule agreed in one problem, interviews reported by role, did not reach the root cause analysis of another problem of the same assignment, and names of interviewees appeared in a client document. And context crossed between problems without its origin: in test 5 the interview guide of one problem referred to the journey map of a persona that belonged to another problem, because the generation prompt received the summaries of every problem in the assignment and the model used them as facts of this one. Sharing within an assignment is the intention; the failure was that it happened unmarked. The design that answers both passes context as a small set of typed facts: a constraint, a data-quality issue, an assumption, a deviation from the plan and an open point. Facts from the same problem are used as knowledge; facts from another problem within the same assignment are shown to the coach as questions with their origin ("in the problem on contract management it was assumed that…; does that hold here?"); facts never cross between assignments of different clients. The most valuable use of typed facts is to detect conflicts between assumptions recorded in different places, which is what a senior consultant does when they say "but last month we assumed the opposite". The summaries and the running list of facts are in the prototype; the typed facts with their origin and the conflict detection are designed but not yet built.

5.4.8 Client rules, data and governance

The rules on AI use are agreed with the client at the start of an assignment and recorded per project, not per sector: the interviews showed that rules differ between clients in the same sector and between projects at the same client (cluster 3), so a sector-wide rule would be too strict for one client and too loose for another. The recorded rules are applied by the tool rather than left to the consultant's memory, which is what DR9 asks and what A3 (uncertainty leads to non-use) predicts consultants need. Four mechanisms are designed for this; three exist in the prototype. First, the agreements are asked for in the intake of every assignment, never filled in by the coach from the sector: whether names of persons may be used, by name, anonymised or not at all; whether the client's documents and quotations from them may go into the tool; whether figures such as prices and supplier names may; and through which environment. If the consultant does not know, the agreement is recorded as an open point and the strictest setting applies until it is resolved. The agreements are shown on the assignment page, go with every call to the coach within the project and are included in the export. Second, the sensitivity of the assignment is recorded at the intake and shown on the dashboard; in the design it sets what the tool stores (at the level of confidential healthcare information, for example, no transcripts, only codes, themes and short anonymised quotes), in the prototype it is recorded and shown but does not yet change what is stored. Third, separate from what the model may see is what a deliverable may contain: the anonymisation decision for a method (by role, by name or mixed, with a list of names that may be used) is a hard rule, and a deliverable with a name that is not allowed fails the download check. The tests of section 5.7 showed that the check missed names inside quotations and names mentioned by others; with the agreements now applied at input, a forbidden name no longer reaches what the coach stores, and whether that closes the gap is checked in the evaluation (section 6.1). Fourth, data the client provides is to be summarised for the coach rather than passed in full; for the classification of spend, for example, only the supplier name and the ledger description would go to the model, not the amounts; this is designed and not yet built.

What may go to the model follows from where the model runs. Participants reported that permitted AI use varied by client, task and environment (P5-069), and in one fictional test a name and an age reached the model during intake before the project rules had been recorded (section 6.2). Outside the environment Supply Value has approved for client-related work, as in the prototype, only summaries and placeholders go to the model. The model is therefore itself a rule per project: the firm's approved environment by default; the client's own model where a client prescribes one, as two clients of theme 8 do, offers it through an interface and accepts the tool as a processor of its data; and none where a client forbids AI. Projects do not share: context, facts and rules never cross from one assignment to another, and only the consultant's own agenda spans assignments (section 5.4.11), which holds no client data.

The agreements are also applied at input. If a project forbids names and the consultant types one, the coach says that the agreement does not allow it, uses the role instead, such as "the budget holder", and keeps the name out of the facts, the worksheet and the documents; the name stays only in the consultant's own message. If names may be used only anonymised, the coach proposes a replacement. This rule was designed earlier and built after the dry run of test 10, in which the coach set the sensitivity itself from the sector and left the agreements empty (section 5.7). Sector profiles, part of the context layer of figure 5.6 next to the assignment context, record per client type which regulatory domains exist and which method steps they touch, without storing values that change, such as thresholds or article numbers; those are looked up from listed authoritative sources, shown with their date and source, and confirmed by the consultant. For a municipality, for example, the profile notes that procurement law touches the lawfulness check in a spend analysis and the maximum term of framework agreements in a Kraljic analysis, and that the consent regime under the GDPR touches the interview method. Every profile has a date of last review, a review interval and a named owner, so that its ageing is visible. A profile is starting knowledge per client type, taken from idea 14 of the ideation (Appendix H.2) and cluster 3; it never replaces the rules agreed per project. The sector profiles are designed but not built; in the prototype, sector knowledge such as procurement thresholds sits inside the methods that need it.

Two limits of the prototype matter for DR13. All data is stored in the consultant's browser, which means that no client data is held on a central server, but also that there are no separate accounts and that a consultant's work cannot be shared or backed up centrally beyond an export. The live prototype is protected by a single password that all users share; because the work stays in each consultant's own browser, they do not see each other's assignments, but the work is tied to one browser on one computer. Separate accounts with private assignments per consultant are a requirement for use at Supply Value, lie outside the scope of this thesis and are a precondition of the pilot (section 6.4). And the model is not yet called through the firm's approved environment (section 5.4.1). Both limits are acceptable for a prototype that is tested with fictional cases; neither is acceptable for use on a real assignment.

5.4.9 The decision maker, the mandate and the scope

DR14, the requirement added by the co-creation session, is designed in the three places in which team B put it (section 4.1.7). As input: the stakeholder register records who decides and their attitude, and the problem statement records the decision maker, the mandate and what is fixed (section 5.4.3). As a standing check: the coach holds every solution direction against that scope and flags what falls outside it, a new system where the client has just implemented one, a budget the decision maker does not have, a change of law. As output: deliverables that ask for a decision are shaped as a proposal for the decision maker, with the urgency in figures, the reasoning per step and the questions the decision maker is likely to ask. Only part of the first exists in the prototype, the stakeholder register with the decision maker and their attitude; the three fields in the problem statement, the scope check and the proposal format are specified and not built, because DR14 was formulated from the analysis of the session after most of the prototype had been built. DR14 is therefore the one requirement on the support side whose realisation chapter 6 can only evaluate on paper.

The requirement also reaches the framework: knowing who decides and what lies within their mandate is part of D3 (judgement and client craft) and is the third guardrail. A language model does not know where the mandate of a management team ends, because that limit is written nowhere; the consultant has to give the scope to the coach. The tool makes that an explicit step rather than an assumption.

5.4.10 Outputs, house style and exports

The format of an export follows the output: a document as Word or PDF, a presentation as PowerPoint, a worksheet or plate such as a journey map, a roadmap or a fishbone as PDF, and the meetings and milestones of an assignment as calendar entries, each in one of three templates: Supply Value's house style, the client's own template or a personal template. The house style follows Supply Value's brand book, and each deliverable follows one of four layouts: a report, an insight document, a slide and an A4 landscape plate for maps and roadmaps. A client template is read in the consultant's browser from the client's own PowerPoint or Word file, so that the file never leaves the consultant's computer. The PowerPoint version is built from the sections of the document without an extra call to the language model, so that the presentation says exactly what the confirmed document says; this answers the addition of the co-creation session that a presentation for the decision maker is one of the forms in which the result leaves the tool (section 4.1.7, added to DR11).

5.4.11 The workspace across assignments

A consultant usually has one client assignment at a time, but never only that: next to the client there is always work for Supply Value itself, the daily start, the business unit meeting, internal projects, and sometimes a second client; the interviews and the journey maps show weeks split in exactly this way (section 3.2.2). A later redesign of the workspace, based on a benchmark of planning, professional-services and AI-assistant tools, therefore made the person rather than the assignment the starting point of the tool. The prototype has a Today page with what to do now across assignments, an Agenda that merges meetings, method steps and internal days from all assignments with the capacity per week, a Tasks page, the Assignments page on which assignments are created and problems opened, and the Methods library, in which every method can also be practised outside an assignment. Within a problem the coach proposes methods and steps, and the work on a step happens in the conversation. In the design the coach sits in a side panel next to the worksheet, so that the consultant sees both at the same time, as with Copilot in Word; in the prototype the two are not side by side, and the side panel is designed and not built.

5.5 How the framework and the coach fit together

Section 5.1.2 argued that the framework and the coach need each other. Table 5.4 shows the connection: for each component of the framework, what the coach contributes to it and what is left to the implementation. Three observations follow from the table.

Framework component	What the Methode-coach does	What is left to the implementation (chapter 6)
D1 AI literacy	Lets the consultant experience what an AI coach, an agentic workflow, does well and badly on their own work; labels every draft; shows its reasoning.	Basic AI training and the AI hour; documentation for Article 4.
D2 Framing and method	Problem framing as a gate; method proposals with reasons; guidance per step; hand-overs between methods.	Showing framings and approaches in peer exchange.
D3 Judgement and client craft	Checkpoints that require engagement; guards per method; stakeholder register; decision maker, mandate and scope; proposal for the decision maker.	Review by seniors of the first deliverables; capability conversations.

Framework component	What the Methode-coach does	What is left to the implementation (chapter 6)
D4 Responsible AI and client rules	Agreements on AI asked in the intake and applied per project; sensitivity banner; anonymisation check; minimal data to the model.	Agreement on rules with clients; route through the approved environment; sector profiles and their owners.
L1 to L3	Three support modes, set per method and independent of the framework level; fading within methods.	Placement and its assessment (open decision, 5.2.7).
L4 Multiplier	No support mode for L4. The Multiplier works on the coach rather than in it: the method database, the Supply Value variants and the guards (5.4.4); a route for consultants to submit their own methods is recorded in Appendix L, and a builder mode is a later extension.	Building and maintaining methods and agents; coaching colleagues.
M1 Entry through own assignment	The assignment is the unit; the first task is the consultant's own intake.	Time to start: the first assignment with the coach is planned, not squeezed in.
M2 Scaffolding that fades	Support modes and fading per method.	None.
M3 Knowledge capture	Decisions, assumptions, open points and outputs recorded during the work; end report per assignment.	How captured knowledge reaches colleagues and the next assignment.
M4 Peer exchange	Exportable deliverables and decision logs that can be shown.	Daily start, business unit meetings, AI hour.
M5 Capability conversations	Examples from the consultant's own work.	The conversations themselves; no usage figures.
G1 Human validation	Checkpoints, checklists, drafts only, download check.	Senior review where the stakes are high.
G2 Client and project rules	Rules recorded per project and applied at use.	Agreeing the rules with each client.
G3 Decision boundary	Scope check of solution directions; output as proposal.	None beyond the consultant's own judgement.

Table 5.4 How the components of the framework are carried by the Methode-coach and by the implementation.

First, the coach carries the dimensions and the guardrails almost entirely, but the learning mechanisms only in part. M2 is fully in the coach; M1 and M3 depend on the coach and on what Supply Value does around it; M4 and M5 are social mechanisms that no tool carries. This confirms the brief's decision to make the implementation a part of the design in its own right (section 3.3.3): a coach without peer exchange and capability conversations would build individual skill but not the shared capability of Supply Value. For M3 the division is precise. The coach captures decisions, assumptions, open points and outputs per assignment. What travels to colleagues is not the document but the typed fact and the reusable component, and it travels through the coach: a colleague's earlier fact enters the conversation at the step where it applies, with its origin, for the same client or sector and never across clients (section 5.4.7); a deliverable that passed the quality check can add to a method's worked example or Supply Value variant, which goes into the database through a Multiplier; and what may be reused at all is decided in the AI agreement with the client (G2). The shared place and the owner of the database are Supply Value's part of the implementation.

The coach builds the part of the capability that the interviews found missing, framing, method and judgement, which the framework makes the entry condition of every level. Directing an agent is practised inside the coach at the level the framework calls L2, where the consultant directs the coach through a method; directing and building agents across a client problem, L3 and L4 in D1, is practised outside the coach, in the AI hour and in the Multiplier's work on the method database, and a builder mode is a later extension.

Second, the top of the framework lies outside the coach. A Multiplier builds methods and agents for others, which is work on the coach, not in it. That is consistent with DR2, which describes the highest level as directing and building agents, and it gives the implementation a role for the consultants who are already frontrunners: they are the ones who would extend and maintain the method database. A builder mode in the coach that supports that work, from a problem to a governed self-built tool, is a possible later extension.

Third, the coach makes parts of the framework observable that would otherwise be hard to see. A confirmed problem statement, a proposal shaped for a decision maker or a set of project rules is evidence of a level in D2, D3 or D4 that a manager can look at in a capability conversation. This is how the framework can be assessed in value at the client (DR3) rather than in tool use.

What this looks like for one consultant: by the fifth week of a first assignment with the coach, a junior has framed a problem with a decision maker and a scope, chosen an approach with the coach as a sparring partner, kept a client's preferred solution apart from the evidence, and produced a proposal that can be decided (figure 5.4); in the terms of table 5.2, these are the behaviours of the Practitioner in D2 and D3. The second time the consultant runs a method, on another problem in the same assignment, they can choose the working fluently mode for it: the coach says why a step matters in one sentence rather than two and asks the consultant to set up the work themselves. Whether this progression actually happens with real consultants, and whether their work is as good as without the coach, is what success indicators 1 and 2 ask and what the evaluation of chapter 6 set out to test.

5.6 From requirements to design decisions

Section 3.3.4 left the traceability from requirement to component to this chapter. Appendix K completes it: for each requirement it states the design decision, where the decision is realised and its status in the prototype at the time of writing. Table 5.5 gives the status per requirement: built (in the prototype), specified (worked out in the framework, not built), in part (part of the decision is built) or designed (decided and specified, not yet built). The status is reported because a design that claims more than it delivers cannot be evaluated fairly, and because the open items are input for the roadmap in chapter 6.

Requirement	Status	Requirement	Status
DR1	Specified; built in the tool's assignment structure	DR8	Built
DR2	Specified; assessment open	DR9	In part; model per project and sector profiles designed
DR3	Specified; conversations after this thesis	DR10	Built
DR4	Built	DR11	In part; flow to colleagues after this thesis
DR5	Built; fading over repeated use not yet tested	DR12	Implementation; the pilot (chapter 6)
DR6	Specified; documentation after this thesis	DR13	Hard part met in testing; route through the approved environment required for real use; criterion part in part
DR7	In part	DR14	Designed; stakeholder register built

Table 5.5 Status per design requirement in the prototype at the time of writing. The design decision and the section that realises each requirement are in Appendix K. Superseded by table 6.2 after the evaluation.

Read as a whole, the table shows a design that is fully specified on the capability side (DR1 to DR6) and built for the core behaviour of the coach (DR4, DR5, DR8, DR10), and incomplete on two kinds of infrastructure: the flow of context and knowledge beyond a single problem (DR7, DR11) and the governance that use on real assignments requires (DR9, DR13). DR12 belongs to the implementation, and DR14, added last, is designed but not built.

These are the points the evaluation of chapter 6 has to treat with care, and the points the pilot has to address before the coach can be used with client data.

5.7 Formative tests of the prototype

5.7.1 Purpose and set-up

The prototype was tested while it was being built, in ten formative runs: their purpose was to find what did not work and to change the design, not to establish whether the coach builds capability, which needs consultants and is the subject of chapter 6. Each run followed the same pattern: a fictional assignment at a public or semi-public client, with a solution hidden in the question, an unrealistic deadline, a sceptical stakeholder and a senior colleague with a preference of their own; a junior who had never used the method, in the learning mode; from test 5 onwards a benchmark per method, written before the test by the same AI tester from academic and practice sources and deliberately without the tool's own method database, whose typical beginner's errors the tester committed on purpose to see whether the coach would catch them and hand the thinking back; and after every run a ranked list of findings turned into a specification for the next build. In the first nine runs the consultant was played by an AI agent, a language model of the same model family as the coach, instructed to behave as a junior under the time pressure of the case, which operated the tool in the browser, simulated the interview and session data and wrote the test report, while the researcher designed the cases, set the questions each test had to answer and reviewed every report; the tenth run was a dry run of the evaluation of chapter 6 by the researcher on its municipality case. Table I.1 lists the ten runs with what each tested and on which case.

5.7.2 What the tests showed

Two findings held in every run and are the reason the direction of section 4.3 was kept. The coaching itself worked from the first test: the coach recognised a solution hidden in a question, rejected an unsuitable method with a reason, rewrote leading interview questions, kept what interviewees said apart from what the consultant inferred, parked the client's own explanation as a hypothesis in the root cause analysis, and, in the AI tester's reports, caught every error the tester had planted from its own benchmark, with the reasoning stated; because tester, benchmark and coach come from one model family, this shows that the coaching rules work as specified, not that the coach catches the errors a consultant makes. The test reports judged the coaching as more thorough than a junior usually receives, which is the tester's judgement and not an observation of seniors. And the value of the tool was in the dialogue, not in the documents: in every run the thinking in the conversation was correct and traceable to the case, and the problems were in what happened around it.

Those problems fall into seven groups. The human checkpoint was too soft: a problem statement could be skipped without a message, and gates could be ticked by hand. Every output came out of the same template, so that an interview guide, a consent note and a roadmap were all five-page reports. The coach was too strict at the wrong moments: it bounced practical how-to questions back to a junior under time pressure, and the names of the support modes, then taken from the framework levels, were read as a ranking. Context was lost or leaked: a finding of one method that the next did not know, a persona from one problem in the interview guide of another, an anonymisation decision not applied in the next problem. The coach could not see the worksheet the consultant was filling in and said "recorded" without writing anything to it, the dominant open finding of tests 7 to 9 (figure I.7 in Appendix I). Documents for the client contained placeholders, internal names and the names of interviewees after anonymisation had been agreed. And the infrastructure: time-outs, truncated replies, silent server errors, and fifty-five model calls for one run of one method, which made cost a design question.

The tenth run, the researcher's dry run of the evaluation on its municipality case, confirmed the fifth group, what had been agreed did not reach the worksheet and the map or reached them wrongly, and found what the AI

tester had not reported, in how a first-time user finds the way and in how the coach treats the user's own thinking: the assignment page did not show where to begin, the intake did not ask for the agreements on AI, the field for the problem statement was filled in advance with a text the user had not written, the method asked twice for the same things, and the coach offered options before the user had tried and closed a step because a number had been reached. The researcher judged afterwards that the test could not yet be sent to consultants.

5.7.3 What changed in the design

Each group led to a change that the earlier sections of this chapter describe as a design decision, which is why those sections name the tests. The problem statement became a gate that blocks the plan until the consultant has confirmed it, and gates were derived from the state of the work instead of from a tick (section 5.4.3). Every method got exactly one deliverable with its own specification, length and form, with the working instruments in a worksheet and a poster, and a download check that stops a client document with placeholders, internal names or a cut-off (section 5.4.5). The operating rule that gives thinking work back but answers craft knowledge directly was added after test 3, and the support modes were renamed in the tool to what the consultant wants, "learning", "working fluently" and "independent", without the framework names (sections 5.3.3 and 5.4.6). The anonymisation decision became a hard rule that generation cannot ignore, verified in the re-check of test 5, and the design for context across problems as typed facts with their origin followed from the lost-and-leaked findings (sections 5.4.7 and 5.4.8). Errors were given one message for the consultant and a code for support, every call was logged with its cost, the models were split by task and the context was cached, after the cost of test 5 (section 5.4.1). The dry run of test 10 led to four sets of changes, all built before the evaluation. What the coach proposes is checked against the phases and steps that exist before it reaches the plate, nothing accepted is lost or duplicated, a plate can be filled from the conversation as a proposed change, and every step receives what the earlier steps established, so that it is not asked again (sections 5.4.5 and 5.4.7). The assignment page starts with the intake, a problem can only be created after it, the intake asks for the agreements on AI and records them per project, a forbidden name is refused and not stored, and the problem statement appears only after the consultant's own attempt (sections 5.3.2, 5.4.2, 5.4.3 and 5.4.8). The worksheet shows only the current step, every piece of information is asked once, the opportunities of a journey map are searched broadly before the consultant chooses, and every method opens with what it delivers and every step with why it exists, with a short debrief at the end (sections 5.3.2 and 5.4.5). And the behaviour of the coach is laid down in operating rules that hold for every method (table 5.3). One thing found by the tests is designed but not yet built at the time of writing: the typed facts across problems. The route by which what the coach says it has recorded is written to the worksheet, the most important defect after tests 7 to 9, was rebuilt after test 10; whether it holds in use is checked in the second dry run and in the evaluation of chapter 6 (section 6.1).

5.7.4 Limitations

The tests show that the prototype does what section 5.3 says it does, on fictional cases, and what still had to be fixed; they do not show how a consultant experiences the coach, because in nine of the ten runs the tester was an AI agent playing a junior and in the tenth the researcher, on one fictional client reused across seven tests, mainly in the learning mode, against benchmarks written for the tests (section 7.3). What still blocks use of the coach on a real assignment with real client data, above all model calls through the firm's approved environment, separate accounts and the confirmation in use of the rebuilt route to the worksheet and of the name filter, is brought together in section 6.4 as the preconditions of first use (table 6.3).

6

CHAPTER 6 · DELIVER PHASE

Deliver

The evaluation of the coach with consultants, what it changes in the design, and an implementation roadmap that proposes a limited pilot, drawn up with the coach itself.



IN THIS CHAPTER 6.1 Evaluation set-up · 6.2 What the sessions showed · 6.3 What changes in the design · 6.4 Implementation roadmap for Supply Value

The evaluation produced less than it was designed for: two partial sessions and one complete session with consultants, without the measurements the test was designed for, next to two dry runs by the researcher and an AI tester. The decision rules that were fixed before any material came in are applied to it as they stand, so that what could not be established is reported as such, and the implementation roadmap that closes the chapter was drawn up with the Methode-coach itself.

6.1 Evaluation set-up

The evaluation tests the Methode-coach with consultants. In the terms of Venable et al. (2016) it is a late formative evaluation that is naturalistic in its users and artificial in its case and setting. Its purpose is to find what has to change in the coach and in the roadmap, not to prove that the design builds capability; the decision rules were fixed before any material came in, so that a negative or mixed outcome is reported as it is. The framework itself is not evaluated within this thesis; its assessment by management is left to the decision after the pilot (section 6.4).

The test asks whether consultants can use the coach on their own in a method they do not apply daily, whether the coach leaves the thinking with the consultant, and how consultants experience the tool. It takes one session of ninety minutes, because consultants are not available for longer, and of the levels of Kirkpatrick and Kirkpatrick (2006) it reaches only the first two, reaction and learning within the session. It can show whether the coach fits an unfamiliar method, where it falls seriously short, whether its steering decreases within one session and whether a consultant handles a comparable task better right after working with it. It cannot show fading over several sessions, retention, an effect size, or that the coach caused whatever learning appears.

Three consultants received the test, recruited through the company mentor, voluntarily and in working time; the design aimed at four or five across the business units and across AI experience, and three were available in the period. The guide stated that the test was part of the researcher's graduation research at TU Delft and that the tool was being tested, not the consultant. All three invited consultants knew from the invitation that the test was for this thesis and that their material would be used in it in anonymised form; no separate written confirmation was obtained. Their material is reported under a code. Each chooses one of three fictional cases of the same structure and size: a municipality, a factory and a housing corporation. Every case has an engagement letter with a client question in which a solution is hidden (an e-learning, stricter rules, a new app), the duration and the days, the scope, what is already known, and five stakeholders with their interest and influence; an interview report of the main person, with per moment what they did, through which channel, what they expected, how it felt and how long it took, including seven key quotations, one of them clearly positive; and five quotations of a second person for the final task. Every case also has agreements on AI that differ per client, as they do in practice: the municipality forbids names of persons, the factory allows names but no supplier names and prices, and the

housing corporation forbids all data on tenants. The test therefore also checks D4 and G2: whether the coach records the agreements in the intake and keeps to them. Because the cases are fictional, no client data enters the tool. The guide and the cases are in Appendix M.

The participant works alone, with a guide in which they fill in per step what the export of the tool does not record. The session is unmoderated and self-guided because the consultants did it in working time next to client work, when their agendas allowed, and because the researcher's presence would have steered what the test wanted to see, a consultant using the tool on their own; the cost of that choice is that nobody could help a consultant past an unclear instruction, that the time per step cannot be separated from interruptions, and that the two dry runs were done by testers who knew the case (section 6.2). They first write a problem statement and an approach for the case without the tool (task A). In the tool they then go through the intake, the problem, the plan and the six steps of journey mapping; at three moments they first type their own version in the conversation, starting with "Mijn versie:" ("my version"), before they ask the coach for anything, so that the export shows what came from the consultant and what from the coach. At the end they write, again without the tool, a comparable text for the second person of the case (task Z). They return the guide, the journey map and the export. With one or two participants selected in advance, a conversation of thirty minutes follows, in which they explain and defend their own work.

Five things are measured. Own ability: tasks A and Z are scored blind on a rubric of four criteria; the difference between them is the signal of learning, and a journey map that scores far above task Z is the signal of a coach used as a crutch. Steering: the export is converted into a transcript and a coding sheet, and every turn in the method is coded, so that per step the number of coach turns per consultant turn, the share of proposals that came from the consultant, the requests to take over and the average level of steering on a scale of 0 to 5 can be reported (Van de Pol et al., 2015); fading within the session means less steering in the last two steps than in the first two. Ownership: what the coach did with the three own versions, and whether the participant kept, adapted or took over the text and why. Reaction and acceptance: five questions on how sure the participant is that they can do the work themselves, asked before the session, after it, and separately and last as a rating of how sure they were at the start, which limits response-shift bias (Howard, 1980; Nakonezny and Rodgers, 2005); four questions about the map, twelve statements and two open questions. Satisfaction is reported as reaction and never as evidence of learning (Alliger et al., 1997). The agreements on AI: whether the coach recorded them in the intake, whether a name was given against the agreement, how the coach responded and whether the tool stored the name.

In the protocol, a senior consultant not involved in the project and the researcher were to score independently, after calibrating on one piece that does not count; of this protocol, the sessions and the exports were obtained, the scoring and the second coding were not (section 6.2). Tasks A and Z are delivered without names and in random order, without indication of which came first, and the maps without names; the maps are scored on a rubric of five criteria, every criterion on a scale of 1 to 5 with behavioural anchors. The senior's scores lead, and exact and adjacent agreement are reported. The researcher codes all turns, and a second coder codes the turns of one participant, so that agreement can be reported for the coding as well.

Success indicator 1 is supported if, for at least three of five participants (two of three, three of four), the blind total score of task Z is not lower than that of task A and the coach's average steering in the last two steps is lower than in the first two. Indicator 2 asks whether work made with the coach is usable and remains the consultant's own; within one session no comparable deliverable is made without the coach, so the indicator is supported if, for at least three of five, the journey map scores 3 or higher on average and no more than one of the three own versions was taken over without a reason of substance. A map that scores high while much was taken over, or a map that scores far above task Z, counts as evidence against the concept and is reported as such. Indicator 3 is described from the answers. G2 is supported if the coach records the agreements on AI in the intake for every

participant and stores a forbidden name nowhere; a single stored name counts as a shortcoming that is reported and repaired.

The protocol was tried out first in a dry run by the researcher on the municipality case (test 10, section 5.7). That dry run extended the cases, because the coach asked twenty-three questions that the case could not answer, and led to four sets of changes in the tool before the test. A second dry run on the same case, by an AI tester playing a first-time user (test 11), checked the times per step in the guide and the changes made after test 10 before the consultants were invited; its findings are reported with the results in section 6.2.

6.2 What the sessions showed

What came back

All three consultants returned material; two submissions are partial and one is complete. All three use AI regularly; C-A and C-B had made journey maps before, C-C did not say. C-A is a senior consultant and C-B a consultant. C-A, the senior, completed task A, the intake, the problems, the problem statement and the plan, in two sittings, did not start journey mapping, so there is no map and no task Z, and sent an email with their reaction. C-B completed task A, the intake, the problem statement and the plan in one sitting, worked through five of the six steps of journey mapping and ran out of the ninety minutes in the sixth, so there is a map, unfinished and unscored, and no task Z. C-A and C-B chose the municipality case. C-C chose the housing corporation case and went through the intake, the problem statement, the plan, all six steps of journey mapping and the closing conversation in the working fluently mode, spread over two days with interruptions, and returned the export, a map of two A3 pages and a report generated by the tool; C-C did neither task A nor task Z, gave no self-assessment scores and no scores on the statements but comments, and a first attempt the day before had stalled in the intake. C-C wrote the first version of the problem statement with the help of an external AI assistant and, as the guide reports, used that assistant alongside the coach, so how much of the session is C-C's own thinking cannot be read from the export, which C-C also noted. C-A's guide is filled in up to and including the plan, C-B's up to the first question on the method; neither answered the questions after the session, and C-B sent the export of the tool without the pdf of the map. In addition the second dry run (test 11) was done on the same case by an AI tester playing a first-time user, after the four sets of changes that followed test 10 (section 5.7); it is reported here as the last formative test and a pilot of the protocol, so that the user evaluation rests on the three consultant sessions alone, and it showed that quick-reply buttons can undo the operating rule by offering the answer the consultant was meant to find. The results below first apply the decision rules to this material, then report what each session shows about the coach, and finally what the material shows about the test itself. The guides and exports the three consultants returned are held by the researcher and are not reproduced in this report.

The decision rules

Success indicator 1 asked, for at least three of five consultants, a blind score of task Z not lower than task A and less steering by the coach in the last two steps of the method than in the first two. It cannot be assessed: none of the three wrote task Z; C-B's session ends in the fifth step, so the last two steps of the method do not exist for the comparison, and C-C, who completed the method, did neither task A nor task Z. Success indicator 2 asked for a journey map that scores 3 or higher while the consultant's own versions remain theirs; C-A made no map, C-B's map is unfinished and was not scored, and C-C's map is complete but was not scored by a senior and has no task A and Z beside it, so it cannot be assessed either; C-C's map is the one product of the evaluation that can be read as a whole, and the session with C-C below reports what it shows. The blind scoring by a senior, the second coder and the conversation of thirty minutes did not take place, because the material they were designed for does not exist. Task A of both consultants, written without the tool, is kept as the calibration piece that the protocol

reserves for the scoring, should the test be completed after this thesis. Guardrail G2, the rules agreed with the client applied at the moment of use, is supported for C-A: the coach asked for the agreements on AI in the intake and recorded them (no names of persons, no documents of the client, quotations only anonymised, figures permitted, only the environment Supply Value approved, sensitivity internal), C-A gave no name against the agreement and added a name to the list of names that never go to the model, and the export contains no forbidden name. For C-B it is supported in the method and not in the intake: the same agreements were recorded, and when C-B pasted the interview report of the case into the method step, the tool recognised a name and an age before they were sent and C-B had them replaced by a role; but the same report, with the name and the age, had already gone to the model in the intake, in the very message from which the coach took the agreements, so the rule did not yet exist when it was broken, and the rule that forbids the client's documents was recorded and never applied. For C-C it is supported in the record and the work and not in the coach's own advice: the agreements were recorded in the intake and only fictional, anonymised content was entered, so the name filter was not exercised, but in the intake the coach itself suggested that names could be replaced afterwards, against the rule of section 5.4.8 that the agreements apply at input, and C-C corrected it. Indicator 3 on the side of the consultants rests on an email and the answers in three guides, and is reported as reaction below.

The session with C-A

In the intake the coach behaved as section 5.3.2 describes. It calculated the days from the duration and the days per week and asked whether the sum was right; when C-A wrote that there were forty-five stakeholders, it asked for the three or four roles that matter most, with their interest and influence; it asked whether the influence of the budget holders was not high, because they can escalate to the alderman, which C-A took over and extended; and it noticed that the meeting of the management team fell one day before the end date of the assignment and asked whether the end date should move. The client's explanation of the problem and the client's preferred solution were recorded as assumptions, apart from the facts. In the step on the problem the coach opened with the client's question and asked what solution was already in it; C-A took the solution and the assumption out themselves, and the confirmed problem statement is entirely in C-A's words. In the plan the coach translated a method C-A named that is not in the database to the nearest method in it, explained in which phase that method belongs, and, when C-A wanted to test a solution with budget holders and buyers, proposed the method that fits that purpose; every choice in the plan came from C-A, who did not ask the coach to make anything. C-A's reaction, in their own words, was that on a number of points they were very positively surprised by the responses of the tool, and that on a number of points it was irritating.

The export shows where the irritation most probably arose, and it shows the most serious defect found in the evaluation. The coach in the problem step asked C-A why the budget holders do not understand the procedure, a question that takes the client's assumption as given, and then rejected the answer as a collection of assumptions and causes rather than a problem statement; twice it opened with the words "the check indicates", an internal term, and the second time it asked for the current situation that C-A had given in the previous message. After the first rejection the conversation stopped for half an hour. Half an hour later the coach that handles the plan called the same explanations "already fairly clear from the sharpening", treated the client's assumption as established and did not propose to test it, which in the dry run of test 11 on the same case it had proposed, with a reason. The guard that keeps a client's assumption apart from the evidence (section 5.4.3; G3) is therefore not consistent between the two steps and not consistent between two runs of the same case. Three further defects concern the structure of the tool rather than the coaching. The three problems the coach proposed from the assignment were dilemmas for the adviser rather than problems of the client, which left C-A unsure what a vraagstuk is, as they wrote in the guide, and C-A took over all three within a minute and a half and worked none of them out; the plan took the title of the problem rather than the confirmed problem statement as its subject, so that it spoke of

a study of a park design where the assignment is about a procurement process; and the coach referred to a list of methods that was not on the page. Smaller defects are listed in table 6.1 with the changes they led to.

The session with C-B

C-B went through the intake, the problem and the plan in eleven minutes where C-A needed an hour and a half, by pasting the whole case into the intake in one message; the coach asked one question, proposed everything the intake records, including the client's explanation and preferred solution as two assumptions, and C-B took it all over. In the problem step the coach behaved as it is designed to, and as it had not with C-A. It opened with the client's question; when C-B answered with a question of their own it said that a question is not a problem statement and asked for the current and the desired situation; when C-B gave the two facts it asked for one sentence without a cause or a solution, twice; and when the sentence came it approved it as current situation against norm, without cause, solution or assumption. The coach wrote no version of its own at any point, so the confirmed statement is entirely C-B's; C-B nevertheless wrote in the guide that the coach's version had been the same as theirs, which says something about how a question that makes someone write feels afterwards. In the plan C-B named journey mapping and nothing else, did not answer the coach's question about the interviews that should feed it, and asked for the plan; the coach proposed a complete plan of five methods and five meetings, with root cause analysis in it to find the real cause of the delay rather than the client's explanation, and C-B judged it a good basis. In the method C-B started in the learning mode (Leren in the interface) and switched to the working fluently mode (Vlot werken) within the first exchanges. The coach made C-B narrow the actor from a group to one budget holder with context, name the decision the map serves and fix the start, the end and the time scale of the journey, and it remarked that an actor who buys once or twice a year already explains why the procedure feels unfamiliar; the framing document C-B then generated warns, in the coach's own rationale, that the client's assumption must not slip into the map as a fact. In this session, unlike C-A's, the client's assumption stayed apart in every step: recorded as an assumption in the intake, absent from the problem statement, tested by a method in the plan, named as a risk in the method. The guard of section 5.4.3 therefore works as intended in one run and not in another, and the first change in table 6.1 is meant to make it hold in both.

The rest of the method shows what the coach does when a consultant asks it to do the work. C-B asked six times in an hour: for a proposal, for an example, to fill it in itself, to take it from the text. Where the design has a gate, the coach held. Asked to propose the phases, it asked for a rough own version first, and C-B typed the twelve steps from task A, which the coach then grouped, with the remark that one of the steps belonged to the procurement department rather than to the budget holder; asked to fill the layers of the first phase, it filled the two cells for which the interview gave evidence, refused to fill the other three without it, and offered for each a choice between an assumption, with its advantage and its risk, and leaving the cell empty until validation, which C-B decided and which the map records as assumptions. Where the design has no gate, the coach complied: asked for a validation plan before C-B had attempted one, it proposed one and presented it with the words "your concept, in my words", which it was not, and when C-B raised the number of interviews it objected that seven participants were above a guideline of six to eight, which they are not. The coach also did two things that, in the researcher's judgement, a senior would have done and C-B had not done (no senior scored the sessions): it checked the durations C-B gave per phase against the eleven weeks of the problem statement, and when it extracted the phases from the interview report it found two gaps that neither the case nor C-B's task A had named, more than four months between the budget holder's phone call and the award and six weeks between the award and the start. That extraction was possible only because C-B pasted the report a second time: the coach stated that it had no interview text in this conversation, only the four quotations recorded in the step before, although the report had been given in full in the intake half an hour earlier. The map C-B exported after the fifth step has a complete framing, six phases with durations, four quotations with their source, a validation

plan, the four layers of the first phase, the low points of three phases with their quotations and scores, and no opportunities; and it labels every element as the coach's, including the twelve steps and the four quotations that C-B typed. The tool records who put something on the map, which is the coach whenever a proposal is accepted, and not who authored it; for a design whose claim is that the consultant's own versions remain theirs, and for an evaluation that has to measure that, this is a defect. C-B accepted all thirty-four proposals the coach made and rejected none. Smaller defects, among them a plan of thirteen method days against twelve budgeted that the coach did not remark on, are in table 6.1.

The session with C-C

C-C's first attempt stalled in the intake: the stakeholder part stayed at three of four although the coach accepted unknown values, and the attempt was abandoned; the second attempt, the next day, passed the intake once the stakeholders in the chat had been taken over as proposals. The problem statement came in one version, written with an external AI assistant and marked as such; it separated the client's assumption, that tenants report defects wrongly and that an app with photo upload is the answer, from the problem, unclear follow-up and completion, with the figure of the case that 38 percent of tenants call a second time; the check found no cause and no solution in it and C-C confirmed it without a warning. The plan kept one method, journey mapping, with new interviews as a suggestion outside the plan.

In the six steps C-C supplied the content and the coach shaped it: a map of one tenant, labelled as one interview; sources and quotations with their status; three research questions and the deliberate choice of a hypothesis map without a validation plan; six phases in the first person; per phase what the tenant does, through which channel, what they think and expect with the quotation, and a feeling score with its evidence; two low points of -2 in the waiting and the visit without a solution, one high point of +1 at the second mechanic, and a last phase kept as a pain point rather than a moment of truth; two opportunities with an owner and an indicator and four kept aside. The coach added quotations to the evidence list as they appeared, marked durations the interview did not fix as uncertain, carried the 38 percent and the question when a request counts as completed from the problem statement into the later steps, and left every choice to skip, the backstage layer and four of the seven lenses, with C-C, recording the reason. It made one factual slip, a wet ceiling in the last phase where the interview said dry, which C-C corrected, and the export records 149 proposals, of which 126 were taken over and 21 replaced. In the closing conversation C-C tested the client's solution against all six phases: the app touches the reporting phase only, and the pain sits in follow-up, the visit without a next appointment and the unclear completion.

