Socio-Technical Boundary Conditions for Chatbot Implementation in Organizations

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy, and Management

by

Martijn van Dis

Student number: 5424887

To be defended in public on **July 8th, 2022**

Graduation committee

Chairperson	: Prof.dr.ir. M.F.W.H.A. Janssen, Engineering Systems and Services
First Supervisor	: Prof.dr.ir. M.F.W.H.A. Janssen, Engineering Systems and Services
Second Supervisor	: Dr. B. Wagner, Multi-Actor Systems
External Supervisor	: MSc J. Taal, Deloitte Consultancy
External Supervisor	: MSc D. Degen, Deloitte Consultancy

Preface

Before you lies the Master Thesis Project entitled “Socio-Technical Boundary Conditions for Chatbot Implementation in Organizations”. The thesis has been written to complete the final stage of the Master’s degree Complex Systems Engineering and Management at the Delft University of Technology. Besides the TU Delft, Deloitte Consultancy has been a supporting organization throughout this research. Deloitte has supported this thesis by providing two personal supervisors whom I could ask for advice, feedback, and make use of their professional network. The upcoming paragraphs describe a personal reflection on the process of writing this thesis.

A personal reflection on the process

Overall, I would describe the complete process as fluent yet dependent upon personal initiative. The process started in August 2021 by finding a first supervisor, a thesis subject, and applying at Deloitte. Fortunately, Marijn Janssen was willing to be my first supervisor and immediately proposed several suggestions for the thesis subject. After reading Designed for Digital on Marijn’s advice, the subject of “socio-technical boundary conditions for Artificial Intelligence implementation” was established. During the subsequent months, I did literature research and established a one-page research proposal to align the expectations of the graduation committee. Then, at the beginning of January, Deloitte assigned me two personal supervisors that were willing to meet weekly. My two supervisors helped in clearly defining my ideas and were proactive in contacting their professional network for potential case studies and interviewees. As a result, when the thesis research started on the 14th of February, all of the necessary foundations had been laid. Overall, I believe that because of the cooperation amongst my graduation committee, the transparent way of communication, and the clearly aligned goals and expectations, the process has been fluent and enjoyable.

A personal reflection on the content

After the kick-off meeting in week 4, I decided to scope down the field of Artificial Intelligence to solely chatbots. Doing so allowed me to delve deeper into the specific application and thereby establish a more grounded knowledge base for the analysis of the boundary conditions. I believe this was the right decision and it has allowed me to demonstrate the line of reasoning throughout the analysis of this research. As a result, the empirical testing of the boundary conditions has been my personal highlight. It was rewarding to have the boundary conditions be validated by industry practitioners and to see a difference in fulfillment between the two case studies.

Also, I loved to explore chatbot technology because I believe chatbots will have a substantial impact on society in the near future through the potential social and business value. Hence, the ethical concerns related to the redistribution of power and transparency should be well-known amongst both designers and users. This hopefully prevents misuse or deception as much as possible.

Learning curve

Whilst working on this 20-week project, I have gained new knowledge on chatbots, conducting qualitative research interviews, and project management. Especially project management, which includes careful time- and stakeholder management, has been of great importance due to the duration of the project. Also, Deloitte has offered me the chance to be part of the Technology, Vision & Architecture team which has enabled me to experience what it is like working within such an organization. This has been extremely valuable for my personal development, especially since I will start my working career after completing this Master's degree. Finally, because I have completed my Bachelor's degree in Industrial Engineering at the University of Groningen, this project has been the perfect opportunity for me to test my individual skills in solving social-technical complex problems.

Having completed this project, there are several things I would do differently next time. First, I would try to scope down the subject as soon as possible. By defining a clear project scope early on, it is easier to align the expectations of the graduation committee, do more specific literature research, and set up the research methodology accordingly. Second, I would aim for a subject that requires a quantitative research approach rather than a qualitative one. This is because I personally prefer to work with numerical and measurable results. On the contrary, conducting the qualitative research throughout this project has forced me to work outside of my comfort zone which has been valuable.

Relevance to CoSEM

The Master's degree Complex Systems Engineering and Management (CoSEM) is about exploring innovations in complex socio-technical environments. Socio-technical complex systems are systems whose behavior is difficult to model and predict because similar inputs will result in different outcomes due to the involvement of people, relationships, or dependencies. This is the opposite of a simple system, where similar inputs will result in similar outcomes; of which a car engine is an example. The socio-technical approach addresses how such systems are explored. Often, a distinction is made between the social, such as the ethics, culture, and legislation, and the technical aspects of the system.

Comparing the aim of the CoSEM degree with the research conducted, I believe there is a strong correlation between the two. The correlation originates from the challenge many Artificial Intelligence solutions face at the moment: the connection between the technical design and the social implications. As a result, the conducted research is split up in a CoSEM manner. At first, literature on the technical design of a chatbot is reviewed, whereafter the social complexity is explored. The analysis of the social complexity includes analyzing the enforced legislation and involved stakeholders on their roles and relationships. After, the information gathered throughout both sections separately is combined to draw new conclusions and gain new insights. Also, a reoccurring conclusion throughout the research is the context-dependency of influential factors in Artificial Intelligence solutions. To me, context-dependency and socio-technical complex systems appear to exhibit similar characteristics in the way that it is hard to draw valid, general conclusions.

Acknowledgment

First of all, I would like to express my gratitude towards the TU Delft, of whom my first supervisor Marijn Janssen especially. Your ongoing involvement and never-ending enthusiasm have been extremely motivating over the course of writing this thesis. Secondly, I would like to thank Seda Gürses. The discussions we have had have stimulated me to think outside of the box, which has contributed greatly to the completeness of this thesis. Also, I am grateful to Ben Wagner for being able to accompany me as my second supervisor during the final stages of my research.

Besides TU Delft, I would like to thank Deloitte Consultancy. The support provided by my two supervisors Jesper Taal and Doris Degen has been exceptional and something I had never imagined beforehand.

Last but not least, I would like to thank my family and friends for the continuous support you have given me over the course of my academic career.

Martijn van Dis
Delft
June 24th, 2022

Management summary

The number of organizations seeking to implement a form of Artificial Intelligence (AI) has never been as high as it is today. AI is known for its potential to revolutionize businesses and add customer value by executing intellectual tasks such as automation, customer engagement, and processing large amounts of data. Besides the possible business advantages the technology offers, there are concerns about the impact it can have on humanity and society. Loss of human autonomy, discrimination of minority groups, and unethical use are examples of the negative manifestations the technology poses. Also, many AI implementations seem to not deliver the promised business value to organizations after being implemented. Therefore, this research focuses on the necessary socio-technical boundary conditions that must be fulfilled before implementation can be successful. In this study, boundary conditions are defined as "the socio-technical constraints that must be satisfied to successfully complete an implementation". Moreover, to scope down the field of AI, this research is mainly focused on the use of AI for customer engagement, of which chatbots in special. The reason for focusing on chatbots especially is because of the "chatbot tsunami" enabled by organizations offering the Software as a Service. By using these services, no in-house understanding of the technology is required, average project time decreases, and control over the technical capabilities is lost. Hence, any organization can deploy a chatbot without considering the social implications it has.

The goal of this research is twofold. At first, the aim is to determine when an AI implementation can be described as "successful" considering a multistakeholder perspective. This is done by exploring two actor perspectives; the industry's and the government's. By conducting eight interviews with industry practitioners, insights are gathered into their view on the necessary success facets that determine if an implementation is deemed successful. After, available literature and legislation are analyzed to determine the government's success facets. By comparing the two results, differences in views are exposed. Second, the boundary conditions for successful chatbot implementation are analyzed. This is done by using a novel framework and the information gathered throughout the previous chapters. After the analysis, the boundary conditions are tested on their empirical presence, validity, and relation through a multiple-case study.