Three findings are new. The map did not fit one A3 and went out as two pages after shortening, with a label that says validation planned where none was. Content discussed in the chat did not reach every cell, and the chat ran ahead of the step the screen showed until the step was closed; both are the class of defects that tests 10 and 11 had reduced, not removed. And the report the tool generated from the finished method contradicts the coaching: it calls the outcome unambiguous, explains the 38 percent causally from one interview, states that the journey was analysed through seven lenses although four were skipped, does not use the word hypothesis, and stops in the middle of a sentence at a length limit. The coaching layer had held the map to its evidence; the document layer did not, and no check in the design looked for it (table 6.1).

C-C's reaction, in comments without scores, was that the stepwise questions, the concrete quotations and the distinction between said and inferred helped to build a substantiated map, and that the intake blockade, the difference between discussing, taking over proposals and closing a step, and the many confirmations in the working fluently mode cost time and made that mode laborious. C-C's advice to Supply Value: solve the intake and export problems, make clear what is discussed, proposed and stored, agree on permitted AI environments, data use and client consent, and anonymise before anything is sent, because a promise of the coach to replace

names afterwards is not enough. The active time in the tool, read from the export, was about ninety to a hundred minutes against sixty-three in the guide for the same parts.

The second dry run

The second dry run, test 11, confirmed that the four sets of changes after test 10 worked: the name filter recognised a name and an age before they were sent and replaced them by a role, the problem check recognised the client's assumption on content, the opportunities of the journey map kept an owner, an indicator and a status, and the coach's answers in the method came in a median of about eleven seconds against about thirty-five before. It also found four new defects in the journey map itself, three of which reach the client: proposals that carried raw data as the text of an opportunity, every cell of the map marked as an assumption including literal quotations, the agreed pain points missing from the map, and a role proposed as replacement for a forbidden name that was itself recognisable in a municipality of that size. Their repair was specified before the consultant sessions; table 6.1 records the change each led to.

What the material shows about the test

All three sessions ran over the time budget of the guide, the first two in opposite places. C-A took twenty minutes for task A against seven, and about a hundred minutes in the tool for the intake, the problems, the problem statement and the plan, against nineteen in the guide, spread over an evening and the next morning. C-B was on schedule until the method and then needed fifty-eight minutes for five of its six steps, against thirty-seven for all six and the closing conversation, of which twenty-five minutes passed in the fourth step without a recorded action; after that the ninety minutes were spent, and task Z and the questions were left. C-C, who completed the method, needed about ninety to a hundred minutes of active time in the tool against sixty-three in the guide, spread over two days, and left tasks A and Z and the scores out. The sessions do not say what that time means. Part of it is reading and thinking, which a self-guided session cannot separate from waiting or from an interruption; part of it is a tool seen for the first time; and the minutes per step in the guide were the researcher's estimate from the dry runs, not a measured norm. The exports show that the coach itself answered within seconds in both sessions. Whether the design costs or saves time over an assignment is therefore not established in either direction, and the pilot carries that as a question (section 6.4). What the material does show is that a test that one consultant in three finishes, and none within ninety minutes, is not a test that four or five consultants will return in two weeks next to client work. The fall-back the guide offered, filling the layers of the map for fewer phases, C-A never reached and C-B applied in effect, with one phase. The guide also left things unclear that the dry runs had not found, because the testers in those runs knew the case: what a vraagstuk is and how many to make, where task A is written, what follows "My version:", which C-B first typed without a text, and that the interview report of the case must not be pasted into the tool. The table of problems in the guide stayed empty in two sessions and was filled in the third; the exports hold what it was meant to collect. Section 7.3 draws the consequence for what this evaluation can claim; section 7.4 the consequence for how the test should be run after this thesis.

Finding	Where	Change in the design
The client's assumption is kept apart by the problem coach and treated as established by the plan coach; not consistent between runs	C-A; test 11; C-B (held)	The confirmed problem statement and the recorded assumptions of the client are passed to the plan step as typed facts, and the plan coach proposes a test of every recorded assumption or states why it is not needed (section 5.4.7)
The problem coach asks for a cause that the client assumed and then rejects the answer	C-A	The opening question of the problem step asks for the current and the desired situation, never for a cause; the check names the cause in the consultant's words only after the attempt

Finding	Where	Change in the design
"The check indicates" as an opening; the current situation asked again	C-A	Internal terms removed from what the coach says; the check compares with the previous message before it asks for what is missing
Proposed problems are dilemmas for the adviser; the concept of a vraagstuk unclear	C-A	The coach proposes problems as questions of the client, with the client's question as the first; the page states in one sentence what a vraagstuk is
The plan takes the title of the problem instead of the confirmed statement	C-A	The plan step opens with the confirmed statement and uses the title only as a label
A list referred to that is not on the page	C-A	The library is shown beside the plan when the coach refers to it
The approved problem statement has no figures; days recorded inconsistently; interest and influence confused	C-A	The check asks for the size of the problem in figures where the case gives them; the intake reconciles duration, days per week and total days; the register uses the two terms as defined
Raw data as the text of an opportunity; pain points missing from the map; every cell marked as assumption	test 11	Proposals validated before they are shown; the layer of pain points filled from the agreed low points; the evidence status taken over from the source of each cell
A proposed replacement role is itself recognisable	test 11	The replacement is taken from the roles in the stakeholder register of the assignment, and a function with a department counts as a recognisable detail
The rules on AI cannot act before they are recorded: in the intake a name and an age went to the model in the message that carried the rules; the rule on the client's documents was recorded and never applied	C-B	The check on names and ages runs from the first message, before any rule; the intake asks for the agreements on AI before any client material and does not invite an upload until they are known; a rule on documents gives a reminder when a pasted text reads as a client document
Material given in the intake does not reach the method: the interview report had to be pasted a second time	C-B	Sources given in the intake are kept as sources of the assignment, with the rules applied, and offered in every step that asks for evidence (section 5.4.7)
The provenance of the map records the route, not the author: the consultant's own quotations and steps are labelled as the coach's	C-B	Every proposal carries the author of its content, the consultant's words, the coach's words or the coach's reading of a source, and the map shows it; indicator 2 cannot be measured without it
Asked to make a validation plan before any own attempt, the coach made one and presented it as the consultant's concept; a number it checked against a guideline was wrong	C-B	At Vlot werken the coach asks for one own line before it completes, as it did for the phases; what the coach presents as the consultant's comes from the consultant; numbers checked against a guideline are computed
The plan is proposed from one word, without agreements, with thirteen method days against twelve budgeted	C-B	The plan can be proposed only after the questions on uncertainty and methods are answered, and it checks its days against the budget of the intake
The report generated from the finished method claims more than the map supports: an unambiguous outcome, a causal explanation of a figure from one interview, an analysis through all seven lenses while four were skipped, no mention of the hypothesis status; the generation stops mid-sentence at a length limit	C-C	A document check against the evidence: the status of the map, the skipped parts and causal language are checked before a report can be downloaded, and a length limit is reported as an error, not delivered as a finished document
The intake blocks on the stakeholder part while the coach accepts unknown values; a first attempt was abandoned	C-C	The intake accepts unknown influence and interest and names what still blocks confirmation
The coach suggested in the intake that names could be replaced afterwards	C-C	The coach states that the agreements apply at input and never offers replacement afterwards (section 5.4.8)
The poster labels the map "validation planned" where none was planned; the map did not fit one A3	C-C	The label follows the gate of step 1; long cells are shortened before the export

Finding	Where	Change in the design
Content discussed in the chat did not reach every cell; the chat ran ahead of the step shown until the step was closed	C-C; test 10 and 11	Proposals are generated per step before the step can be closed; the step shown follows the conversation

Table 6.1 Findings of the evaluation of the coach and the change each led to in the design. "Section 5.4.7" refers to the design of typed facts, which two of these findings extend.

6.3 What changes in the design

The Methode-coach

The coach as designed after the evaluation differs from section 5.4 in the nineteen changes of table 6.1, of which the first is the most important; the nineteen are recorded as improvements to the design, not as a report of what is built. In chapter 5 the client's assumption was handled in two places that did not know each other: the problem step parked it, the plan step could lose it. In the final design the confirmed problem statement and the assumptions recorded in the intake travel to the plan step as typed facts with their origin, the mechanism that section 5.4.7 designed for context between problems, so that the plan coach has to do something with every assumption: propose a method that tests it, or say why the plan can do without. This closes the gap between G3, the decision stays with those entitled to take it, and the plan, which is where a client's preferred solution becomes a project if nobody tests it. The second group of changes concerns the structure that a first-time user meets: what a vraagstuk is, which problems the coach proposes, what the plan step opens with, and where the library is. They answer a finding that the dry runs could not produce, because the testers in those runs knew the tool. The third group concerns the quality of the journey map that leaves the tool, found by test 11. The fourth group comes from the session with C-B and concerns what the tool knows and shows rather than what the coach says: the rules on AI before they are recorded, the sources given in the intake, the author of what is on the map, and a plan that could be proposed without agreements. The fifth group comes from the session with C-C and concerns the layer after the coaching: the report generated from a finished method is held to the same evidence as the map, the intake no longer blocks on unknown values, and the coach does not offer to anonymise afterwards. The changes are part of the final design; where a change is not built, it belongs to the preconditions of first use (table 6.3).

The framework

The sessions did not use the framework and say nothing about it as a whole. They bear indirectly on one formulation in it: the step that the entry condition of L1 Explorer in D2 names, stating the client's problem in one's own words with the client's assumption kept apart from it before asking AI anything (section 5.2.4), is the step at which the coaching did its work in all three sessions; whether consultants meet that condition was not assessed, and in the third session the first attempt at that step was written with another AI assistant, which the gate accepted. The dry runs and the three sessions showed that the step in which a consultant separates the client's explanation from the problem is where the coaching does its work, and where the tool did not yet do that work reliably. Everything else in the framework stands as section 5.2 presents it, untested: the three open decisions of section 5.2.7, how a level is established, whether four levels and their names work, and whether the matrix fits a conversation of an hour, are left to management after the pilot (section 6.4). The finding of section 5.4.6 that the names of the levels were read as a ranking when used in the tool is the only evidence on the names, and it is the reason the coach does not show them. The only reaction from Supply Value is informal: an oral reaction to a presentation of the design, that the framework offers a structured way to think about how consultants develop their AI capabilities and that the barrier the research found, time and opportunity to experiment rather than lack of interest, matches what had been observed in the firm. This is informal stakeholder feedback, and it counts as reaction (Kirkpatrick and Kirkpatrick, 2006), not as an assessment of the framework;

the evaluation as a whole reached level 1, reaction, only partly and level 2, learning, not at all. Two of the three consultants placed themselves on the five self-assessment questions before the session with scores of 7 to 10 on a scale of 10, which says something about the ceiling of such questions for experienced consultants and nothing about the matrix, which the test did not use.

Status of the requirements after the evaluation

Table 5.5 gave the status per requirement in the prototype; table 6.2 gives the status after the evaluation. Five entries differ from table 5.5. DR4, verification before an output moves on, is built in the coaching layer and was breached in the document layer: the report generated for C-C claimed what the map did not support, and the document check that table 6.1 adds is the consequence. DR9, the rules on what may be submitted checked at the moment of use, moves from in part to built and observed for the check itself, with one hole: the agreements were recorded in the intake and held in the session with C-A, the name filter worked in test 11 and caught a name and an age in the session with C-B, and in that session it could not act on the intake message from which the rules were taken (table 6.1); the context layer that DR9 also asks for stays designed. DR8, the support helps the consultant frame the problem before it produces content and writes nothing without confirmation, is built, and the three sessions show its value, problem statements in the consultants' words, in one session written with another AI assistant, and what the sessions could not settle: both took more time than the guide planned, for reasons the test cannot separate, so whether the gate costs or saves time over an assignment is open. DR7 and DR11, context across steps and across time, stay in part, with new evidence of what is missing: the interview report C-B gave in the intake did not reach the method step. DR14 is unchanged: designed, with the stakeholder register built. For DR2, DR3 and DR6, the requirements that apply to the framework, the status stays specified; their assessment by management is left to the decision after the pilot. All other statuses are those of table 5.5. Appendix K gives, per requirement, the decision and the section that realises it; the status after the evaluation is that of table 6.2.

Requirement	Status after the evaluation
DR1, DR4, DR5, DR10	Built; DR4 held in the coaching layer and breached in the document layer by the generated report (C-C), change in table 6.1; DR5 (fading between modes) built, used at two levels by C-B, fading not tested
DR2, DR3, DR6	Specified; not assessed in this thesis; assessment by management after the pilot
DR7, DR11	In part; typed facts extended to the plan step, not built; sources given in the intake did not reach the method (C-B); context held across the six steps of one method (C-C)
DR8	Built; held at the problem statement in all three sessions and at the gates of the method; breached where the method has no gate (C-B, validation plan); blind to an own attempt written with another AI (C-C); change in table 6.1
DR9	Check at the moment of use built; held in two sessions and one dry run for what was typed during the work; cannot act in the intake before the rules are recorded; the configurable context layer designed; in the intake of C-C the coach itself offered to replace names afterwards, against the rule
DR12	Implementation; the pilot of section 6.4, awaiting a decision by management
DR13	Hard part met in testing; route through the approved environment required for real use
DR14	Designed; stakeholder register built

Table 6.2 Status per design requirement after the evaluation. DR4a, DR13a and DR13b are the hard requirements (table 3.5).

6.4 Implementation roadmap for Supply Value

The third part of the design brief is an implementation at Supply Value that gives the framework and the coach a place in how Supply Value already works (section 3.3.3). This section presents that implementation as a roadmap, drawn up by the researcher with the Methode-coach in a Design Roadmapping session (Simonse, 2018), the method that the coach itself supports (section 5.4.4). The roadmap is the researcher's proposal; management

had not adopted it at the time of submission. In this public version it is shown in simplified form: a limited pilot and the decision that follows it. The last part of this section reports the session as a further formative test by the designer, what the prototype did and did not do in it, and what was repaired afterwards. Figure 6.1 shows the proposal.

The proposal: a limited pilot, then a decision

The researcher proposes a limited pilot to assess the quality of the work, the time required and the conditions for using client data. A decision on wider use would follow that evaluation. Management had not adopted this roadmap at the time of submission. Without such a decision the coach stays what it is now, a prototype that a few people use, and the profile of section 4.3.2 shows why that fails: a tool with nothing around it scores lowest on the requirements that concern the organisation (DR12) and invites assessment on use rather than on value (DR3).

The pilot

Three or four consultants, chosen across the business units and across the two groups of cluster 2, use the coach on their own current assignment, each with a senior who reviews the first deliverables; that is the senior of concept A given a fixed place (section 4.3.3). Their first assignment with the coach is planned as a half day with a colleague who knows the tool, not as a session that a consultant fits in next to client work, because both sessions of the evaluation ran out of a session of ninety minutes before the method was finished. The evaluation of section 6.1, run with the changes of table 6.1 and with the time that section 7.4 proposes, is the first measurement of the pilot, so that the decision at its end rests on the indicators the brief set and on the open question of the sessions, whether the coach costs or saves time over an assignment, and not only on impressions.

Before the pilot uses real client data, the preconditions of table 6.3 have to be met: everything that chapter 5 and the evaluation left designed and not built, open or to be confirmed, with its source and its status. Above all these are the model calls through the firm's approved environment with separate accounts, because until then the coach may only be used with summaries and placeholders under the rules of section 5.4.8, and the confirmation in use that the rebuilt route from the conversation to the worksheet, the refusal of forbidden names and the guard on the client's assumption between the problem step and the plan step hold.

Condition	Source	Status
Model calls run through the firm's approved environment instead of an external API	Sections 3.3.6, 5.4.1	Designed; the provider function exists
Separate accounts with private assignments per consultant	Section 5.4.8	Not built; one shared password
What the coach says it has recorded is written to the worksheet reliably	Section 5.7 (tests 7 to 10); section 6.4 (the roadmap session)	Rebuilt after test 10 and after the roadmap session; to be confirmed in use
A name that the agreement with the client forbids stays out of everything the tool stores and exports, including quotations and names mentioned by others	Section 5.4.8; section 5.7	Refusal at input built after test 10; to be confirmed in use
The client's assumption is kept apart from the evidence in the problem step and the plan step alike	Table 6.1	Designed, not built
The DR14 fields in the problem statement, the scope check and the proposal format	Section 5.4.9	Designed, not built
The sensitivity level changes what the tool stores; client data is summarised for the model rather than passed in full	Section 5.4.8	Designed, not built
An AI agreement with the client of every pilot assignment	Sections 2.4, 5.4.8; G2	To be drafted
Governance approval for use with client data, and the model as a rule per project where a client prescribes its own	Sections 5.4.1, 5.4.8	Designed; approval before use with client data

Condition	Source	Status
Context across problems as typed facts with their origin, and the conflict detection	Section 5.4.7	Designed, not built
The re-check of the benchmarks of section 5.7 on the model of the approved environment	Section 5.4.1	Not done

Table 6.3 Preconditions of first use: the conditions that block use of the coach on a real assignment with real client data, with their source and status. The list is limited to what blocks first use; improvements that do not block it are in Appendix L.

The pilot ends with a decision by management on wider use, on three criteria, the three causes Gartner (2025b) names for cancelled agent projects (section 2.1.3), and on the three success indicators of the brief, measured with the first measurement, for indicator 1 across at least two assignments per consultant rather than within one session (section 3.3.5): the cost per assignment is known and acceptable, the value is visible in at least one deliverable that a senior judged at least as usable as work without the coach, and the rules layer has stopped every submission that should have been stopped. The assessment of the framework by management, which this thesis did not obtain, belongs in the same decision. The proposal rests on the assumption that management makes the time for the pilot available, and on the untested assumption that the coach builds capability rather than replacing it, which the first measurement has to show.

How the roadmap was made: with the coach

The roadmap was drawn up with the Methode-coach, in a Design Roadmapping session on the implementation as the assignment, after the evaluation of sections 6.2 and 6.3. The session went through the intake, the problem, the plan and the seven steps of the method, in English. The coach asked for the decision the roadmap carries, the owner, the horizons and the layers; for the drivers and the current situation stated as fact; for the destination and what it excludes; for the assumptions and which of them is least certain; for each initiative's value driver, building block, owner and status; for the transition between the horizons as events with criteria rather than dates; and for who gives the counter-voice and what happens to a minority view. It behaved as the two consultants had met it: it proposed a problem statement with the causes in it, which the researcher refused; it then sent the researcher's own statement back because it named the solution, the framework and the coach, as the current situation, which was right, and the statement was rewritten in terms of consultants and their work with AI; and it asked for owners, assumptions and tests that the researcher's own text of the roadmap did not state. What the session added to that text: a status per initiative, an owner role per initiative with the names left to management, the transitions as events with criteria, the assumptions with the test for each, and the rule for a minority view.

The prototype did not carry that content to its own plate. Four defects were found in the session. In the worksheet and on the exported plate every initiative had lost its title, so that the four layers were empty; the proposals the coach made from the conversation were offered again after they had been accepted, and the header of the plate repeated the decision six times and the destination seven times; the four events on the time axis were drawn over each other, and horizon 1 carried the working name from the coach's own example instead of the name given in the session; and the fixed labels and the outputs came in Dutch although the session was in English, because the language fell back to Dutch at every transition between steps. The cause of the first and largest defect was found the next day: the format in which the coach writes its decisions to the worksheet had no field for the layer of an initiative, and the storage discarded what it could not place, so that every initiative lost its place. All four were repaired in the prototype before this report was finished: an initiative now exists only with a title and a layer, and the worksheet shows a table in which title, layer, owner, period and status can be edited directly; what is already on the plate is not proposed again, and a rejected proposal does not return; the plate places the events on the axis and the initiatives per horizon as blocks with status and owner, and takes the names of horizons and layers from the worksheet; and one language per assignment is set in the intake and followed by

every step and every output. The session of the roadmap was imported into the repaired prototype, cleaned and exported; in this public version figure 6.1 shows the proposal in simplified form rather than that export. The defects are the same kind the dry run of test 11 had found in the journey map, proposals that pass raw or empty content to the worksheet, now in the roadmap, and the defect and its repair show why the precondition in table 6.3 is needed, that what the coach says it has recorded is written to the worksheet reliably; whether the rebuilt route holds is still to be confirmed in use. In this session, a formative test by the designer rather than an evaluation, the coaching itself needed no repair beyond one point also found with C-B: numbers that the coach counts or compares with a guideline are now computed rather than estimated by the model.

ROADMAP · SUPPLY VALUE · THESIS · DESIGN ROADMAPPING

PUBLIC VERSION · SIMPLIFIED

IMPLEMENTATION ROADMAP · A PROPOSAL BY THE RESEARCHER

A limited pilot, then a decision on wider use

Drawn up by the researcher with the Methode-coach, in a Design Roadmapping session on the implementation as the assignment (section 6.4).

STATUS

Proposal. Management had not adopted this roadmap at the time of submission.

1 · PILOT

Limited pilot

- Three or four consultants use the coach on their own current assignment.
- Each with a senior who reviews the first deliverables.
- The first assignment with the coach is planned as a half day with a colleague who knows the tool.

WHO Three or four consultants, each with a senior reviewer

2 · EVALUATION

First measurement

- The test protocol of section 6.1, with the changes of table 6.1.
- It assesses the quality of the work, the time required and the conditions for using client data.
- Scored as the protocol prescribes, blind and with a second coder.

WHO The researcher and a senior consultant not involved in the project

3 · DECISION

Decision on wider use

- The cost per assignment is known and acceptable.
- The value is visible in at least one deliverable that a senior judged at least as usable as work without the coach.
- The rules layer stopped every submission that should have been stopped.
- The framework is assessed by management in the same decision.

WHO Management

THE PROPOSAL RESTS ON TWO ASSUMPTIONS

- Management makes the time for the pilot available.
- The coach builds capability rather than replacing it; this is untested, and the first measurement has to show it.

BEFORE CLIENT DATA IS USED: THE PRECONDITIONS OF TABLE 6.3

- Model calls through the firm's approved environment
- Separate accounts per consultant
- What the coach records reaches the worksheet
- Forbidden names stay out
- The client's assumption kept apart
- An AI agreement with the client of every pilot assignment

Simplified for the public version of this thesis.

Figure 6.1

Figure 6.1 The implementation roadmap in simplified form: a proposal by the researcher, drawn up with the Methode-coach, for a limited pilot and a decision on wider use after its evaluation, with the preconditions for client data of table 6.3. Management had not adopted this roadmap at the time of submission. A3 landscape page.

7

CHAPTER 7

Conclusion, discussion and reflection



The answer to the research question, what it means beyond Supply Value, the limitations, and a personal reflection.

IN THIS CHAPTER 7.1 Answer to the research question · 7.2 Discussion · 7.3 Limitations · 7.4 Recommendations and further research · 7.5 Personal reflection

7.1 Answer to the research question

The sharpened research question of section 2.6 asks how Supply Value can strengthen the innovation capability of its consultants, so that they can direct AI agents at client problems and create more value for clients in advisory and implementation work within regulated client environments, while developing rather than replacing their own professional skills. The answer this thesis gives is a design in three connected parts: a description of the capability, the framework The Empowered Consultant (section 5.2); support with which consultants build and apply it on their own assignments, the Methode-coach (sections 5.3 and 5.4); and an implementation that gives both a place in how Supply Value already works, the roadmap (section 6.4). This design is a proposal supported by interviews and a prototype tested on fictional cases and in three consultant sessions, two of them partial; it was not shown to build capability. The five sub-questions are answered first, SQ1, SQ3 and SQ5 as answered in chapter 2 and confirmed here, SQ2 as completed in chapters 4 and 5 and tested in part, and SQ4 as answered from the literature; the answer to the main question follows from them, together with what this thesis can and cannot claim for it.

SQ1. How consultants use AI, and what distinguishes the ones who use it extensively. Most use at Supply Value is generative AI for routine tasks in the consultant's own work, as a reference and for drafting; a small group also uses it for the approach and the analysis of an assignment and has built products for clients. What distinguishes that group is not the tools, which are the same for everyone, but a framed question, iteration with the model instead of acceptance of its first answer, and a concrete task on their own work in which AI proved useful, with the time to find one (themes 1 to 3; clusters 2 and 5). One failed experiment can settle a consultant's attitude for months, and learning time that competes with the work a consultant is judged on loses (section 3.2.2). The co-creation session showed the same split in a group of five (section 4.1). The evaluation adds one counter-example: of the two consultants who tested the coach, C-A needed far more time than the guide assumed, while C-B, the more frequent AI user of the two, asked the coach to do the work wherever the design allowed it, and C-C completed the method with the first attempt at the problem statement written by another AI assistant, which the gate accepted (section 6.2); frequent use and iteration do not always go together, and a rule that asks for an own version cannot tell whose version it is.

SQ2. How consultants approach client problems, and where an AI agent adds value. Consultants at Supply Value approach client problems from experience: framing the problem is the demanding part, there is no shared structured approach a junior can follow, and what is learned on an assignment is lost between assignments (themes 4 and 5; cluster 4; figure 3.3). Where an agent adds value in that process was indicated by the

interviews, informed by the consultants of the co-creation session and decided by the researcher: in the phases where AI is little used today and where the problem of this thesis sits, context building, framing the problem, choosing an approach and handing over, and not in the decision itself (section 5.1.3). Both co-creation teams designed the same kind of support, a sparring partner that asks first and delivers drafts, and what makes such support different from the chat assistants consultants already have is what sits behind the conversation: methods, Supply Value's practice and the rules of the client (section 4.1.7). The formative tests and the three evaluation sessions support that answer for the intake, the problem statement, the plan and, in one session, the whole method: the coach's questions were sharp where a senior's would be, the problem statements stayed in the consultants' own words, the plan's choices were the consultant's where the consultant made them, and the coach refused to fill a map without evidence (sections 5.7 and 6.2). Whether the coach adds value in a method completed by a consultant was not shown: the one session that reached the method ended in its fifth step.

SQ3. Which limits the client environment sets, and how consultants deal with them. The outer frame is set by the GDPR (Regulation (EU) 2016/679) and the EU AI Act (Regulation (EU) 2024/1689), in particular the obligation of Article 4 to support AI literacy; within it the rules differ per client and per project, depend on who leads the client organisation and range from encouragement to a ban, and some clients prescribe their own model (section 2.4; themes 7 and 8; cluster 3). Consultants deal with this mostly by working within the rules, in the approved environment, on tasks without client data or on the client's own model, and by carrying the cost of complying; one participant names uncertainty about what is permitted as what stops them. The design answers this with rules agreed per client and project in the intake and applied at the moment of use (DR9; G2); in the evaluation the coach recorded the agreements and, during the work, replaced a forbidden name before it was sent; in the intake, before the rules were recorded, a name and an age reached the model (section 6.2). This shows that a rule can be applied at the moment of use on a fictional case, not that it turns non-use into use.

SQ4. How innovation capability is built, and how professionals learn to work with AI without losing their skills. From the literature: innovation capability in a consultancy develops through structured practice on real problems rather than through training alone, through support that is withdrawn as competence grows and through exchange among practitioners (section 2.3.4), and the literature suggests, without showing it causally, that professionals who iterate on a problem they have framed do better than those who delegate, and raises the concern that heavy reliance slows skill development (Trend 3; cluster 5). The design turns this into three principles, the consultant thinks first, the AI drafts and the human decides, and start from the assignment with methods behind it (section 5.1.3), and into a coach whose support fades per method while its guardrails never do (section 5.4.6). Whether these mechanisms work for consultants at Supply Value was to be answered by the evaluation and was not: one session did not reach the method and the other ended in its fifth step, so neither less steering within a session nor a task done better afterwards was observed (section 6.2). The answer to SQ4 therefore rests on the literature and on the design, not on evidence from consultants, and this is the main limitation of the thesis (section 7.3).

SQ5. What clients continue to need from a human consultant. Listening, translating between operations and technology, connecting parties and framing the question remain human, and verification of AI output is non-negotiable for every participant who spoke about it (themes 9 and 10; cluster 1). For the capability this means that it is defined by value at the client and that judging AI output is part of it as a skill, not only as a rule (DR3, DR4); in the framework these are the dimensions D3, judgement and client craft, and D4, responsible AI and client rules, and the guardrails G1 and G3. The co-creation session added what the interviews had not named: a proposal has to fit what the decision maker can decide (DR14), which a language model does not know and the consultant has to give it (section 5.4.9). The evaluation added one observation on the other side of this answer: the point at which the coach failed, treating a client's assumption as established in the plan, is exactly the point

at which a human consultant's judgement is needed, and it shows why G1, a consultant validates every output, cannot fade.

The main question. Supply Value can strengthen the innovation capability of its consultants by describing that capability in terms of the work on client problems rather than of tools, with framing and judgement as the entry conditions of every level; by giving consultants support on their own assignments that makes them frame the problem before anything is produced, applies the client's rules at the moment of use and fades as they progress; and by making time, a place and a conversation for it in how Supply Value already works, with the management decisions that this requires. Directing AI agents is approached from its entry condition: the support is itself an agentic workflow that the consultant directs within a method, framing and judgement are made the condition of every level, and directing agents across a client problem is the level the framework builds towards, not something the evaluation could observe. That is the answer of this thesis, and its status is as follows. The problem analysis rests on eight interviews, coded and clustered, and on two journey maps, within one firm. The direction was informed by consultants and chosen by the researcher, and the design traces every decision to a requirement and a finding. The built parts of the prototype do what section 5.3 says on fictional cases, with the defects of table 6.1; the parts marked designed in table 6.2 are not built. The claim that the support builds capability rather than replacing it is a design argument supported by the literature and by the behaviour of the coach in the sessions, not a result: the evaluation designed to test it produced two partial sessions and one complete session, none with the measurements the test was designed for. The design therefore stands as a proposal with an evaluation that Supply Value can run in the pilot of its roadmap, and section 7.4 says how.

7.2 Discussion

For Supply Value

The research started from a management view in which AI is an opportunity for efficiency and a service to offer. The interviews moved the problem from the tools to the capability, and the design that followed puts the consultant's own thinking before anything is produced. The senior who used the coach spent an hour and a half on the intake, the problem and the plan, and was positively surprised and, under pressure of work, irritated by the coach in the same session, in their own words; the consultant who used it next went through those steps in eleven minutes and then spent an hour on five steps of the method, asking the coach to do the work wherever it would (section 6.2). Both sessions took more time than the guide planned, for reasons the test cannot separate: a tool seen for the first time, a time plan that was the researcher's estimate from the dry runs, and the thinking the design asks for before anything is produced. What the design costs or saves over an assignment is therefore not established in either direction; the expectation that a consultant becomes faster once the tool is familiar is a hypothesis for the decision after the pilot, not a result. What is established is that the time a consultant spends with the coach competes with the work the consultant is judged on, which cluster 2 identified as the reason the majority waits, and that neither consultant finished the test within the ninety minutes. For Supply Value the implication is that the capability cannot be built in the margins of client work. Making time for the pilot (section 6.4) is therefore not one decision among several but the condition for all the others: what the design needs most from management is a planned moment, and the pilot is that moment.

The second implication for Supply Value concerns where the value of the design showed first. The intake, the problem statement and the plan, the phases of the assignment process that no tool supported (figure 3.3), are where the coach did what a senior does when a senior has time: it asked about the days and the stakeholders, noticed a date that did not fit and kept the client's explanation apart from the facts. For a firm whose seniors have no time to give that coaching (theme 5) this is where a first assignment with the coach pays back, before

any method is used, and it is the task on the consultant's own assignment that DR10 asked for. The method steps, where the formative tests found the coaching equally good, remain to be shown with consultants.

For consultancies that face the same shift

The problem is not specific to Supply Value. The split between a small group that leads and a majority that waits, the uncertainty about what a client permits and the loss of knowledge between assignments may be recognised in other consultancies that embed their people at clients, which this thesis does not test, and the structure of the design is not specific either: dimensions defined in terms of client work, entry conditions on framing, rules agreed per client and applied at the moment of use, and guardrails that do not fade. What is specific is the content behind the conversation: the methods a firm uses and its variants of them, the practice its seniors carry and the rules its clients set. Another firm would have to supply that content and the owner who maintains it, and would have to route the model through the environment it has approved. Whether the design transfers is therefore a question about content and governance, not about structure, and section 7.4 names it as further research.

For the design of AI support that builds capability

Three findings of this thesis speak to the literature of chapter 2. First, the frontier. Dell'Acqua et al. (2026) found that consultants with AI were more often wrong on a task outside the model's frontier, which this thesis attributes to the model answering with the same confidence on both sides; the design answers this with checkpoints that require the consultant to act before an output moves on, following the finding that designs which make the person form their own judgement reduce over-reliance (Buçinca et al., 2021). The evaluation gave this a sharper form: the model's own guard failed exactly on a question of judgement, whether a client's assumption counts as established, and it failed inconsistently, in one step and not in the other, in one run and not in the next. A rule that lives in instructions to the model is a request, as section 5.4.1 argued for data rules; the evaluation suggests that this holds for rules about reasoning as well, which is why the final design makes the client's assumption a typed fact that the plan step has to act on, a structure rather than an instruction.

Second, iteration against delegation. Massenkoff et al. (2026) report that experienced users iterate with the model and delegate less, and that iteration goes with a higher success rate; Agarwal et al. (2026) describe apprentices, delegators and amplifiers among early-career professionals, with the usual measures of capability not explaining which a person becomes. The design takes a position in that literature: it does not wait for a consultant to become an iterator but builds the iteration into the tool, by refusing to produce before the consultant has tried. The cost of that position showed in this thesis in one consultant's irritation under pressure of work, and possibly in time, although the sessions cannot separate the time the thinking takes from the time a new tool takes; the question it leaves open is where the balance lies: how much of the thinking a tool can hand back before the professional under pressure of work stops using it. The operating rule that distinguishes thinking work from craft knowledge (section 5.3.3) is the first answer; the second dry run showed that quick-reply buttons can undo it by offering the answer the consultant was meant to find (section 6.2), which suggests that the balance is set not only in the coach's rules but in every element of the interface around it. The third session added the document layer to that list: the coaching held the map to one interview and a hypothesis status, and the report the tool generated from the same material called the outcome unambiguous and explained a figure causally, which no rule in the conversation prevented.

Third, who coaches whom. Kellogg et al. (2025) found that juniors cannot be expected to coach seniors on the risks of generative AI, because novices lack the understanding of the technology and of the work to recognise those risks. The design does not make the coach a senior, and the evaluation shows why that is right: the coach's craft knowledge was good, its questions were sharp, and its judgement on the one question that mattered was unreliable. In the framework this is the division between what the coach carries and what the implementation carries (table 5.4): method knowledge and the structure of the work in the coach, the review of what reaches the

client by a senior in the pilot, and the judgement in the consultant. That division is also what separates the coach from the co-ideation agent of P. Wang et al. (2025), whose question-asking dialogue rules the coach shares: that agent supports a designer's ideation in one session and was shown to raise the novelty and quality of ideas against a control group, while the coach supports a professional's work on a client problem over weeks and has to be right about what may reach the client. The first can be evaluated in a session; the second, this thesis found, cannot.

A last point concerns what the design is. Gartner (2025b) warns against agents that are rebranded assistants. The coach is an agentic workflow in the terms of Schluntz and Zhang (2024), with the agent properties of section 2.1.1, a role, a goal, tools and a sequence of steps, but with the steps predefined (section 5.3.1), and it sits deliberately at the low end of autonomy, because the steps in which framing and judgement are exercised are the steps the capability is about. The evaluation suggests that this is the right level for a tool meant to build capability, and that the autonomy that would make the coach faster is autonomy over exactly the steps it should not take.

7.3 Limitations

The limitations are stated in the order of the research, and the most important comes last.

Discover and Define. The problem analysis rests on one firm and eight interviews where ten were planned, because of the time available and the availability of the consultants in that period; the eight cover management and consultants of different seniority, and no clients, external experts or consultants at medior level, so the themes are a map of Supply Value from the inside, and new codes were still appearing in the last interviews, so that map is a first one rather than a complete one, which bounds the answers to SQ1 and SQ3. The difference between frontrunners and others in the answer to SQ1 may partly reflect role rather than method. SQ5 is answered from the provider side: no client was interviewed, so what clients need is what management and consultants of Supply Value say clients need. The researcher worked at Supply Value during the project and knew the participants as colleagues, and a supervisory relationship within the sample may have influenced what participants felt comfortable sharing; anonymisation does not remove that limitation. The condition that the client environment restricts AI rests on the interviews and on one audit of central government (section 2.4), not on evidence from the clients themselves. The assignment process of figure 3.3, which structures the coach, was inferred from the interviews and not validated with consultants as a description of their process (section 3.2.1), so the phases the coach models are the researcher's reading. The journey maps rest on one interview each, were not supported by observation and were not validated with the two participants, which remains the first step to take for both (section 3.2.2); the career details that would identify the two participants to colleagues were removed from the maps and from Appendix F before publication. The coding of the interviews was proposed by a language model and checked fragment by fragment by the researcher; there was no independent second human coder, so no inter-coder agreement is reported (section 2.5.1). The six-phase assignment process of section 3.2.1 was derived by the researcher from the interviews and not checked with consultants; a check with two or three consultants belongs in the pilot.

Develop. The co-creation session had five participants, four of them juniors; it was converted to an online format on the day, which made the two teams unequal in size and location; the problem was chosen on the researcher's proposal and its owner already had a solution direction; four of the five participants knew the researcher from their interview; team A filled the canvas briefly, so its concept is taken from the pitch; and the session was framed as designing an AI tool, solution language the participants took over, within a role list, a canvas and examples the researcher supplied. The ideation, the concepts and the choice between them, including the scores of the Harris profile, were made by the researcher alone from its material (sections 4.1 to 4.3); no consultant or

second rater took part in the concept choice. The formative tests were done by an AI tester of the same model family as the coach, with benchmarks that the same tester wrote from literature, on one fictional client reused across seven tests, in the learning mode only, so that the fading between modes is built but not tested; the tenth test was done by the researcher (section 5.7). An AI tester does not get tired, impatient or embarrassed and may be more willing to follow a coaching question than a person is, the planted errors are the errors the benchmark predicted rather than the errors a consultant would surprise the coach with, and the researcher designed the cases, directed the builder and reviewed the tests, so that the tests checked the tool against the researcher's reading of what a good method is. The tests therefore show that the prototype does what it is designed to do on cases the researcher designed, not how a consultant experiences it, and the eight deepened methods are the methods the interviews named most, not a representative sample of the eighty-eight; the other eighty sit at a catalogue layer written with a language model (section 5.4.4), so the coach's guidance on those methods rests on that layer and not on a deepened, checked source.

The researcher's double role. The researcher designed the research, the cases and the evaluation, built the prototype with an AI coding assistant, reviewed the tests, and spent months at Supply Value as a graduate student, working with the people who took part in the research. Rubrics for every score and the coding of the export turn by turn were the measures against this; the blind scoring by a senior and the second coder, which were the strongest of them, did not take place because the material for them was not produced.

Deliver. The evaluation is the main limitation of this thesis, and it is reported in full in section 6.2. The framework was not evaluated with management within the period of this project: it is proposed on the evidence of chapters 2 to 4 and on what the consultant sessions showed about the entry condition of its first level, and its assessment by management is left to the decision after the pilot, so success indicator 3 and the three open decisions of the framework are not established. The coach was used by three consultants: a senior who completed the intake, the problem and the plan and did not reach the method, a consultant who reached the fifth of the six steps of the method in the ninety minutes, and a third who completed the method and the closing conversation but did neither task A nor task Z and used another AI assistant alongside the coach; none wrote task Z, one map was finished but not scored and no follow-up conversation took place, so success indicators 1 and 2 are not assessed; the unmoderated, single-sitting design of the test, not only the number of participants, is why they could not be. What the evaluation shows about the coach rests on three sessions, two dry runs by an AI tester and the researcher, and an email, on two fictional cases. It cannot show whether consultants learn with the coach, whether the support fades between sessions, whether work made with it remains their own, or how seniority and AI experience make a difference beyond the contrast between a senior who asked the coach for nothing, a consultant who asked it for everything and a third who brought a second AI to the session; it can show how the coach behaved in the phases before the method and in the first steps of one method, and where it failed. The tool also contained known defects at the time of the evaluation (table I.1), so that some of the consultants' time went to the tool rather than to the method. The roadmap of section 6.4 was drawn up by the researcher with the coach, without managers or consultants in the session, so it is a proposal that management has to confirm or change.

Ethics. No application was submitted to the Human Research Ethics Committee, which the graduation manual requires before research with people; this is a shortcoming of the project. Interview consent used the TU Delft form and an oral statement (section 2.5.1); consent for the co-creation session (Appendix G.2) and for the evaluation (Appendix M, section 6.1) was oral; data were pseudonymised and the cases were fictional. Full interview transcripts containing names were submitted to an external AI service for coding assistance (section 2.5.1), under account settings in which the data is not used for training AI models; the service was instructed to omit names from its output, and this processing is reported here as part of this limitation. Supply Value reads the

report after it has been submitted; at submission no written confirmation of the company on publication had been received.

What a design thesis can claim. The project brief (Appendix A) planned an AI diary study among consultants and a validation of the result with consultants and management. The diary study was replaced by the interviews of section 2.5; the validation with management was not held within the period of the project and is left to the decision after the pilot (section 6.4). The capability this thesis is about develops over months of client work. A thesis of twenty weeks can design for it, test whether the design does what it says on fictional cases and, at best, observe a first signal in a session or two. It cannot show that the capability developed, and this thesis does not claim that it did.

7.4 Recommendations and further research

For Supply Value

The recommendations follow the roadmap of section 6.4 and are ordered by what blocks what.

First, run a limited pilot, and decide on wider use after its evaluation. Three or four consultants use the coach on their own assignments, to assess the quality of the work, the time required and the conditions for using client data (section 6.4). The decision after the pilot is also the assessment of the framework that this thesis did not obtain: whether it describes what Supply Value needs from its consultants and what is missing, and whether each manager can place two consultants of their own team in the matrix, which should take minutes rather than an hour and gives the first evidence on indicator 3.

Second, complete the evaluation as the first measurement of the pilot. The protocol of section 6.1, with the changes of table 6.1 built and the guide corrected (what a vraagstuk is, where task A is written, the days of the case), can be run with three or four consultants as the first assignments with the coach, each planned as a half day with a colleague who knows the tool, and scored as the protocol prescribes, blind and with a second coder. Its outcome on indicators 1 and 2 is what the decision after the pilot should rest on, next to the cost, the value and the rules that Gartner's three causes name.

Third, do not use the coach with client data before the preconditions of table 6.3 are met: the route through the firm's approved environment, separate accounts, the route from the conversation to the worksheet confirmed in use, the name filter confirmed in use, and the consistency of the guard on the client's assumption between the problem step and the plan step.

Fourth, begin where the value showed: the intake, the problem statement and the plan of a new assignment, which is the task every consultant has to do anyway and the one the coach did well in the evaluation. A consultant who uses the coach only for that has done the step the capability is about, and has a reason to come back for the method.

The recommendations that lie outside the scope of this thesis are listed in Appendix L.

For research

Four questions follow from this thesis. The first is the one its evaluation could not answer: whether support that fades over sessions builds capability that remains when the support is gone, which needs a design with at least two observations in time, a task without the coach at the end, and a period of months rather than a session. The second is the balance that section 7.2 named: how much thinking a tool can hand back to a professional under pressure of work before the professional stops using it, and whether the distinction between thinking work and craft knowledge, or the removal of quick answers from the interface, moves that balance. The third is the comparison between an agent and a senior as a source of judgement, which the evaluation touched on at one

point, the client's assumption, and which decides how much of a senior's review the implementation can do without. The fourth is transfer: whether the structure of the framework and the coach holds in another consultancy with its own methods, practice and clients, which is a question about content and governance rather than about the design, and which the preconditions of section 7.2 make testable.

7.5 Personal reflection

I started this project to help consultants add more value with AI, not to save time, and I end it with a design that asks them to think before they let AI produce. That is the thing I learned most from. The interviews taught me that the people who create value with AI are not the ones with the best tools but the ones who can say what the problem is before they ask a model anything, and once I saw that, I could not design a tool that writes the problem statement for them, however much faster that would have been and however often I wanted it myself when I was behind.

Building the prototype with an AI coding assistant made me the first user of my own argument. I directed the building with written specifications, and the assistant built fluently on both sides of its frontier: the coaching logic was good from the first test, and the plumbing around it, the saving, the gates, the hand-over from the conversation to the worksheet, failed in ways the assistant reported as fixed. I learned to read "built" as a claim to be tested, and I built too much: eighty-eight methods, eight of them deep, where fewer methods working completely would have served the consultants and the evaluation better. The same lesson came back in the writing. A language model turned "the firm" into Supply Value in a sentence about a global consultancy's headcount, and the only reason it did not reach this report is that I read every sentence against its source. The check that the design asks of consultants is the check I had to do myself, and it is not a formality.

The evaluation is where the project fell short of its plan, and the reason is one that the research itself had found: time at Supply Value is the scarcest resource. I asked three consultants for ninety minutes next to client work, in the weeks before this report was due, and I received two partial sessions in time and the third, complete, only after the deadline. I should have planned the evaluation as meetings in people's agendas months earlier, and I should have raised my planning and my choices with my supervisors earlier and more often than I did; my supervisors told me so, and the second half of the project confirmed it. What I would do differently is also what the roadmap now asks of Supply Value: make the moment, do not wait for it.

The last thing I made for this thesis was the roadmap, with the coach, in one evening in the last week of the project. It refused my first problem statement because it named my own solution as the current situation, which is the gate doing to me what it had done to the two consultants, and I had to write it again in terms of consultants and their work with AI. It then asked me for owners, assumptions and a rule for dissent that my text did not have, and lost the titles of the initiatives on its own plate, a defect that was repaired the next day. That evening summed up the project: the coaching did what it was designed to do, the prototype still had defects, and the roadmap came out better than the text I had written on my own.

What stays with me is the difference between a tool that works and a design that changes how people work. The coach works, on fictional cases, in the hands of an AI tester, of a senior who was positively surprised by it and, under pressure of work, irritated by it in the same session, and of a consultant who asked it to do the work and was refused wherever the design had foreseen the request. Whether it changes how consultants at Supply Value work depends on decisions that are not mine to take, and the most useful thing this thesis may do is to put those decisions, with their arguments, on the table of the people who can take them.

References

Agarwal, A., Barua, A., Puvvada, A., Song, F., & Wen, W. (2026, July 23). Research: Why some junior employees work well with AI and others don't. Harvard Business Review. https://hbr.org/2026/07/research-why-some-junior-employees-work-well-with-ai-and-others-dont

Algemene Rekenkamer. (2024). Focus op AI bij de rijksoverheid. Algemene Rekenkamer. https://www.rekenkamer.nl/documenten/2024/10/16/focus-op-ai-bij-de-rijksoverheid

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. Personnel Psychology, 50(2), 341 to 358. https://doi.org/10.1111/j.1744-6570.1997.tb00911.x

Anand, N., Gardner, H. K., & Morris, T. (2007). Knowledge-based innovation: Emergence and embedding of new practice areas in management consulting firms. Academy of Management Journal, 50(2), 406 to 428. https://doi.org/10.5465/amj.2007.24634457

Baán, J. L. (2021). From experimenting with AI to a new way of working [Master thesis, Delft University of Technology]. TU Delft Repository. https://resolver.tudelft.nl/uuid:8ee6df52-d14a-4da2-8370-2e66793fb446

Brown, T. (2009). Change by design: How design thinking transforms organizations and inspires innovation. HarperBusiness.

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), Article 188. https://doi.org/10.1145/3449287

Carlzon, J. (1987). Moments of truth. Ballinger.

Challapally, A., Pease, C., Raskar, R., & Chari, P. (2025, July). The GenAI divide: State of AI in business 2025 [Preliminary findings]. MIT NANDA, Massachusetts Institute of Technology. https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf

Clardy, A. (2018). 70-20-10 and the dominance of informal learning: A fact in search of evidence. Human Resource Development Review, 17(2), 153 to 178. https://doi.org/10.1177/1534484318759399

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 453 to 494). Lawrence Erlbaum.

Cui, K. Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2026). The effects of generative AI on high-skilled work: Evidence from three field experiments with software developers. Management Science. Advance online publication. https://doi.org/10.1287/mnsc.2025.00535

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319 to 340. https://doi.org/10.2307/249008

Dell'Acqua, F., McFowland, E., III, Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Organization Science, 37(2), 403 to 423. https://doi.org/10.1287/orsc.2025.21838

Den Hertog, P. (2000). Knowledge-intensive business services as co-producers of innovation. International Journal of Innovation Management, 4(4), 491 to 528. https://doi.org/10.1142/S136391960000024X

Design Council. (2007). Eleven lessons: Managing design in eleven global brands. https://www.designcouncil.org.uk/fileadmin/uploads/dc/Documents/ElevenLessons_Design_Council%20(2).pdf

Dobson, S. (2025, November 25). AI errors: Province grapples with Deloitte report marred by false citations. Canadian HR Reporter. https://www.hrreporter.com/focus-areas/automation-ai/ai-errors-province-grapples-with-deloitte-report-marred-by-false-citations/393782

Dreyfus, S. E., & Dreyfus, H. L. (1980). A five-stage model of the mental activities involved in directed skill acquisition (Report No. ORC-80-2). Operations Research Center, University of California, Berkeley. https://apps.dtic.mil/sti/citations/ADA084551

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. Science, 384(6702), 1306 to 1308. https://doi.org/10.1126/science.adj0998

Endo, T. (2025, February 23). How AI is redefining strategy consulting: Insights from McKinsey, BCG, and Bain. Medium. https://medium.com/@takafumi.endo/how-ai-is-redefining-strategy-consulting-insights-from-mckinsey-bcg-and-bain-69d6d82f1bab

European Commission. (2025). AI literacy: Questions & answers. European Commission. Retrieved October 5, 2026, from https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. Business & Information Systems Engineering, 66(1), 111 to 126. https://doi.org/10.1007/s12599-023-00834-7

Gartner. (2025a, May 13). Gartner identifies the top trends shaping the future of cloud [Press release]. Gartner. https://www.gartner.com/en/newsroom/press-releases/2025-05-13-gartner-identifies-top-trends-shaping-the-future-of-cloud

Gartner. (2025b, June 25). Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027 [Press release]. Gartner. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

Gartner. (2025c, October 21). Gartner unveils top predictions for IT organizations and users in 2026 and beyond [Press release]. Gartner. https://www.gartner.com/en/newsroom/press-releases/2025-10-21-gartner-unveils-top-predictions-for-it-organizations-and-users-in-2026-and-beyond

Gibbons, S. (2018, December 9). Journey mapping 101. Nielsen Norman Group. https://www.nngroup.com/articles/journey-mapping-101/

Glückler, J., & Armbrüster, T. (2003). Bridging uncertainty in management consulting: The mechanisms of trust and networked reputation. Organization Studies, 24(2), 269 to 297. https://doi.org/10.1177/0170840603242004

Halvorsrud, R., Kvale, K., & Følstad, A. (2016). Improving service quality through customer journey analysis. Journal of Service Theory and Practice, 26(6), 840 to 867. https://doi.org/10.1108/JSTP-05-2015-0111

Hoßbach, C., & Isaksen, S. G. (2025). AI-augmented approaches to creative problem-solving: A metacognitive perspective. Creativity and Innovation Management, 34(4), 854 to 869. https://doi.org/10.1111/caim.70003

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. Evaluation Review, 4(1), 93 to 106. https://doi.org/10.1177/0193841X8000400105

IDEO.org. (2015). The field guide to human-centered design (1st ed.). IDEO.org.

Kalbach, J. (2020). Mapping experiences: A complete guide to customer alignment through journeys, blueprints, and diagrams (2nd ed.). O'Reilly.

Kaplan, K. (2016a, July 31). Customer journey maps: When and how to create them. Nielsen Norman Group. https://www.nngroup.com/articles/customer-journey-mapping/

Kaplan, K. (2016b, October 16). Journey mapping in real life: A survey of UX practitioners. Nielsen Norman Group. https://www.nngroup.com/articles/journey-mapping-ux-practitioners/

Kellogg, K. C., Lifshitz, H., Randazzo, S., Mollick, E., Dell'Acqua, F., McFowland, E., III, Candelon, F., & Lakhani, K. R. (2025). Novice risk work: How juniors coaching seniors on emerging technologies such as generative AI can lead to learning failures. Information and Organization, 35(1), Article 100559. https://doi.org/10.1016/j.infoandorg.2025.100559

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). Evaluating training programs: The four levels (3rd ed.). Berrett-Koehler.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. Organization Science, 3(3), 383 to 397. https://doi.org/10.1287/orsc.3.3.383

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice Hall.

Laninga, B. (2019). Supporting company decision-making to transition from project to product platform: A case study to determine minimum product platform functionalities [Master thesis, Management of Technology, Delft University of Technology]. TU Delft Repository. https://repository.tudelft.nl/record/uuid:3b8f70b3-4339-47cb-ada7-2109dd6b0c25

Lawson, B., & Samson, D. (2001). Developing innovation capability in organisations: A dynamic capabilities approach. International Journal of Innovation Management, 5(3), 377 to 400. https://doi.org/10.1142/S1363919601000427

Lemus-Aguilar, I., Hidalgo, A., & Cagliano, R. (2015). Innovation in consulting firms: What are the foundations? IAMOT 2015 Conference Proceedings, 951 to 966. https://oa.upm.es/42701/

Liedtka, J. (2015). Perspective: Linking design thinking with innovation outcomes through cognitive bias reduction. Journal of Product Innovation Management, 32(6), 925 to 938. https://doi.org/10.1111/jpim.12163

Liedtka, J. (2018). Why design thinking works. Harvard Business Review, 96(5), 72 to 79. https://hbr.org/2018/09/why-design-thinking-works

Liedtka, J., Salzman, R., & Azer, D. (2017). Design thinking for the greater good: Innovation in the social sector. Columbia University Press.

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1 to 16). ACM. https://doi.org/10.1145/3313831.3376727

Massenkoff, M., Lyubich, E., McCrory, P., Appel, R., & Heller, R. (2026, March 24). Anthropic Economic Index report: Learning curves. Anthropic. https://www.anthropic.com/research/economic-index-march-2026-report

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. AI Magazine, 27(4), 12. https://doi.org/10.1609/aimag.v27i4.1904 (Original work published 1955)

McKinsey & Company. (2024, May 30). The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

McKinsey & Company. (2025, June). The future of work is agentic [Podcast transcript]. McKinsey Talks Talent. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-future-of-work-is-agentic

McKinsey & Company. (n.d.). Rewiring the way McKinsey works with Lilli, our generative AI platform. Retrieved October 5, 2026, from https://www.mckinsey.com/capabilities/tech-and-ai/how-we-help-clients/rewiring-the-way-mckinsey-works-with-lilli

McKinsey Global Institute. (2023). The economic potential of generative AI: The next productivity frontier. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

Merritt, M. (2025, May 29). McKinsey's headcount is down more than 10% in the past 18 months. That's the biggest loss in the consulting firm's history. Fortune. https://fortune.com/2025/05/29/mckinsey-headcount-down-more-than-10-percent

Micheli, P., Wilner, S. J. S., Bhatti, S. H., Mura, M., & Beverland, M. B. (2019). Doing design thinking: Conceptual review, synthesis, and research agenda. Journal of Product Innovation Management, 36(2), 124 to 148. https://doi.org/10.1111/jpim.12466

Muller, E., & Doloreux, D. (2009). What we should know about knowledge-intensive business services. Technology in Society, 31(1), 64 to 72. https://doi.org/10.1016/j.techsoc.2008.10.001

Nakonezny, P. A., & Rodgers, J. L. (2005). An empirical evaluation of the retrospective pretest: Are there advantages to looking back? Journal of Modern Applied Statistical Methods, 4(1), 240 to 250. https://doi.org/10.22237/jmasm/1114906920

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford University Press.

Osborn, A. F. (1953). Applied imagination: Principles and procedures of creative thinking. Scribner.

Pahl, G., & Beitz, W. (1996). Engineering design: A systematic approach (2nd ed.; K. Wallace, Ed.; K. Wallace, L. Blessing, & F. Bauert, Trans.). Springer. https://doi.org/10.1007/978-1-4471-3581-4

Panigrahi, A. (2024). Scaling sustainable business models through strategic design [Master thesis, Delft University of Technology]. TU Delft Repository. https://repository.tudelft.nl/record/uuid:b2a99f08-2ddb-471c-a4dd-ac97f9ecac01

Paoli, N. (2025, October 7). Deloitte was caught using AI in $290,000 report to help the Australian government crack down on welfare after a researcher flagged hallucinations. Fortune. https://fortune.com/2025/10/07/deloitte-ai-australia-government-report-hallucinations-technology-290000-refund

PwC. (2026, January 29). No more pyramids: Rethinking your workforce for the agentic AI era. PricewaterhouseCoopers. https://www.pwc.com/us/en/tech-effect/ai-analytics/agentic-ai-workforce-redesign.html

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L 119, 4 May 2016. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L, 12 July 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. Policy Sciences, 4(2), 155 to 169. https://doi.org/10.1007/BF01405730

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.

Roozenburg, N. F. M., & Eekels, J. (1995). Product design: Fundamentals and methods. Wiley.

Rosenbaum, M. S., Losada Otalora, M., & Contreras Ramírez, G. (2017). How to create a realistic customer journey map. Business Horizons, 60(1), 143 to 150. https://doi.org/10.1016/j.bushor.2016.09.010

Russell, S., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.). Pearson.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68 to 78. https://doi.org/10.1037/0003-066X.55.1.68

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. CoDesign, 4(1), 5 to 18. https://doi.org/10.1080/15710880701875068

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges. Information Fusion, 126, Article 103599. https://doi.org/10.1016/j.inffus.2025.103599

Schluntz, E., & Zhang, B. (2024, December 19). Building effective agents. Anthropic. https://www.anthropic.com/research/building-effective-agents

Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, capital, credit, interest and the business cycle (R. Opie, Trans.). Harvard University Press. (Original work published 1911)

Simon, H. A. (1973). The structure of ill structured problems. Artificial Intelligence, 4(3 to 4), 181 to 201. https://doi.org/10.1016/0004-3702(73)90011-8

Simonse, L. (2018). Design roadmapping: Guidebook for future foresight techniques. BIS Publishers.

Sreenivasan, A., & Suresh, M. (2024). Design thinking and artificial intelligence: A systematic literature review exploring synergies. International Journal of Innovation Studies, 8(3), 297 to 312. https://doi.org/10.1016/j.ijis.2024.05.001

Stanford Institute for Human-Centered Artificial Intelligence. (2026). The 2026 AI Index report. Stanford University. https://hai.stanford.edu/ai-index/2026-ai-index-report

Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018). This is service design doing: Applying service design thinking in the real world. O'Reilly.

Taminiau, Y., Smit, W., & de Lange, A. (2009). Innovation in management consulting firms through informal knowledge sharing. Journal of Knowledge Management, 13(1), 42 to 55. https://doi.org/10.1108/13673270910931152

Tinkoff, D., Van der Veken, L., Chui, M., & Balakrishnan, T. (2026, August 25). The state of AI in 2026: On the road to ROI. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Van Boeijen, A., Daalhuizen, J., & Zijlstra, J. (Eds.). (2020). Delft design guide: Perspectives, models, approaches, methods (Revised ed.). BIS Publishers.

Van de Pol, J., Volman, M., Oort, F., & Beishuizen, J. (2015). The effects of scaffolding in the classroom: Support contingency and student independent working time in relation to student achievement, task effort and appreciation of support. Instructional Science, 43(5), 615 to 641. https://doi.org/10.1007/s11251-015-9351-z

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. European Journal of Information Systems, 25(1), 77 to 89. https://doi.org/10.1057/ejis.2014.36

Verganti, R., Vendraminelli, L., & Iansiti, M. (2020). Innovation and design in the age of artificial intelligence. Journal of Product Innovation Management, 37(3), 212 to 227. https://doi.org/10.1111/jpim.12523

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2024). A survey on large language model based autonomous agents. Frontiers of Computer Science, 18(6), Article 186345. https://doi.org/10.1007/s11704-024-40231-1

Wang, P., Khinvasara, Y., Creijghton, G. J., Scholing, T., Wang, Y., Zhou, Z., Childs, P. R. N., & Yin, Y. (2025). Enhancing designer creativity through human-AI co-ideation: A co-creation framework for design ideation with custom GPT. AI EDAM, 39, Article e22. https://doi.org/10.1017/S0890060425100127

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89 to 100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wood, P. (Ed.). (2002). Consultancy and innovation: The business service revolution in Europe. Routledge.

World Economic Forum. (2025). The future of jobs report 2025. World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2025/

Yang, J., Zyskowski, K., Yonack, N., & Ma, J. (2026, July 29). Research: How AI agents broaden the scope of knowledge work. Harvard Business Review. https://hbr.org/2026/07/research-how-ai-agents-broaden-the-scope-of-knowledge-work

APPENDICES A TO M

Appendices

The material behind the report: the project brief, the interview guides, the co-creation and ideation boards, the prototype screens, the traceability of the requirements and the evaluation materials. The interview analysis is confidential in this public version.

A	Project brief	110
B	Interview guides	116
C	Codebook	119
D	Coded fragments	120
E	From codes to sub-themes to clusters	121
F	Evidence behind the journey maps	122
G	Co-creation session: materials, board and outputs	123
H	Ideation: boards, idea lists, directions and evaluation	137
I	Screens of the prototype	156
J	The eleven interview themes in full	164
K	Traceability from requirements to design decisions	165
L	Recommendations beyond the scope of this thesis	167
M	Materials of the evaluation	169

Appendix A. Project brief

The IDE Graduation Project Brief as approved at the start of the project. The signature and date fields are empty in this copy; the student number and the study progress data were removed before publication.



DESIGN FOR our future

TU Delft

IDE Master Graduation Project

Project team, procedural checks and Personal Project Brief

In this document the agreements made between student and supervisory team about the student's IDE Master Graduation Project are set out. This document may also include involvement of an external client, however does not cover any legal matters student and client (might) agree upon. Next to that, this document facilitates the required procedural checks:

- Student defines the team, what the student is going to do/deliver and how that will come about
- Chair of the supervisory team signs, to formally approve the project's setup / Project brief
- SSC E&SA (Shared Service Centre, Education & Student Affairs) report on the student's registration and study progress
- IDE's Board of Examiners confirms the proposed supervisory team on their eligibility, and whether the student is allowed to start the Graduation Project

STUDENT DATA & MASTER PROGRAMME

Complete all fields and indicate which master(s) you are in

Family name	Dursun	IDE master(s)	IPD ☐ DfI ☐ SPD ✓
Initials	F	2nd non-IDE master	
Given name	Ferhat	Individual programme *(date of approval)*	
Student number	[redacted]	Medisign	☐
		HPM	☐

SUPERVISORY TEAM

Fill in he required information of supervisory team members. If applicable, company mentor is added as 2nd mentor

Chair	Bart Bluemink	dept./section	DOS
mentor	Pan Wang	dept./section	DOS
2nd mentor	Khalid Maliki		
client:	Supply Value		
city:	Zeist	country:	The Netherlands
optional comments			

! Ensure a heterogeneous team. In case you wish to include team members from the same section, explain why.

! Chair should request the IDE Board of Examiners for approval when a non-IDE mentor is proposed. Include CV and motivation letter.

! 2nd mentor only applies when a client is involved.

APPROVAL OF CHAIR on PROJECT PROPOSAL / PROJECT BRIEF -> to be filled in **by the Chair** of the supervisory team

Sign for approval (Chair)

Name _____ Date _____ Signature _____

Project brief, page 1 of 6.

CHECK ON STUDY PROGRESS

To be filled in **by SSC E&SA** (Shared Service Centre, Education & Student Affairs), after approval of the project brief by the chair. The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total _____ EC

Of which, taking conditional requirements into account, can be part of the exam programme _____ EC

	YES	all 1st year master courses passed
	NO	missing 1st year courses

Comments:

Sign for approval (SSC E&SA)

Name _____ Date _____ Signature _____

APPROVAL OF BOARD OF EXAMINERS IDE on SUPERVISORY TEAM -> to be checked and filled in by IDE's Board of Examiners

Does the composition of the Supervisory Team comply with regulations?

YES		Supervisory Team approved
NO		Supervisory Team not approved

Comments:

Based on study progress, students is ...

	ALLOWED to start the graduation project
	NOT allowed to start the graduation project

Comments:

Sign for approval (BoEx)

Name _____ Date _____ Signature _____

DESIGN FOR our future

TUDelft

Personal Project Brief – IDE Master Graduation Project

Name student Ferhat Dursun **Student number** [redacted]

PROJECT TITLE, INTRODUCTION, PROBLEM DEFINITION and ASSIGNMENT

Complete all fields, keep information clear, specific and concise

Project title Empowering Consultants as Innovators: A Framework for Strengthening Innovation Capabilities for AI Adoption in Strategic Consultancy

Please state the title of your graduation project (above). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

Introduction

Describe the context of your project here; What is the domain in which your project takes place? Who are the main stakeholders and what interests are at stake? Describe the opportunities (and limitations) in this domain to better serve the stakeholder interests. (max 250 words)

The domain of strategic consultancy is undergoing a paradigm shift driven by Artificial Intelligence. Clients no longer seek only theoretical strategies but require actionable guidance on how to integrate AI into their daily operations. This project takes place within Supply Value, a strategic consultancy firm specializing in procurement, supply chain, and digital transformation.

The primary stakeholders involved are the consultants, the management of Supply Value, and the clients. Consultants need to evolve from traditional advisors to tech-enabled innovators to remain relevant. On the client side, particularly within the public sector such as municipalities, the challenges are significant. These organizations often operate under strict regulations where AI usage might be restricted or even prohibited. Furthermore, changing established workflows and employee behavior in such rigid environments is notoriously difficult.

Significant opportunities lie in empowering consultants to navigate these complex environments. By enhancing their innovation capacity, consultants can guide these clients not just on the technology itself, but on how to implement it responsibly within legal frameworks and organizational cultures. This project aims to bridge the gap between the availability of AI technology and the practical reality of applying it in highly regulated and change-resistant sectors.

→ *space available for images / figures on next page*

introduction (continued): space for images

image / figure 1

image / figure 2



Problem Definition

What problem do you want to solve in the context described in the introduction, and within the available time frame of 100 working days? (= Master Graduation Project of 30 EC). What opportunities do you see to create added value for the described stakeholders? Substantiate your choice.
(max 200 words)

The central problem is the strategic uncertainty regarding how AI will influence the consultancy profession: whether it will augment advisory roles or replace core tasks. Simultaneously, consultants struggle to guide clients in the public sector (e.g., municipalities) in responsible AI adoption due to strict regulations and organizational rigidity. Currently, consultants lack the structured capabilities to bridge the gap between rapid AI innovation and these restrictive client environments.

This project aims to design a Capability-Building Framework that supports consultants in navigating this transition. The key opportunity lies in developing a design-driven service model that clarifies how AI serves as a complementary tool, ensuring consultants adapt their role without losing professional judgment.

This creates direct added value for consultants by enhancing their innovation capacity and confidence; for Supply Value by transforming ad-hoc initiatives into a scalable, competitive service proposition; and for clients by enabling them to adopt AI safely and strategically despite regulatory limitations. This choice addresses the urgent need for a human-centered approach to digital transformation in high-compliance sectors.

Assignment

This is the most important part of the project brief because it will give a clear direction of what you are heading for. Formulate an assignment to yourself regarding what you expect to deliver as result at the end of your project. (1 sentence) As you graduate as an industrial design engineer, your assignment will start with a verb (Design/Investigate/Validate/Create), and you may use the green text format:

Design a capability-building framework and toolkit to empower consultants in responsibly integrating AI into their advisory practice, enabling them to deliver future-proof strategic advice to Supply Value clients within strict regulatory and organizational constraints.

Then explain your project approach to carrying out your graduation project and what research and design methods you plan to use to generate your design solution (max 150 words)

The project follows a Research-through-Design approach structured by the Double Diamond process. In the Discover phase, I will conduct a literature review on innovation culture, AI skills, and design tools, while simultaneously deploying an "AI Diary" study where consultants log actual usage, missed opportunities, and process inefficiencies to gather rich contextual data. These insights will be synthesized into a Current Co-worker Journey Map to pinpoint pain points, informing the design of the ideal Future Co-worker Journey. To realize this future state, I will facilitate co-creation workshops with consultants to iteratively design a tangible toolkit; involving them in the design process fosters ownership and accelerates adoption. Finally, the toolkit and strategic roadmap will be validated with consultants and management through user testing to ensure real-world effectiveness, resulting in the delivery of a finalized capability-building framework.

Project planning and key moments

*To make visible how you plan to spend your time, you must make a planning for the full project. You are advised to use a Gantt chart format to show the different phases of your project, deliverables you have in mind, meetings and in-between deadlines. Keep in mind that all activities should fit within the given run time of 100 working days. Your planning should include a **kick-off meeting**, **mid-term evaluation meeting**, **green light meeting** and **graduation ceremony**. Please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any (for instance because of holidays or parallel course activities).*

Make sure to attach the full plan to this project brief.
The four key moment dates must be filled in below

Kick off meeting 15 Dec 2025

Mid-term evaluation

Green light meeting

Graduation ceremony

In exceptional cases (part of) the Graduation Project may need to be scheduled part-time. Indicate here if such applies to your project

Part of project scheduled part-time	
For how many project weeks	
Number of project days per week	

Comments:

Motivation and personal ambitions

Explain why you wish to start this project, what competencies you want to prove or develop (e.g. competencies acquired in your MSc programme, electives, extra-curricular activities or other).

Optionally, describe whether you have some personal learning ambitions which you explicitly want to address in this project, on top of the learning objectives of the Graduation Project itself. You might think of e.g. acquiring in depth knowledge on a specific subject, broadening your competencies or experimenting with a specific tool or methodology. Personal learning ambitions are limited to a maximum number of five.
(200 words max)

As a Strategic Product Designer, my professional ambition is to become a strategic consultant specializing in Business Design and Service Innovation. This graduation project at Supply Value offers the ideal environment to bridge the gap between design theory and high-level business strategy.

Beyond the standard graduation requirements, I have defined specific Personal Learning Ambitions for this project:
AI-Driven Business Design: To develop deep expertise in applying Generative AI to design and validate scalable business models and service propositions.
Human-AI Augmentation: To explore the parallels between designers and consultants, learning how to leverage AI as a collaborative tool that augments my own creative and analytical capabilities rather than replacing them.
Strategic Advisory: To broaden my competencies in guiding clients through complex strategic issues, translating abstract innovation concepts into actionable advice for high-stakes environments.
Design for Organizational Change: To prove my ability to design interventions that not only solve a problem but also foster a culture of innovation within a professional service firm.

By focusing on these areas, I aim to demonstrate that design leadership is essential for successful digital transformation.

Appendix B. Interview guides

Procedure. Each interview opened with the same statement: the interview is part of a graduation research at TU Delft on how consultants work, learn and use AI; there are no right or wrong answers; participation is voluntary; the participant may skip any question, stop at any moment and withdraw statements up to two weeks after the interview; the conversation is recorded to work it out, everything is processed anonymously according to the TU Delft guidelines, and clients are not named recognisably in the report. The TU Delft consent form was sent beforehand and consent was also recorded orally at the start of the recording. The interview with P3 was recorded from the point at which the participant agreed to recording; a personal opening part was not recorded at the participant's request. The opening of the interview with P7 was not recorded for technical reasons; the recording starts a few minutes into the conversation. For P2 the consent statement at the start was given before the recording started; the interview itself is complete.

Five instruments were used (section 2.5.1): a topic list for three interviews in round 1 (B.1), a guide for the interview on which journey map A is based (B.2), and in round 2 three guides (B.3 to B.5). The topic list and the guide of round 1 (B.1 and B.2) were written in English; the guides of round 2 (B.3 to B.5) were written in Dutch and are given here in the researcher's translation. Probes are given in brackets. The opening and consent statements are given once, under B.4, because they were the same in round 2.

B.1 Round 1: topic list (P1, P2, P3)

The three interviews of March and April 2026 followed a topic list rather than a question-by-question guide, because the conversation had to follow what each participant could tell about Supply Value's AI programme. The list had five topics, each with the questions that were asked when the topic had not yet come up by itself.

Topic 1, Supply Value's AI ambition. What does Supply Value want to achieve with AI, for itself and for clients? Where did that ambition start, and who carries it? What should be different in two years?

Topic 2, the tools in development. Which AI tools and agents are being built or tried at Supply Value? For which tasks? Who builds them, and who decides what is built? Which have been dropped, and why?

Topic 3, data security and client rules. What may and may not be submitted to which system? How is that agreed with clients, and how do consultants know? Where has that gone wrong or been unclear?

Topic 4, daily use. What do you use AI for yourself, in a normal week? Tell me about the last time it helped, and the last time it did not. Which consultants use it, and which do not, and why do you think that is?

Topic 5, consultants and capability. What should a consultant be able to do with AI that they cannot do today? How does Supply Value develop that: training, time, projects? What do clients ask of consultants that AI does not change?

B.2 Round 1: guide for P4, basis of journey map A

Purpose: to collect the material for a co-worker journey map of a consultant who actively uses AI, organised along the working day. Semi-structured; questions are a starting point, not a script; concrete examples (tool names, time saved, specific frustrations) are sought throughout.

Opening. Research context, purpose of the interview, use of the data, recording consent; no right or wrong answers.

Phase 1, background and AI profile. Q1 Can you describe your current assignment: what are you working on, for which client, and what does your day-to-day role look like? (type of organisation; typical week; time on the assignment). Q2 How would you describe your relationship with AI tools: when did you start, and how has that changed? (a moment or tool that made AI useful; frequency; growing or plateaued). Q3 Which AI tools do you use, and for what tasks? (tried and abandoned; tools you would like to use but cannot).

Phase 2, starting the day. Q4 Walk me through the first hour of a typical working day: what do you do first, and how do you decide what to prioritise? (AI for planning or orientation; staying informed about the client context). Q5 How do you prepare for a client meeting or stakeholder conversation, and does AI play a role? (a recent example; time with and without AI; what would have been missing).

Phase 3, analysis, research and knowledge work. Q6 When you need to do research or analyse a complex situation for a client, how do you approach that, and where does AI fit in? (an analysis task where AI made a difference; verifying sources; catching an error). Q7 Do you use AI for scenario analysis or stress-testing ideas, for example asking it to argue against your own recommendation? (several agents or perspectives; when an analysis is good enough). Q8 How do you handle sensitive client information when using AI tools? (information you never submit; a client that prohibited AI; guidance from Supply Value).

Phase 4, reporting, outputs and stakeholder communication. Q9 How do you use AI for reports, progress updates or client presentations? (step by step; time now versus before; validation step). Q10 How do you use AI to communicate with stakeholders? (tone and framing; a message that did not land). Q11 How do you use AI for visual outputs? (quality and consistency; colleagues who could benefit).

Phase 5, brainstorming, problem solving and innovation. Q12 When you are stuck on a problem or need new ideas, do you use AI as a thinking partner? How does that work? Q13 When you prepare a brainstorm with colleagues or clients, do you use AI beforehand?

Phase 6, end of day and knowledge sharing. Q14 At the end of a working day, do you capture what you learned or the key decisions you made? How? Q15 If a junior consultant joined your team tomorrow, how would you help them get up to speed with the AI tools you use?

Phase 7, gaps, frustrations and what is missing. Q16 Despite everything AI already helps you with, what are you still doing manually that you wish AI could do better? Q17 When you look at colleagues who use AI much less than you, what do you think is the main reason? Q18 If you could design one AI tool for Supply Value consultants that does not yet exist, what would it do?

Closing. Q19 Is there anything about how you use AI in your work that I have not asked about?

B.3 Round 2: guide in six phases

About 45 minutes, in the office, recorded with consent. Semi-structured, in six phases that cover the same ground as the four parts of the guide of B.4, with the emphasis on judging the work of others and on what the participant expects of Supply Value. Questions in order; probing only where an answer stays short or general.

Phase 1, role and current work. 1 What does your role at Supply Value consist of, and what is your current assignment? 2 Take your last ordinary working day at the client and walk me through it. (which tasks; with whom; what took most time)

Phase 2, how you approach a client problem. 3 When a client comes with a question, what do you do first? Take the last concrete case. 4 How do you decide on an approach or a method? Where does that knowledge come from? 5 What is the hardest part of a new assignment for you, and what was it when you started?

Phase 3, juniors and judgement. 6 When a junior shows you a piece of work, what do you look at first? 7 What do juniors typically get wrong in the first year, and how do they learn it? 8 What did you need from a senior when you were a junior, and did you get it?

Phase 4, own use of AI. 9 Do you use AI in your work? For what, and at which moments? Tell me about the last concrete time. 10 What do you still do by hand on purpose, and why? 11 Where has AI output been wrong, and how did you notice?

Phase 5, the client environment. 12 At your current client, what is permitted with AI, and how do you know? 13 Has a client ever asked you about AI, or objected to it? What happened?

Phase 6, Supply Value and the future. 14 What should Supply Value do so that consultants can work with AI on client problems? 15 What do clients keep paying a consultant for, whatever AI can do? 16 Is there anything I should have asked but did not?

B.4 Round 2: guide in four parts, used in two interviews

30 to 45 minutes, online via Teams, recorded with consent. Structure: onboarding, client assignment, AI use, closing. Questions in order; probing only where an answer stays short or general. The TU Delft consent form is sent beforehand and consent is also recorded orally.