Investigating the industry's perspective on "successful AI implementation" has led to two findings. First, the industry appears to be largely concerned with the technical capabilities and social adoption of the technology to achieve the intended business value. Second, by evaluating the causes of failure from an industry perspective as well, a hierarchy amongst the success facets is suggested. Hence, a strategy map portraying the interrelation amongst the success facets is proposed. Furthermore, the government's perspective appears to be concerned with protecting society and preventing (unintended) misuse yet values the business opportunities greatly as well. Therefore, the two-actor views seem to be complementary rather than opposing.

The second research goal results in 33 proposed boundary conditions for chatbot implementation in organizations. By testing the conditions on two case studies, in which case A was a more successful chatbot implementation than case B, the effect of satisfying the boundary conditions has been evaluated. As a result, it is concluded that by fulfilling the applicable boundary conditions, a successful implementation can be established. Also, it is suggested that not all boundary conditions apply to every chatbot implementation and that there might be a relation between the conditions. The relation indicates that not fulfilling condition X might lead to not being able to fulfill condition Y.

To correctly interpret the results, one should be aware of several limitations to this research. First, the research is conducted in the Netherlands. As a result, Dutch legislation is analyzed to deduce the government's perspective on success facets, and the interviewees were employed in the Netherlands also. Because the literature suggests a high context-dependency of influential factors in AI implementation, it remains unknown if the results are applicable to different contexts as well. Second, only eight interviews have been conducted to explore the industry's perspective on "success". Because of the limited amount, it can be argued whether information saturation was reached. Third, the research was conducted by a single researcher which could imply an unintended researcher's bias in certain aspects of the analysis. Fourth, the boundary conditions have been tested on two chatbot implementations supported by the same organization. Therefore, the boundary conditions should be validated by different case studies in different organizations as well to strengthen their validity.

Table of contents

List of tables

List of figures

List of abbreviations

AI	=	Artificial Intelligence
BCs	=	Boundary Conditions
BLEU	=	Bilingual Evaluation Understudy
CBOW	=	Continuous Bag of Words
CSF	=	Critical Success Factor
DL	=	Deep Learning
DNN	=	Deep Neural Network
DOI	=	Diffusion of Innovations
FSC	=	Food Supply Chain
GloVE	=	Global Vectors
HAL	=	Hyperspace Analogue to Language
HCI	=	Human-Computer-Interaction
HHI	=	Human-to-Human Interaction
IS	=	Information System
IT	=	Information Technology
LSA	=	Latent Semantic Analysis
LSTM	=	Long Short Term Memory
MC	=	Markov Chain
ML	=	Machine Learning
NB	=	Naïve Bayes
NLG	=	Natural Language Generation
NLP	=	Natural Language Processing
NLU	=	Natural Language Understanding
NN	=	Neural Networks
RNN	=	Recurrent Neural Networks
SaaS	=	Software as a Service
Seq2Seq	=	Sequence-to-sequence
SVM	=	Support Vector Machine
TAM	=	Technology Acceptance Model
TOE	=	Technology, Organization, Environment
WE	=	Word Embedding
Word2Vec	=	Word-to-Vector

1 Introduction

1.1 Thesis motivation

Since its introduction during the late 1950s, Artificial Intelligence (AI) has been considered a promising but challenging technology that can revolutionize businesses and their offerings (Brock & von Wangenheim, 2019; Reim et al., 2020). Because AI is capable of executing intellectual tasks such as automation, customer engagement, decision making, and innovation based on input data, it can offer tremendous potential for gaining a competitive advantage and adding customer value (Benbya et al., 2021; McCarthy, 2007). But the implementation of the technology appears prone to failure.

Gartner has recently published a report that states 85% of AI projects do not reach their promised potential after being implemented in an organization (Nimdzi Insights & Edge, 2019). As shown by Reim et al. (2020), the literature suggests several implementation challenges including transparency of the technology, lack of trust among employees, the presence of analog processes, and the misunderstanding of AI. Hence, the difficulties of AI implementation stretch further than solely the technological requirements. Rather, "exploring the connections between an AI's technical design and its social implications will be key in ensuring feasible and sustainable AI systems that benefit society and that people want to use" (Dahlin, 2021, p.3).

Because the use of AI is expected to increase over the coming years, it is desired to enhance the success rate of implementation projects (Nimdzi Insights & Edge, 2019). Over the last years, research has been conducted attempting to identify the prohibitors and enablers of AI implementation in organizations (Ahmad et al., 2020; Bérubé et al., 2021; Hamm & Klesel, 2021; S. Kumar et al., 2021; Miller, 2021; Radhakrishnan & Chattopadhyay, 2020). Often, the proposed factors are socio-technical, regard the misalignment between the technology and the organization, and are formulated in a broad sense without commenting on the interrelationships, change over time, or degree of importance. Also, the literature suggests these factors to be highly dependent on the sector, organization, and application for which the AI is used; referred to as context-dependency (Bérubé et al., 2021; H. Chen et al., 2021; S. Kumar et al., 2021). Due to these uncertainties, it remains unknown which of the factors affecting the implementation success are essential, which are good to have, and which are redundant.

Because the current literature suggests a high context-dependency of the influential factors, the scope of this study is limited to the implementation of chatbot technology. The reason for exploring chatbot technology especially is because of the enormous number of implementations over the recent years, which are often facilitated by external companies offering the technology in the form of Software as a Service (SaaS) (Grudin & Jacques, 2019; Rapp et al., 2021). To deploy AI-based software in the form of a service rather than developing it on-premise affects the required in-house understanding, implementation time, project costs, and causes a loss of technological control (Link, 2013). Hence, enhancing the ease with which organizations can implement chatbots can cause the connections between the technical design and social implications to be unaddressed. Therefore, to ensure the alignment between the technical and the social aspects of the technology, the essential factors for the implementation of chatbots in organizations are explored throughout this thesis. The research question defined for this study is: *How do the socio-technical boundary conditions affect the success of chatbot implementation in organizations?* In which boundary conditions (BCs) is defined as "the socio-technical constraints that must be met before a project can be successful". By narrowing the scope to a specific subfield of AI, the goal is to establish BCs that are application-specific rather than sector or organization specific. This intends to enhance the generalizability of the results.

1.2 The research goals

The introduction outlined in the section above directly leads to the goal of this research; to investigate how the BCs affect the success of chatbot implementation in organizations. To maintain a structured approach during the research, a set of research questions is used as guidance. The main research question is formulated as: *How do the socio-technical boundary conditions affect the success of chatbot implementation in organizations?* To answer the research question in a stepwise manner, additional sub-research questions are formulated. These questions are:

1. What are the factors affecting AI implementation?
    a. What are the technical factors influencing the implementation?
    b. What are the social factors influencing the implementation?
2. When is AI implementation deemed successful from different actor perspectives?
3. What are the socio-technical boundary conditions for chatbot implementation?
    a. What boundary conditions can be deducted by analysis?
    b. Are the boundary conditions supported by empirical evidence?
    c. How are the boundary conditions related?

As indicated by the sub-research questions, the research explores the broad field of AI at first (divergence of the scope) whereafter the scope is set to solely chatbot implementation (convergence of the scope). The divergence is used to mitigate the risk of acquiring a tunnel vision at the start of the research, thereby not stimulating idea generation. Also, including the broad field of AI at first is expected to enhance the generalizability of results obtained later on. Therefore, the broad scope is used to acquire a proficient knowledge base on AI and the socio-technical environment it is in. Then, different actor perspectives are investigated on their definition of a "successful AI implementation". The motivation for this research question originates from the subjective nature of the meaning of "success", especially in situations where many stakeholders are involved having (potentially) opposing interests. Finally, the acquired knowledge is used to investigate the BCs for the specific use of AI for chatbot technology in organizations. Future research is suggested to follow a similar approach for different application domains.

1.3 The research contributions

By answering the research questions, the research contributions of this thesis consist of three parts. Firstly, the definition of a “successful AI implementation” from different actor perspectives is explored. Being aware of the key success facets that different actors consider during an implementation project is expected to enhance the adoption and acceptance of the technological changes. This can be achieved by considering a multistakeholder approach in the design and evaluation of the technology. Secondly, by investigating the industry’s perspective on “successful AI implementation”, a hierarchy amongst success facets is exposed. The surprising contribution suggests a hierarchy and interrelationship between these success facets and is presented in the form of a strategy map. Thirdly, BCs for chatbot implementation in organizations are proposed. The proposed BCs serve as hard requirements that must be fulfilled before an implementation project can be successful. Therefore, the proposed BCs can be used for future chatbot implementations as guidelines for the design and development, or to identify neglected aspects in troublesome implementations.

1.4 Thesis outline

The remainder of this thesis is structured as follows. At first, the theoretical background is provided on AI and chatbot technology to gain a deeper understanding of the discussed technology. Thereafter, a literature review is performed on the enablers of AI implementation and the prohibitors to AI implementation. The literature review is meant to provide background knowledge on the current understanding regarding the factors influencing the implementation process, therewith answering the first sub-research question. After, the social environment is analyzed in chapter 3 to identify the involved stakeholders and to explore the social complexity. Chapter 4 investigates the definition of a “successful implementation” from a multi-actor perspective, thereby answering the second sub-research question. This is achieved by conducting eight qualitative research interviews with industry practitioners and performing a literature analysis. After chapter 4, all of the previous chapters are used to analyze the implementation process of chatbots in chapter 5. By doing so, a list of BCs is proposed which is thereafter tested on two empirical cases through a case study. Finally, chapter **Error! Reference source not found.** reflects on the research by discussing its conclusions, research contributions, and limitations. To provide a high-level overview of the research structure, a visual representation of the research framework is presented in Figure 1.

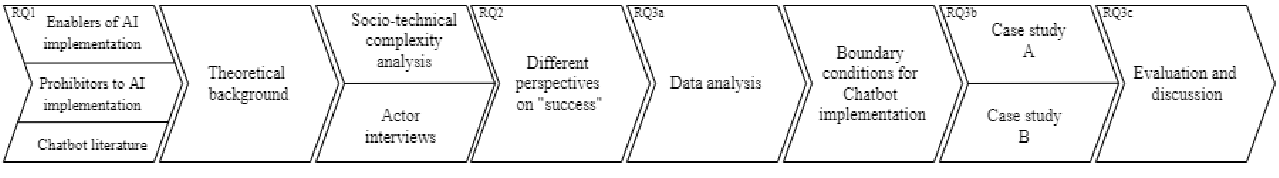

Figure 1: A visual overview of the research framework

2 Theoretical background

2.1 Introduction

The whole field of science and engineering related to building intelligent machines, of which intelligent computer programs especially, is AI (McCarthy, 2007). To scope down the field of AI, it is split into four business applications. The division of the four business capabilities of AI is proposed by Benbya et al. (2021) and consists of: automation, engagement, decision-making, and product exploration. Other typologies proposed in the literature distinguish AI on characteristics such as the type of intelligence, the type of technology embedded, or the function performed (Benbya et al., 2021). The reason for not using one of these typologies is that AI technologies seem to overlap in certain aspects regularly. Also, AI technologies are becoming increasingly embedded within organizational applications (Davenport, 2018). And, because the focus of this research is on the implementation of AI in organizations, the division into four business capabilities is considered to be the most logical way to split the field of AI.

AI-enabled **automation** includes automating structured or semi-structured tasks within an organization. The automated tasks can be both physical or cognitive in nature and are most often characterized as being labor-intensive and repetitive. Also, AI-enabled robotics can sense their environment, learn, and act accordingly. Being able to do so differentiates them from other automation robotics often used in the manufacturing industry. Secondly, AI-enabled **engagement** describes the capability to understand and engage with humans using natural human language. This can be based upon text-based technologies such as chatbots, as well as voice-controlled systems that make use of Machine Learning (ML) and Natural Language Processing (NLP). Thirdly, AI-enabled **insights and decisions** revolve around the exploration of data using algorithms to provide faster, better, or novel insights to decision-makers. Such tasks generally consist of classification, recognition, or prediction capabilities. Fourthly, AI-enabled **product exploration** relates to the enhancement of the exploration process for new products or services. This can either be done by providing AI data-driven insights that support decision-making throughout the innovation process, or by developing new products. Nowadays, the latter is predominantly present in the pharmaceutical industry for the discovery of new drugs (Fleming, 2018).

The remainder of this section on the theoretical background is focused on the business application of AI-enabled engagement; of which chatbot technology especially. The theoretical background on chatbots suffices to gain a proficient understanding of chatbot technology to follow the line of reasoning used during the analysis for a successful implementation in the consecutive chapters.

2.2 Chatbots

"AI-enabled engagement refers to the general capability of computers to understand, respond, engage, and converse with humans using natural human language" (Benbya et al., 2021, p.6). This can be in the form of text-based and voice-based technologies. The technologies used for AI-enabled engagement are under continuous development and have become exceptionally popular in numerous application domains over the last years due to advances in AI, ML, and NLP (Gupta et al., 2020; Rapp et al., 2021). Examples of applications that make use of AI-enabled engagement are chatbots, digital humans, and intelligent agents. In special, the Human-Computer Interaction (HCI) through natural written language by the use of chatbots receives much attention. The “chatbot tsunami” refers to the use of chatbots in different application domains facilitated by flexible platforms that support the design and integration of the software as a service (Grudin & Jacques, 2019; Rapp et al., 2021). The modularity of the technology, the ability to have multiple conversations in parallel, the full-time availability of the service, and the memorization of previous conversations are some advantages that make chatbots appealing business cases for the industry (Adamopoulou & Moussiades, 2020). But even though the development is ongoing, the technology is not able to have full human-level language abilities yet (Benbya et al., 2021). This causes the technology to not meet the user’s expectations and can make the user feel frustrated, misunderstood, or powerless (Rapp et al., 2021).

The tension between the user’s expectation and the machine’s capabilities leads to the discussion on *Humanlike vs. Machinelike Conversations* (Benbya et al., 2021). Humanlike conversations are imitated by assigning human-like attributes to the chatbot such as a name and an avatar; this process is called anthropomorphism. Doing so is suggested to enhance the customer’s experience in an effective and enjoyable manner (Crolic et al., 2022). But literature also suggests that more humanlike conversations should not always be the end goal due to increased expectations and undesirable perceptions of anthropomorphism (Hill et al., 2015). A situation in which this is the case is presented by Crolic et al. (2022). The authors’ research results show that “when customers enter a chatbot-led service interaction in an angry emotional state, chatbot anthropomorphism has a negative effect on customer satisfaction, overall firm evaluation, and subsequent purchase intentions”(Crolic et al., 2022, p.132). The negative effect is suggested to be caused by inflated pre-encounter expectations of chatbot efficacy.

The previous example illustrates the context-dependency and state-dependency that shape the HCI and its potential consequences. In the following sections, chatbot technology is explored to gain a deeper understanding of the underlying concepts and design choices that must be made whilst designing conversational agents. The technological knowledge base will later be used to describe and explore the far-reaching social ripple effect that can be caused by the use of a chatbot.