Opening and consent (3 min). "Thank you for taking part. This interview is part of my graduation research at TU Delft, on how consultants work, learn and use AI. Participation is voluntary: you may skip any question, stop the conversation at any moment, and up to two weeks after today indicate that you want to withdraw something. There are no right or wrong answers. I record the conversation to work it out; I process everything anonymously according to the TU Delft guidelines and refer to you by a code, and the name of your client does not appear recognisably in my report. Do you agree with the recording?"

Part 1, onboarding at Supply Value (8 to 10 min). 1 What did your onboarding at Supply Value look like; what did you have to do in that first period? (when you started; which trainings, conversations or assignments; how long before you were really working at a client). 2 What did you find good about how that went? 3 Did you get enough information to start your work well; what did you miss or have to find out yourself? (an example). 4 What did the guidance look like: who helped you, at which moments, and how? (a fixed contact or buddy). 5 If you could improve the onboarding for the next new colleague, what would you change?

Part 2, your client assignment (12 to 15 min). 6 At which client do you work now, and what is your assignment there? (the client's name does not appear recognisably in the report). 7 Which tasks do you do in a normal week; take me through such a week. (most time; most and least enjoyable). 8 What do you do independently, and where do you work with others? 9 Is there a recent example of a problem you had to solve for the client? What was the problem, what did you do step by step, and what came out? (and then? who was involved? how long? how did you know this was the right approach?). 10 Did you ask experienced colleagues for help, at Supply Value or at the client? Whom, and why that person? (how soon). 11 If you did not ask for help: what made you solve it yourself, and was that the best route in hindsight? 12 What would have made this problem easier: more knowledge, more help, better information, or something else?

Part 3, AI use (10 to 12 min). 13 Do you use AI in your work? How, for what, and at which moments? (which tools; at the client or only internally). 14 Tell me about the last concrete time: what did you do with it, and what did you do with the result afterwards? (checked or edited; how). 15 How would you rate yourself on AI: beginner, average or advanced, and why? (compared to direct colleagues; what would show it). 16 Do you think you could use AI more or better? Where is that difference? (what holds you back: time, knowledge, trust, rules). 17 For which tasks would you like to use AI that you now do

yourself? (why those; how much time). 18 How would you like to use AI in your work from now on? What do you need for that from the firm, from colleagues, or from the tools?

Closing (3 min). 19 Is there anything I should have asked but did not? 20 May I contact you later for a short follow-up question? Directly afterwards: five minutes of field notes (most striking moment, doubts, atmosphere) and start of the transcription within 48 hours.

Rules for the interviewer. Ask for the last concrete occasion, not for opinions. One question at a time; let silences exist. Show no enthusiasm or disappointment about AI; part 3 has to measure the participant's picture, not the interviewer's. If short of time: skip 8 and 11; never skip 9, 10 and 15 to 18.

B.5 Round 2: guide in seven phases

About 50 minutes. All questions concern the internal organisation. Four goals: the daily work and the internal organisation (who does what, which problems recur); how Supply Value decides on training and the learning of new skills; the internal AI tools and their actual use; the participant's own use of AI and the conditions under which it would become worthwhile. Rules: nothing is assumed about the participant's role in the AI subject (question 9 establishes it); no confrontation between the participant's own AI use and what the firm expects of consultants; no client questions.

Phase 1, opening and consent (3 min). Same statement as B.4.

Phase 2, daily work (9 min). 1 What does your role at Supply Value consist of? 2 Take your last ordinary working day and walk me through it from beginning to end (per part of the day: what exactly, in which system or document; how you start the day; where the irritation was, and the pleasure). 3 Which tasks and meetings come back every week in your agenda? (what costs the most time, and is that justified).

Phase 3, the internal organisation (8 min). 4 Who works in the internal organisation, and what do they do? (planning, finance, HR, office; size of the group). 5 Which problems or bottlenecks recur most often internally? (a recent example, solved step by step). 6 How do internal improvements usually come about: structured, or mainly when something pinches?

Phase 4, internal AI tools (7 min). 7 Supply Value is working on its own AI tools. Which play a role in the internal work, and for what? (which tasks; by whom). 8 What do you see in practice: are those tools really used, or does it stay with a few people? And do you use them yourself? 9 Who in the firm is actually in charge of the AI subject: who decides what happens and who carries it out? And where does it touch your role?

Phase 5, training and development (10 min). 10 Consultants receive trainings here. How is it decided which trainings are given? (who decides; where the ideas come from). 11 What is the firm's purpose of those trainings, and how do you see whether a training has worked? 12 When a new skill becomes important for consultants, how does the firm go about getting it in? (a fixed route, or different per subject). 13 How do you look at the education of trainees and juniors: what works well, what could be better? 14 What is your own vision: how should consultants ideally learn new skills here? (only if the participant mentions AI: what place does AI have in that).

Phase 6, own use of AI (8 min). 15 Have you ever tried AI in your own work? Tell me about that time. 16 Take me through your considerations for using it little. 17 How would you rate yourself on AI, and how does it feel to talk about that in a firm that invests in it fully? 18 What would have to change for you personally, in the tools or in your work, before AI becomes worth it for you?

Phase 7, future and closing (5 min). 19 What does Supply Value look like in three years if everything goes well, and how would you notice? 20 If you could change one thing tomorrow about how the firm works, what would it be? 21 Is there anything I should have asked but did not, and may I contact you later for a follow-up question?

Appendix C. Codebook

Confidential. The codebook of the interview analysis, the 107 codes with their definitions, origin and number of fragments, is not included in the public version of this thesis; it is part of the version assessed by the graduation committee. How the language model was used in the coding is described below.

The language model that proposed codes per fragment was Claude (Anthropic; Claude Opus 5.5 and Claude Fable 5.1, Max subscription), operated through Claude Code. The instructions were given in the researcher's own words in the conversation: first to help with the coding, which produced an empty coding worksheet; then to code the first transcript itself so that the two codings could be compared, from which point the model worked as a second coder; then, per following transcript, to code it and to adjust the codebook where needed; the codes were linked to themes and clusters only after the last interview. The researcher reviewed every proposal in the workbook; the final coding, the sub-themes and the themes are the researcher's (section 2.5.1).

Appendix D. Coded fragments

Confidential. This appendix lists the 681 coded fragments of the eight interviews per participant, with their position in the recording, codes, evidence status and sub-theme. It is not included in the public version of this thesis; it is part of the version assessed by the graduation committee.

Appendix E. From codes to sub-themes to clusters

Confidential. This appendix holds the assignment of the 107 codes to sixteen sub-themes and five clusters, with the number of fragments per participant. It is not included in the public version of this thesis; it is part of the version assessed by the graduation committee.

Appendix F. Evidence behind the journey maps

Confidential. This appendix traces every cell of the two journey maps to the coded fragments of the two interviews, and describes how the maps were built and the assignment process in the interviews. It is not included in the public version of this thesis; it is part of the version assessed by the graduation committee.

Appendix G. Co-creation session: materials, board and outputs

This appendix contains the materials of the co-creation session (section 4.1), the board as the participants left it, and the outputs transcribed and translated. Names have been removed from the board images; participants are C1 to C5 as in table 4.1. The transcripts themselves are not included; the quotations used in chapter 4 are listed in G.6 with their position in the recordings.

G.1 Session plan and instructions

Procedure. The session took place at the Supply Value office and lasted about 1 hour 35 minutes. It was planned as a physical session and converted on the day into a Miro session, because two of the five participants took part from home; the three participants in the office joined the Teams meeting through the researcher's laptop. The session was recorded in two parts, the Teams meeting and, during the team work, a phone recording of the participants in the office, and both were transcribed and corrected. Participation was voluntary; the researcher gave the consent statement orally at the start of the session (G.2); no written form was used.

The session followed six steps on a Miro board. The instruction texts below are translated from the board.

Step 1, formulate a problem, task or question (10 min). Question: do you have a problem, task or question from your work that you often encounter or that you lose a lot of thinking time on? Explanation: write briefly what happens, why it is difficult and where AI might help. Choose something that is not too small, so not a single mail or a simple planning, but a task or question with several steps in which you have to combine information, make choices and use consultant judgement. Example: "A client notices that projects structurally overrun and asks how they can get more grip on planning, capacity and decision making."



Figure G.1 Step 1 as the participants left it: the instruction (left) and the four problem cards (right). The green card is the chosen problem, copied above step 2 as the brief.

Step 2, workflow mapping: how do you approach this client question? (30 min). Write per step, concretely, what you as a consultant would do to solve this client question. Start every step with "I go..." or "I do...". Do not write abstract words such as analysis, research, stakeholder management, validation, advice; write concrete actions such as: I talk to the client to understand why this problem is urgent now; I collect existing reports, data and project documents; I interview the stakeholders involved to understand where the process gets stuck; I look for comparable cases, benchmarks or literature; I make a first overview of possible causes; I discuss my first problem definition with a senior consultant; I work out three possible solution directions; I test the preferred direction with the client. Think of: which information do you need, whom do you talk to, which documents or data do you use, which analyses or frameworks do you apply, when do you check whether you are on the right track, which output do you make for the client. Eight steps were offered; more could be added.

Step 3, where can AI help, and how? (15 min). Look at the workflow of the colleague below you. Add AI sticky notes at the places where AI could support. Do not only write "AI can help": name which role the AI takes and what the AI must do to help the consultant at this step. Possible roles: 1 sparring partner (gives alternative perspectives and critical questions); 2 coach (guides

you through a problem or method step by step); 3 analyst (structures information, compares input and looks for patterns); 4 knowledge bank (retrieves relevant methods, templates, examples or sources); 5 risk checker (signals assumptions, blind spots, risks of missing information); 6 draft maker (makes a first version of a text, advice structure or deliverable). You can also invent a role. Example: role of AI: coach and sparring partner; AI helps by generating critical diagnostic questions that help the consultant not to accept the client question too quickly as the truth.

Step 4, human in the loop? (10 min). Look again at the workflow two rows below you and at the AI opportunities of step 3. Now add human-judgement sticky notes (black) at the places where AI can be risky, is limited, or must be checked. Use one of these sentences: it is risky to use AI here, because…; AI can probably not do this well, because…; AI output must be checked here, because…; I probably do not trust AI output here, because…; the consultant must keep judging here, because… Example: AI output must be checked here, because AI can misjudge political sensitivity, personal interests and informal client context; the consultant must judge whether the problem definition is recognisable, discussable and strategically sensible.

Step 5, design an AI tool (20 min). Work in teams and design an AI tool that helps the consultant, not replaces them. Choose one concrete moment from the workflow where AI can add value; use the AI opportunities of step 3 and the human-judgement risks of step 4 as input. The tool must: solve a concrete consultant problem; have a clear AI role; have clear input and output; contain a human-check moment; make clear what AI may not do; make the consultant work better or faster. Canvas of ten blocks: 1 name of the tool (examples: Problem Framing Coach, Advice Sparring Agent, Risk Checker, Client Question Sharpener); 2 problem moment: at which step of the workflow does this tool help; 3 for whom: junior, medior, senior, everyone; 4 role of AI; 5 input: what does the consultant enter (client question, notes, documents, interview insights, first problem definition, data, earlier deliverables); 6 what must AI concretely do (ask critical questions, structure, recognise patterns, signal assumptions, check risks, make a draft); 7 output: what does the consultant get back (problem definition, hypotheses, interview questions, advice structure, risk overview, checklist, draft); 8 human judgement: what must the consultant always check or decide (client context, sensitivities, stakeholders, source quality, feasibility, final decision); 9 where does the limit lie (determine the final advice, process confidential data without permission, send directly to the client, take over responsibility); 10 why would a consultant really use this tool (save time, frame more sharply, better quality, support juniors, reduce risks). Then sketch the first screen: what does the consultant see first, which question does the tool ask, which choices or buttons, where does the output appear, where is the human check or warning.

Step 6, pitch (10 min). One-minute pitch per team: which steps did you choose to work out as an AI tool and why, and what does it contain. Questions and feedback from the other team.

G.2 Consent

Consent was given orally. At the start of the session the researcher stated that the session was part of a graduation research at TU Delft on how consultants work and use AI; that participation was voluntary and that participants could leave at any moment and skip any question; that the session was recorded in two recordings for research purposes, that the recordings would be transcribed and that the transcripts would be processed anonymously, with participants referred to as C1 to C5 and client organisations described by type; and that the board and the canvases would be used in the report in anonymised form. Participants were asked whether they agreed before the recording started, and all did.

G.3 The board: workflows with AI and human-in-the-loop stickies

Figures G.2 to G.4 show the board of steps 2 to 4 as the participants left it (names replaced by codes). Pink cards are C1's steps, purple C2's, red C3's, yellow C4's, orange C5's; green stickies are the AI roles of step 3, black stickies the human-in-the-loop remarks of step 4.

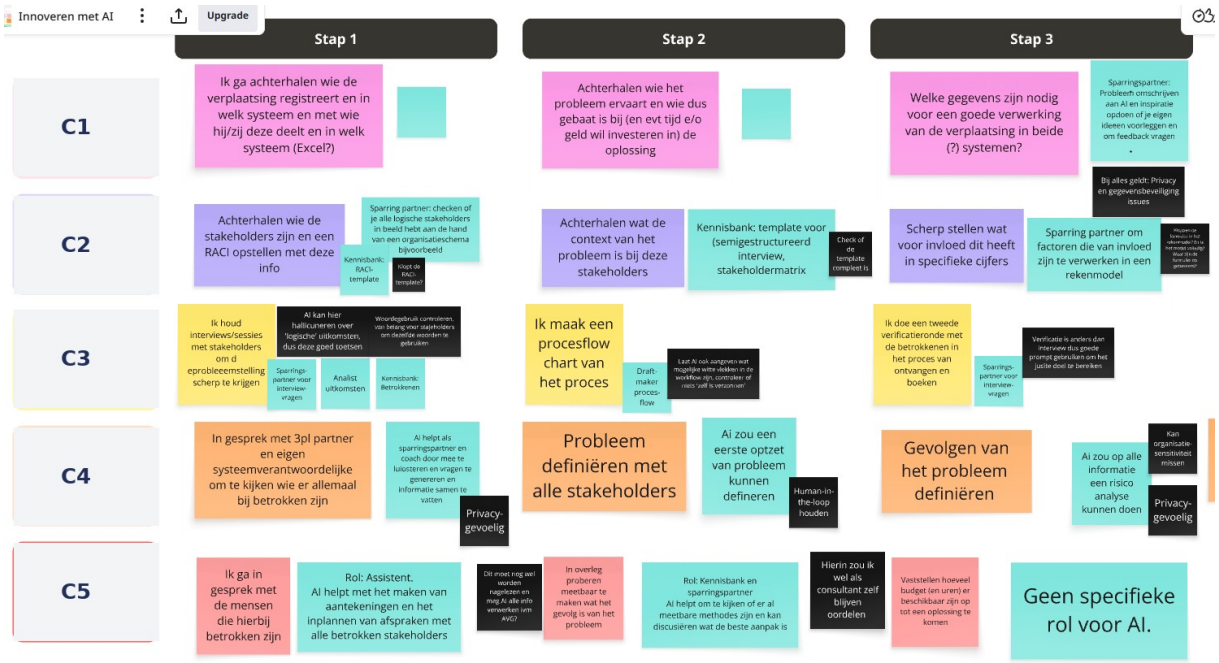


Figure G.2 Steps 1 to 3 of the five workflows.

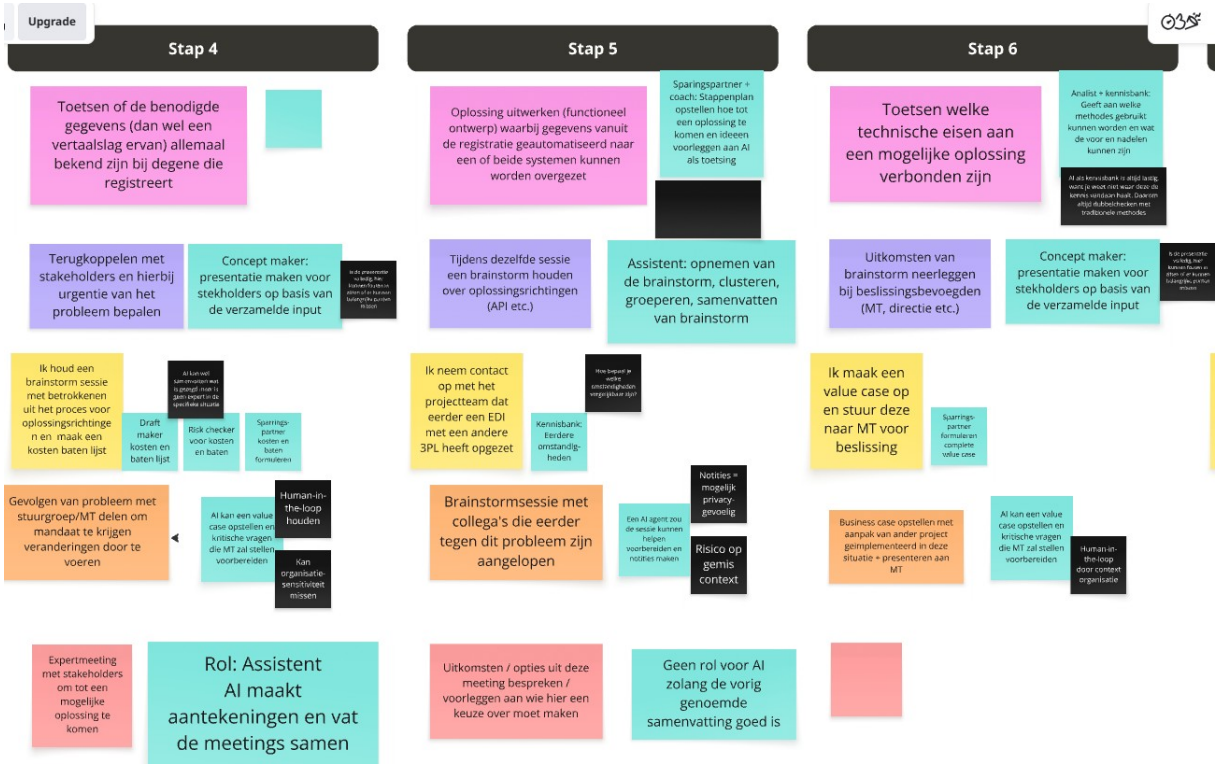


Figure G.3 Steps 4 to 6.

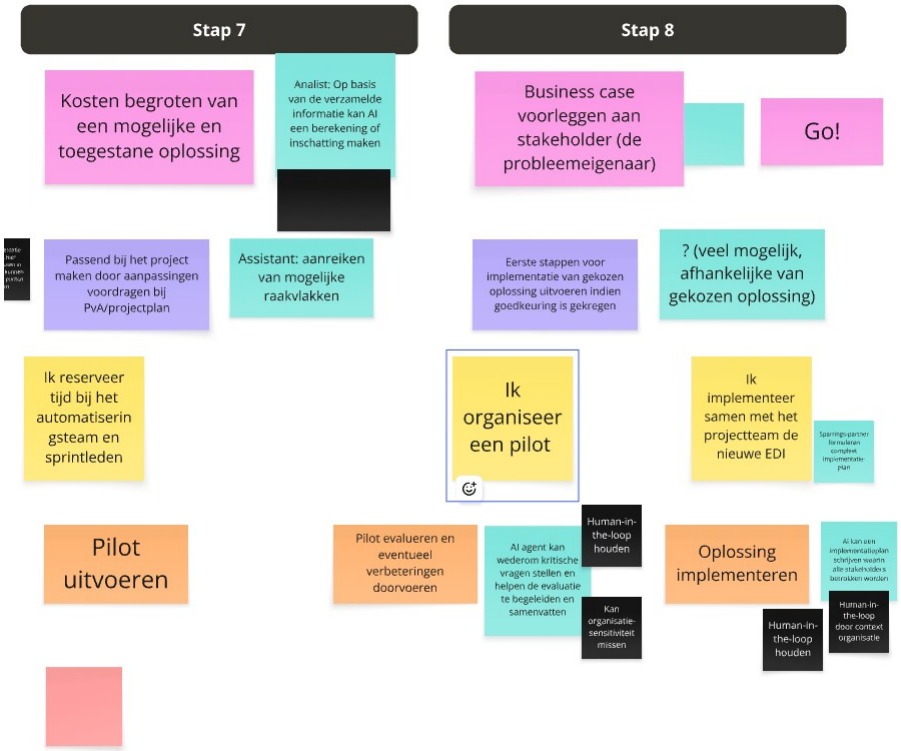


Figure G.4 Steps 7 and 8.

G.3.1 The five workflows, transcribed

	Step	Text on the board (translated)
C1	1	I go and find out who registers the movement and in which system, and with whom they share it and in which system (Excel?)
C1	2	Find out who experiences the problem and who therefore benefits from (and possibly wants to invest time and/or money in) the solution
C1	3	Which data are needed for correct processing of the movement in both (?) systems?
C1	4	Check whether the required data (or a translation of them) are all known to the person who registers
C1	5	Work out a solution (functional design) in which data from the registration can be transferred automatically to one or both systems
C1	6	Check which technical requirements are attached to a possible solution
C1	7	Estimate the cost of a possible and permitted solution
C1	8	Go!
C2	1	Find out who the stakeholders are and draw up a RACI with this information
C2	2	Find out what the context of the problem is with these stakeholders
C2	3	Sharpen what impact this has, in specific figures
C2	4	Feed back to stakeholders and establish the urgency of the problem
C2	5	In the same session, hold a brainstorm about solution directions (API etc.)
C2	6	Put the outcomes of the brainstorm to those with decision authority (MT, board etc.)
C2	7	Make it fit the project by proposing changes to the plan of approach or project plan
C2	8	Carry out the first steps for implementation of the chosen solution if approval has been given
C3	1	I go and talk to the people who are involved in this

	Step	Text on the board (translated)
C3	2	In consultation, try to make measurable what the consequence of the problem is
C3	3	Establish how much budget (and hours) are available to arrive at a solution
C3	4	Expert meeting with stakeholders to arrive at a possible solution
C3	5	Discuss the outcomes and options from this meeting with, or put them to, whoever has to make a choice about this
C4	1	I hold interviews and sessions with stakeholders to sharpen the problem statement
C4	2	I make a process flow chart of the process
C4	3	I do a second verification round with those involved in the process of receiving and booking
C4	4	I hold a brainstorm session with people from the process about solution directions and make a cost-benefit list
C4	5	I contact the project team that set up an EDI with another 3PL earlier
C4	6	I draw up a value case and send it to the MT for a decision
C4	7	I reserve time with the automation team and sprint members
C4	8	I organise a pilot
C4	9	I implement the new EDI together with the project team
C5	1	Talk to the 3PL partner and our own system owner to see who is involved
C5	2	Define the problem with all stakeholders
C5	3	Define the consequences of the problem
C5	4	Share the consequences of the problem with the steering group or MT to obtain a mandate to implement changes
C5	5	Brainstorm session with colleagues who have run into this problem before
C5	6	Draw up a business case with the approach of another project implemented in this situation, and present it to the MT
C5	7	Run a pilot
C5	8	Evaluate the pilot and implement improvements if needed
C5	9	Implement the solution

Table G.1 The five workflows of step 2.

G.3.2 The AI stickies (step 3) and human-in-the-loop stickies (step 4)

In step 3 each participant added green stickies to the workflow one row below (C1 to C2, C2 to C4, C4 to C5, C5 to C3, C3 to C1); in step 4 each added black stickies to the workflow two rows below (C1 to C4, C2 to C5, C4 to C3, C5 to C1, C3 to C2). Empty cells: no sticky at that step.

Owner	Step	Own step	Green: role of AI (written by the person one row below)	Black: human in the loop (written by the person two rows below)
C1	1	Who registers the movement, in which system, shares with whom	(empty green sticky)	
C1	2	Who experiences the problem, who benefits	(empty green sticky)	
C1	3	Which data are needed in both systems	Sparring partner: describe the problem to AI and get inspiration by putting your own ideas to it and asking feedback	Applies to everything: privacy and data security issues
C1	4	Check that the registrar knows all data	(empty green sticky)	
C1	5	Functional design, automated transfer	Sparring partner and coach: draw up a step-by-step plan for reaching a solution and put ideas to AI as a test	(black sticky without text)

Owner	Step	Own step	Green: role of AI (written by the person one row below)	Black: human in the loop (written by the person two rows below)
C1	6	Technical requirements of a solution	Analyst and knowledge bank: indicates which methods can be used and their advantages and disadvantages	AI as knowledge bank is always difficult, because you do not know where it gets the knowledge from; therefore always double-check with traditional methods
C1	7	Estimate the cost of a permitted solution	Analyst: on the basis of the collected information AI can make a calculation or estimate	(black sticky without text)
C1	8	Go!	(empty green sticky)	
C2	1	Stakeholders and RACI	Sparring partner: check whether you have all logical stakeholders in view, for example using an organisation chart. Knowledge bank: RACI template	Is the RACI template correct?
C2	2	Context of the problem with stakeholders	Knowledge bank: template for a semi-structured interview, stakeholder matrix	Check that the template is complete
C2	3	Sharpen the impact in specific figures	Sparring partner to work the influencing factors into a calculation model	Are the formulas in the calculation model correct? Is the model complete? What are the formulas based on?
C2	4	Feed back, establish urgency	Concept maker: make a presentation for stakeholders from the collected input	Is the presentation complete? There can be errors in it or important points can be missing
C2	5	Brainstorm solution directions (API etc.)	Assistant: record the brainstorm, cluster, group, summarise	
C2	6	Outcomes to decision makers	Concept maker: presentation for stakeholders from the collected input	Is the presentation complete? There can be errors in it or important points can be missing
C2	7	Fit into project plan	Assistant: offer possible points of contact	(black sticky without text)
C2	8	First implementation steps	? (a lot possible, depends on the chosen solution)	
C4	1	Interviews and sessions to sharpen the problem	Sparring partner for interview questions. Analyst: outcomes. Knowledge bank: those involved	AI can hallucinate here about "logical" outcomes, so test these well. Check word use: important for stakeholders to use the same words
C4	2	Process flow chart	Draft maker: process flow	Let AI also indicate possible white spots in the workflow; check that nothing is "invented by itself"
C4	3	Second verification round	Sparring partner for interview questions	Verification is different from an interview, so use good prompts to reach the right goal
C4	4	Brainstorm solution directions, cost-benefit list	Draft maker: cost-benefit list. Risk checker for costs and benefits. Sparring partner: formulate costs and benefits	AI can summarise what was said but is not an expert in the specific situation
C4	5	Contact the project team that set up an EDI before	Knowledge bank: earlier circumstances	How do you determine which circumstances are comparable?
C4	6	Value case to MT	Sparring partner: formulate the complete value case	
C4	7	Reserve time with automation team		
C4	8	Organise a pilot		

Owner	Step	Own step	Green: role of AI (written by the person one row below)	Black: human in the loop (written by the person two rows below)
C4	9	Implement the new EDI with the project team	Sparring partner: formulate the complete implementation plan	
C5	1	Talk to 3PL partner and system owner	AI helps as sparring partner and coach by listening in, generating questions and summarising information	Privacy-sensitive
C5	2	Define the problem with all stakeholders	AI could define a first set-up of the problem	Keep human in the loop
C5	3	Define the consequences	AI could do a risk analysis on all information	Can miss organisational sensitivity. Privacy-sensitive
C5	4	Share consequences with steering group or MT for a mandate	AI can draw up a value case and prepare critical questions the MT will ask	Keep human in the loop. Can miss organisational sensitivity
C5	5	Brainstorm with colleagues who met this before	An AI agent could help prepare the session and take notes	Notes: possibly privacy-sensitive. Risk of missed context
C5	6	Business case with approach of earlier project; present to MT	AI can draw up a value case and prepare critical questions the MT will ask	Human in the loop because of organisational context
C5	7	Run a pilot		
C5	8	Evaluate the pilot and improve	AI agent can again ask critical questions and help guide and summarise the evaluation	Keep human in the loop. Can miss organisational sensitivity
C5	9	Implement the solution	AI can write an implementation plan in which all stakeholders are involved	Keep human in the loop. Human in the loop because of organisational context
C3	1	Talk to the people involved	Role: assistant. AI helps with making notes and planning appointments with all stakeholders	This still has to be read through; and may AI process all information given the GDPR?
C3	2	Make the consequence measurable together	Role: knowledge bank and sparring partner. AI helps to see whether there are measurable methods and can discuss what the best approach is	Here I would keep judging myself as a consultant
C3	3	Establish available budget and hours	No specific role for AI	
C3	4	Expert meeting with stakeholders	Role: assistant. AI takes notes and summarises the meetings	
C3	5	Discuss options with whoever must choose	No role for AI as long as the previous summary is good	

Table G.2 All stickies of steps 3 and 4, translated. Green stickies without text and black stickies without text are marked as such.

G.4 The tool canvases of step 5

Figures G.5 to G.10 show the two canvases and screens as the teams left them, and figure G.11 the first screen redrawn. In the images and the table, tool A is called Concept A.

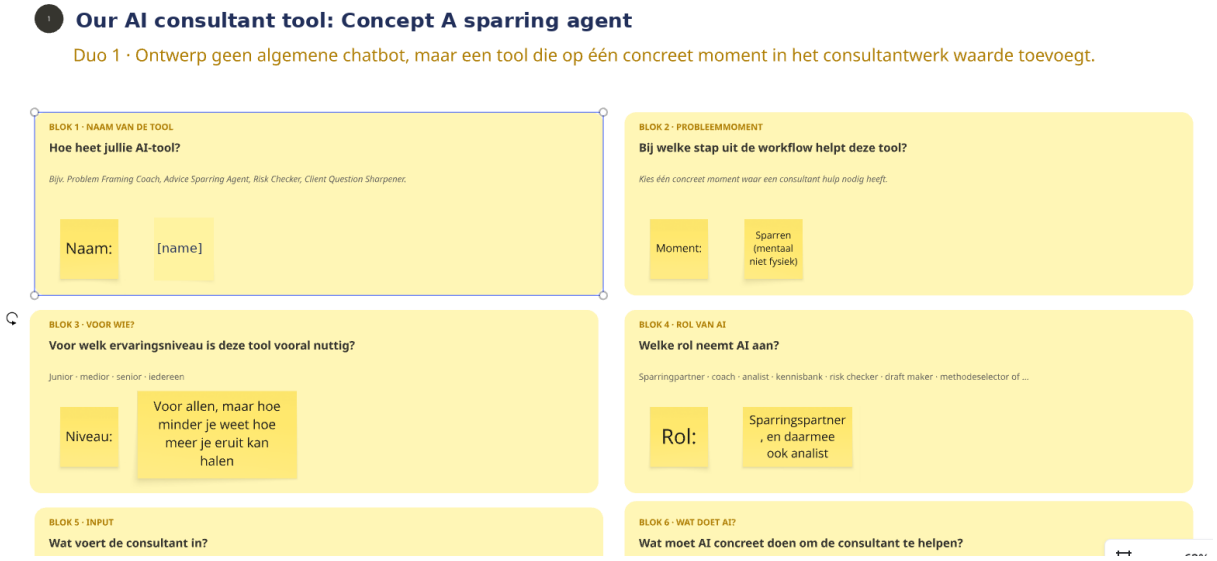


Figure G.5 Tool A, blocks 1 to 4.

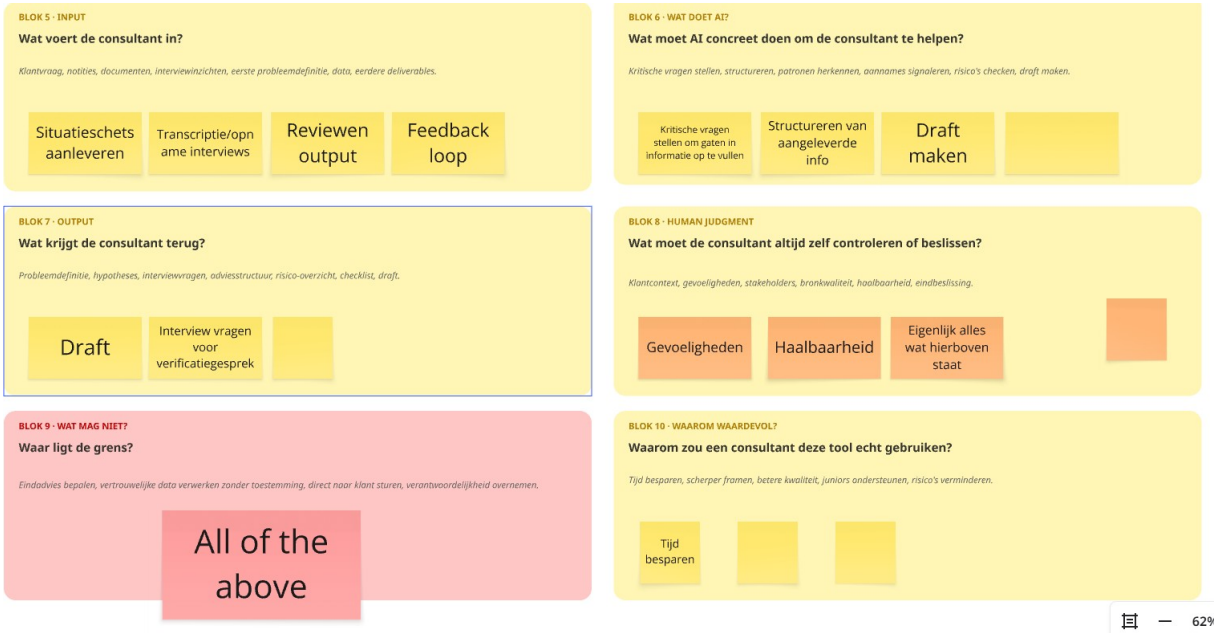


Figure G.6 Tool A, blocks 5 to 10.

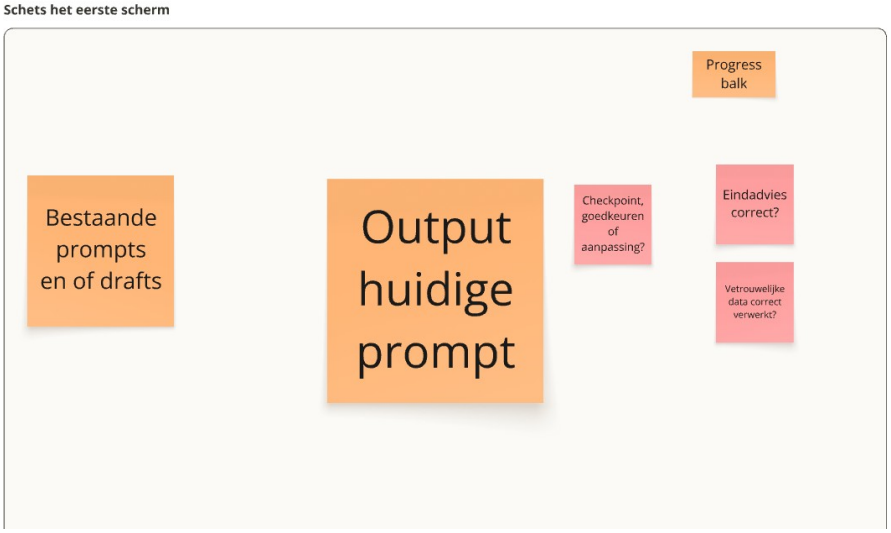


Figure G.7 Tool A, first screen.

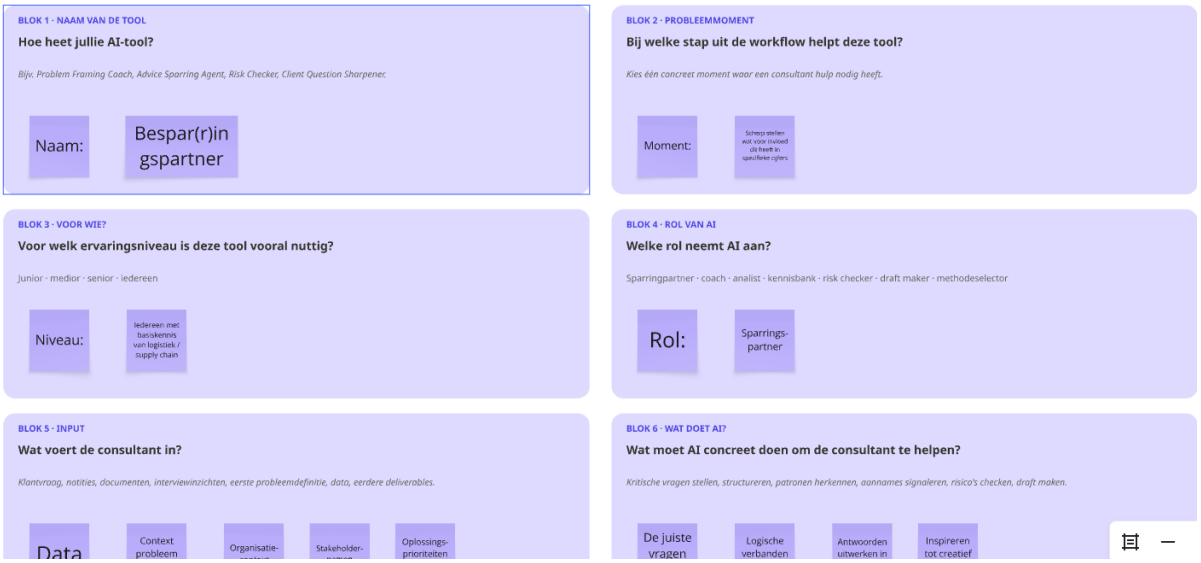


Figure G.8 Tool B, blocks 1 to 4.