2.2.1 Chatbot typology

There are diverse characteristics on which chatbots can be classified. Examples of such characteristics are the chatbot’s capabilities, ease of use, or underlying technology. For now, the proposed typology by Gupta et al. (2020) is presented, which categorizes chatbots into three distinct types.

Menu/Button-Based Chatbots

The first type describes chatbots that present the user with button options or top-down menus. These chatbots are the simplest form and thereby the most commonly used. The technology supporting chatbot interaction is based upon the principle of a decision tree. A decision tree is a decision support tool that guides the user toward an answer (a leaf node) by making consecutive choices. Even though decision tree-based chatbots are simple to understand and interpret, "these menu-based chatbots are comparatively slower in terms of performance and cannot be completely reliable to get the desired answer" (Gupta et al., 2020, p.255).

Keyword Recognition-Based Chatbots

The second type relates to chatbots that allow the user to input text rather than selecting one of the button options. By doing so, the user has more freedom and is enabled to ask advanced questions which allows for a more natural conversation. The chatbot then recognizes keywords (the lexical form) within the question based on keyword matching to determine the required response. To decide on the required response, a rule-based system is used that leads to a set of humanly hand-coded pre-defined answers. Despite a more natural conversation is frequently desired by the industry, allowing users to input text can lead to situations in which the chatbot is misinterpreting the question or not being capable of providing the desired answer (Crolic et al., 2022; Gupta et al., 2020).

Contextual Chatbots

The final chatbot type is technologically the most complex yet can provide the best quality of user experience if designed correctly (Gupta et al., 2020). The contextual chatbot makes use of NLP, ML, and Deep Learning (DL) to determine the user's request and answer accordingly. By memorizing the user's previous likings or requests and by being connected to multiple data sources, personalized answers can be generated that show improvement in quality over time. Unfortunately, the difficulty for this type of chatbot is the building and training of the algorithms used (Adamopoulou & Moussiades, 2020).

2.2.2 Chatbot implementation

The implementation process of a chatbot is highly characterized by the context and the use of the technology. Organizations must clearly define the *who*, *what*, *when*, *where*, and *why* the chatbot is developed and implemented. Several other common challenges related to the design are addressed by Castro et al. (2018): how to design interactions that maximize customer satisfaction? How to shape the conversational environment; will the bot provide text-based responses, hyperlinks, or visuals? Which services are taken over by the bot and which need human assistance? How is the provided data managed? And will the conversations including the user data be logged and stored? At first, organizations must be able to answer such questions on a conceptual level before moving to the chatbot's architecture design. The architecture design will determine the final technological competencies of the bot.

2.2.3 Chatbot architecture design

There are four main components to the technical architecture of a chatbot design: the front-end, the back-end, the knowledgebase, and the corpus (R. Kumar & Ali, 2020). These four components are supported by four different characteristic base elements of modern chatbots: the knowledge domain, response generation, text processing, and ML model (Lokman & Ameedeen, 2019).

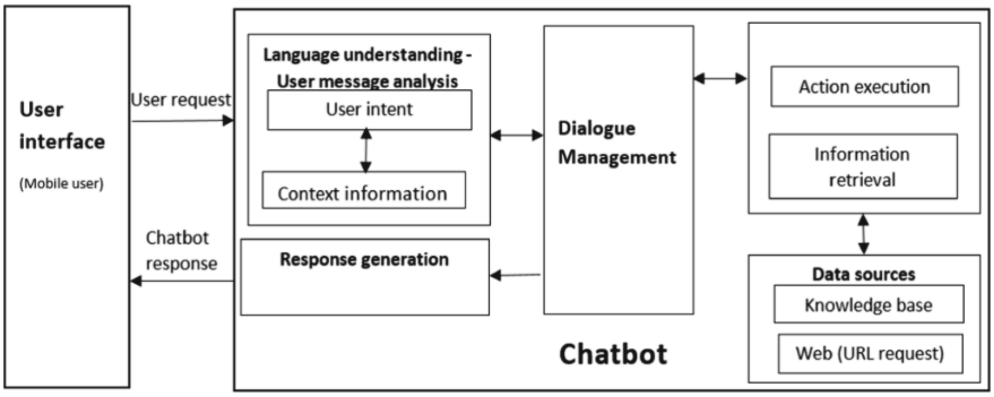

Figure 2: A general chatbot architecture (from Adamopoulou & Moussiades, 2020, p.380)

Understanding the use case of the chatbot will help to set up the requirements and make design choices for developers. At the same time, it helps users to understand what to expect from the chatbot (Adamopoulou & Moussiades, 2020). As guidance to the technical architecture design, Figure 2 is used. The HCI starts with a user request at the front end, shown in the upper left side of the figure. The user input consists of a request of which the form is dependent on the type of chatbot (such as menu/button-, text-, voice-based). This leads to the knowledge base and the first base element: **the knowledge domain**. The knowledge domain of a chatbot suggests the knowledge it covers and can be described as either open or closed. An open domain means the chatbot must know general knowledge including recent news and basic human understanding. A closed domain means the chatbot is knowledgeable on a specific topic such as customer service or simple task execution. Current research suggests closed domain chatbots are easier to build and are capable of producing sufficient results, whereas this is more troublesome for open-domain chatbots (Lei et al., 2017; Serban et al., 2017).

The type of domain influences the requirements for the language understanding and response generation of the chatbot which often uses NLP and AI to determine user intent (middle left side of Figure 2) and is described as the back-end (R. Kumar & Ali, 2020). The **response generation** can be based upon retrieval of pre-defined responses (as in Keyword Recognition-Based Chatbots), upon generative responses that are generated on trained classifiers (as in Contextual Chatbots), or a hybrid of the two types (Lokman & Ameedeen, 2019). The most popular model used for both retrieval and generative methods is called Sequence-to-Sequence (Seq2Seq). In the following sections, the Seq2Seq and other ML concepts are discussed.

2.2.4 Machine Learning for chatbots

AI is the broad field of building intelligent machines, of which ML is a subset including the machines that can learn and improve without being explicitly programmed. NLP is the scientific field that covers linguistics, computer science, and AI and regards the understanding of natural language interaction between humans and machines. NLP can be divided into Natural Language Understanding (NLU) and Natural Language Generation (NLG), which describe the understanding of human input and generation of machine output, respectively. A generic visual representation of these fields is presented in Figure 3.

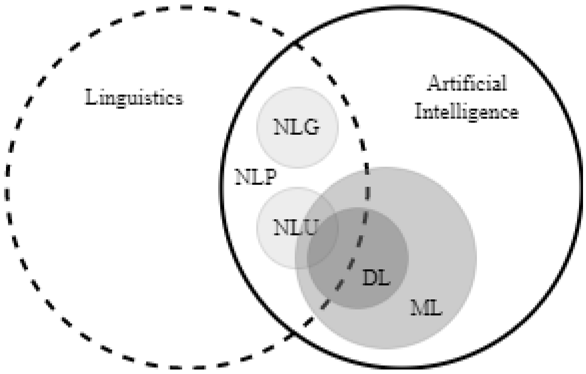

Figure 3: A visual representation of AI and the related subfields

The basics of ML

ML is the general term to describe machines with the ability to learn from data. This is achieved by approximating algorithm parameters (most commonly applied is the Method of Least Squares, see Abdi (1974) for an elaborate explanation) that are thereafter used to “predict” new outcomes. To approximate the algorithm parameters, part of the available data is used which is referred to as *the training data*. For ML parameters, more training data results in better parameters (Lokman & Ameedeen, 2019). The estimated parameters are then tested on *the test data* to determine the model’s accuracy. For an elaborate explanation of ML, the types of ML, and its challenges please refer to Chowdhary (2020) chapter 13: Machine Learning. Within ML, there is a sub-field called Deep Learning (DL). DL is inspired by the human brain and makes use of artificial Neural Networks (NN). Sections 2.2.4.3 and 2.2.4.4 require a basic understanding of these concepts, therefore a brief introduction is presented in Appendix A: Deep Learning introduction.

NLP for chatbots

To correctly understand the user’s text input, NLU is of particular importance to the chatbot performance (Vijayaraghavan et al., 2020). NLU helps the bot to understand the text data, sentiment, intent, and grammar. Examples of NLU tasks include slot filling and intent classification, which is used to determine essential (contextual) information about the user and his/her request (Q. Chen et al., 2019). Slot filling can be used to save or adjust user information and stimulate future conversations such as a user's "name", "e-mail", or "telephone number". These slots can be requested by the bot and completed by the user.

Intent classification describes the categorization of a user's input to a classified intent. For example, "I would like to book a table for tonight" and "I want to make a reservation" must both be classified under the intent "make a reservation" even though the word choice is different. To determine the correct intent, Word Embedding (WE) is used to extract the semantic and syntactic meaning. "Word Embedding (WE), or vector representation of word, are real numbers in vector space that can denote a semantic relationship (by distributional hypothesis) between words within a specific vocabulary" (Lokman & Ameedeen, 2019, p.1014). This means that words are translated into a real number vector, of which the vector numbers of word A are influenced by its context (words B and C in the same context window). As a result, semantic and syntactic relationships among words are captured which are represented by similar word vectors in similar contexts. A common example of this is the vector computation of the word *queen*. Manipulating the vector representation of *king* by computing the vectors *king – man + woman* results in the vector representation of *queen.* Such relations are captured for semantics (such as a country and its capital) and syntactic (such as verb tenses) also.

Overall, there are two methods to generate WE: the count-method and the predictive-method. On the one hand, the count-method counts the co-occurrence of words within a specific context, thereby forming clusters of words that mean the same (e.g. LSA; a count of word frequency in a context, HAL; a count of pair distances in a set window). On the other hand, the predictive-method predicts the co-occurrence based on approximated ML parameters. When compared, the predictive-method greatly outperforms the count-method (Baroni et al., 2014).

The upcoming four sections introduce ML models that are used in modern chatbots. The first two sections cover WE models, and the final two sections briefly describe the LSTM and the Seq2Seq models. These models are discussed to provide a basic understanding of the technology and procedures enabling the capabilities of modern chatbots. Regularly, a combination of these four ML models is used for the NLP of a chatbot (e.g., chatbot DeepProbe uses LSTM-Seq2Seq and chatbot SuperAgent uses GloVE and LSTM-Seq2Seq) (Lokman & Ameedeen, 2019).

2.2.4.1 GloVE

Global Vectors (GloVE) is a word representation scheme that is aimed at extracting semantic relations between words into their embeddings and was introduced in 2014 by Jeffrey et al. (2014). The model counts the co-occurrence of words (denoted by $\mathcal{X}$), whose values $\mathcal{X}_{ij}$ translate to the number of times j occurs in the context of i. Then, $\mathcal{X}_i = \sum_k \mathcal{X}_{ik}$, is the number of times the word appears in the context. Finally, the probability is computed by: $P_{ij} = P(j|i) = \mathcal{X}_{ij} / \mathcal{X}_i$. An example probability matrix for the target words *ice* and *steam* is shown in Table 1.

Probability and ratio	K = solid	K = gas	K = water	K = fashion
P(k | ice)	$1.9 * 10^{\wedge -4}$	$6.6 * 10^{\wedge -5}$	$3.0 * 10^{\wedge -3}$	$1.7 * 10^{\wedge -5}$
P(k | steam)	$2.2 * 10^{\wedge -5}$	$7.8 * 10^{\wedge -4}$	$2.2 * 10^{\wedge -3}$	$1.8 * 10^{\wedge -5}$
P(k | ice)/P(k | steam)	8.9	$8.5 * 10^{\wedge -2}$	1.36	0.96

Table 1: Co-occurrence probabilities for ice and steam (from Jeffrey et al., 2014, p.1534)

As can be seen in Table 1, the number shown in the last row depicts the ratio of the probabilities of *ice* over *steam*. It can be expected that the word *solid* (column 2) is mentioned more often in combination with the word *ice* rather than *steam;* this shows in a high ratio (final row, column 2). The opposite is true for the third column, related to the combination with *gas*. Finally, if *ice* and *steam* are either related or unrelated to another word (*water* and *fashion*), the ratio must be close to 1. Therefore, probability is an important expression for the GloVE vector representation.

Besides the word-word statistics and co-occurrence probabilities, GloVE considers the contextual distance of words within the context window. This means that if words are adjacent, the value is 1. If words are one word apart, the value is ½. Three words apart, ⅓, and so on. The co-occurrence probabilities in combination with the contextual distance determine the weighting function (f($\mathcal{X}_{ij}$)). Having calculated the weighting function, the loss function can be calculated. The loss function, shown in Equation 1, formally represents GloVE and is used as the optimization parameter.

$$J = f(X_{ij})(w_i^T u_j + bw_i + bu_j — \log(X_{ij}))^2$$

Equation 1: The loss function for GloVE Word Embedding

For the loss function parameters, $f(X_{ij})$ is the weighing function, $w_i^T u_j$ the dot product of the input and output vectors, $bw_i + bu_j$ the bias terms, and X_{ij} the number of occurrences of *j* in context *i*. For further reading on GloVE, please refer to Jeffrey et al. (2014).

2.2.4.2 Word2Vec

Word2Vec is another WE representation model that is often used to translate text input into vector representations. However, whereas GloVE is based on words' (global) co-occurrences, Word2Vec captures whether words appear in similar (local) contexts by training a Neural Network (NN) (Rong, 2014). The model is used in two different ways, which are algorithmically the same. The Continuous Bag-of-Words (CBOW) predicts the target word (e.g., *ice*) from the context (e.g., *Frozen water turns into*). The second model, The Skip-Gram, predicts the surrounding context from the target word, which in essence is the inverse of CBOW. Normally, the CBOW model performs better on smaller datasets and the Skip-Gram model on larger ones. To increase the accuracy of the model, the right type should be used and the training data, vector dimension, and window size can be increased. For an elaborate overview of the Word2Vec learning, please refer to Rong (2014).

2.2.4.3 LSTM

Whereas the two previously discussed methods are used to transform a text input into a WE, Long-Short Term Memory (LSTM) is a type of NN model that is used to classify the input and understand the user request. For a brief introduction to NNs, please see Appendix A: Deep Learning introduction. LSTM is, just at Gated Recurrent Unit (GRU), a RNN designed to solve the Vanishing/Exploding Gradient problem (Lecun et al., 2015). Similar to RNN, a LSTM cell combines the output from the previous iteration with the novel input as the inputs for the next step. In addition, a LSTM cell has a *state,* which is memorized over multiple time steps. The state is used as the third addition to the input and is updated according to the output also. The *state* of a LSTM cell is shaped by its *gates,* which are shown by the gates i_t, f_t, and o_t in Figure 4. The gates control the flow of information in a node and function as weights and biases to determine the degree of importance of the current input. For the cell shown below, the input gate dictates the degree of saved information being recalled, the forget gate decides the amount of current information to be stored, and the output gate determines how much the output is influenced by the new calculations. By solving the Vanishing/Exploding Gradient problem, the use of LSTM can drastically improve the analysis of text input by providing a solution to the long-term dependency of texts (Shewalkar et al., 2019). A long-term dependency describes situations in which the desired output is dependent on the input presented a long number of inputs before.