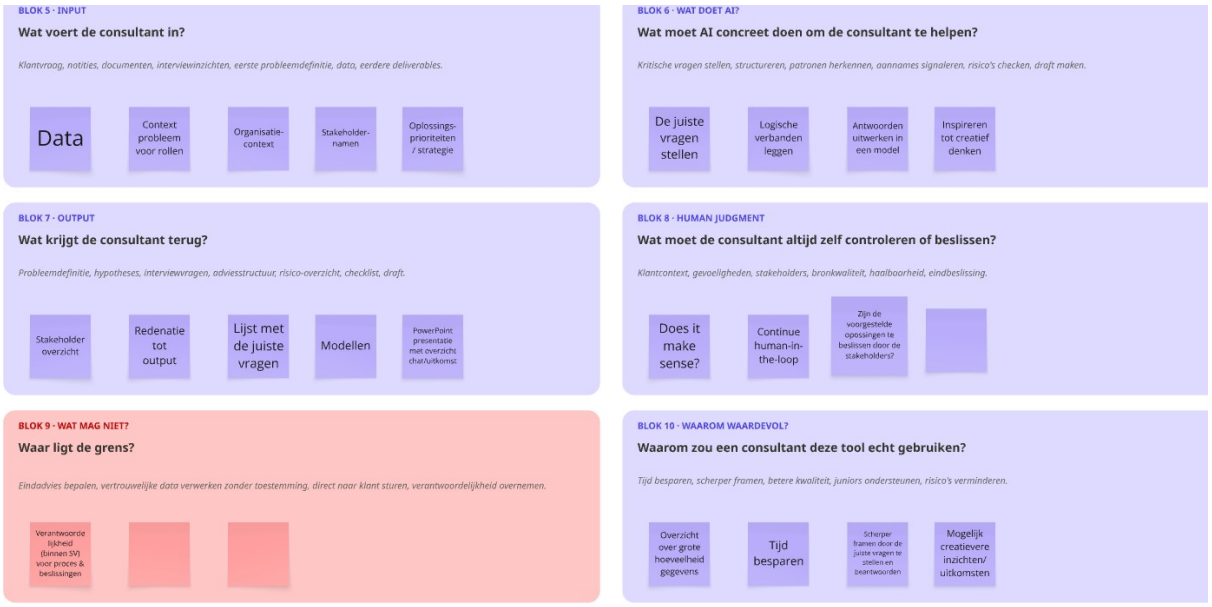


Figure G.9 Tool B, blocks 5 to 10.

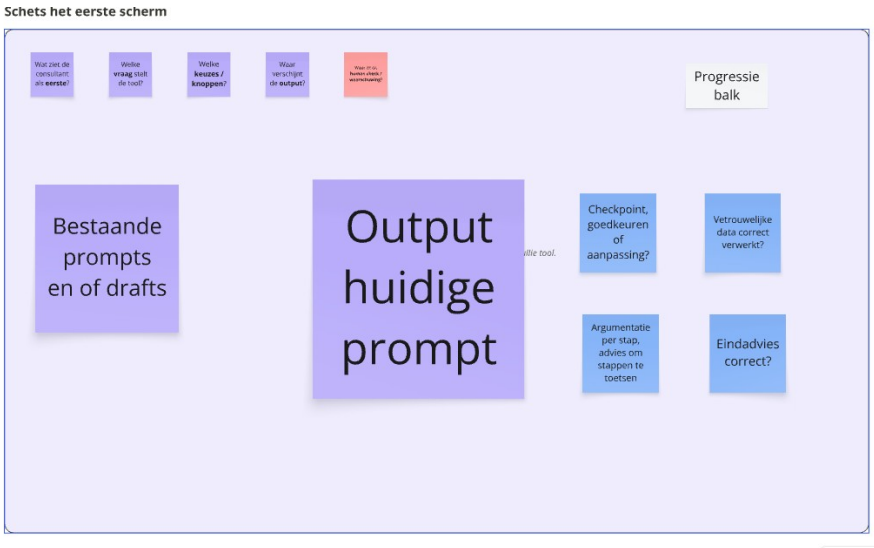


Figure G.10 Tool B, first screen.

Block	Tool A (team A: C4, C5)	Tool B (team B: C1, C2, C3)
1 Name	Concept A "as a service"	Bespar(r)ingspartner
2 Moment	Sparring (mentally, not physically)	Sharpening what impact this has, in specific figures
3 For whom	For all, but the less you know the more you can get out of it	Everyone with basic knowledge of logistics or supply chain
4 Role	Sparring partner, and thereby also analyst	Sparring partner
5 Input	Provide situation sketch; transcription or recording of interviews; review output; feedback loop	Data; context of the problem for the roles; organisational context; stakeholder names; solution priorities or strategy
6 What AI does	Ask critical questions to fill gaps in information; structure the information provided; make a draft; (one empty sticky)	Ask the right questions; make logical connections; work answers out in a model; inspire creative thinking
7 Output	Draft; interview questions for the verification conversation; (one empty sticky)	Stakeholder overview; reasoning to output; list of the right questions; models; PowerPoint presentation with overview of chat or outcome
8 Human judgement	Sensitivities; feasibility; actually everything written above; (one empty sticky)	Does it make sense?; continuous human in the loop; are the proposed solutions ones the stakeholders can decide on?; (one empty sticky)
9 Limit	All of the above (the canvas examples: determine the final advice, process confidential data without permission, send directly to the client, take over responsibility)	Responsibility (within SV) for process and decisions; (two empty stickies)
10 Why	Save time; (two empty stickies)	Overview over a large amount of data; save time; sharper framing by asking and answering the right questions; possibly more creative insights or outcomes
First screen	Progress bar; existing prompts and/or drafts (left); output of current prompt (centre, large); checkpoint: approve or adjust?; final advice correct?; confidential data processed correctly?	As tool A, plus: argumentation per step, advice to test steps

Table G.3 The two canvases transcribed.

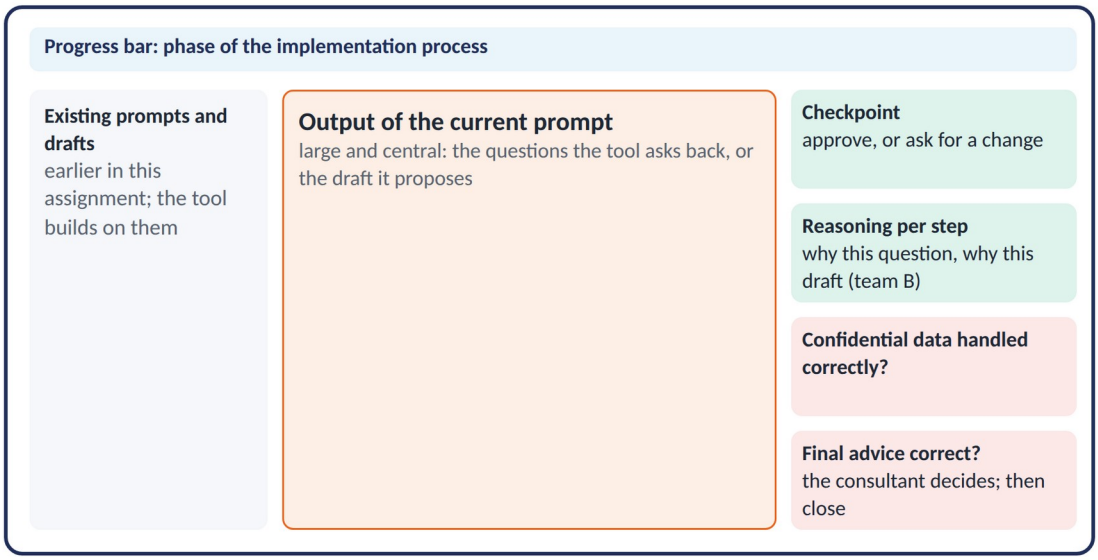


Figure G.11 The first screen of the tool, redrawn from the sketches of both teams. Team B adopted team A's screen and added the reasoning panel.

G.5 The four problems of step 1

	Text on the board (translated and lightly shortened to protect anonymity); C1 orally
C2	At the different organisations I have started at, I notice that it costs me much effort to define and find out the urgency and background of various circumstances. Finding out is a rather tedious and time-consuming task. As a consultant you change environment, background and stakeholders relatively often, so structuring and finding these out would be a nice task to hand over to an AI.
C3	At my client it costs me a lot of time to find out exactly how the processes run. There are many people involved who all store their information and mail exchanges in a different place. This makes it difficult to create an overview of exactly where we are in the process and who is working on what.
C4	Client works with a 3PL partner with a stock management system. There is no link between our own system and this system. At the moment movements are registered in Excel and shared, after which the actions are adjusted or booked by hand in our own system. The problem is that these bookings experience much delay and/or go wrong, so that stock differences arise between our system and reality. (Written twice on the board; the copy in green is the brief for step 2.)
C1	Orally (the board did not work for this participant at that moment): many different things and many stakeholders, each kept in a different way by the predecessor (several OneNotes, a logbook, action lists), from which a monthly programme report with qualitative assessment and risks and issues has to be produced; costs much time.

Table G.4 The problems of step 1.

G.6 Quotations used in chapter 4

Who	Quotation (translated)	Recording and position
C2	"You are not really being creative, it is really just finding things out."	Part 1, step 1 (before 19:42)
C4	"The basic problem is the difference between reality and the bookings in our system."	Part 1, 19:42
C4	"The 3PL does neatly what we ask; it is about how we handle the export internally and the manual checks."	Part 1, 37:32
C4	"I see that at several points here it says sparring partner, so we could also make one tool as sparring partner for the whole process."	Part 1, 1:10:38
C4	"I am really a novice when it comes to AI, I have no idea."	Part 1, 1:09:19
C5	"I can just about use Excel."	Part 1, 1:09:38

Who	Quotation (translated)	Recording and position
C5	"The less you know, the more you get out of it."	Part 1, 1:13:15
C5	"In the end, nine times out of ten, you still know more about the context."	Part 1, 1:21:01
C5	"It will always be a draft and never the final version."	Part 2, 25:53
C5	"The final advice will never be determined by an AI alone; confidential data is never processed without permission and the right systems; it will not send things to clients itself; it will not take over responsibility. You cannot use as an excuse: the AI did it wrong."	Part 2, 26:19 to 26:49
C3	"I think you then very quickly get something too general, ChatGPT-like; then you just get information, but I think that is not what you want with this tool."	Part 2, 0:02
C1	"The knowledge bank is the most usable, with what we already have."	Part 2, 0:44
C2	"You can say very nicely, we have this problem and we must solve it, but it works much better if you say: we have this problem, it costs this much, it has this much influence, it is urgent; then you have a much better case to present to the management team."	Part 2, 1:43 to 2:00
C1	"That is the essence of the sparring partner: it draws it out of you and thereby goes from qualitative to quantitative."	Part 2, 2:41 to 2:51
C2	"The nicest thing about AI is that you no longer have to start from a blank page."	Part 2, 2:53
C2	"A difference in how the back end of the model is set up, so that you get a sincerely good sparring partner."	Part 2, 3:53 to 4:13
C1	"If the user keeps saying inkoop, we should not come with procurement because that is what our model says."	Part 2, 9:33
C2	"So that it becomes as little of a black box as possible, and why it asks certain questions."	Part 2, 12:45
C2	"So where does that factor of a thousand come from?" (reacting to C3's example of a model that multiplies by a thousand without saying why)	Part 2, 15:31
C3	"Not that you just go along with whatever the AI proposes and think, it will be fine."	Part 2, 15:59
C1	"Are the proposed solutions ones the stakeholders can decide on? The law has to change, nice, but we are not going to do that. We need more money, so we print money."	Part 2, 16:49 to 17:02
C2	"I do not want the AI to make an end product that goes to the client."	Part 2, 18:24
C2	"So that you can think creatively yourself and get to the solution yourself, instead of getting a ready-made answer."	Part 2, 30:12
C5	"The part that it can use confidential information, and actually not much more."	Part 2, 27:15
C2	"Not all the data that can be found on the internet, but much more focused on that expertise."	Part 2, 33:25

Table G.5 Quotations with their position in the recordings (part 1: Teams recording; part 2: phone recording of the office team and the pitches). Positions are those of the transcript block in which the sentence occurs.

G.7 The five workflows aligned on eight phases

Figures G.12 and G.13 show the five workflows of step 2 aligned on the eight phases A to H, one consultant per row, with the AI roles of step 3 as green tags and the human-in-the-loop remarks of step 4 as black tags; section 4.1.4 reads them. The texts of the stickies are in G.3.1.

Five workflows for the same client problem, phases A to D

AI role proposed by a colleague (step 3) · Human-in-the-loop remark by a second colleague (step 4)

	A Talk to the people involved	B Define the problem and context	C Quantify impact, data, budget	D Verify, secure a mandate
C1 senior	**1 Find out who registers the movement, in which system, shared with whom**	**2 Find out who experiences the problem and who benefits**	**3 Which data are needed to process the movement in both systems** Sparring partner: describe the problem, get inspiration and feedback Privacy and data security, at every step	**4 Check that the registrar knows all required data**
C2 junior	**1 Find out who the stakeholders are, draw up a RACI** Sparring partner: all logical stakeholders in view? Knowledge bank: RACI template Is the RACI template correct?	**2 Find out the context of the problem with these stakeholders** Knowledge bank: interview template, stakeholder matrix Is the template complete?	**3 Sharpen what impact this has, in specific figures** Sparring partner: work the factors into a calculation model Are the formulas correct and complete? Based on what?	**4 Feed back to stakeholders, establish urgency** Concept maker: presentation from the collected input Is the presentation complete?
C3 junior	**1 Talk to the people involved** Assistant: notes, planning appointments Must be read through; may AI process all information (GDPR)?	**2 Together, make the consequence of the problem measurable** Knowledge bank and sparring partner: measurable methods, best approach Here I keep judging myself as a consultant	**3 Establish available budget and hours** No specific role for AI	
C4 junior	**1 Interviews and sessions with stakeholders to sharpen the problem** Sparring partner: interview questions Analyst: outcomes Knowledge bank: those involved AI can hallucinate "logical" outcomes: test them Same words as stakeholders	**2 Make a process flow chart** Draft maker: process flow Let AI flag white spots; check nothing is invented		**3 Second verification round with those who receive and book** Sparring partner: interview questions Verification differs from interviewing: prompt for the goal
C5 junior	**1 Talk to the 3PL partner and the system owner: who is involved** Sparring partner and coach: listen in, generate questions, summarise Privacy-sensitive	**2 Define the problem with all stakeholders** First set-up of the problem definition Keep human in the loop	**3 Define the consequences of the problem** Risk analysis on all information Can miss organisational sensitivity Privacy-sensitive	**4 Share consequences with steering group or MT for a mandate** Value case and the critical questions the MT will ask Keep human in the loop Can miss organisational sensitivity

Steps numbered as written on the board.

Figure G.12 The five workflows aligned on eight phases, phases A to D, with the AI roles a colleague proposed at each step (green) and the human-in-the-loop remarks a second colleague added (black); the legend is on the figure. Texts translated and shortened from the board; the full texts are in Appendix G.

Five workflows for the same client problem, phases E to H

☐ AI role proposed by a colleague (step 3) ■ Human-in-the-loop remark by a second colleague (step 4)

	E Generate solution directions	F Decide: case to the management team	G Plan, cost, prepare	H Pilot and implement
C1 senior	**5 Functional design: registration data transferred automatically** Sparring partner and coach: step plan, test ideas		**6 Check technical requirements** Analyst and knowledge bank: methods with pros and cons Knowledge of unknown origin: double-check with traditional methods **7 Estimate the cost of a permitted solution** Analyst: calculation or estimate	**8 Go!**
C2 junior	**5 Brainstorm solution directions in the same session** Assistant: record, cluster, summarise	**6 Put the outcomes to decision makers (MT, board)** Concept maker: presentation from the collected input Is the presentation complete?	**7 Propose changes to the project plan** Assistant: offer possible points of contact	**8 First implementation steps once approved** ? Depends on the solution
C3 junior	**4 Expert meeting with stakeholders towards a solution** Assistant: notes and summaries	**5 Discuss options with whoever has to choose** No role for AI as long as the summary is good		
C4 junior	**4 Brainstorm solution directions; cost-benefit list** Draft maker: cost-benefit list Risk checker Sparring partner: formulate costs and benefits AI can summarise but is no expert in this situation **5 Contact the team that set up an EDI with another 3PL** Knowledge bank: earlier circumstances Which circumstances are comparable?	**6 Value case to the MT for a decision** Sparring partner: formulate the value case	**7 Reserve time with the automation team**	**8 Organise a pilot** **9 Implement the new EDI with the project team** Sparring partner: implementation plan
C5 junior	**5 Brainstorm with colleagues who met this problem before** Agent: prepare the session, take notes Notes are privacy-sensitive Risk of missed context	**6 Business case from an earlier project; present to MT** Value case and the critical questions the MT will ask Human in the loop: organisational context		**7 Run a pilot** **8 Evaluate the pilot and improve** Agent: critical questions, guide and summarise the evaluation Keep human in the loop Can miss organisational sensitivity **9 Implement the solution** Implementation plan involving all stakeholders Human in the loop: organisational context

Steps numbered as written on the board.

Figure G.13 The five workflows, phases E to H.

Appendix H. Ideation: boards, idea lists, directions and evaluation

This appendix contains the method of the ideation of section 4.2, the boards (H.1), the full idea lists (H.2), and the directions and the ideas set aside per question (H.3).

Method

A how-might-we question reformulates an insight as an open design question (IDEO.org, 2015, p. 85); the Delft Design Guide describes the same step as "how-tos" (Van Boeijen et al., 2020). The three words each do something: "how" assumes that a solution exists, "might" keeps the question open to more than one answer, and "we" makes it a shared task. The source of a question is always an insight from the Define phase, never an idea; the question has to be open enough to allow several concepts and narrow enough to give direction. A question that contains its own solution ("how might we build a tool that asks before it answers?") closes the space it is meant to open, which is why the design requirements of section 3.3 were deliberately not used as the source: a requirement is already an answer. The requirements return in section 4.3 as the criteria against which the concepts are judged.

Four sources from the Define phase were used, and each question in table 4.2 of section 4.2 names its origin: the five insight clusters of section 3.1; the moments of truth and the opportunities of the journey maps (tables 3.3 and 3.4), which locate the clusters in the week of a consultant; the assignment process of figure 3.3, which locates them in the phases of an assignment; and the additions of the co-creation session (section 4.1.7), in particular the decision maker and the mandate, and the question whether a structured method should be the backbone of the support, which the session left open and which is carried here as a question rather than a decision. For each source several candidate questions were written; they were then held against four criteria, which are also the criteria that explain why some candidates were set aside. The question must trace back to a cluster, a moment of truth or an addition of the session (evidence); it must fall within the design goal and the constraints of section 3.3 (scope); it must concern the research question, that is, the capability of a consultant rather than the organisation or the market (relevance); and it must concern something Supply Value can influence (room to act). Two examples of candidates that failed: "how might we convince municipalities to permit AI?" is a real problem but lies outside Supply Value's room to act and outside the scope; "how might we replace hours by repeatable solutions?" (A5, theme 11) is a question about the business model rather than about the capability of a consultant and lies outside the scope. The twelve questions that remain are ordered by the phase of the assignment process in which they play, because that order shows at a glance where the problem of this thesis sits: nine of the twelve fall in phases 2 to 4 of figure 3.3 or in the conditions around them, which is the answer to sub-question SQ2 that section 3.3.7 left open.

For each question an unfiltered list of ideas was generated in two steps. The researcher first wrote down their own answers, as they came, without judging them; the list was then extended, question by question, to at least twenty ideas, with the explicit rule that every type of solution had to be present: people (colleagues, seniors, managers), the process at the client (agreements and moments with the client), learning (training and courses), aids without AI (templates, checklists, dossiers, formats) and AI. The rules of brainstorming were applied as Osborn (1953) formulated them and as the Delft Design Guide repeats them: defer judgement, go for quantity, welcome bold ideas, build on each other's ideas (Van Boeijen et al., 2020). Bold and impractical ideas were kept on purpose, because they mark the edges of the space and because a poor idea often contains a good one: "go straight to the client's CEO with all your questions" is unworkable, but "a senior calls the sponsor to hear the question behind the engagement letter" sits inside it. The round produced 262 ideas, 42 for the first question and 20 for each of the others. The twelve boards are in H.1 and the full lists in H.2.

H.1 The ideation boards

One board per how-might-we question, as used in section 4.2.3. The boards show a selection of the ideas for each question; the full lists are in H.2.

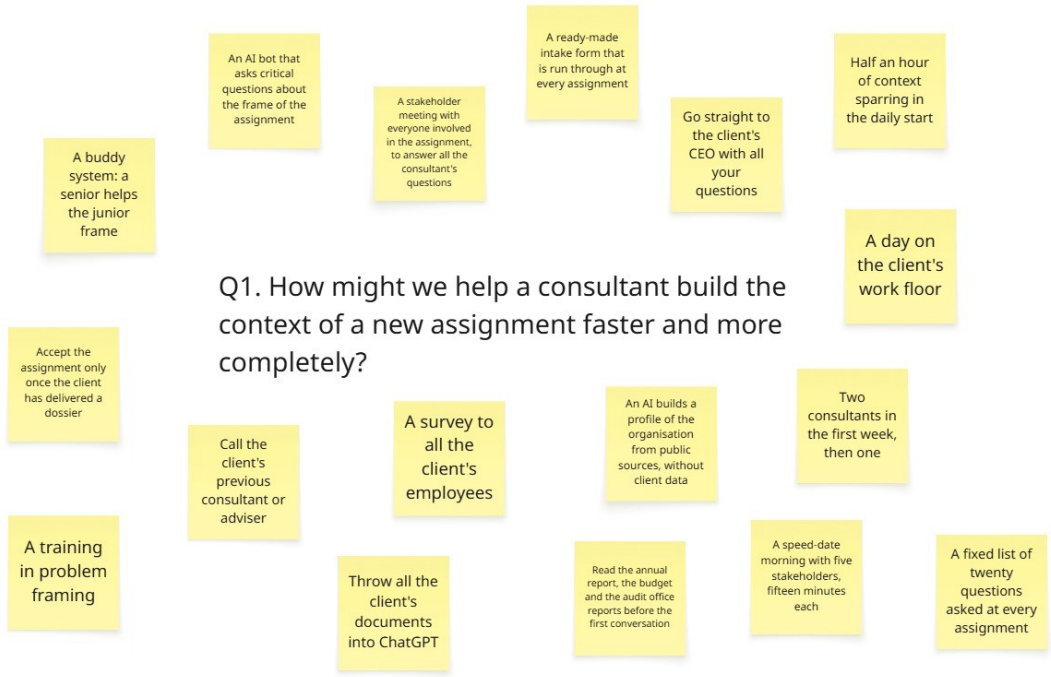


Figure H.1 Ideation board, question 1.

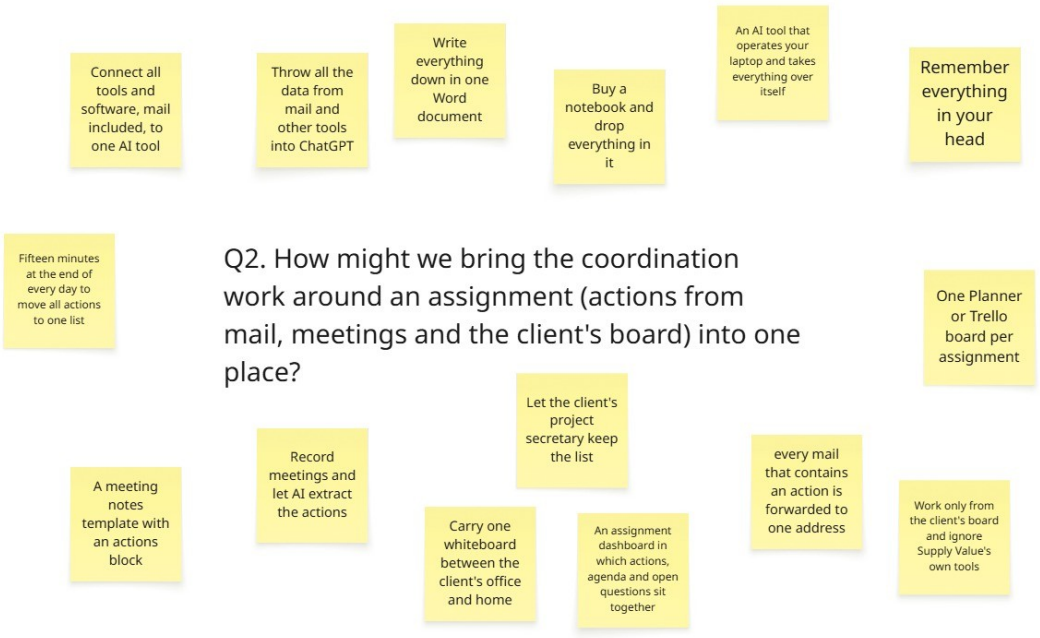


Figure H.2 Ideation board, question 2.

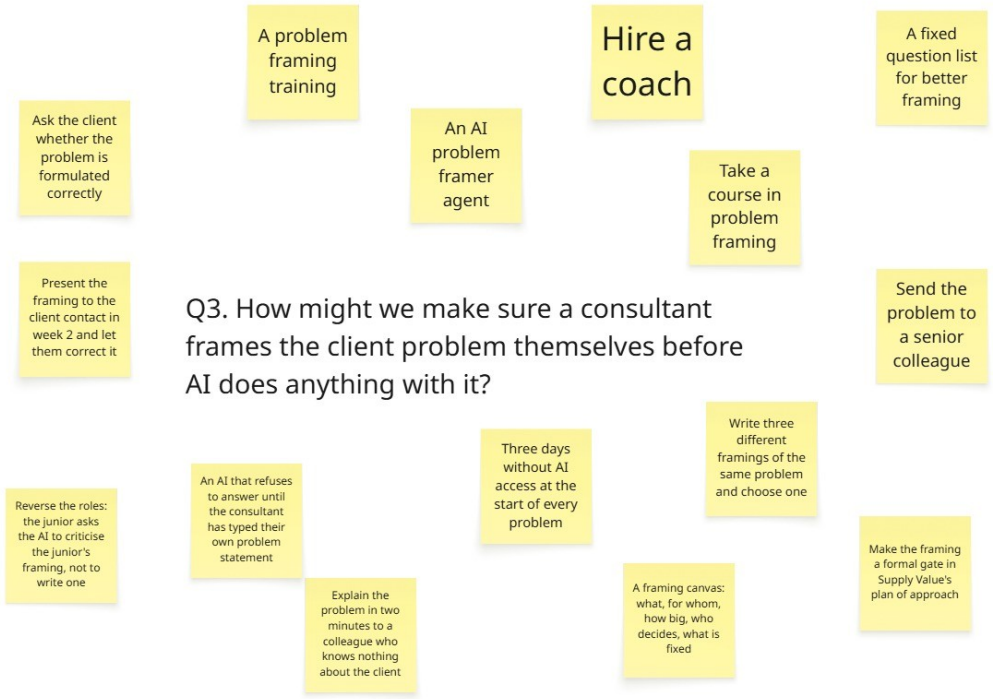


Figure H.3 Ideation board, question 3.

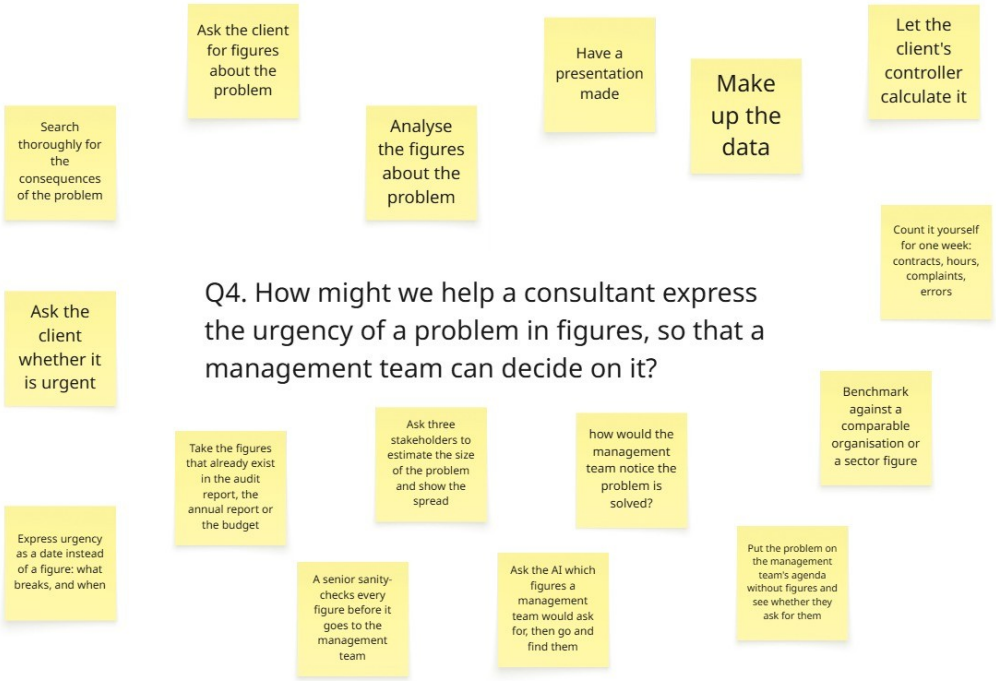


Figure H.4 Ideation board, question 4.



Figure H.5 Ideation board, question 5.

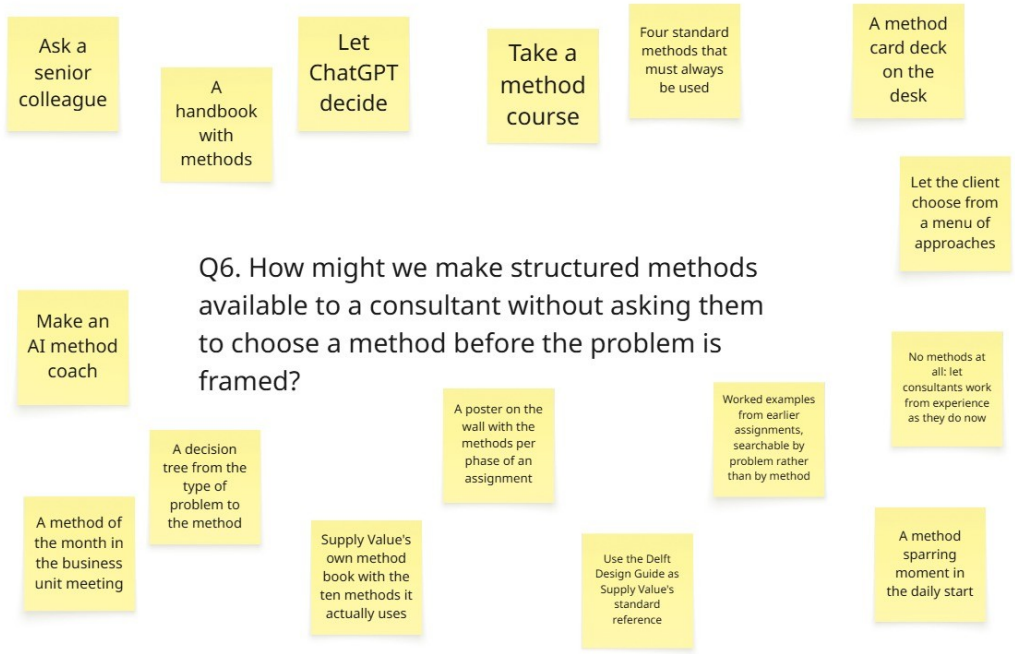


Figure H.6 Ideation board, question 6.

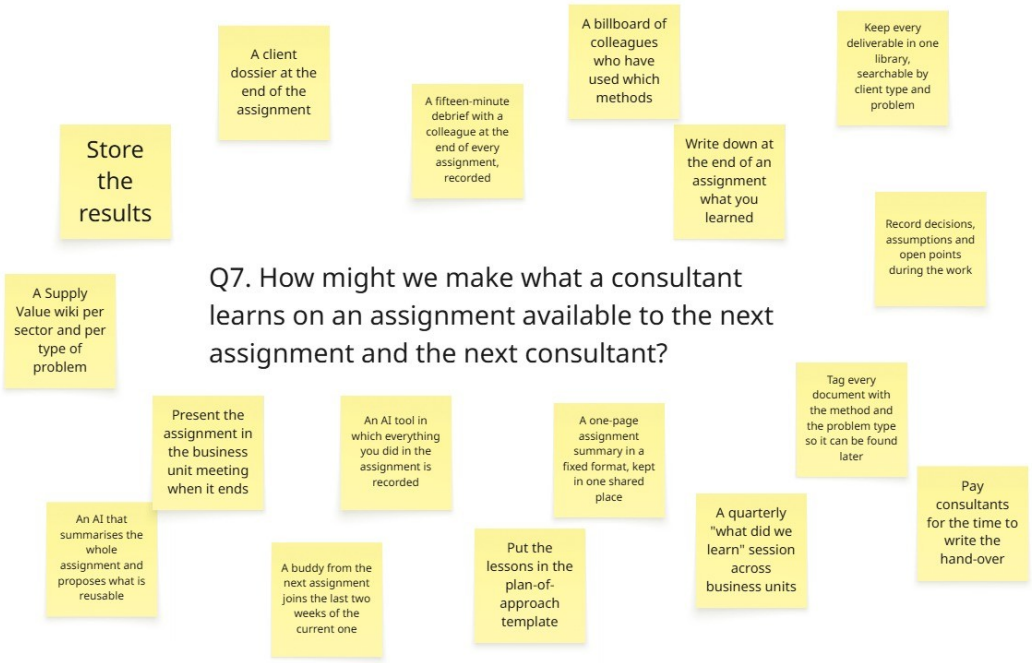


Figure H.7 Ideation board, question 7.

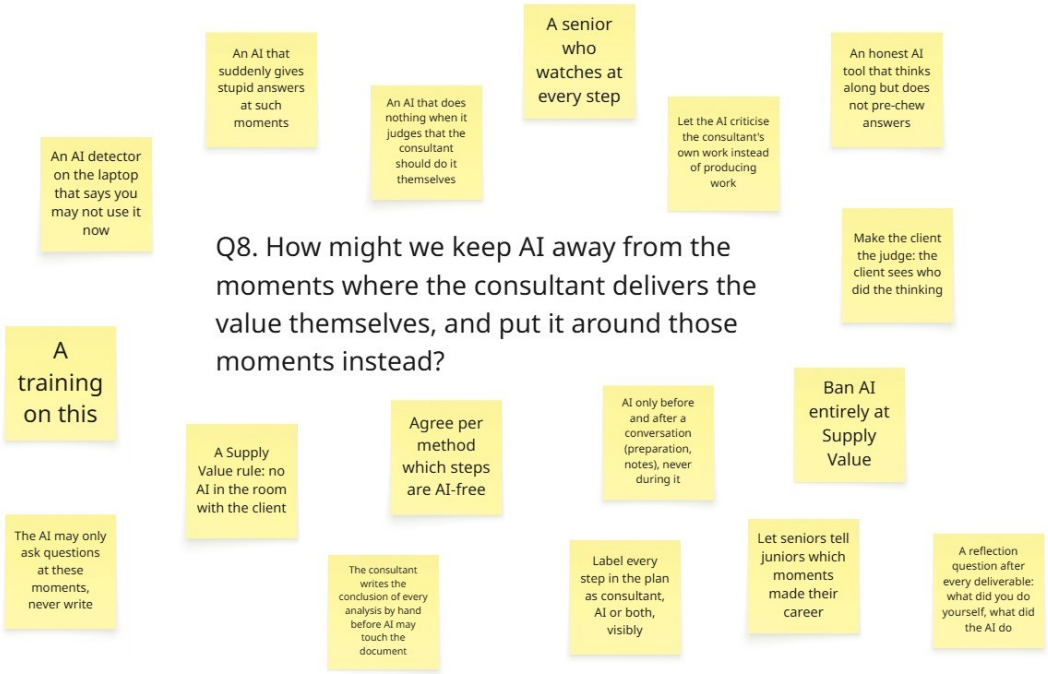


Figure H.8 Ideation board, question 8.

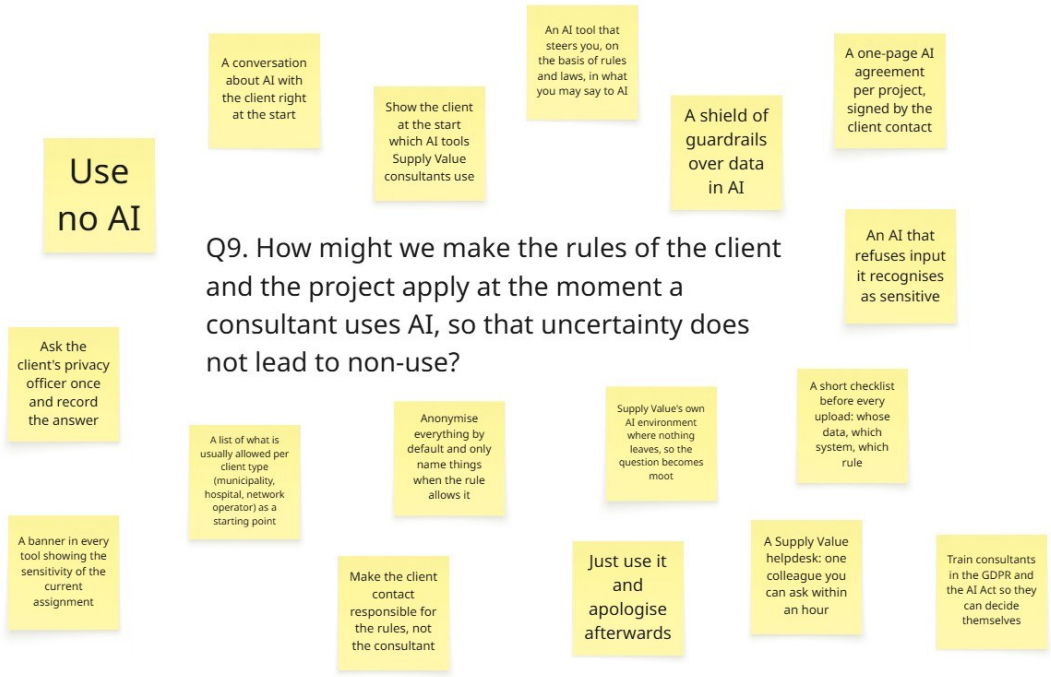


Figure H.9 Ideation board, question 9.

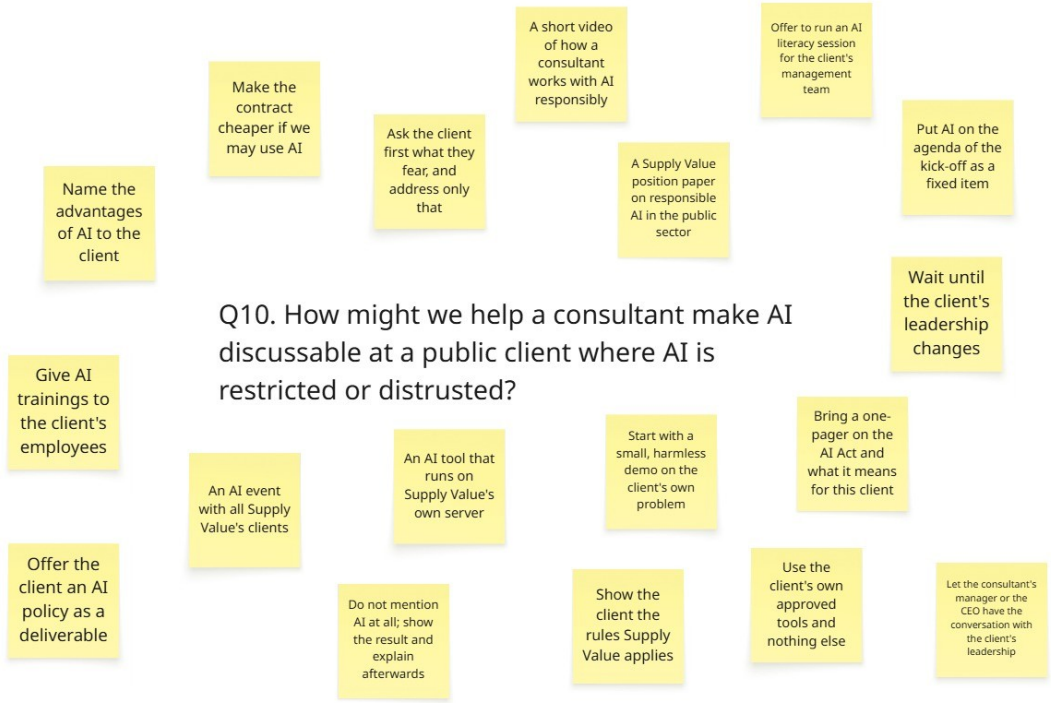


Figure H.10 Ideation board, question 10.



Figure H.11 Ideation board, question 11.

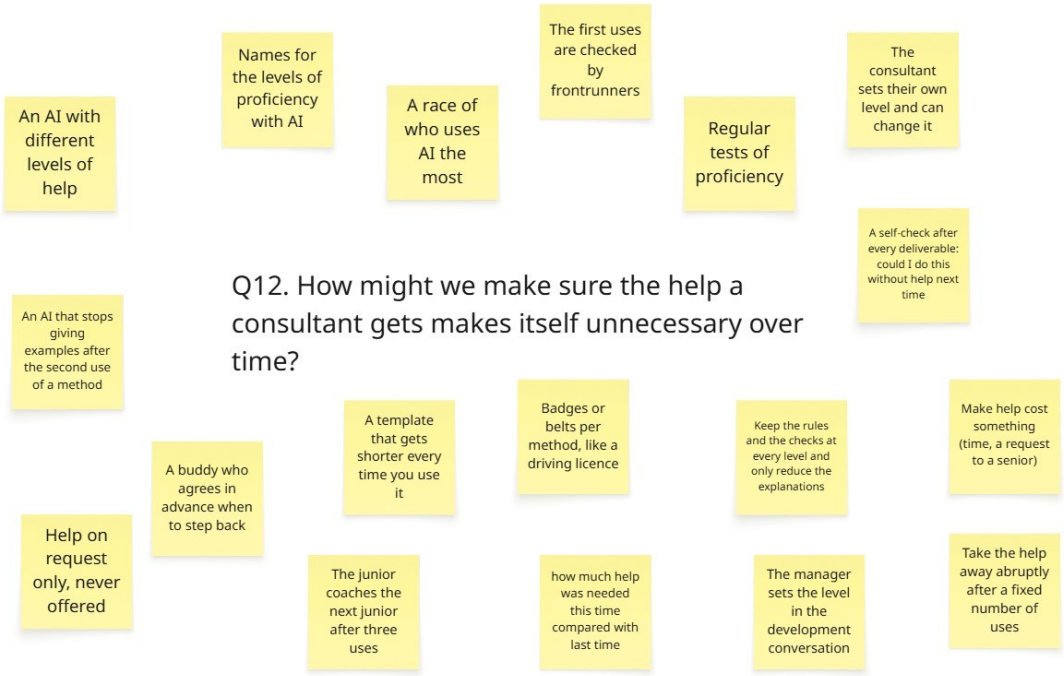


Figure H.12 Ideation board, question 12.

H.2 Full idea lists

All 262 ideas as generated, per question, in the order in which they were written down. F numbers are the researcher's own first answers; the other numbers are the extension of the list to at least twenty. No idea was removed from the list; the selection is made in H.3.

Q1 How might we help a consultant build the context of a new assignment faster and more completely?

F1 An AI bot that asks critical questions about the frame of the assignment; F2 A stakeholder meeting with everyone involved in the assignment, to answer all the consultant's questions; F3 A training in problem framing; F4 A buddy system: a senior helps the junior frame; F5 A ready-made intake form that is run through at every assignment; 6 Go straight to the client's CEO with all your questions; 7 Sit in the corridor at the client for the first week and talk to everyone who passes; 8 A survey to all the client's employees; 9 Call the client's previous consultant or adviser; 10 Accept the assignment only once the client has delivered a dossier; 11 A senior walks along for the first two weeks; 12 A kick-off pack that Supply Value sends to the client before the start; 13 Throw all the client's documents into ChatGPT; 14 Ask for a tour of the organisation; 15 Read the annual report, the budget and the audit office reports before the first conversation; 16 An AI that reads the engagement letter and asks questions; 17 An organisation chart with faces on the wall; 18 Use the daily start to put your questions to colleagues; 19 A video or podcast by the client sponsor: why this assignment exists; 20 A day on the client's work floor; 21 Lunch with the client's team in the first week; 22 A fixed list of twenty questions asked at every assignment; 23 A shared Supply Value dossier per client; 24 A who-knows-what list within Supply Value: who has worked at which client or sector; 25 The client contact writes a "what you need to know before you start" memo; 26 An introduction mail from the sponsor to all stakeholders; 27 A stakeholder map drawn with the client contact on a whiteboard; 28 Two consultants in the first week, then one; 29 The first week is not billed; 30 An AI builds a profile of the organisation from public sources, without client data; 31 An AI asks after every conversation: what did you learn, what do you still not know; 32 A checklist per sector: what you always need to know at a municipality, a care institution, a network operator; 33 A speed-date morning with five stakeholders, fifteen minutes each; 34 Daily start: who knows someone at this client; 35 A plan-of-approach template with the context questions built in; 36 Write down your own assumptions before the first conversation and test them after two weeks; 37 A senior calls the sponsor to hear the question behind the engagement letter; 38 Ask for the previous advisory report on the same subject; 39 A hand-over from the colleague who worked at this client before; 40 Half an hour of context sparring in the daily start; 41 A fixed context week in the plan of approach; 42 Preparation per stakeholder conversation, as a template or by AI.

Q2 How might we bring the coordination work around an assignment (actions from mail, meetings and the client's board) into one place?

F1 Connect all tools and software, mail included, to one AI tool; F2 Throw all the data from mail and other tools into ChatGPT; F3 Write everything down in one Word document; F4 Buy a notebook and drop everything in it; F5 An AI tool that operates your laptop and takes everything over itself; F6 Remember everything in your head; 7 One Planner or Trello board per assignment, agreed with the client at the start; 8 Fifteen minutes at the end of every day to move all actions to one list; 9 Ask the client to give the consultant an account in its own task system; 10 A fixed weekly action list sent to the client contact after every progress meeting; 11 A meeting notes template with an actions block, used in every meeting; 12 Record meetings and let AI extract the actions; 13 An AI that reads your mailbox and proposes actions, which you accept or reject; 14 Let the client's project secretary keep the list; 15 Work only from the client's board and ignore Supply Value's own tools; 16 Agree with the client that every action goes through one channel only; 17 Carry one whiteboard between the client's office and home; 18 Daily stand-up of ten minutes with the client team, actions on the board; 19 An assignment dashboard in which actions, agenda and open questions sit together; 20 Replace the failing link by a manual rule: every mail that contains an action is forwarded to one address.

Q3 How might we make sure a consultant frames the client problem themselves before AI does anything with it?

F1 A problem framing training; F2 An AI problem framer agent; F3 Hire a coach; F4 Take a course in problem framing; F5 A fixed question list for better framing; F6 Send the problem to a senior colleague; F7 Ask the client whether the problem is formulated correctly; 8 Write the problem statement by hand on one page before opening any AI tool; 9 An AI that refuses to answer until the consultant has typed their own problem statement; 10 Explain the problem to a colleague who knows nothing about the client, in two minutes; 11 A framing canvas: what, for whom, how big, who decides, what is fixed; 12 Present the framing to the client contact in week 2 and let them correct it; 13 Give the consultant three days without AI access at the start of every problem; 14 Write three different framings of the same problem and choose one; 15 A senior reviews every problem statement before the plan is made; 16 Reverse roles: let the junior ask the AI to criticise the junior's framing, not to write one; 17 The framing becomes the first slide of every deliverable, so it has to exist; 18 Use the five-why or problem tree method before any analysis; 19 Ask the AI only closed questions about the framing ("is the decision maker named?"), never open ones; 20 Make the framing a formal gate in Supply Value's plan of approach: no approach without a signed-off problem statement.

Q4 How might we help a consultant express the urgency of a problem in figures, so that a management team can decide on it?

F1 Analyse the figures about the problem; F2 Have a presentation made; F3 Make up the data; F4 Search thoroughly for the consequences of the problem; F5 Ask the client for figures about the problem; F6 Ask the client whether it is urgent; 7 Take the figures from the audit report, the annual report or the budget that already exist; 8 A one-page business case template: cost of the problem, cost of doing nothing, cost of the measure; 9 Ask three stakeholders to estimate the size of the problem and show the spread; 10 Count it yourself for one week: how many contracts, hours, complaints, errors; 11 An AI that asks "how would the management team notice the problem is solved?" until there is a number; 12 Benchmark against a comparable organisation

or sector figure; 13 Let the client's controller calculate it; 14 Express urgency in a date instead of a figure: what breaks, and when; 15 Use the client's own KPIs and show where the problem shows up in them; 16 A senior sanity-checks every figure before it goes to the management team; 17 Build a simple spend or process model in Excel from the client's data; 18 Ask the AI to list which figures a management team would ask for, then go and find them; 19 Tell a story of one case (one contract, one citizen) and then scale it up with figures; 20 Put the problem on the management team's agenda without figures and see if they ask for them.

Q5 How might we make sure that every solution a consultant proposes falls within what the client's decision maker can actually decide?

F1 Investigate who the decision makers are; F2 Determine who takes which decisions; F3 Make a stakeholder analysis; F4 A checklist of how realistic a decision is; F5 Make a requirement list; 6 Ask the decision maker at the start: what can you decide alone, what needs the board or the council, what is off the table; 7 Record what is fixed on one page: law, budget, systems, contracts, policy; 8 A RACI per problem, filled in with the client contact; 9 Read the mandate scheme or the delegation rules of the organisation; 10 Test every solution direction against the fixed list before working it out; 11 An AI that flags solutions outside the recorded scope ("this needs a new system; the client has just implemented one"); 12 Present three options to the decision maker, all within their mandate, instead of one ideal; 13 Ask a senior who has worked with this type of client what a management team will never decide; 14 Rehearse the management team meeting with a colleague playing the sceptical director; 15 Propose the solution that falls outside the mandate anyway, and let the decision maker reject it; 16 Ask the client's controller or legal adviser to check the options before they go up; 17 Phase the solution: what can be decided now within the mandate, what needs a later decision; 18 Write the proposal as the decision text the decision maker would sign; 19 Keep a list of solutions rejected by this client before and why; 20 Ignore the mandate and propose what is best for the organisation.

Q6 How might we make structured methods available to a consultant without asking them to choose a method before the problem is framed?

F1 Make an AI method coach; F2 A handbook with methods; F3 Let ChatGPT decide; F4 Ask a senior colleague; F5 Four standard methods that must always be used; F6 Take a method course; 7 A decision tree: from the type of problem to the method; 8 A poster on the wall with the methods per phase of an assignment; 9 A method of the month in the business unit meeting; 10 Attach a method to every deliverable template, so the method comes with the output; 11 An AI that proposes a method from the confirmed problem statement and explains why; 12 Supply Value's own method book with the ten methods it actually uses; 13 Let the client choose from a menu of approaches; 14 Worked examples from earlier assignments, searchable by problem, not by method; 15 A method card deck on the desk; 16 Pair every junior with a senior for the first use of a method; 17 No methods at all: let the consultant work from experience as now; 18 A plan-of-approach template in which each phase lists the methods that fit; 19 Use the Delft Design Guide as Supply Value's standard reference; 20 A method sparring session in the daily start: "I have this problem, what would you do".

Q7 How might we make what a consultant learns on an assignment (about the client, the problem and what worked) available to the next assignment and the next consultant?

F1 Store the results; F2 A client dossier at the end of the assignment; F3 A billboard of colleagues who have used which methods; F4 Write down at the end of an assignment what you learned; F5 An AI tool in which everything you did in the assignment is recorded; 6 A fifteen-minute debrief with a colleague at the end of every assignment, recorded; 7 Record decisions, assumptions and open points during the work, not afterwards; 8 A one-page assignment summary in a fixed format, kept in one shared place; 9 Present the assignment in the business unit meeting when it ends; 10 The consultant who starts at a client calls the one who left; 11 A Supply Value wiki per sector and per type of problem; 12 Keep every deliverable in one library, searchable by client type and problem; 13 Nothing: accept that knowledge leaves with people; 14 Make the end report for the client also the internal hand-over; 15 Tag every document with the method and the problem type so it can be found later; 16 An AI that summarises the whole assignment and proposes what is reusable; 17 A buddy from the next assignment joins the last two weeks of the current one; 18 A quarterly "what did we learn" session across business units; 19 Put the lessons in the plan-of-approach template, so the next plan starts from them; 20 Pay consultants for the time to write the hand-over.

Q8 How might we keep AI away from the moments where the consultant delivers the value themselves, and put it around those moments instead?

F1 An AI that does nothing when it judges that the consultant should do it themselves; F2 An AI that suddenly gives stupid answers at such moments; F3 An AI detector on the laptop that says you may not use it now; F4 A senior who watches at every step; F5 An honest AI tool that thinks along but does not pre-chew answers; F6 A training on this; 7 Agree per method which steps are AI-free (the interview itself, the conversation with the decision maker); 8 AI only before and after a conversation (preparation, notes), never during it; 9 A Supply Value rule: no AI in the room with the client; 10 The consultant writes the conclusion of every analysis by hand before AI may touch the document; 11 A "do it yourself first" button that hides the AI panel for a step; 12 Let the AI ask questions only and never write at these moments; 13 Label every step as consultant / AI / both, visibly in the plan; 14 Reward in capability conversations the work done without AI, not the volume of AI use; 15 Let seniors tell juniors which moments made their career, and keep those AI-free; 16 Make the client the judge: the client sees who did the thinking; 17 Ban

AI entirely at Supply Value; 18 A reflection question after every deliverable: what did you do yourself, what did the AI do; 19 Let the AI criticise the consultant's own work instead of producing work; 20 Nothing: trust consultants to know when to use it.

Q9 How might we make the rules of the client and the project apply at the moment a consultant uses AI, so that uncertainty does not lead to non-use?

F1 A conversation about AI with the client right at the start; F2 Show the client at the start which AI tools Supply Value consultants use; F3 Use no AI; F4 An AI tool that steers you, on the basis of rules and laws, in what you may say to AI; F5 A shield of guardrails over data in AI; 6 A one-page AI agreement per project, signed by the client contact; 7 A traffic light per data type: green may go in, orange anonymised, red never; 8 A banner in every tool showing the sensitivity of the current assignment; 9 Anonymise everything by default and only name things when the rule allows it; 10 Supply Value's own AI environment where nothing leaves, so the question becomes moot; 11 Ask the client's CISO or privacy officer once and record the answer; 12 A short checklist before every upload: whose data, which system, which rule; 13 An AI that refuses input it recognises as sensitive; 14 A list of what is allowed per client type (municipality, hospital, network operator) as a starting point; 15 Use only the client's own AI tool when it has one; 16 Make the client contact responsible for the rules, not the consultant; 17 A Supply Value helpdesk: one colleague you can ask within an hour; 18 Train consultants in the GDPR and the AI Act so they can decide themselves; 19 Agree the rules in the engagement letter; 20 Just use it and apologise afterwards.

Q10 How might we help a consultant make AI discussable at a public client where AI is restricted or distrusted?

F1 Name the advantages of AI to the client; F2 Make the contract cheaper if we may use AI; F3 An AI tool that runs on Supply Value's own server; F4 Give AI trainings to the client's employees; F5 An AI event with all Supply Value's clients; 6 Start with a small, harmless demo on the client's own problem; 7 Bring a one-pager on the AI Act and what it means for this client; 8 Do not mention AI at all; show the result and explain afterwards how it was made; 9 Let a peer organisation (another municipality) tell what it does with AI; 10 Offer the client an AI policy as a deliverable; 11 Ask the client first what they fear, and address that only; 12 Show the client the rules Supply Value applies, before they ask; 13 Use the client's own approved tools and nothing else; 14 A Supply Value position paper on responsible AI in the public sector; 15 Let the consultant's own manager or management have the conversation with the client's leadership; 16 Put AI on the agenda of the kick-off as a fixed item; 17 Accept the no and work without AI at that client; 18 Make a short video of how a consultant works with AI responsibly; 19 Offer to run an AI literacy session for the client's management team; 20 Wait until the client's leadership changes.

Q11 How might we make a first attempt with AI safe and worthwhile for a consultant who is fully booked at a client, so that one failure does not end it?

F1 An AI coach that solves a consultant's problem or question together with AI; F2 An AI builder agent that helps with the use of AI; F3 Fixed, mandatory moments to do something with AI; F4 Send consultants to clients less; F5 Oblige consultants without an assignment to learn more AI; 6 A colleague sits next to you for the first hour, on your own task; 7 Pick the first task together: something that costs you an hour every week; 8 Half a day per month, booked and not billable, for a first AI task; 9 An AI hour that moves to the client site or online, instead of the office; 10 A list of ten first tasks that worked for other consultants, with the prompt; 11 Share failed attempts in the daily start, and reward the sharing; 12 Retry the failed task of P7 together, on a current model, and show the difference; 13 Make the first attempt a pair task: a frontrunner and a waiter; 14 A thirty-minute "bring your own task" session every two weeks; 15 Nothing mandatory: wait until the consultant asks; 16 A WhatsApp or Teams channel where you can ask a frontrunner a quick question; 17 Put the first AI task in the onboarding of every new consultant; 18 Let the manager ask in the development conversation what the consultant tried; 19 Give every consultant a budget for an AI tool of their own choice; 20 Make the first attempt about the consultant's own hours, not the client's problem.

Q12 How might we make sure the help a consultant gets makes itself unnecessary over time?

F1 An AI with different levels of help; F2 Give names to the levels of proficiency with AI; F3 A race of who uses AI the most (Ferhat doubts this one); F4 The first uses are checked by frontrunners; F5 Regular tests of proficiency; 6 A buddy who agrees in advance when to step back; 7 A template that gets shorter every time you use it; 8 An AI that stops giving examples after the second use of a method; 9 The third time, the consultant does the step alone and has it checked afterwards; 10 Badges or belts per method, like a driving licence; 11 The consultant sets their own level and can change it; 12 Nothing: let people decide for themselves when they no longer need help; 13 Help on request only, never offered; 14 Let the junior coach the next junior after three uses; 15 A self-check after every deliverable: could I do this without help next time; 16 Measure independence, not use: how much help was needed this time compared with last time; 17 Take the help away abruptly after a fixed number of uses; 18 The manager decides the level in the development conversation; 19 Make help cost something (time, a request to a senior) so people stop asking when they can; 20 Keep the rules and the checks at every level and only reduce the explanations.

H.3 Directions and ideas set aside

Per question, the directions into which the ideas of H.2 were grouped (section 4.2.4), with their type, the ideas they contain and what they cover, followed by the ideas that were set aside and the reason.

Q1 How might we help a consultant build the context of a new assignment faster and more completely?

Source: Figure 3.3, phase 2; cluster 4. 42 ideas.

Direction	Type	Ideas	What it covers
1a Senior and peer support in the first weeks	People	F4, 11, 18, 34, 37, 39, 40	Buddy, senior walks along, hand-over from the predecessor, context sparring and questions in the daily start.
1b Structured intake and self-framing instruments	Aid	F5, 17, 22, 27, 32, 35, 36	Intake form, twenty questions, sector checklist, plan template with context questions, own assumptions written down, stakeholder map on a whiteboard.
1c Agreements with the client at the start	Process at client	F2, 12, 14, 19, 20, 21, 25, 26, 33, 41	Kick-off with all stakeholders, kick-off pack, memo and introduction mail from the sponsor, tour, work-floor day, speed-dates, a context week in the plan.
1d Desk research before the first conversation	Aid / AI	9, 15, 30, 38	Public documents first; a profile from public sources; the previous report; the previous adviser.
1e AI that structures and questions the assignment context	AI	F1, 16, 31, 42	Reads the engagement letter, asks what is missing, prepares and debriefs conversations.
1f Knowledge about the client within Supply Value	Aid / People	23, 24	Client dossier and who-knows-what list; taken further in Q7.
1g Learning	Learning	F3	Training in problem framing; taken further in Q3.

Table H.1 Directions for question 1.

Set aside	Reason
6 straight to the CEO; 7 the corridor; 8 a survey to all employees	Burden the client and the relationship before anything has been delivered; outside what a junior can do in week 1.
10 accept only with a dossier	Not in Supply Value's power; clients do not have one.
13 all documents into ChatGPT	Cluster 3 and DR13: client data to an external system without agreement.
28 two consultants; 29 first week unbilled	Cost decisions for management; not a design direction.

Table H.2 Ideas set aside for question 1.

Q2 How might we bring the coordination work around an assignment into one place?

Source: Journey map A, moment of truth MT-A1, opportunity A/02. 20 ideas.

Direction	Type	Ideas	What it covers
2a One agreed channel or board with the client	Process at client	7, 9, 14, 15, 16, 18	One board per assignment, an account in the client's system, one channel only, a daily stand-up, the client's secretary.
2b Personal discipline and formats	Aid	F3, F4, 8, 10, 11, 20	One document or notebook, fifteen minutes a day, a weekly action list, a notes template with an actions block, a forwarding rule.
2c An assignment workspace that holds actions, agenda and open questions together	Aid / AI	19	The place where the actions of all assignments meet; later the workspace of the support.
2d AI that extracts actions from mail and meetings	AI	F1, 12, 13	Proposes actions from mail and recordings; the consultant accepts or rejects.

Table H.3 Directions for question 2.

Set aside	Reason
F2 everything into ChatGPT	Cluster 3 and DR13.

Set aside	Reason
F5 an AI that operates the laptop	Autonomy over the consultant's work beyond DR8; data of several clients on one machine.
F6 remember everything; 17 carry a whiteboard	The situation of map A today.

Table H.4 Ideas set aside for question 2.

Q3 How might we make sure a consultant frames the client problem themselves before AI does anything with it?

Source: Cluster 1; A4; Trend 3. 20 ideas.

Direction	Type	Ideas	What it covers
3a Learning: training, course, coach	Learning	F1, F3, F4	Framing as a skill taught outside the work.
3b Senior and peer check of the framing	People	F6, 10, 15	A senior reviews every problem statement; explain it to an outsider in two minutes.
3c Framing instruments	Aid	F5, 8, 11, 14, 18	Question list, framing canvas (what, for whom, how big, who decides, what is fixed), three framings, five-why, one page by hand.
3d Framing as a formal gate in the process	Process	F7, 12, 17, 20	No approach without a signed-off problem statement; the framing confirmed with the client; the framing as the first slide.
3e AI that asks and criticises but does not write the framing	AI	F2, 9, 16, 19	Refuses to answer until the consultant has typed a statement; criticises the consultant's framing; closed questions only.

Table H.5 Directions for question 3.

Set aside	Reason
13 three days without AI access	Cannot be enforced and punishes rather than teaches; its intent returns in 3d and 3e.

Table H.6 Ideas set aside for question 3.

Q4 How might we help a consultant express the urgency of a problem in figures, so that a management team can decide on it?

Source: Session, tool B; DR3. 20 ideas.

Direction	Type	Ideas	What it covers
4a Use the figures that already exist at the client	Process at client	F5, 7, 13, 15, 17	Audit report, annual report, budget, KPIs, the controller, the client's own data in a simple model.
4b Gather or estimate figures yourself	Aid	F1, F4, 9, 10, 12	Count for a week, three estimates and the spread, a benchmark.
4c Urgency formats	Aid	8, 14, 19	One-page business case, urgency as a date, one case scaled up.
4d Check and challenge	People	F6, 16, 20	A senior checks every figure; ask the client; test it on the agenda.
4e AI that asks for the number and for what the management team will ask	AI	F2, 11, 18	Keeps asking until there is a figure; lists the questions a management team asks; builds the presentation from the confirmed figures.

Table H.7 Directions for question 4.

Set aside	Reason
F3 make up the data	Integrity; guardrail G1 and the client's trust.

Table H.8 Ideas set aside for question 4.

Q5 How might we make sure that every solution a consultant proposes falls within what the client's decision maker can actually decide?

Source: Session (C1 examples); DR14; cluster 3. 20 ideas.

Direction	Type	Ideas	What it covers
5a Map the decision makers and their mandate at the start	Process at client	F1, F2, F3, 6, 8, 9	Stakeholder analysis, RACI, the mandate scheme, ask the decision maker what they can decide alone.
5b Record what is fixed and test every direction against it	Aid	F4, F5, 7, 10, 19	One page of law, budget, systems, contracts, policy; a realism checklist; the list of what this client rejected before.
5c Shape the proposal for the decision maker	Aid	12, 17, 18	Three options within the mandate; phased decisions; the decision text itself.
5d Senior and client checks	People	13, 14, 16	A senior who knows the client type; a rehearsal with a sceptical director; the controller or legal adviser.
5e AI that flags directions outside the recorded scope	AI	11	Needs 5a and 5b as input; cannot know the mandate by itself.

Table H.9 Directions for question 5.

Set aside	Reason
15 propose it anyway and let them reject it; 20 ignore the mandate	Contradicts DR14 and the session's main finding; 15 can be a learning moment but not a design.

Table H.10 Ideas set aside for question 5.

Q6 How might we make structured methods available to a consultant without asking them to choose a method before the problem is framed?

Source: 4.1.7; theme 4; DR1. 20 ideas.

Direction	Type	Ideas	What it covers
6a Method knowledge in documents	Aid	F2, 8, 12, 15, 19	Handbook, Supply Value method book, poster, card deck, the Delft Design Guide.
6b Methods linked to the problem or the phase, not browsed	Aid	7, 10, 14, 18	Decision tree from problem type, deliverable templates that carry a method, worked examples searchable by problem, plan template per phase.
6c People: senior, sparring, method of the month, pairing	People	F4, 9, 16, 20	Method knowledge from colleagues, which is how it travels today (theme 5).
6d Learning	Learning	F6	A method course.
6e AI that proposes a method from the framed problem and explains why	AI	F1, 11	The method follows the framing; the library stays behind the conversation.

Table H.11 Directions for question 6.

Set aside	Reason
F3 let ChatGPT decide	A general model gives the average of what is written about a method, without Supply Value's practice (4.1.6).
F5 four standard methods always	Contradicts DR1: the problem decides the method.
13 let the client choose	Theme 4: the client wants results, not method terms.
17 no methods at all	The current situation; keeps method knowledge with people (cluster 4).

Table H.12 Ideas set aside for question 6.

Q7 How might we make what a consultant learns on an assignment available to the next assignment and the next consultant?

Source: Cluster 4; theme 5; figure 3.3 phase 6; DR11. 20 ideas.

Direction	Type	Ideas	What it covers
7a Capture during the work, not afterwards	Aid / AI	F1, F5, 7, 16	Decisions, assumptions and open points recorded as they happen; an AI summary of what is reusable.
7b Hand-over at the end	Aid	F2, F4, 6, 8, 14, 20	Client dossier, lessons, debrief, one-page summary, the end report as hand-over, paid time to write it.
7c A shared, findable library	Aid	11, 12, 15	Wiki per sector and problem type; deliverables tagged by method and problem.
7d People and moments	People	F3, 9, 10, 17, 18	Present at the business unit meeting, call the predecessor, a buddy in the last weeks, a quarterly session, the billboard of who used what.
7e Feed lessons into the next plan	Aid	19	The plan-of-approach template starts from recorded lessons.

Table H.13 Directions for question 7.

Set aside	Reason
13 nothing, accept the loss	The current situation (cluster 4).

Table H.14 Ideas set aside for question 7.

Q8 How might we keep AI away from the moments where the consultant delivers the value themselves, and put it around those moments instead?

Source: Moments of truth MT-A2, MT-B2; cluster 1. 20 ideas.

Direction	Type	Ideas	What it covers
8a Agree which steps are AI-free, per method and in the plan	Process	7, 8, 9, 13	The interview, the conversation with the decision maker, the room with the client; AI before and after.
8b AI that asks, criticises and withholds, but does not produce at these moments	AI	F1, F5, 11, 12, 19	Thinks along without pre-chewing; a do-it-yourself-first step; questions only.
8c The consultant does it by hand first	Aid	10	The conclusion written before AI touches the document; links to Q3.
8d People as the judge	People	F4, 15, 16	A senior watches; seniors name the moments that made their career; the client sees who did the thinking.
8e Reflection and what is rewarded	Learning	F6, 14, 18	Training on this; the development conversation rewards own work, not AI volume; a reflection question per deliverable.

Table H.15 Directions for question 8.

Set aside	Reason
F2 an AI that gives stupid answers on purpose	Deceptive; destroys the trust the tool needs (theme 9).
F3 a detector on the laptop	Surveillance; cluster 5 says learning comes from guided practice, not control.
17 ban AI; 20 do nothing	The two ends the assignment rules out.

Table H.16 Ideas set aside for question 8.

Q9 How might we make the rules of the client and the project apply at the moment a consultant uses AI, so that uncertainty does not lead to non-use?

Source: Cluster 3; A3; themes 7, 8; DR9. 20 ideas.

Direction	Type	Ideas	What it covers
9a Agree the rules with the client at the start and record them	Process at client	F1, F2, 6, 11, 16, 19	A conversation at the start, a one-page agreement, the privacy officer asked once, rules in the engagement letter.
9b Make the rules visible and simple at the moment of use	Aid	7, 8, 12, 14	Traffic light per data type, sensitivity banner, upload checklist, a starting list per client type.

Direction	Type	Ideas	What it covers
9c Technical guardrails	AI / tech	F4, F5, 9, 10, 13, 15	Anonymise by default, an AI that refuses sensitive input, Supply Value's own environment, the client's own tool.
9d People and knowledge	People / Learning	17, 18	A helpdesk colleague within the hour; training in GDPR and the AI Act.

Table H.17 Directions for question 9.

Set aside	Reason
F3 use no AI	The non-use the question wants to prevent.
20 use it and apologise	A breach of cluster 3 and of the client's trust.

Table H.18 Ideas set aside for question 9.

Q10 How might we help a consultant make AI discussable at a public client where AI is restricted or distrusted?

Source: Moment of truth MT-A3; theme 8; section 2.4. 20 ideas.

Direction	Type	Ideas	What it covers
10a Be transparent about Supply Value's own AI practice	Process at client	F2, 7, 12, 18	Show which tools and rules Supply Value applies, a one-pager on the AI Act, a short video.
10b Start small and show, rather than argue	Process at client	F1, 6, 8	A harmless demo on the client's own problem; the result first, the explanation after.
10c Make AI part of the agenda and of the offer	Process / Offer	F4, 10, 16, 19	AI on the kick-off agenda; an AI policy or literacy session as a deliverable; training for client staff.
10d Peers and leadership	People	F5, 9, 15	Another municipality tells its story; a client event; management talks to the client's leadership.
10e Work inside what the client allows	Tech	F3, 13	The client's approved tools; a tool on Supply Value's own server.

Table H.19 Directions for question 10.

Set aside	Reason
F2 a cheaper contract if AI may be used	A commercial decision for management (A5), not a capability direction.
17 accept the no; 20 wait for new leadership	Fallbacks, not designs; 17 remains the honest answer at some clients (theme 8).

Table H.20 Ideas set aside for question 10.

Q11 How might we make a first attempt with AI safe and worthwhile for a consultant who is fully booked at a client, so that one failure does not end it?

Source: Cluster 2; moments of truth MT-B1, MT-B3; assumptions A1, A2. 20 ideas.

Direction	Type	Ideas	What it covers
11a The first task is the consultant's own, chosen together, from tasks that worked for others	Aid	7, 10, 12, 20	Something that costs an hour a week; a list of ten proven first tasks with the prompt; P7's failed task retried on a current model.
11b A colleague beside you	People	6, 13, 16	A frontrunner and a waiter as a pair; a channel for quick questions.
11c Time and moments that reach fully booked consultants	Process	F3, 8, 9, 14	Booked time, the AI hour at the client site or online, bring-your-own-task sessions.
11d Failure is normal	People	11	Failed attempts shared in the daily start and rewarded.
11e Manager and onboarding	Process	17, 18	The first AI task in onboarding; the manager asks what was tried.
11f AI support for the first use	AI	F1, F2	A coach that solves the consultant's own question together with them; a builder agent.

Table H.21 Directions for question 11.

Set aside	Reason
F4 send consultants to clients less	Not possible for a consultancy that bills by the hour (A5).
F5 oblige the bench to learn AI	Compulsion; cluster 2 says the trigger is a use case on own work, not an order.
15 nothing; 19 a tool budget	The current situation; a budget is tool-first, which A2 corrected.

Table H.22 Ideas set aside for question 11.

Q12 How might we make sure the help a consultant gets makes itself unnecessary over time?

Source: Cluster 5; section 2.3.4; DR5. 20 ideas.

Direction	Type	Ideas	What it covers
12a Named levels of help, set by the consultant or the manager	Aid / AI	F1, F2, 11, 18	Levels of help with names; the consultant sets and changes their own level; the manager in the development conversation.
12b Fading built into the instrument	Aid / AI	7, 8, 17, 20	A template that gets shorter; no example the second time; explanations fade, rules and checks stay.
12c People who step back	People	F4, 6, 9, 14	Frontrunners check the first uses; a buddy agrees when to step back; the third time alone; the junior coaches the next junior.
12d Reflection and measuring independence rather than use	Learning	F5, 15, 16	A self-check per deliverable; how much help was needed compared with last time; tests of proficiency.
12e Help on request, or at a cost	Process	13, 19	Never offered, only asked for; asking a senior costs time.

Table H.23 Directions for question 12.

Set aside	Reason
F3 a race of who uses AI most; 10 badges per method	Reward use rather than capability (DR3, cluster 5); 10 is kept as a variant of 12a if it rewards independence.
12 nothing	The current situation.

Table H.24 Ideas set aside for question 12.

H.4 Motivation of the scores in the Harris profile

For each criterion of figure 4.3 the score per concept and the reason. Scores were first given by the researcher and then challenged against the evidence of chapters 2 to 4; the reason given is the one that stood after that pass.

Criterion	Scores and reasons
O Develops judgement rather than replacing it	A ++: guided practice on real problems is how the capability is learned (cluster 5), and a senior beside you is its most direct form; it happens only where a senior has time, and juniors at a client do not ask (P6), which counts under DR10 and DR12. B −: instruments make a step explicit but give no feedback, and judgement develops through feedback (section 2.3.4). C +: feedback on the consultant's own attempt at every step, but no knowledge of the organisation (cluster 1) and, without people around it, the delegation risk of Trend 3.
DR1 Work process, not tool adoption	A ++: described entirely in the work. B ++: the instruments are the work process made explicit, without an AI tool. C +: starts from the assignment and the problem, but the capability is described only through the tool; who does not open it has no description of what they should be able to do.
DR2 Levels, framing as entry condition	A −−: no levels; progression depends on the senior, whose own AI level varies (table 2.1). B −: instruments are the same for everyone and a gate is binary; only framing as a condition is present. C +: named levels of help that fade and framing as a hard gate, but the top level, building agents, lies outside a tool and is not described.

Criterion	Scores and reasons
DR3 Value at the client, not use	A +: seniors judge work by whether the client can decide on it, but there is no fixed assessment. B −: what is visible is whether the instruments were used, which is compliance. C −: a tool logs use and "how much is it used" is the obvious measure, which DR3 excludes and cluster 5 warns against; the tool does produce evidence of value, such as a confirmed problem statement, so not −−.
DR4 Validation as a skill (the rule is a hard requirement, checked in the first step)	A +: the senior validates the deliverable, not every output, and teaches judgement of content rather than where the model fails. B −: a checklist is a rule, not a skill, and nothing checks that it is followed. C +: every output is a draft that needs confirmation, which no other concept has; but passively presented AI output does not reduce over-reliance, while designs that make the person act first do (Buçinca et al., 2021), and whether it becomes a skill depends on how the checkpoint is built, which the concept does not yet decide.
DR5 Support fades	A +: the buddy agrees when to step back; depends on the senior, not recorded. B −: a template that gets shorter and help on request exist, but the consultant chooses and nothing tests whether they can. C ++: named levels per method that fade with repeated use; rules and checks do not fade, only explanations and examples.
DR6 AI literacy documented	A −: only if the senior has it, and AI use is not tied to seniority; nothing on paper. B −: the traffic light and the AI agreement teach what may go where and are documents, but no description of what a consultant must know. C −: experience on own work of what AI does well and badly, but nothing on paper and nothing for who does not open the tool.
DR7 Context across steps and time	A −: a weekly buddy holds context as people do today (theme 5), but has other assignments and is not there in week 6. B −: instruments record context but do not carry it; the consultant must re-read and connect. C ++: the differentiator from Copilot that both session teams named (section 4.1.6).
DR8 Asks before it answers	A +: depends on the senior; a senior under time pressure tells how it is done, and nothing in the concept enforces that the consultant frames first. B +: a canvas asks and never fills in, the purest form of asking first, and nothing enters the work the consultant did not write; but it does not help when the consultant is stuck, and theme 6 asks for a sparring partner. C ++: a rule in the concept, independent of a person, with both additions of the session.
DR9 Rules checked at use	A −: the senior may not know the rules and is not there at the moment of upload. B −: the agreement per project and the traffic light cover "agreed per client and project" fully, but nothing checks at use or stops a submission (cluster 3). C ++: checks and stops at use, with rules per client type as a starting point; the agreement itself is a conversation that no tool has, which is B's strength to carry into the design.
DR10 Own task first, no configuration	A +: a colleague beside you for the first hour on your own task is threshold-free, but depends on senior time. B +: the intake form on the own assignment and booked time. C +: the intake of the own assignment is the first task and pays back, but it is a new application to open and learn, creating an assignment feels like configuration to the waiting majority, and a tool gives nobody time (cluster 2).
DR11 Knowledge flows back	A −: knowledge flows through people as today; not recorded, not reusable (cluster 4). B +: hand-over format, library and lessons into the next plan are reusable form for Supply Value, but only at the end, and cluster 4 says knowledge is lost before the end. C +: captured during the work, with reusable outputs such as the presentation for the decision maker; but the flow back to Supply Value is not something a tool arranges.
DR12 Existing moments, time to learn	A +: daily start, business unit meeting and development conversation are the existing moments; senior time is billable too. B +: booked time, the AI hour at the client or online, onboarding; the instruments touch the moments less. C −−: a tool with nothing around it. None of the three redesigns one workflow per business unit.
DR13 Criterion part: no significant IT investment, usable at client sites (the agreements and data part is hard and was checked in the first step, where A passed on one condition and C on two)	A +: no technology, but the buddy arrangement needs senior time that is otherwise billable. B ++: templates and agreements cost nothing to deploy and work anywhere. C −: an own web application is an IT effort to deploy and maintain, and a client that restricts tools may block it; the route to + is a browser application without installation, data in the browser, and the model route through the firm's approved environment (section 4.3.3).