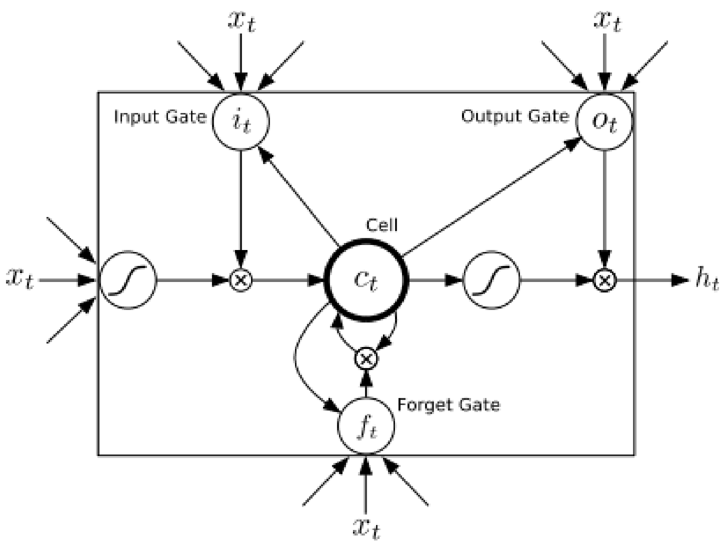

Figure 4: A schematic overview of a LSTM cell (from Graves et al., 2013, p.2)

2.2.4.4 Seq2Seq

A Sequence-to-sequence (Seq2Seq) model is a type of ML architecture often used for language processing, text summarization, and image captioning. The architecture uses RNN, which are often in the form of the GRU or LSTM, and was introduced by Google in 2014. A schematic overview of the architecture is presented in Figure 5. The Seq2seq model consists of three key components: the encoder, the intermediate (encoder) vector, and the decoder. At first, the inputs (denoted as x1, x2, x3) are embedded using a WE model. Thereafter, the vector representations are encoded by the RNNs. By encoding the vector inputs, variable-length vectors are transformed into an intermediate encoder vector of a similar length (Palasundram et al., 2019). This allows the model to map fixed-length inputs with a fixed-length output of which the vector lengths may differ. The third step involves the RNNs in the decoder to decode the intermediate vector and rightly predict the corresponding output of a variable length.

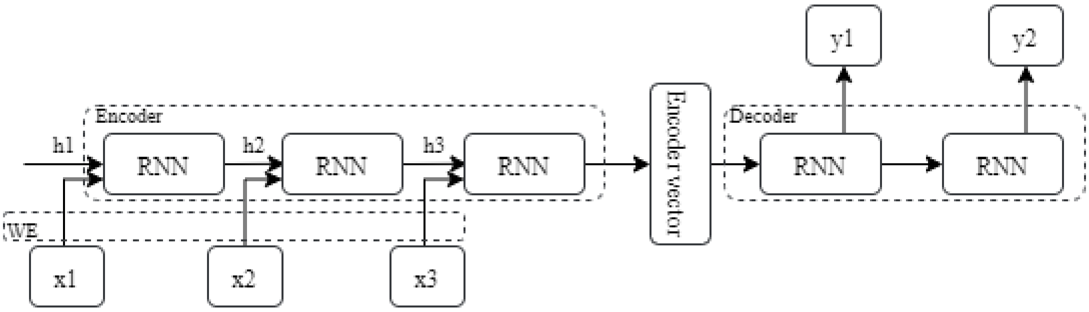

Figure 5: A schematic overview of the Seq2Seq model architecture (from Kostadinov, 2019)

Besides the most simplistic Seq2Seq model shown above, some enhancements need to be made to strengthen the model's performance for more complex tasks. Examples of these enhancements can be: reversing the input sequence, adding an Attention mechanism, or using character embedding instead of WE. This leads to the following statement by Palasundram et al. (2019, p.58): "One key challenge with Seq2Seq (as in other NN models) is that there are so many settings and hyperparameters that need to be tuned in order to get a good performing working model".

The statement points to the conclusion of this subsection on Machine Learning for chatbots. Several concepts and models have been briefly introduced to highlight the complexity of the technology and to indicate the considerable number of architectural design choices that must be made to correctly design a chatbot. It must be noted, apart from the models discussed above, there are many more of which each can be of use for a particular use case. As a starting point for further reading, one may look for Retrieval-based NN, Generation-based NN, Multilayer perceptron, and Convolutional NN.

2.2.5 Chatbot evaluation

There are many ways in which the performance of a chatbot can be represented. Thorat and Jadhav (2020) suggest the most important parameters are: multitasking, multi-channel support, flexibility, price, total users, engaged users, scalability, retention rate, and fallback rate. Opposingly, Lokman and Ameedeen (2019, p.1017) state: "There is strong evidence that automated evaluation metrics (evaluation without human judgment) commonly used in chatbot literature are not correlated strongly with human judgment". Therefore, the authors suggest using human evaluation metrics to assess the quality of the system's output toward human perception. Proposed metrics include a human assessment of the system performance, A/B testing, and BLEU. Human assessment describes the evaluation techniques in which humans assess the chatbot's performance. This is often measured with relative scales such as 1 to 5 (1 indicating an unhelpful conversation and 5 being an excellent conversation) or assessing the capabilities as *bad*, *fair*, *good*, or *excellent*. Besides such methods, A/B testing deploys two versions (version A and version B) of the chatbot with the same central system to compare the user's feedback and results on both versions. Lastly, BLEU (bilingual evaluation understudy) is an algorithm that is used to assess the quality of machine-translated texts. The quality is determined by comparing the machine-written output to that of a human.

Unfortunately, the human evaluation techniques suffer from two drawbacks: the difficulty in obtaining sufficient feedback data and the lack of explainability of users' (dis)satisfaction (Vijayaraghavan et al., 2020). Therefore, algorithm inspection techniques proposed by Vijayaraghavan et al. (2020) are meant to analyze the inner working of the system and are used by developers to identify the bot's strengths and weaknesses. The techniques specifically help developers to pinpoint where mistakes originate and where adjustments must be made. In Table 2, popular chatbot algorithms are presented with a brief introduction and their testing methodology. For a more elaborate explanation, please refer to Vijayaraghavan et al. (2020); the article on which Table 2 is based.

Chatbot algorithm	Description	Testing methodology
Naïve Bayes (NB)	NB aims to classify words based on probabilities to identify the user intent. By training, the algorithm learns the commonality between words and certain categories thereby assigning more weight to the corresponding words.	Since intent classification is often the first step in chatbot conversation, high accuracy is required. This is often tested with the help of k-fold cross-validation (see Refaeilzadeh et al. (2016) for an elaborate explanation of cross-validation).
Support Vector Machines (SVM)	SVM also aims to classify words, yet is based upon structural risk minimization. It offers the opportunity to handle large amounts of text data and often outperforms NB and K-NNs.	K-fold, precision, and recall are often used. Precision describes how good the model is at predicting a category (=TP/(TP+FP)). Recall describes how many times the model was able to detect a category (=TP/(TP+FN)). *TP = True Positive, TN = True Negative,* *FP = False Positive, FN = False Negative.*
Deep Neural Networks (DNN)	DNNs are described in the previous section on 2.2.4.3. The required volume of training data and explainability of DNNs remains a problem, especially for the chatbot domain.	The Turing test is used to assess the performance of the DNN. The Turing test describes the ability of a machine to mimic human behavior. Passing the test means that the machine is indistinguishable from a human.
Markov Chains (MC)	MC is often used for text generation. By knowing the probability of co-occurrences among words, the text of n-order chains can be constructed. N-order addresses the number of words in the same group. Multiple groups can make up one generated sentence. The larger the n-order, the more likely the text will be to the training data.	The performance can be tested by grammatical parsing, output analysis, and user feedback testing. All three techniques rely on human evaluation of the generated text.
Natural Language Processing (NLP)	NLP includes a wide range of functionalities such as sentiment analysis, entity recognition, masking, and text summarization.	Variations of the Turing test can be used to assess the performance of the NLP algorithm. Also, human assessment by imposing difficult or error-prone text inputs should be considered.

Table 2: Chatbot algorithms including their testing methodology

2.2.6 The social effects of chatbots

The HCI and the human-to-human Interaction (HHI) is changing because of the extensive use of chatbots. The deployed bots facilitate many of the conversations we have online today and pose potential for social good but do not come without any risks. Nowadays, the most frequent use cases for the technology remain to be customer service support, media and content distribution, and marketing (Følstad et al., 2018). The technology is claimed to enhance a service's availability, accessibility, and affordability due to the automated system (Følstad et al., 2018; Nguyen, 2019). But some of these claims can be argued. For example, by making use of a chatbot for customer support, organizations externalize the required actions to the user rather than internalize these to their customer service department. If the user is able to execute the necessary actions and fulfill his/her request, the bot may have been a more efficient and effective solution. However, if the user is not able to do so, it may pose risks for society or increase the costs for the organization in the long term. As an example, chatbots are widely used for healthcare applications to provide healthcare advice or redirect users to the specialized hospital department (e.g., Symptomate, TARS). If the user is advised or redirected incorrectly due to: the bot not understanding the request, the bot not providing personalized answers, or the user not being able to collaborate with the bot (by being minority groups or edge cases), the negative externalities can be far-reaching and more costly than before.

Besides the commercial use cases for the technology, there are chatbot applications that can serve society well. Følstad et al. (2018) describe three areas in which chatbots can be used for social good: chatbots for autonomy, chatbots for competence, and chatbots for social relatedness. Chatbots for autonomy relates to the possibility for users to digitally empower themselves by being able to easily gain access to external digital systems or services. Chatbots for competence describes the possibility for users to be assisted by chatbots full-time. For example, for education and support training. Chatbots for social relatedness describes the possibility for chatbots to decrease social isolation and disconnectedness. Two example cases of such chatbot uses are described by Zhou et al. (2020) and Oh et al. (2017). Zhou et al. (2020) describe the design and implementation of XiaoIce. "XiaoIce is uniquely designed as an artificial intelligence companion with an emotional connection to satisfy the human need for communication, affection, and social belonging" (Zhou et al., 2020, p.53). Also, Oh et al. (2017) describe the use of a chatbot that uses high-level NLU and emotion recognition for on-demand psychiatric counseling. Hence, if such bots can achieve high levels of user experience and user satisfaction, they can offer a promising solution for social isolation, disconnectedness, and psychiatric counseling thereby stimulating the social readiness of the technology.

Ethical risks

The technological advances pictured by the previous two examples open up ethical challenges specifically related to this field of AI. Firstly, there is an ethical risk related to the asymmetrical redistribution of power in HCI (Murtarelli et al., 2021). Nowadays organizations are able to collect, integrate, and aggregate all sorts of user information. The chatbots are able to analyze all of this data and take actions according to the goals and wishes of the organization. These actions are used within the HCI and thereby shape the conversation with the human agent. This information asymmetry leads to a redistribution of power between the involved parties and should therefore be carefully evaluated on moral and ethical values. Secondly, the humanization of chatbots is often considered to improve the user experience and is therefore strived after by organizations (Crolic et al., 2022). Chatbots are designed according to a persona, are equipped with a profile picture, and make use of dynamic response delays to mimic human behavior (anthropomorphism) (Crolic et al., 2022; Gnewuch et al., 2018). By doing so, there is a risk of users not knowing they are in contact with a machine rather than a human. Discovering this at a later stage might lead to the feeling of regret, disappointment, or being deceived, thereby not be willing to interact with the service anymore (Yang et al., 2022). Thirdly, there is an ethical risk regarding the management of online security and users' privacy. Closely related to the previous two concerns, users might be tricked into revealing personal information to unknown people making use of chatbots. The personal information can be stored and aggregated after which it might be sold or used by people with bad intentions. Therefore, security, transparency, privacy, and confidentiality remain core values for these ethical concerns.

2.2.7 Conclusion

The theoretical background introduced throughout this chapter is meant to set the scope for the remainder of this research. The chapter started on a high-level exploring AI and four of the common business-applications of AI, of which AI for engagement has been elaborated on in more detail. Thereafter, a chatbot typology consisting of three main categories has been discussed. Strongly related to chatbot typology are the architecture design choices that must be made during the design phase of a chatbot implementation. One of the conclusions is that there are many choices to be made and parameters to be optimized in order to design a well-performing chatbot; this conclusion was supported by the statement by Palasundram et al. (2019) on page 25. At last, chatbot evaluation techniques, the social effects, and the ethical risks are investigated. It is argued that human assessment is needed to properly evaluate the performance and use of chatbots, because automated quantitative measures are limited in reviewing the holistic picture. Also, even though chatbots have proven to be of social use to counteract loneliness or depression, the misuse due to the redistribution of power and humanization of the bot must be actively managed to protect society at large.

2.3 Enablers of Artificial Intelligence implementation

Enablers of AI implementation can be described as factors that allow, ease, or enhance the implementation of the technology thereby potentially increasing the project's success. In the field of Project Management, the understanding of project success is widely discussed and changes over time (Rodney Turner & Müller, 2005). During the 1980s and 1990s, Critical Success Factor (CSF) frameworks "were developed on the basis that success is stakeholder-dependent and involves interaction between project supplier and recipient" (Rodney Turner & Müller, 2005, p.56). Success factors are described as (critical) factors or project dynamics that have a significant impact on the project's success (Pinto & Slevin, 1988).

In the past, research has been done on CSFs across the project life cycle (Pinto & Slevin, 1988), on the project management of Information Systems (IS) and Information Technologies (IT) industries (Hartman & Ashrafi, 2002), as well as on other disruptive IT-technologies such as Enterprise Resource Planning (ERP) systems (Tarhini et al., 2015). However, because the implementation of AI affects society and individuals' lives to a greater extent than preceding technologies, "the success factors for AI projects are important and dramatically more expansive than those for a typical information systems project" (Miller, 2021, p.388). Therefore, the following section explores the suggested enablers of AI implementation.

2.3.1 Search strategy

To find relevant articles on success factors, the Scopus database is used with the search string: TITLE((success factor OR critical success factor OR csf OR enabler OR enablers OR promoter OR facilitator OR supporter) AND (artificial intelligence OR ai)). The search string results in 10 articles which are presented in Table 3. The table highlights the reference, study design, key findings and whether or not the article is included within the synthesized literature (the quality appraisal).

#	Reference	Study design	key findings	Quality appraisal: Include / exclude
1	Miller, G. J. (2021). Artificial Intelligence Project Success Factors: Moral Decision-Making with Algorithms. Proceedings of the 16th Conference on Computer Science and Intelligence Systems, FedCSIS 2021, 25, 379–390. https://doi.org/10.15439/2021F26. (Miller, 2021)(Miller, 2021)(Miller, 2021)	A systematic literature review is used to identify 71 success factors in 14 groups related to moral decision-making with algorithms.	71 success factors divided over 14 groups, that are all divided within the 3 categories: Management, Procedures, or Product Qualities.	Include.
2	Hamm, P., & Klesel, M. (2021). Success factors for the adoption of artificial intelligence in organizations: A literature review. 27th Annual Americas Conference on Information Systems, AMCIS 2021, (August).	A literature review is performed to categorize all of the identified success factors for implementing AI.	A structured overview of 36 factors, categorized by the Technology, Organization, Environment (TOE) framework.	Include.