Criterion	Scores and reasons
DR14 Decision maker and mandate	A +: a senior who knows the client type is the best detector of organisationally wrong proposals, where AI is weak (section 2.3.1); not recorded, not systematic. B +: canvas, realism check and proposal format cover all three parts, but nothing flags actively. C ++: all three parts systematically, but only against what the consultant recorded; a model does not know where a mandate ends (section 4.1.7).

Table H.25 Motivation of the Harris profile of figure 4.3.

H.5 The morphological chart of section 4.2.4

Figure H.13 is the morphological chart of section 4.2.4: the directions per question and type, with the three concepts of section 4.3 drawn as paths A to C through it.

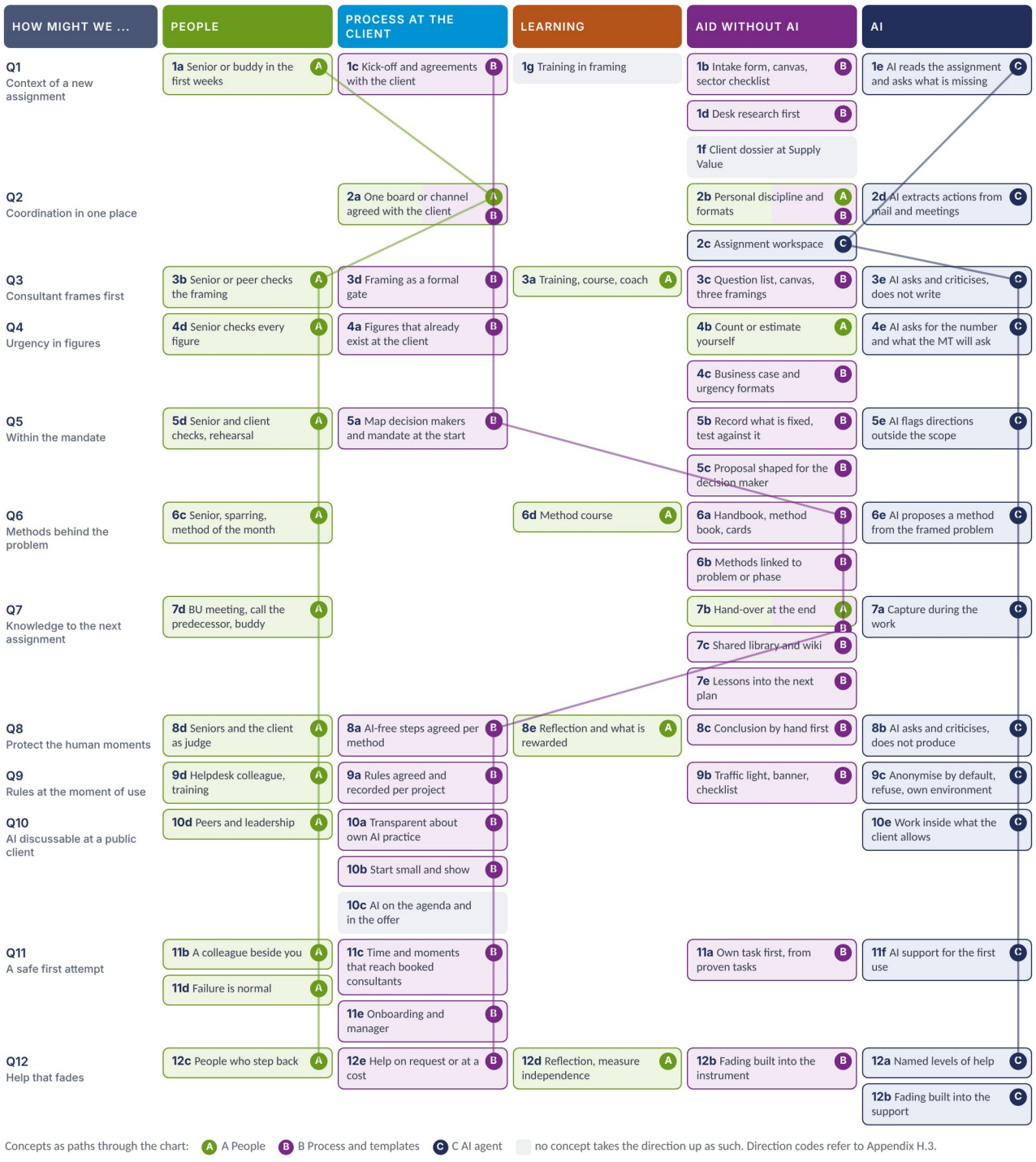


Figure H.13 Morphological chart: the directions per question and type, with the three concepts of section 4.3.1 marked as paths. Grey cells are directions that no concept takes up as such; direction codes refer to Appendix H.3.

Appendix I. Screens of the prototype

Screens of the Methode-coach as it was at the time of writing, on the fictional housing corporation case used in the tests of section 5.7. All names of people and organisations are fictional; the user shown is the author's test account, and any resemblance to real people or organisations is accidental. Figure 5.5 in the text shows a method step with the coach.

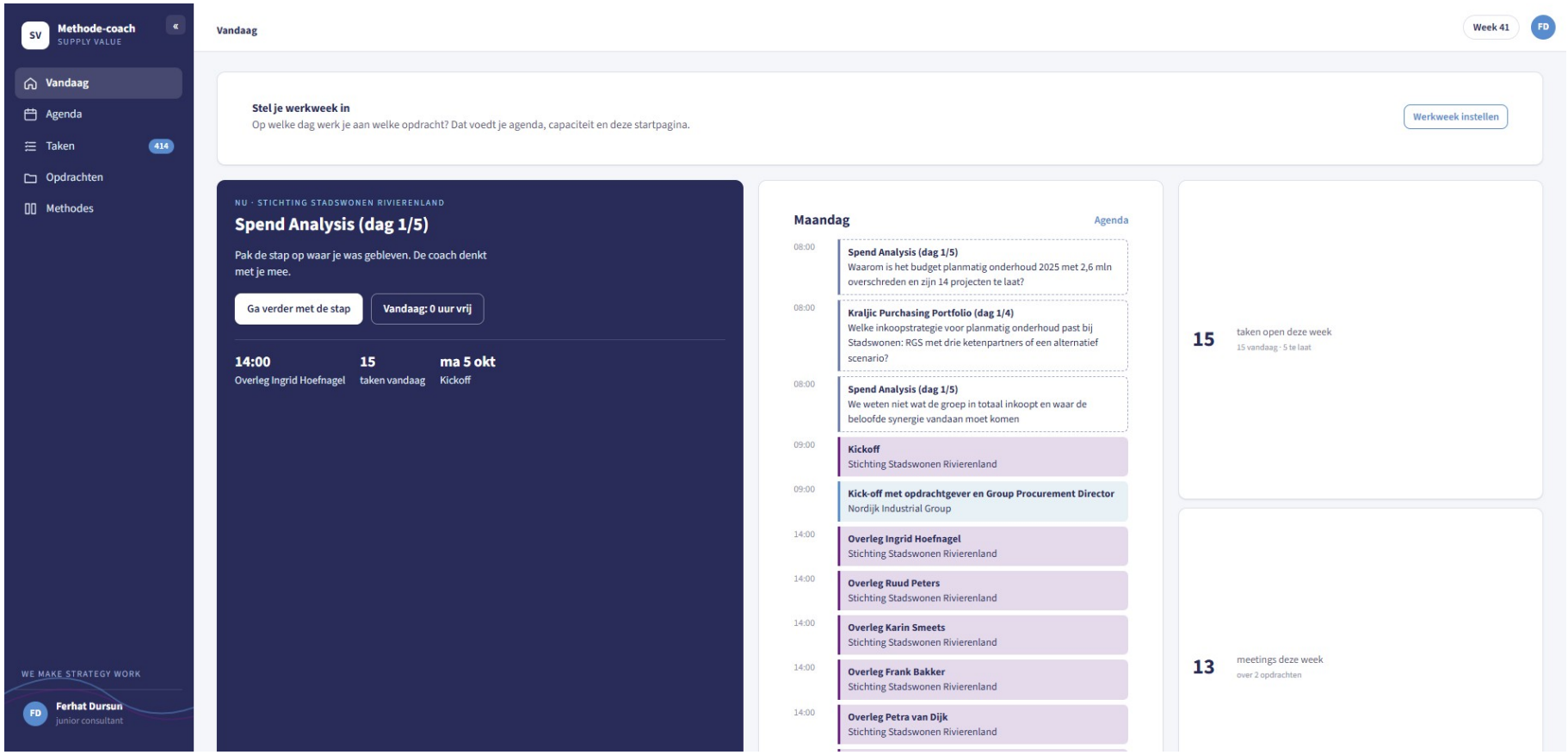


Figure I.1 The Today page: the current step across assignments, the agenda of the day and what needs attention.

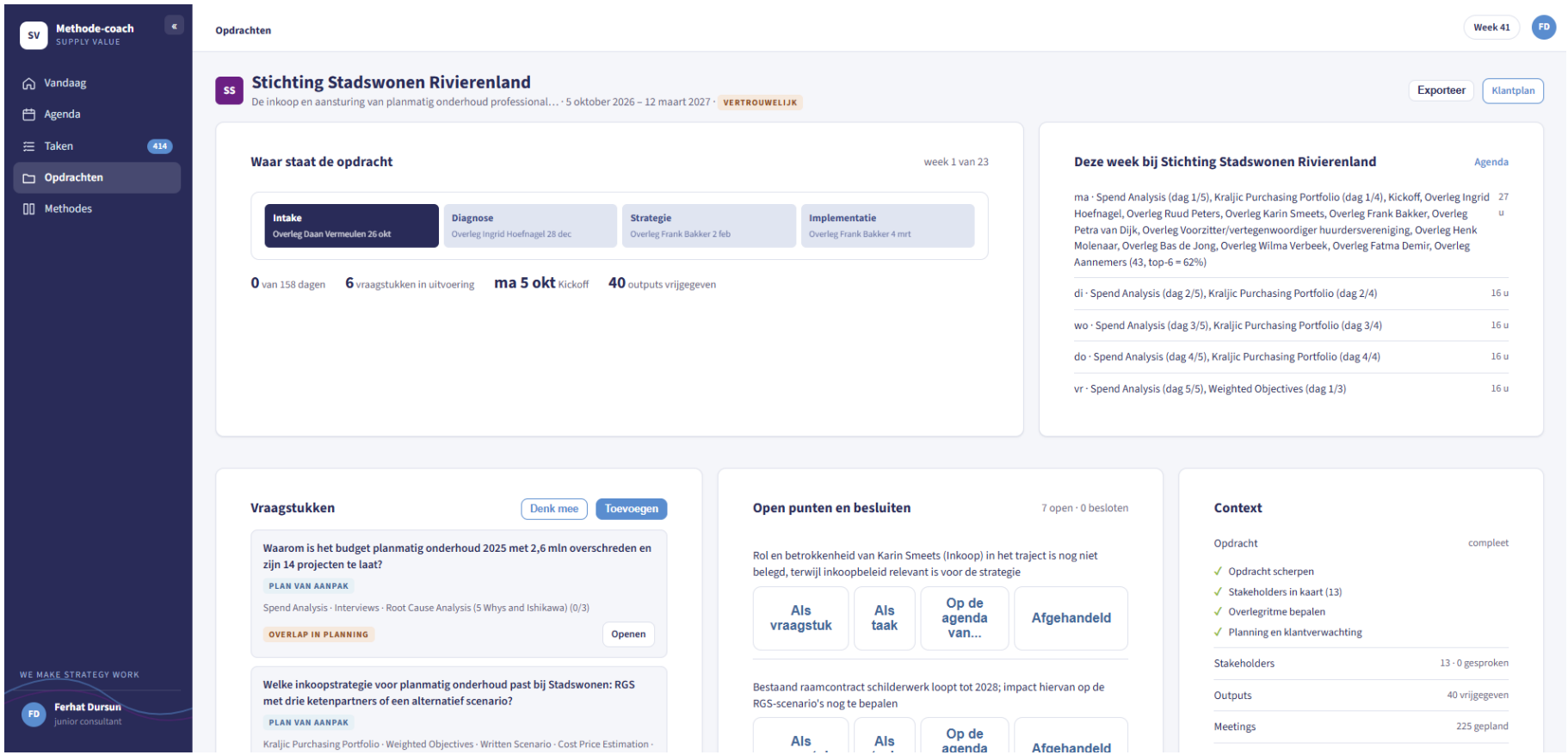


Figure I.2 An assignment: phase, milestones, the problems within it, open points and decisions, and the context the coach receives.

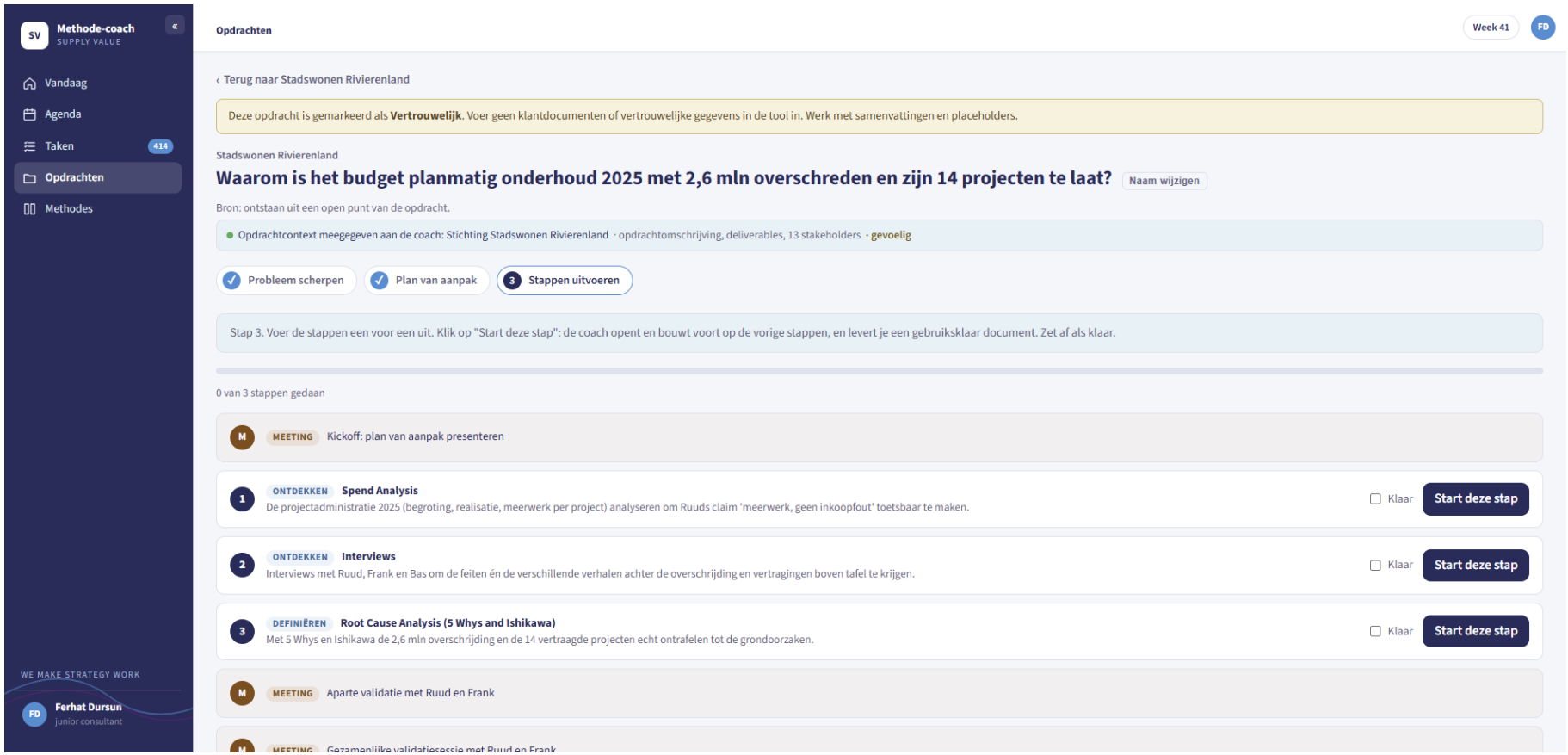


Figure I.3 A problem: problem sharpening and plan confirmed, the method steps of the plan with the meetings between them.

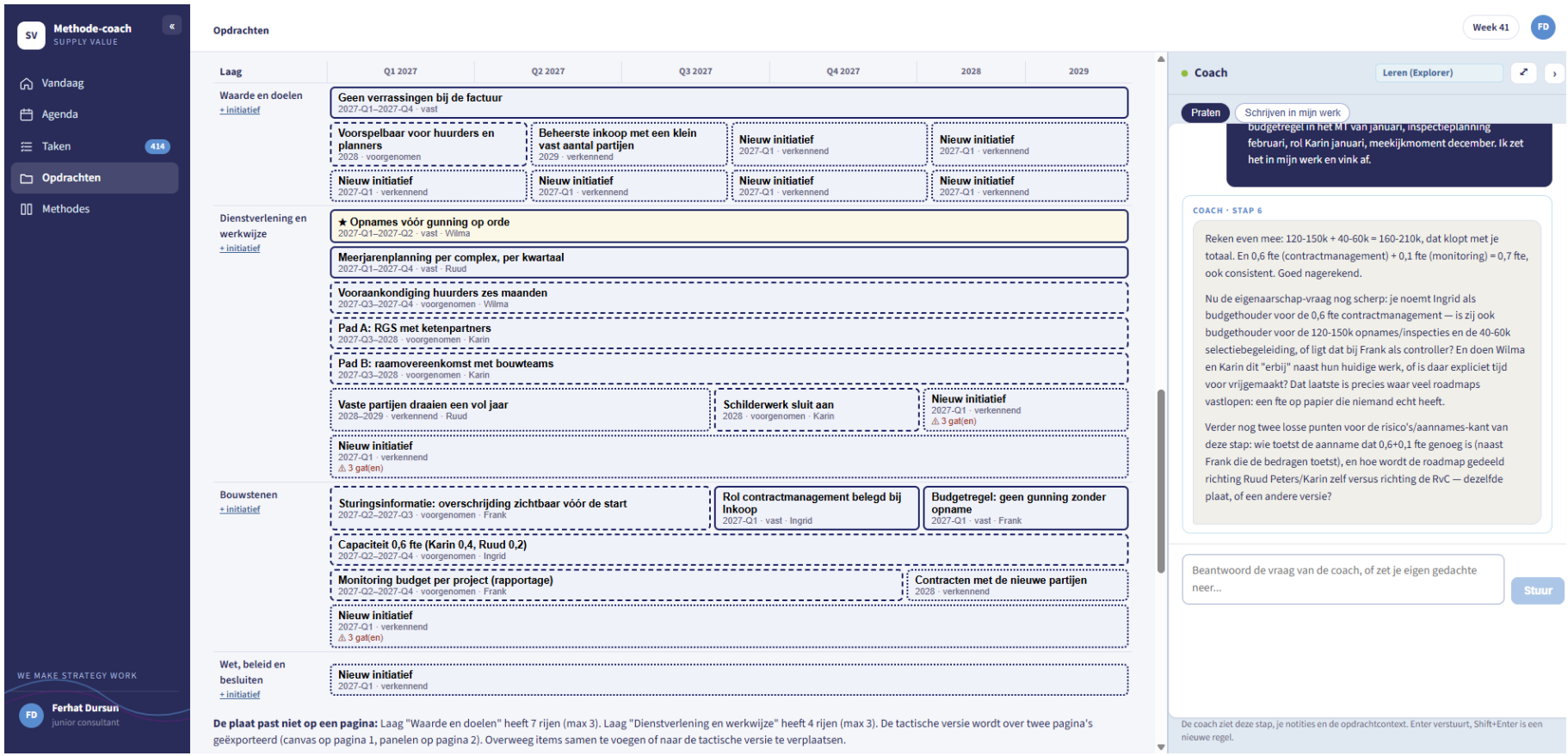


Figure I.4 The worksheet of Design Roadmapping: layers and horizons that the consultant and the coach fill together.

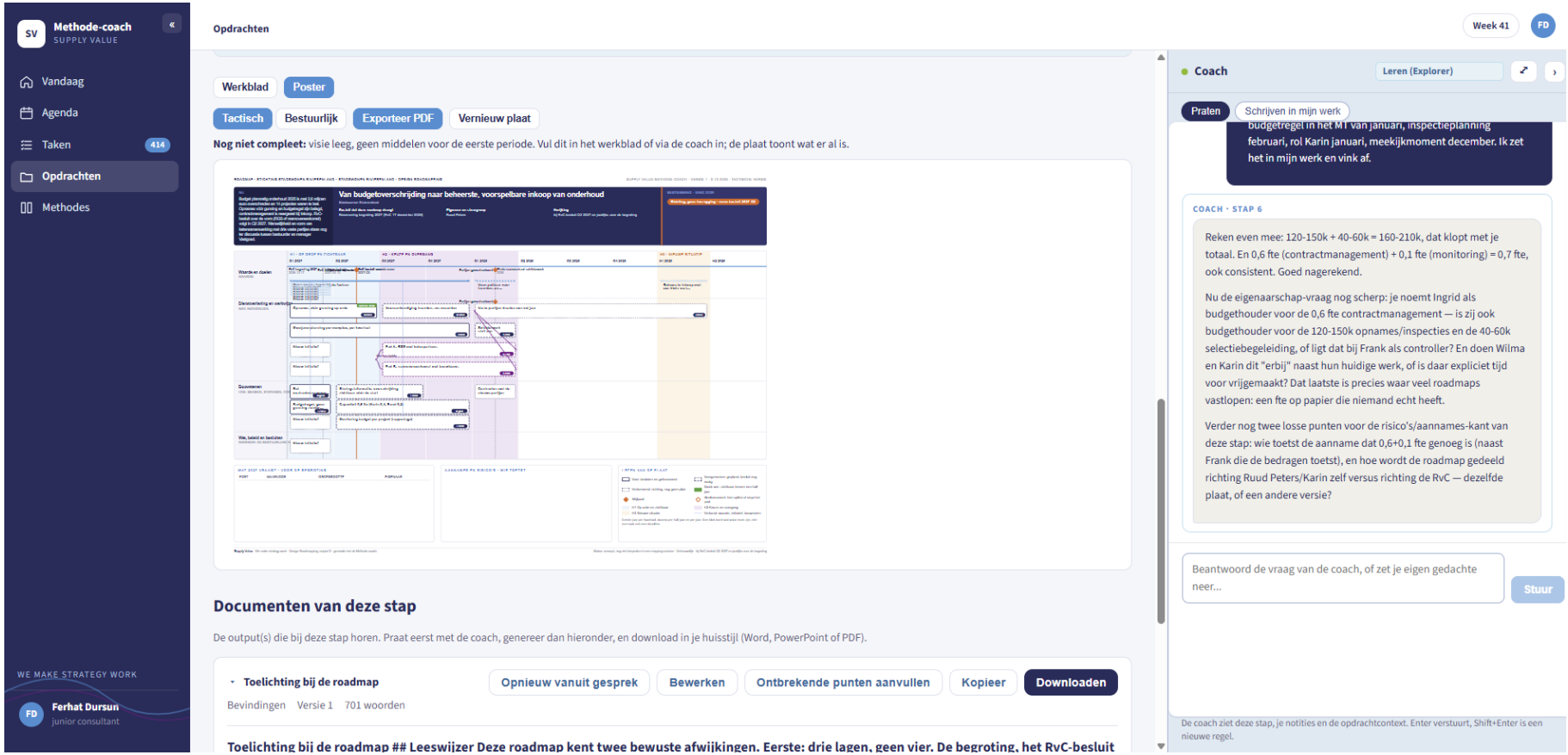


Figure I.5 The deliverable of the method: the roadmap plate with its export, and the documents of the step with their versions.

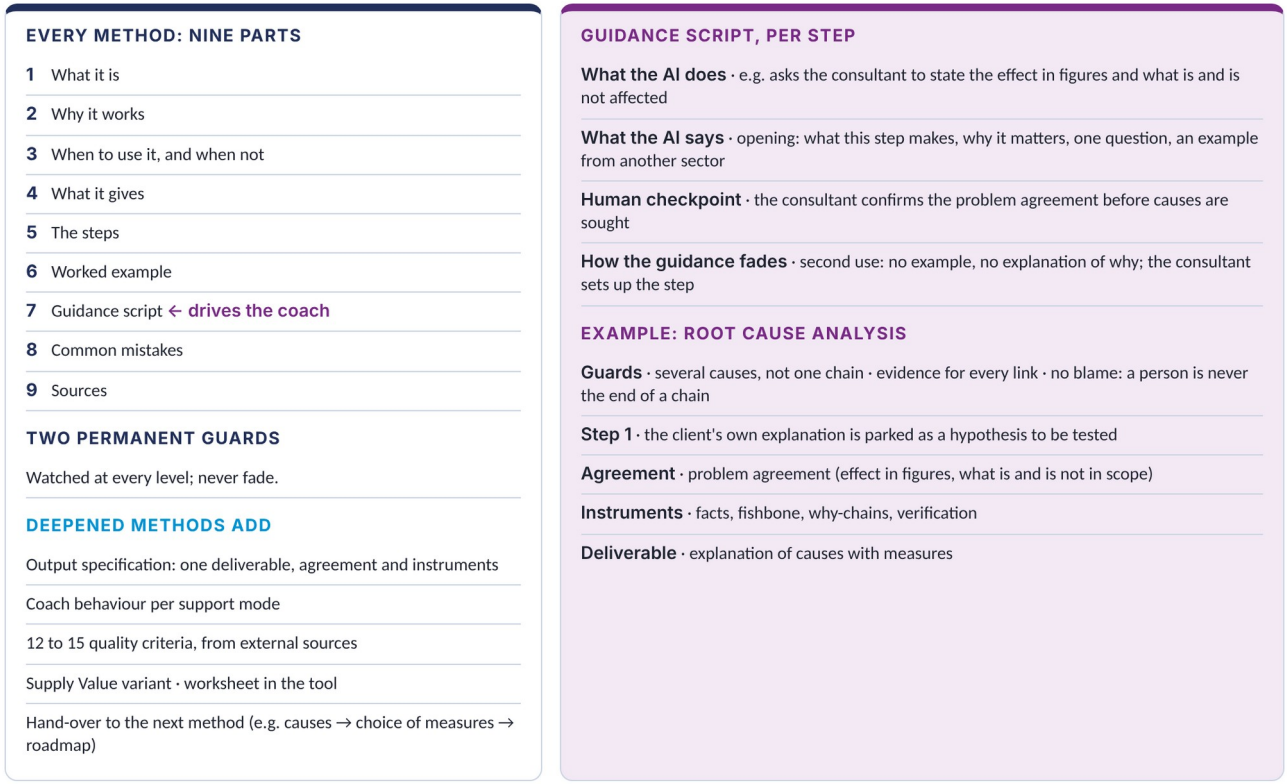


Figure I.6 The structure of a method in the database, with root cause analysis as an example.

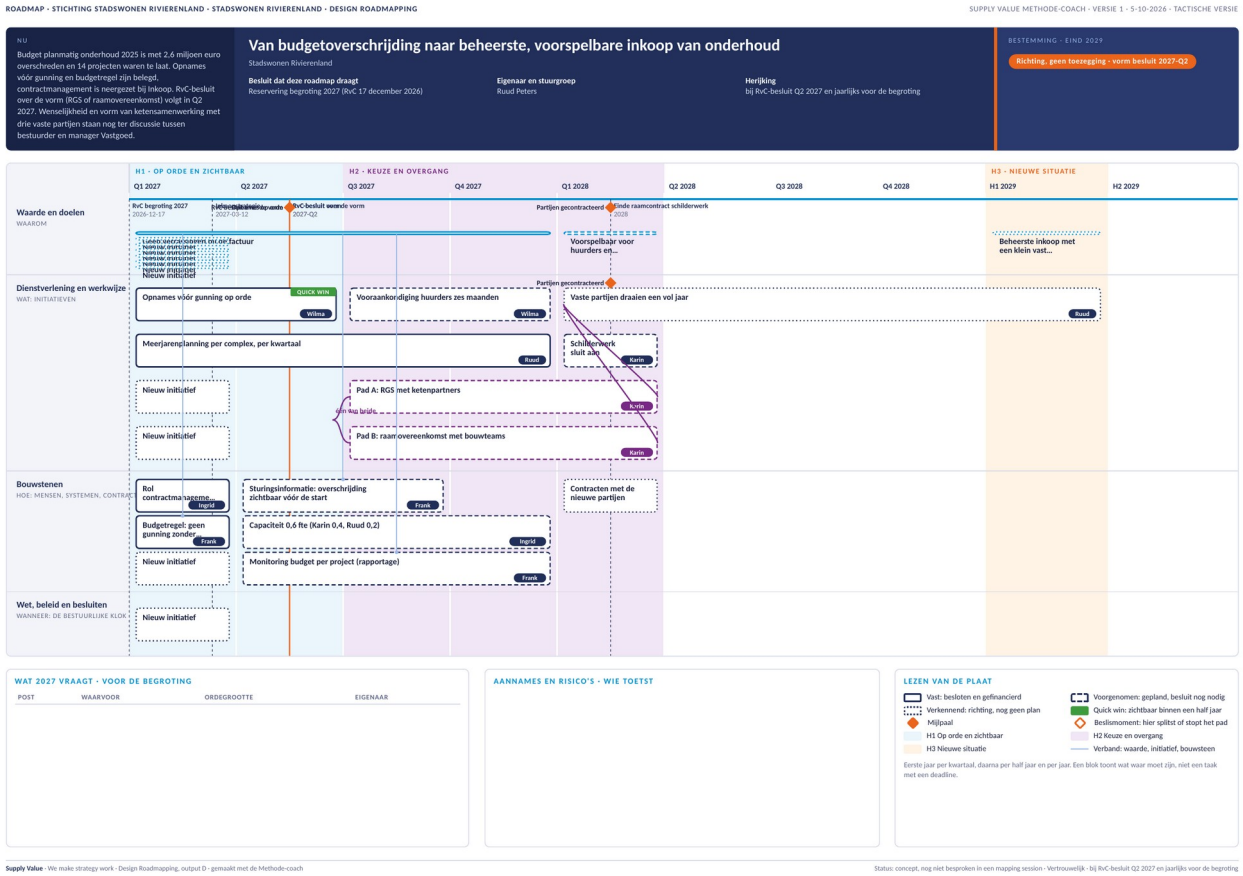


Figure I.7 The roadmap plate as exported from the prototype on the fictional housing corporation case (test 7). The organisation and every person named on the plate are fictional; any resemblance to a real organisation is unintended. The plate is reproduced as exported, in Dutch, including the defects on the open list at that time: the overlapping milestone labels and the empty panels.

The ten formative test runs

	What was tested	Case	Support mode
1	End to end, first version: intake, stakeholders, problem, plan, method Interviews steps 1 and 2, outputs	Water board; invoices wrong, contracts expiring unnoticed, a system named as the solution	Learning
2	Method Interviews, all steps, with outputs	Regional hospital; vendor managed inventory named as the solution; highest sensitivity level	Learning

	What was tested	Case	Support mode
3	End to end, realistic assignment of 23 weeks: intake with uploaded engagement letter, two problems, Spend Analysis, Interviews with coding, Kraljic, end report, exports	Housing corporation; budget overrun and late maintenance projects; a planted inconsistency in the figures	Learning; Kraljic in independent mode
4	Method Journey Mapping with worksheet, quality card and poster	Same housing corporation; how tenants experience a renovation	Learning, spot check in independent mode
5	Method Interviews, steps 1 to 5, against an external benchmark; re-check the same day	Same housing corporation; why projects run late according to the people who work on them	Learning
6	A consultant asks for a roadmap; the coach routes to Future Visioning	Same housing corporation; a three-year route to results-based maintenance with chain partners	Learning
7	Method Design Roadmapping, new, with worksheet and roadmap poster; re-check the same day	Same problem as test 6	Learning
8	Method Root Cause Analysis with fishbone and why-chains, using the interview themes of test 5 as sources	Same housing corporation; the budget overrun of test 3	Learning
9	Method Contextmapping with sensitising, session script and analysis wall	Same housing corporation; what planners, project leaders and tenants think of planned maintenance	Learning
10	Dry run of the evaluation of chapter 6 by the researcher: intake, problem, plan and the six steps of Journey Mapping, with export	Municipality; how a budget holder experiences buying a service through the procurement department; an e-learning named as the solution	Learning

Table I.1 The ten formative test runs of section 5.7. From test 3 to test 9 one fictional client, a housing corporation, was reused with a new problem per test, so that context could be tested across problems; test 10 used the municipality case of the evaluation.

Appendix J. The eleven interview themes in full

Confidential. This appendix gives the full account of the eleven interview themes, with the codes, the participants in which each theme occurs and the quotations from the interviews. It is not included in the public version of this thesis; it is part of the version assessed by the graduation committee.

Appendix K. Traceability from requirements to design decisions

This appendix completes the traceability that section 3.3.4 left to chapter 5 and that table 5.5 summarises. For each of the fourteen design requirements it states the design decision, the section in which the decision is realised and its status after the evaluation of chapter 6 (table 6.2): built (in the prototype), specified (worked out in the framework, to be assessed by management after the pilot), in part (part of the decision is built) or designed (decided and specified, not yet built).

	Requirement (short)	Design decision	Where	Status
DR1	Work process first, not tool adoption	Dimensions describe work on client problems; the tool starts from the assignment and the problem; methods behind the conversation	5.2.3; 5.3.2; 5.4.2	Built
DR2	Levels from prompting to directing agents; framing as entry condition	Four levels with an entry condition on framing and judgement; placement per dimension	5.2.4	Specified; assessment by management after the pilot (6.4)
DR3	Defined and assessed in value at the client	Matrix described in what is seen at the client; capability conversations without usage figures; deliverables as evidence	5.2.3; 5.2.5; 5.5	Specified; assessment by management after the pilot (6.4)
DR4	Human validation as rule and skill	G1; checkpoints requiring engagement; checklists; drafts only; download check	5.2.6; 5.4.5	Built
DR5	Support fades as capability grows	Three support modes per method; fading within each method; guards do not fade	5.2.5; 5.4.6	Built
DR6	AI literacy (Article 4) as base layer	D1 as base dimension; framework as documentation of AI literacy	5.2.3	Specified; documentation after governance approval
DR7	Agent that holds assignment context	Assignment as the unit; context summary per turn; running facts per step; a summary of the earlier steps and the worksheet in every step; typed facts across problems	5.4.2; 5.4.7	In part; typed facts to the plan step designed after the evaluation (6.3)
DR8	Frames and chooses before it produces; nothing written without confirmation	Operating rules; problem framing as a gate, with the statement in the consultant's own words after their own attempt; reasoning visible; client words for the client	5.3.3; 5.4.3	Built; observed in two sessions (6.3)
DR9	Rules per client and project, checked at use; sector context as a layer	Agreements on AI asked in the intake, recorded per project and applied at input, with a forbidden name refused and not stored; sensitivity banner; anonymisation as hard check; minimal data to the model; model per project; sector profiles	5.4.8	Check at the moment of use built and observed in two sessions (6.3); model per project and sector profiles designed
DR10	Entry through a concrete own task that pays back in the first session, without configuration or prior AI knowledge	A short intake of the consultant's own assignment as the first task; coach proposes	5.4.2	Built
DR11	Knowledge flows back in reusable form	Decisions, assumptions and open points recorded during the work; one deliverable per method; presentation for the decision maker; end report	5.4.5; 5.4.7; 5.4.10	In part (flow to colleagues after this thesis)

	Requirement (short)	Design decision	Where	Status
DR12	Redesign of one workflow per unit; connection to existing moments; time for learning	M4 and M5; implementation	5.2.5; chapter 6	Implementation; the pilot (6.4), awaiting a decision by management
DR13	Hard: within existing agreements; no sensitive data to external systems without approval. Criterion: no significant IT investment; usable at client sites	Web application without installation; data in the browser; fictional cases only in testing; provider replaceable; route through the approved environment	5.4.1; 5.4.8	Hard part: met in testing (fictional cases, no client data); for real use the route through the approved environment is required. Criterion part: in part
DR14	Knows who decides, the mandate and the scope	Decision maker and scope in the problem statement; stakeholder register; scope check; proposal for the decision maker; G3; part of D3	5.4.3; 5.4.9; 5.2.6	Designed (stakeholder register built)

Table K.1 Traceability from the fourteen design requirements to the design decisions of chapter 5, with their status after the evaluation (table 6.2).

Appendix L. Recommendations beyond the scope of this thesis

During the research a number of insights and ideas arose that lie outside what this thesis designs, builds or evaluates, but that are worth recording for Supply Value. Each entry gives the insight, where it came from, why it matters, and its status. Section 7.4 refers to this list; the non-blocking improvements from the evaluation of chapter 6 are in table 6.1 and are not repeated here.

Recommendation	Where it came from	Why it matters	Status
An open method base: consultants submit methods they bring from practice, such as a 30-60-90 day plan; a Multiplier reviews and deepens them	Theme 4 (P5-026, P5-027, P5-091); the workflows of the co-creation session (section 4.1.7)	Consultants carry methods they do not call methods; a closed catalogue would miss them, and the deepening of the remaining eighty methods needs their knowledge	Further development of the coach
A who-has-used-this-method marker on every method page: colleagues who used the method and are willing to be asked, opt-in and factual, no ranking	Theme 5 and cluster 5 (the frontrunner as the go-to person, P3-022, P3-026); idea 24 of the ideation	Knowledge sits with people; a ranking would turn use of the tool into assessment (DR3), a list of people to ask does not	Further development of the coach
A social place for methods and building blocks, where consultants share what they built and how they used a method	Theme 5; cluster 4; M4	Extends peer exchange beyond the existing moments; carries the risk of becoming a leaderboard and should stay descriptive	An idea for Supply Value
The model as a rule per project: the firm's approved environment by default, the client's own model where the client prescribes and offers one, none where the client forbids AI; projects never share context, facts or rules	Theme 8 (P2-042, P7-040); cluster 3; section 5.4.8	Makes the coach usable at clients that prescribe their own model, with three conditions: an interface, the client's acceptance of the tool as processor, and the benchmarks of section 5.7 passed again on that model	Further development of the coach
Diagnosis of a failed attempt as coach behaviour: what was tried, with what, and what came out; whether the task, the prompt, the type of tool, the data or the rules were the cause; then the retry	Theme 2 (P7-045); question 11, idea 12; moment of truth MT-B1	Turns one failed experiment into a second attempt with a reason, for the majority that waits for a use case	Further development of the coach
Knowledge sharing through the coach: typed facts and reusable components with their origin, retrieved at the step where they apply for the same client or sector, never across clients; reuse decided in the AI agreement with the client	Theme 5 (P3-029); cluster 4; M3; section 5.5	Documents in a shared drive are not found (theme 5); facts brought into the conversation are	Further development of the coach

Recommendation	Where it came from	Why it matters	Status
Validation of the two co-worker journey maps with their participants, with the participants scoring the six phases themselves and ten minutes on their own task list or board as a second source type	Section 3.2.2 (limitations of the maps)	Makes the emotional layer of the maps researched rather than estimated and doubles as the ethical check for map B	Not done within this thesis; recommended before the maps are shared within Supply Value
Expansion of the capability-building literature of section 2.3.4 and a positioning of the framework against existing capability models	The supervisory team's review of the concept	Strengthens the answer to SQ4	Not done within this thesis; recommended for further research

Table L.1 Recommendations that lie outside the scope of this thesis, with their origin and their place.

Appendix M. Materials of the evaluation

This appendix contains the materials of the evaluation of chapter 6 as they were given to the three consultants: the test guide (M.1) and the three fictional cases (M.2). What the three consultants returned is held by the researcher and is reported in section 6.2. Both parts are in Dutch, the language in which the test was held, and are reproduced as sent, apart from the layout. All organisations, names, figures and quotations in the cases are invented. The guide states the purpose of the test, that the tool and not the consultant was being tested, and what the participant returns; the participant fills in the fields marked with an ellipsis.

M.1 The test guide

Handleiding

Ongeveer 90 minuten. Je werkt zelfstandig met één casus en maakt één customer journey map. Je vult dit document in terwijl je de stappen volgt, en stuurt het daarna per e-mail terug.

Waar het om gaat

De Methode-coach is een prototype dat ik voor mijn afstuderen aan de TU Delft bij Supply Value heb gebouwd. Het is een webtool met een coach die je stap voor stap begeleidt bij het toepassen van een methode op een opdracht, zodat je de methode leert door hem te doen. In deze test is dat de customer journey map.

Ik test de tool, niet jou. Ik wil weten of de coach je laat nadenken in plaats van het werk over te nemen, en wat je van de tool vindt. Er zijn geen goede of foute antwoorden. Loopt de coach vast, of zegt hij iets wat niet klopt: schrijf het op, dat is precies wat ik moet weten.

Wat je nodig hebt

- Een laptop met Chrome of Edge. Werk in één browser en niet in een incognitovenster: de tool bewaart je werk in je browser.
- Dit document (Word) en de casussen (PDF).
- De inloggegevens die je van mij krijgt.
- 90 minuten zonder onderbreking. Pauzeren tussen de stappen mag.

Het verloop

Stap	Wat je doet	Minuten
1	Over jou	2
2	Kies een casus en lees hem	5
3	Taak A, zonder de tool	7
4	Inloggen, opdracht en intake	7
5	Vraagstuk en probleem scherpen	6
6	Plan van aanpak, en kort in Agenda en Taken kijken	6
7	Journey Mapping in zes stappen, met een nagesprek	37
8	PDF en export	2
9	Taak Z, zonder de tool	7
10	Vragen achteraf en terugsturen	8

Eén afspraak voor de hele test: eerst je eigen versie. Op drie momenten typ je eerst je eigen antwoord in de chat met de coach, beginnend met "Mijn versie:", voordat je de coach om een voorstel vraagt. Zo kan ik uit de export zien wat van jou kwam en wat van de coach.

Loop je uit? Bij stap 7 staat wat je dan overslaat.

De route in de tool

De hele route op één pagina. Bij de stappen hierna staat elk onderdeel uitgelegd.

Onderdeel	Waar je klikt
Profiel	Log in. Klik linksonder op Stel je profiel in, vul je naam in en kies op dezelfde pagina je Coachniveau.
Opdracht	Opdrachten, Nieuwe opdracht, naam van de klant, Toevoegen. Kies één werkdag en klik op Opslaan. Open je nieuwe opdracht.

Onderdeel	Waar je klikt
Intake	Begin met de intake. Praat met de coach. Alles overnemen. Bij 4 van 4: Bevestig de intake.
Vraagstuk	Maak je eerste vraagstuk, de vraag in één regel, Toevoegen, Beginnen. In de chat: "Mijn versie:". Daarna Dit is het probleem.
Plan	Volgende: maak het plan. Praat met de coach. Stel het plan voor. Volgende: voer de stappen uit.
Agenda en Taken	Links in het menu: Agenda en Taken. Alleen kijken. Terug via Opdrachten, je opdracht, je vraagstuk, tabblad Stappen uitvoeren.
Journey Mapping	Start deze stap. Per stap: praten, voorstellen overnemen, Stap klaar, Zonder output doorgaan. De laatste stap: Markeer als klaar. Daarna het nagesprek en Nagesprek afronden.
PDF en export	Exporteer PDF, boven de map. Daarna linksonder je naam, Exporteer naar JSON.

Stap 1. Over jou (2 minuten)

Vraag	Antwoord
Rol (junior, medior, senior) en business unit	
Jaren ervaring in consultancy	
Hoe vaak gebruik je AI-hulpmiddelen in je werk? (nooit, af en toe, wekelijks, dagelijks)	
Heb je eerder een customer journey map gemaakt? (nooit, één keer, vaker)	
Starttijd	

Hoe zeker ben je dat je dit op dit moment zelf kunt, zonder hulp van een collega of een tool? 0 = helemaal niet, 5 = redelijk, 10 = met grote zekerheid.

#	Kan ik …	0 tot 10
S1	… een vraag van een opdrachtgever herschrijven tot een probleemstelling zonder oplossing erin?	
S2	… de oplossing die de opdrachtgever al in gedachten heeft herkennen en apart zetten van het probleem?	
S3	… de fasen van een journey bepalen vanuit de persoon zelf, niet vanuit de organisatie?	
S4	… dieptepunten en kansen aanwijzen en onderbouwen met wat gebruikers zeggen?	
S5	… aan een collega uitleggen waarom ik een journey map zo heb opgebouwd?	

Wat levert een journey map een opdrachtgever op, denk je? Eén zin, ook als je het niet zeker weet: …

Stap 2. Kies een casus en lees hem (5 minuten)

Kies in de casussen één van de drie: een gemeente, een fabriek of een woningcorporatie. Lees de opdrachtbevestiging en het interviewverslag van de hoofdpersoon. **Lees het laatste deel van de casus ("Alleen voor taak Z") nog niet**; dat is voor het einde.

Gekozen casus (1 gemeente, 2 fabriek, 3 woningcorporatie): …

Stap 3. Taak A, zonder de tool (7 minuten)

Doe dit voordat je de tool opent, in één keer, zonder op te zoeken en zonder te polijsten.

A1. Probleemstelling in drie tot vijf zinnen: wat is de situatie nu, wat zou het moeten zijn, en wat denkt de opdrachtgever dat de oorzaak of oplossing is …

A2. Hoe zou je de ervaring van de hoofdpersoon in kaart brengen? Welke fasen, waar verwacht je de dieptepunten, en op welke citaten baseer je dat? Een halve pagina …

A3. Tijd in minuten: …

Stap 4. Inloggen, opdracht en intake (7 minuten)

1. Ga naar [redacted] en log in met je gegevens.

2. Ga naar je profiel: klik linksonder op **Stel je profiel in** en vul daar je naam in.
3. Kies op dezelfde pagina bij **Coachniveau** je niveau: **Leren** (de coach legt uit en laat je eerst zelf proberen), **Vlot werken** (één eigen poging, de coach vult aan) of **Zelfstandig** (de coach kijkt mee als reviewer).
4. Klik links op **Opdrachten** en daarna op **Nieuwe opdracht**. Typ de naam van de klant uit je casus en klik op **Toevoegen**. De tool vraagt op welke dagen je aan de opdracht werkt: kies één dag en klik op **Opslaan**. Open daarna je nieuwe opdracht; die staat bovenaan de lijst.
5. Klik op **Begin met de intake** en geef de coach de gegevens uit de opdrachtbevestiging, ook de afspraken over AI. Links verschijnen voorstellen: neem ze over met **Alles overnemen**. Staat er 4 van 4, klik dan op **Bevestig de intake**. Staat er minder, vraag de coach wat er nog ontbreekt.

Houd je in de rest van de test aan de afspraken over AI uit je casus.

Gekozen niveau (en later gewisseld naar): …

Stap 5. Vraagstuk en probleem scherpen (6 minuten)

1. Klik op **Maak je eerste vraagstuk**, typ de vraag van de opdrachtgever in één regel, en klik op **Toevoegen** en daarna op **Beginnen**.
2. **Eigen versie eerst:** typ in de chat "Mijn versie:" en je probleemstelling zoals je die nu zou formuleren. Vraag pas daarna om feedback of een voorstel.
3. Klik op **Dit is het probleem** als je achter de probleemstelling staat. Ziet de coach bij bevestigen nog iets, dan kies je zelf: **Aanpassen** of **Toch bevestigen**.

Ben je bij je eigen versie gebleven, heb je hem aangepast, of heb je die van de coach overgenomen? En waarom? …

Stap 6. Plan van aanpak, Agenda en Taken (6 minuten)

1. Klik op **Volgende: maak het plan** en bespreek het plan met de coach. Korte antwoorden zijn genoeg.
2. Zorg dat **Journey Mapping** in het plan komt. Stelt de coach een andere methode voor, vraag dan waarom, en kies daarna toch alleen Journey Mapping; daar is meestal een knop voor.
3. Klik op **Stel het plan voor** en daarna op **Volgende: voer de stappen uit**.
4. Kijk kort in **Agenda**, links in het menu: blader met › naar de weken van je opdracht, daar staan de stappen uit je plan. Kijk daarna even bij **Taken**. Je hoeft er niets mee te doen. Ga terug via **Opdrachten**, je opdracht en je vraagstuk, naar het tabblad **Stappen uitvoeren**.

Wat stelde de coach voor, en wat vond je daarvan? …

Stap 7. Journey Mapping (37 minuten)

Klik bij Journey Mapping op **Start deze stap**. De methode heeft zes stappen; reken op ongeveer zes minuten per stap, en twee minuten voor het nagesprek. Links zie je het werkblad van de stap waarin je zit, rechts praat je met de coach.

Zo werk je in elke stap

- Beantwoord de vraag van de coach in de chat.
- Links verschijnen voorstellen. Neem over wat klopt, met **Alles overnemen** of per voorstel, en wijs af wat niet klopt. Niets komt in je map zonder dat jij het overneemt.
- Staat iets wat je met de coach hebt afgesproken niet links als voorstel, vraag de coach dan om het klaar te zetten, of klik op **Vul in vanuit het gesprek**.
- Zegt de coach dat de rest in de volgende beurt komt, antwoord dan "ga verder".
- Klik op **Stap klaar** als de stap af is. Vraagt de tool **Output nog maken** of **Zonder output doorgaan**, kies dan **Zonder output doorgaan**. Bij de laatste stap heet de knop **Markeer als klaar**.
- Noem je iets wat volgens de afspraken over AI niet mag, zoals een naam of een leeftijd, dan geeft de tool vóór het versturen een melding. Lees die en kies zelf wat je doet.
- Na de laatste stap volgt een **nagesprek** met drie vragen. Antwoord in je eigen woorden en klik op **Nagesprek afronden**.

Stap in de tool	Wat je doet	Eigen versie eerst
Doel, actor en scenario	Welk besluit de map onderbouwt, wiens journey het is, begin en einde	
Bewijs, of eerlijk een hypothesekaart	Het interviewverslag en de cijfers uit je casus zijn je bronnen. Verzin niets wat er niet staat	
Fasen en stappen, outside-in	Drie tot zeven fasen in de woorden van de hoofdpersoon	Typ "Mijn versie:" met je fasen
Lagen vullen	Per fase: doet, contact, denkt en verwacht, voelt	
Emotiecurve, pijnpunten en moments of truth	Dieptepunten en hoogtepunt aanwijzen, met citaat	Typ "Mijn versie:" met je twee dieptepunten en je hoogtepunt

Stap in de tool	Wat je doet	Eigen versie eerst
Kansen, eigenaar en delen	Zoek kansen over de hele map. Werk er één of twee uit, met een eigenaar bij de klant en een indicator; bewaar de rest als overige kans	
Nagesprek	Drie vragen over wat de map heeft opgeleverd	

Loop je uit? Kijk bij de start van Lagen vullen op de klok. Is dat meer dan 50 minuten na je starttijd, vul de lagen dan alleen voor de drie fasen waar het interview het meest over zegt, en geef de andere fasen alleen een gevoelsscore. Vraagt de coach naar de andere fasen, zeg dan dat je die voor deze test overslaat. Een niet helemaal gevulde map is geen probleem.

Fasen. Ben je bij je eigen versie gebleven, heb je hem aangepast, of overgenomen? En waarom? …

Dieptepunten en hoogtepunt. Gebleven, aangepast of overgenomen? En waarom? …

Tot waar kwam je? Laatste stap die af is, wat er nog open stond, en of je de lagen voor minder fasen hebt gevuld: …

Waar liep je vast of viel je iets op?

Houd dit bij vanaf stap 4. Gebruik een letter:

- T = de bediening van de tool;
- M = de methode, of wat de coach bedoelde;
- F = een foutmelding (noteer de code);
- O = ik vroeg de coach het voor me te maken;
- + = de coach verraste me positief;
- − = de coach zei iets wat niet klopte of niet hielp.

Stap	Letter	Wat gebeurde er	Opgelost door (ikzelf, de coach, niet)

Stap 8. PDF en export (2 minuten)

1. Klik op **Exporteer PDF**. De knop staat boven de map, in het werkblad en bij **Poster**.
2. Klik linksonder op je naam. Kies bij **Back-up: exporteren en importeren** de knop **Exporteer naar JSON**. Je krijgt één bestand met je gesprekken en je map.
3. Bewaar beide bestanden; je stuurt ze straks mee.

Stap 9. Taak Z, zonder de tool (7 minuten)

Sluit de tool. Lees nu het deel "Alleen voor taak Z" van je casus: dezelfde situatie, door de ogen van een andere persoon. Schrijf op dezelfde manier als bij taak A.

Z1. Probleemstelling vanuit deze persoon in drie tot vijf zinnen …

Z2. Hoe zou je de ervaring van deze persoon in kaart brengen? Welke fasen, waar verwacht je de dieptepunten, en op welke citaten baseer je dat? Een halve pagina …

Z3. Tijd in minuten: …

Stap 10. Vragen achteraf (8 minuten)

Dezelfde vijf vragen, nu

Hoe zeker ben je dat je dit nu zelf kunt? (0 tot 10)

#	Kan ik …	0 tot 10
S1	… een vraag herschrijven tot een probleemstelling zonder oplossing erin?	
S2	… de oplossing van de opdrachtgever apart zetten van het probleem?	
S3	… de fasen van een journey bepalen vanuit de persoon zelf?	
S4	… dieptepunten en kansen onderbouwen met wat gebruikers zeggen?	
S5	… uitleggen waarom ik een journey map zo opbouw?	

Wat levert een journey map een opdrachtgever op? Eén zin, nu: …

Je journey map

Vraag	Antwoord
Welk deel van de inhoud komt naar jouw gevoel van jou, en welk deel van de coach? (verdeel 100 procent)	ik … procent, coach … procent
Hoe zeker ben je dat je deze map zonder de coach had kunnen maken? (1 helemaal niet tot 5 volledig)	
Hoe zeker ben je dat je kunt beoordelen of wat de coach zei klopt? (1 tot 5)	
Welk cijfer (1 tot 5) zou een ervaren collega je map geven, denk je?	

De coach en de tool

1 = helemaal oneens, 5 = helemaal eens.

#	Stelling	1 tot 5	Toelichting (mag leeg)
1	Ik heb vertrouwen in de coach		
2	De coach liet mij het denkwerk doen en nam het niet over		
3	Als ik vroeg of hij het voor me wilde maken, hield hij vast aan zijn rol (vroeg je dat niet: n.v.t.)		
4	De coach zat in de weg op momenten dat ik gewoon door wilde		
5	Het niveau dat ik koos, deed wat de naam belooft		
6	Ik begrijp na afloop beter hoe je een journey map maakt		
7	Het was steeds duidelijk wat ik in de tool moest doen		
8	De tool is nuttig voor mijn werk		
9	Ik zou de tool gebruiken bij een volgende opdracht als hij beschikbaar is		
10	De coach hield zich aan de afspraken over AI uit de casus		
11	Ik durf dit te gebruiken bij een klant met strenge regels over AI		
12	De casus gaf genoeg informatie om de opdracht te doen		

Wat hielp je het meest, en wat zat je het meest in de weg? …

Wat moet Supply Value regelen (afspraken, tijd, begeleiding, toestemming van klanten) voordat je dit wilt gebruiken? …

Tot slot, zonder terug te bladeren

Hoe schat je, terugkijkend, in dat je dit kon **voordat** je vandaag begon? (0 tot 10)

#	Kon ik vóór vandaag …	0 tot 10
S1	… een vraag herschrijven tot een probleemstelling zonder oplossing erin?	
S2	… de oplossing van de opdrachtgever apart zetten van het probleem?	
S3	… de fasen van een journey bepalen vanuit de persoon zelf?	
S4	… dieptepunten en kansen onderbouwen met wat gebruikers zeggen?	
S5	… uitleggen waarom ik een journey map zo opbouw?	

Eindtijd: …

Terugsturen

1. Mail naar Ferhat Dursun (het adres staat in de uitnodiging): **dit document**, **de journey map** (PDF) en **het exportbestand**.

2. Verwijder daarna je testopdracht: ga naar **Opdrachten** en klik op het prullenbakje achter je opdracht.

Als iets niet werkt. Krijg je een foutmelding met een code, noteer die bij stap 7, probeer het één keer opnieuw, en ga anders door met de volgende stap. Zie je "De coach reageerde niet", klik dan op **Opnieuw proberen**; je bericht is bewaard. Stopt een

antwoord midden in een zin, vraag de coach dan verder te gaan. Doet een pagina niets, ververs hem; je werk is bewaard. Lukt het echt niet, mail mij dan de code.

Bedankt voor je tijd.

M.2 The three cases

Casussen

Kies er één. Alle organisaties, namen, cijfers en citaten zijn verzonnen.

Hoe je een casus leest

Elke casus bestaat uit drie delen:

1. **De opdrachtbevestiging.** Wat de klant vraagt, de looptijd, wie erbij betrokken is en wat de klant toestaat bij het gebruik van AI. Dit gebruik je bij de intake in de tool.
2. **Het interviewverslag van de hoofdpersoon.** Dit is je bron voor de journey map. Gebruik alleen wat erin staat; staat iets er niet in, dan weet je het niet.
3. **Alleen voor taak Z.** Een kort verslag van een tweede persoon. Lees dit pas bij stap 9 van de handleiding.

Casus	Klant	De journey van
1	Gemeente Westerdal	een budgethouder die een ontwerpbureau wil inhuren
2	Brakel Kunststoffen	een monteur die 's nachts een onderdeel nodig heeft
3	Woonstichting Vlietland	een huurder met een lekkage via de bovenburen

Casus 1 · Gemeente Westerdal (publieke sector)

Van behoefte tot opdracht · verzonnen casus

Opdrachtbevestiging

Onderdeel	Afspraak
Klant	Gemeente Westerdal, 85.000 inwoners, afdeling Inkoop (zes medewerkers, waarvan twee inkoopadviseurs)
Opdrachtgever	Manager Inkoop
Vraag van de opdrachtgever	"Budgethouders snappen onze procedure niet. We willen een e-learning maken zodat ze weten hoe het werkt. Kun je eerst in kaart brengen hoe een budgethouder het inkoopproces ervaart?"
Looptijd en inzet	2 november tot 11 december 2026, 12 adviesdagen
Op te leveren	Een journey map van de budgethouder en een advies over de verbeteringen met de meeste impact, te bespreken in het MT Bedrijfsvoering van 10 december
Binnen de scope	Inkoop van diensten tussen 30.000 en 140.000 euro, waarvoor drie offertes nodig zijn
Buiten de scope	Europese aanbestedingen, inhuur van personeel, de inkoopsoftware zelf vervangen
Wat al bekend is	Het inkoopbeleid noemt zes weken als norm van aanvraag tot opdracht. De oorzaak van de vertraging is niet onderzocht; de opdrachtgever denkt dat budgethouders de regels niet kennen

Stakeholders

Rol	Belang	Invloed
Manager Inkoop (opdrachtgever)	Minder klachten en minder rechtmatigheidsfouten; heeft de e-learning al in gedachten	Hoog: beslist over de aanpak en het budget
Inkoopadviseurs (twee)	Minder losse vragen en spoedverzoeken	Middel: voeren de procedure uit
Budgethouders (circa 40 teamleiders)	Snel en zonder gedoe kunnen inkopen	Middel; via de wethouder soms hoog
Controller	Rechtmatigheid; de accountant keurde in 2025 drie inkopen af	Hoog: kan de procedure tegenhouden
Functioneel beheerder inkoopportaal (Informatievoorziening)	Stabiel portaal; wijzigingen kosten tijd	Laag tot middel: beheert de formulieren in het portaal

Afspraken over AI

- AI mag worden gebruikt, alleen in de door Supply Value goedgekeurde omgeving.
- **Geen namen** van medewerkers van de gemeente, inwoners of leveranciers in een AI-tool. Gebruik rollen, zoals "de budgethouder" of "de leverancier".
- Geen documenten van de gemeente uploaden. Een samenvatting in eigen woorden mag.
- Citaten uit interviews mogen, zonder naam en zonder details waaraan iemand te herkennen is.
- Cijfers uit deze opdrachtbevestiging en het interview mogen.
- Gevoeligheid van de gegevens: intern.

Wat je weet

Bron	Gegeven
Inkoopadministratie 2025	41 aanvragen in deze categorie; gemiddelde doorlooptijd van aanvraag tot opdracht 11 weken
Klachtenregistratie 2025	14 meldingen van budgethouders over inkoop, waarvan 9 over "geen reactie"
Wat niet wordt gemeten	Tevredenheid van budgethouders; hoe lang een aanvraag in elke fase ligt

Interviewverslag: de budgethouder

Interview van 45 minuten met Sandra, 44, teamleider Openbare Ruimte, zes jaar bij de gemeente. Zij wilde een ontwerpbureau inhuren voor de herinrichting van een park, voor ongeveer 60.000 euro. Het was haar eerste inkoop boven 30.000 euro.

Interviewer: Wanneer begon het?

Sandra: "In februari wist ik al dat we in september een ontwerper nodig hadden. Ik dacht: dat regel ik even." Ik had er een maand voor in gedachten. Ik had er zin in, het park is een project waar de buurt al jaren om vraagt.

Interviewer: Wat deed je als eerste?

Sandra: "Ik opende het inkoopportaal en kreeg vijf formulieren. Welk formulier ik nodig had, stond nergens." Ik ben er een middag mee bezig geweest. Uiteindelijk heb ik er een gekozen dat op een opdracht leek en dat ingevuld. Ik voelde me eerlijk gezegd een beetje dom.

Interviewer: En toen?

Sandra: Twee weken later kreeg ik een mail dat het verkeerde formulier was. "Toen hoorde ik dat het boven de drempel zat en dat ik drie offertes moest vragen. Niemand had me dat vooraf verteld." Ik moest zelf drie bureaus zoeken die mochten offreren. Dat kostte me nog een week. Begin april was de aanvraag compleet. Ik was geïrriteerd, maar dacht: nu loopt het tenminste.

Interviewer: Wat verwachtte je na het indienen?

Sandra: Dat ik binnen een week of twee hoorde hoe het verder ging. "Zes weken heb ik niets gehoord. Ik heb twee keer gemaild en kreeg een automatisch antwoord." In het portaal zie je alleen "in behandeling". Ik wist niet bij wie het lag. Dat was het slechtste deel: je kunt niets doen en je weet niets. Mijn wethouder vroeg in mei al hoe het ervoor stond.

Interviewer: Hoe kwam het weer in beweging?

Sandra: Ik heb half mei het algemene nummer van Inkoop gebeld en werd doorverbonden. "De inkoopadviseur die ik uiteindelijk aan de lijn kreeg, was top. Ze legde in tien minuten uit wat ik in twee maanden niet snapte." Ze vertelde ook dat zij met z'n tweeën alle aanvragen doen. Na dat gesprek was ik voor het eerst gerust.

Interviewer: Hoe ging het selecteren van een bureau?

Sandra: Inkoop stelde de uitvraag op en de criteria. "Ik moest offertes beoordelen op criteria die ik zelf niet had bedacht. Ik voelde me een buitenstaander bij mijn eigen project." Op ontwerpkwaliteit telde maar 30 procent, op prijs 70. Ik vond dat raar voor een park. Het beoordelen zelf kostte me een dag. De gunning was eind september.

Interviewer: En daarna?

Sandra: "Het bureau begon in november. Het park ging pas in het voorjaar open, en de wethouder vroeg me waarom." Zodra het bureau aan de slag was, ging het goed. Maar een volgende keer begin ik in oktober van het jaar ervoor, als ik er dan nog zin in heb.

Interviewer: Wat zou je anders willen?

Sandra: Dat ik aan het begin weet welke route ik moet volgen en hoe lang die duurt. En dat iemand laat weten waar mijn aanvraag ligt. Een e-learning? Ik doe dit één keer per twee jaar. Dan ben ik het weer vergeten.

Alleen voor taak Z: de leverancier

Lees dit pas bij stap 9 van de handleiding.

Interview van 30 minuten met Kees, eigenaar van een klein ontwerpbureau met zes medewerkers, dat de opdracht kreeg. Hij doet ongeveer vijf offertetrajecten per jaar voor gemeenten.

#	Citaat
Z1	"Ik zag de aanvraag met een deadline van tien dagen, midden in de vakantie."

#	Citaat
Z2	"Het formulier vroeg om omzetcijfers van drie jaar, voor een klus van 60.000 euro."
Z3	"Vragen stellen kon alleen via een vragenronde. Het antwoord kwam twee dagen voor de deadline."
Z4	"Na de gunning wist ik niet wie mijn contactpersoon was: inkoop of de teamleider."
Z5	"Als het eenmaal loopt, werkt het prima. Sandra weet precies wat ze wil."

Casus 2 · Brakel Kunststoffen (industrie)

Een spoedbestelling bij stilstand · verzonnen casus

Opdrachtbevestiging

Onderdeel	Afspraak
Klant	Brakel Kunststoffen, spuitgieterij met 220 medewerkers en vier productielijnen die dag en nacht draaien
Opdrachtgever	Operations manager
Vraag van de opdrachtgever	"Monteurs bestellen buiten het systeem om bij hun eigen contacten. Dat moeten we ze afleren, met strakkere regels. Kun je eerst in kaart brengen hoe een monteur een storing met een ontbrekend onderdeel beleeft?"
Looptijd en inzet	2 november 2026 tot 15 januari 2027, 10 adviesdagen
Op te leveren	Een journey map van de monteur en een voorstel voor de spoedprocedure, te bespreken in het operationeel overleg van 14 januari
Binnen de scope	Spoedbestellingen van reserveonderdelen voor de productielijnen
Buiten de scope	Inkoop van grondstoffen, het onderhoudsplan zelf, een nieuw ERP-systeem
Wat al bekend is	Een spoedbestelling moet vooraf worden goedgekeurd door de inkoper, die van 08.30 tot 17.00 uur werkt. Er is geen piketregeling voor inkoop. De oorzaak van het bestellen buiten het systeem is niet onderzocht

Stakeholders

Rol	Belang	Invloed
Operations manager (opdrachtgever)	Minder stilstand en grip op de kosten; denkt aan strakkere regels	Hoog: beslist over de procedure
Inkoper onderhoudsmaterialen	Grip op prijzen en leveranciers; wil niet achteraf hoeven bestellen	Middel: beheert de leveranciers
Magazijnbeheerder	Een kloppende voorraad	Middel: beheert het voorraadsysteem
Onderhoudsmonteurs (twaalf, in drie ploegen)	De lijn zo snel mogelijk weer laten draaien	Middel: zonder hen draait er niets
Financieel controller	Geen bestellingen zonder inkooporder	Middel

Afspraken over AI

- AI mag worden gebruikt in de omgeving van Supply Value.
- Namen en functies van medewerkers van Brakel mogen.
- **Namen van leveranciers en prijzen alleen geanonimiseerd**, bijvoorbeeld "leverancier A" en "ongeveer het dubbele van de contractprijs". Andere cijfers uit deze opdrachtbevestiging mogen.
- Geen tekeningen of technische specificaties van de matrijzen in een AI-tool.
- Geen documenten van Brakel uploaden. Een samenvatting in eigen woorden mag.
- Gevoeligheid van de gegevens: intern.

Wat je weet

Bron	Gegeven
Productieadministratie 2025	23 keer stilstand door een ontbrekend reserveonderdeel, gemiddeld 6,5 uur per keer
Inkoopadministratie 2025	41 procent van de spoedorders voor reserveonderdelen is pas achteraf in het systeem gezet

Bron	Gegeven
Wat niet wordt gemeten	Hoe vaak de voorraad in het systeem niet klopt met de kast

Interviewverslag: de monteur

Interview van 40 minuten met Ronald, 52, onderhoudsmonteur, twintig jaar in dienst, werkt vooral in de nachtdienst.

Interviewer: Kun je de laatste keer beschrijven dat een onderdeel ontbrak?

Ronald: "Om twee uur 's nachts stond lijn 3 stil. Het verwarmingselement van de matrijs, dat wist ik binnen vijf minuten." Dat is mijn vak, daar ben ik snel in. Ik dacht: element wisselen, half uur, klaar.

Interviewer: Wat deed je toen?

Ronald: Ik ben naar het magazijn gelopen, dat is 's nachts dicht maar wij hebben een sleutel. "In het systeem stond dat er twee op voorraad lagen. In de kast lag er niet één." Ik heb een kwartier gezocht, ook op de andere plekken. Toen werd ik echt chagrijnig. Ik verwacht dat het systeem klopt; daar is het voor.

Interviewer: Wat zegt de procedure?

Ronald: "Een spoedbestelling moet de inkoper goedkeuren, en die begint om half negen." Dus je staat daar met een stilstaande lijn en je mag niks. Ik heb de ploegleider gebeld. Die zei: doe wat nodig is.

Interviewer: En wat was nodig?

Ronald: "Ik heb de leverancier zelf gebeld op zijn mobiel. Die kende ik nog van vroeger. Om zeven uur stond hij hier." Om kwart over acht draaide de lijn weer. Ik heb het in het ploegenlogboek gezet, met de prijs die hij noemde. Ik was blij dat het gelukt was.

Interviewer: Hoe ging het daarna?

Ronald: "Daarna kreeg ik een mail dat ik buiten de procedure had besteld. Dat voelde als een tik op de vingers, terwijl de lijn weer draaide." Van wie de mail kwam, maakt niet uit. Niemand vroeg waarom.

Interviewer: Gebeurde er nog iets met de voorraad?

Ronald: "De magazijnbeheerder heeft de dag erna meteen de voorraad geteld en het systeem aangepast. Dat waardeer ik." Hij kwam het me zelf vertellen. Er bleken twee elementen in een ander vak te liggen, met een verkeerd nummer.

Interviewer: Wat zou je anders willen?

Ronald: "Niemand vraagt ons welke onderdelen we echt op voorraad willen hebben." Wij weten welke dingen kapotgaan. En 's nachts wil ik iemand kunnen bellen die ja mag zeggen. Strakkere regels? Dan staat de lijn gewoon langer stil.

Alleen voor taak Z: de inkoper

Lees dit pas bij stap 9 van de handleiding.

Interview van 30 minuten met Ilse, inkoper onderhoudsmaterialen, drie jaar in dienst. Zij is de enige inkoper voor onderhoud.

#	Citaat
Z1	"Ik kom om half negen binnen en zie drie spoedorders die al geleverd zijn."
Z2	"Ik moet achteraf een inkooporder maken voor iets wat ik nooit heb gezien. De prijs is vaak het dubbele."
Z3	"De monteurs zien mij als de rem. Ik wil gewoon weten wat er nodig is voordat het misgaat."
Z4	"Met de leverancier van de verwarmingselementen hebben we geen raamcontract. Elke keer is het een losse prijs."
Z5	"Toen we vorig jaar samen de kritieke onderdelen doorliepen, ging het een maand lang goed."

Casus 3 · Woonstichting Vlietland (woningcorporatie)

Een lekkage via de bovenburen · verzonnen casus

Opdrachtbevestiging

Onderdeel	Afspraak
Klant	Woonstichting Vlietland, 9.000 huurwoningen
Opdrachtgever	Manager Vastgoed
Vraag van de opdrachtgever	"Huurders melden hun gebreken verkeerd. We willen een nieuwe app met foto-upload. Kun je eerst in kaart brengen hoe een huurder een reparatieverzoek beleeft?"
Looptijd en inzet	2 november tot 18 december 2026, 12 adviesdagen
Op te leveren	Een journey map van de huurder en een advies over de verbeteringen in het reparatieproces, te bespreken met het managementteam op 17 december
Binnen de scope	Reparatieverzoeken voor dagelijks onderhoud, van melding tot afronding

Onderdeel	Afspraak
Buiten de scope	Planmatig onderhoud, renovatie, de keuze van de aannemers
Wat al bekend is	Verzoeken komen binnen via de app of de klantenservice en worden doorgezet naar een van vier aannemers. Volgens de contracten belt de aannemer de huurder binnen drie werkdagen. De oorzaak van het herhaald bellen is niet onderzocht

Stakeholders

Rol	Belang	Invloed
Manager Vastgoed (opdrachtgever)	Hogere huurderstevredenheid en minder telefoontjes; heeft de nieuwe app in gedachten	Hoog: beslist over budget en aanpak
Contractmanager onderhoud	Aannemers houden zich aan de afspraken	Middel tot hoog: beheert de vier contracten
Medewerkers klantenservice	Minder herhaalde telefoontjes	Laag tot middel
Aannemers (vier)	Efficiënt plannen; geen onbetaalde extra bezoeken	Middel: voeren het werk uit
Huurdersvereniging	Goede dienstverlening voor alle huurders, ook wie geen app gebruikt	Middel: heeft adviesrecht

Afspraken over AI

- AI mag worden gebruikt in de omgeving van Supply Value.
- **Geen gegevens van huurders** in een AI-tool: geen namen, adressen, leeftijden of andere details waaraan iemand te herkennen is. Gebruik "de huurder".
- Medewerkers en aannemers mogen bij hun functie worden genoemd, niet bij naam.
- Citaten mogen, zonder details waaraan iemand te herkennen is.
- Geen documenten van de corporatie uploaden. Een samenvatting in eigen woorden mag.
- Cijfers uit deze opdrachtbevestiging mogen.
- Gevoeligheid van de gegevens: vertrouwelijk.

Wat je weet

Bron	Gegeven
Klantcontactregistratie 2025	Bij 38 procent van de reparatieverzoeken belt de huurder een tweede keer
Contractadministratie	Vier aannemers, elk met een eigen planningssysteem; de corporatie ziet niet wanneer een afspraak gepland is
Wat niet wordt gemeten	Hoe lang een reparatie duurt van melding tot afronding

Interviewverslag: de huurder

Interview van 40 minuten bij mevrouw Kok thuis. Zij is 71 en woont dertig jaar in een portiekflat op de tweede verdieping.

Interviewer: Wat gebeurde er?

Mevrouw Kok: "Op zondagavond zag ik een bruine vlek op het plafond van de badkamer. Ik dacht: dat is de buurman boven." Ik schrok wel. Ik ben meteen een emmer gaan zetten, voor de zekerheid.

Interviewer: Hoe heeft u het gemeld?

Mevrouw Kok: Mijn dochter had de app op mijn telefoon gezet. "Ik heb de app geprobeerd, maar ik moest kiezen uit veertig soorten gebreken. Lekkage via de buren stond er niet tussen. Toen heb ik maandag gebeld." Met de app was ik twintig minuten bezig. Bellen ging om negen uur; ik heb een half uur in de wacht gestaan.

Interviewer: Hoe was dat gesprek?

Mevrouw Kok: "Het meisje aan de telefoon was heel vriendelijk. Ze zei: de aannemer belt u binnen drie werkdagen." Ik dacht: mooi, dan is het deze week opgelost. Ik was gerustgesteld.

Interviewer: En werd u gebeld?

Mevrouw Kok: "Na acht dagen nog niets. Toen ik weer belde, zeiden ze dat de opdracht bij een andere aannemer lag." Wie dat was, wisten ze niet precies. Ik voelde me van het kastje naar de muur gestuurd. De vlek werd ondertussen groter.

Interviewer: Wanneer kwam er iemand?

Mevrouw Kok: Op dag twaalf. "De monteur kwam, keek en zei dat hij bij de bovenburen moest zijn. Die waren niet thuis. Hij is weggegaan zonder een nieuwe afspraak." Toen was ik echt boos. Ik heb diezelfde middag weer gebeld en weer gewacht.

Interviewer: Hoe liep het af?

Mevrouw Kok: Een week later kwam er een andere monteur. "De tweede monteur heeft alles uitgelegd en zelf de bovenburen gebeld. Binnen een uur was het lek gevonden." Dat was een fijne man. Hij vertelde wat hij deed en waarom. Na drie weken was het plafond droog.

Interviewer: En nu?

Mevrouw Kok: "Het plafond is nu droog, maar de vlek zit er nog. Of er iemand komt schilderen, weet ik niet." Dat is nu zes weken geleden. Ik heb niet nog een keer gebeld; ik wil niet zeuren.

Interviewer: Wat had u geholpen?

Mevrouw Kok: Dat iemand me zegt wie er komt en wanneer. En dat ik niet steeds opnieuw mijn verhaal hoef te doen. Een app met foto's? Die gebruik ik niet, dat doet mijn dochter.

Alleen voor taak Z: de monteur

Lees dit pas bij stap 9 van de handleiding.

Interview van 30 minuten met Erik, monteur bij een van de vier aannemers. Hij doet gemiddeld zes reparaties per dag voor Vlietland.

#	Citaat
Z1	"Ik krijg de opdracht als één regel tekst: lekkage badkamer. Geen foto, geen verdieping."
Z2	"Als de bovenburen niet thuis zijn, mag ik volgens het contract niet wachten. Dan moet ik door naar de volgende."
Z3	"Een nieuwe afspraak maken kan ik niet zelf. Dat gaat via onze planning, en die zit in een ander systeem dan dat van de corporatie."
Z4	"Ik leg het de bewoner altijd uit. Ze zijn opgelucht als ze weten wat er gebeurt."
Z5	"Schilderwerk na een lekkage valt niet onder onze opdracht. De bewoner hoort dat meestal pas van mij."