3	Welsing, Martin & Maetschke, Jan & Ays, Julian & Gützlaff, Andreas. (2021). KI und Data Mining in der Produktion sinnvoll nutzen. VDI-Z. 163. 56-59. 10.37544/0042-1766-2021-07-08-56.	Artificial intelligence in small and medium-sized enterprises - Five factors decide on the success of implementation: Putting AI and data mining to good use in production | [Ki und data mining in der produktion sinnvoll nutzen]		Exclude.
4	Dora, M., Kumar, A., Mangla, S. K., Pant, A., & Kamal, M. M. (2021). Critical success factors influencing artificial intelligence adoption in food supply chains. International Journal of Production Research. https://doi.org/10.1080/00207543.2021.1959665	Using TOEH theory with rough-SWARA to determine the CSFs in a Food Supply Chain (FSC).	CSFs: technology readiness, security, privacy, customer satisfaction, perceived benefits, demand volatility, regulatory compliance, competitor pressure and information sharing among partners.	Include.
5	Chen, H., Li, L., & Chen, Y. (2021). Explore success factors that impact artificial intelligence adoption on telecom industry in China. Journal of Management Analytics, 8(1), 36–68. https://doi.org/10.1080/23270012.2020.1852895	The paper proposes a framework to explore the impacts of success factors on AI adoption in telecom industry by integrating the technology, organization, and environment (TOE) framework and diffusion of innovation (DOI) theory.	Direct managerial capability impact on AI adoption: Compatibility, relative advantage, complexity, managerial support, technical capability. Indirect impact: Government involvement, market uncertainty, competitive pressure, vendor partnership.	Include.
6	Mir, U. B., Sharma, S., Kar, A. K., & Gupta, M. P. (2020). Critical success factors for integrating artificial intelligence and robotics. Digital Policy, Regulation and Governance, 22(4), 307–331. https://doi.org/10.1108/DPRG-03-2020-0032	Determine CSFs in developing IASs using AI and robotics.	the most important CSFs in integrating AI with robotics in India: Emerging economy multinational enterprises (EMNEs), governance, utility, manpower, capital, software, data, and hardware.	Include.
7	Adha, T. J., Henuk, Y. L., & Supriana, T. (2020). Evaluation of factor influencing the success of Artificial Insemination (AI) of beef cattle through UPSUS SIWAB program in Deli Serdang Regency, Sumatera Utara Province, Indonesia. IOP Conference Series: Earth and Environmental Science, 454(1). https://doi.org/10.1088/1755-1315/454/1/012055	The study attempts to increase the Artificial Insemination (AI) of beef cattle in the Deli Serdan Regency of Sumatera Utara.	Not applicable.	Exclude.
8	Alhashmi, S. F., Salloum, S. A., & Mhamdi, C. (2019). Implementing Artificial Intelligence in the United Arab Emirates Healthcare Sector: An Extended Technology Acceptance Model. International	Explores CSFs by applying the Technology Acceptance Model (TAM) in the healthcare sector. Qualitative study	Most common CSFs of TAM suggests: Perceived usefulness (PU), Perceived ease of use (PEU), Attitudes towards use (ATU), Behavioral intention to use (BIU).	Include

	Journal of Information Technology and Language Studies (IJITLS), 3(3), 27–42. Retrieved from http://journals.sfu.ca/ijitls	among 53 employees working in Dubai IT and healthcare sectors.	Extended TAM (ETAM) includes: managerial, organizational, operational, and IT infrastructure factors have a positive effect on PU and PUE.	
9	Yoon, S. N., & Lee, D. H. (2019). Artificial intelligence and robots in healthcare: What are the success factors for technology-based service encounters? International Journal of Healthcare Management, 12(3), 218–225. https://doi.org/10.1080/20479700.2018.1498220	Developing a framework for the success factors of a technology-based service encounter (TBSE).	TBSE needs to be patient-friendly and easy to use so that patients reduce their rejection to changes and actively participate in the technologies. Service providers need to be acquainted with the use of the technologies, their advantages, and disadvantages through training.	Include.
10	Ybañez, A. P., Ybañez, R. H. D., Fojas, A. J. G., Malate, P. L. T., Abela, J. V., Nuñez, E. S., … Lopez, I. F. M. (2017). Retrospective analysis of selected artificial insemination (AI) related practices, maternal factors, and success rate of AI in water buffaloes in three rural areas in the Philippines (1998-2015). Livestock Research for Rural Development, 29(4).	Study aims to retrospectively analyze Artificial Insemination (AI) in Philippine water buffaloes.	Not applicable.	Exclude.

Table 3: The search results for enablers of AI implementation

2.3.2 Selection criteria

The quality appraisal criteria is based on the degree of reusability of the results presented by the literature. For example, references 1 and 2 both conduct a literature review whereafter an objective summary is presented in the form of a table. Even though both sources focus on AI from a different perspective (reference 1 focuses on moral decision making in AI whilst reference 2 on the adoption of AI in organizations) the results are considered to be valuable to this research. Also, both literature reviews do not include other references mentioned in Table 3, nor do they include the same literature in their study. Therefore, both articles are included.

Secondly, the study design of references 4, 6, and 9 attempt to identify the CSFs in a specific industry. It is assumed that industry-specific CSFs are related to the more general CSFs described in the literature reviews of references 1 and 2. To support ground for the assumption, the differences and similarities between the more general CSFs and the industry-specific CSFs will be explored later. Therefore, references 4, 6, and 9 are included also.

Following, references 5 and 8 both make use of existing frameworks (TOE, DOI, TAM) to a specific industry to determine the CSFs. These references are included since they can serve as an example on how to expose field-specific CSFs. Also, both articles clearly elaborate on the applied frameworks which can be valuable during a later stage in this research.

Finally, references 7 and 10 are excluded. The study design of both articles is on the CSFs of Artificial Insemination of animals, abbreviated as AI. Because of the abbreviation, the articles satisfied the search term yet are not applicable to this research.

2.3.3 Synthesized literature

In Table 4, the success factors for AI adoption as described by Hamm & Klesel (2021) (ref.2) is presented in column 3. The results of this study are presented in particular because the "results can help scholars and practitioners to include those factors in theory development and to implement AI projects more successfully" (Hamm & Klesel, 2021, p.1). Also, the authors state the number of times the success factor is mentioned in the reviewed literature (column 4, Table 4). Knowing this helps to assess how well-established the factor is by academia. In the second column of Table 4, the success factors related to moral decision-making by Miller (2021) (ref.1) are presented. The two studies' results are presented in parallel to show the contrast in suggested success factors caused by the different scope and context of the studies (the context-dependency).

The success factors in Table 4 are divided according to the Technology, Organization, and Environment (TOE) framework developed by Tornatzky et al. (1990) because this framework has been widely applied to describe technology adoption and has been found useful in the past (Bérubé et al., 2021). The framework helps to make a distinction in factors related to the technology (e.g., capabilities and characteristics), the organization (e.g. (in)formal structures and culture), and the environment (e.g., regulation and industry characteristics).

Type of success factor	Description of success factors in moral decision-making (ref.1)	Description of success factors for AI implementation in organizations (ref.2)	Number of times mentioned in literature (ref.2)
Technological	Source Data Qualities	Compatibility/ IT infrastructure	10
	Training Data Qualities	Relative advantage	7
	Models & algorithms Qualities	Availability and quality of data	5
	User Interface qualities	Tool availability	3
	System configuration	Identified business needs	2
	Data privacy & confidentiality	Security / reliability	2
	Decision quality	Complexity	2
	Transparency & understandability	Perceived barriers	1
	Usage control	Generalizability/ scalability	1
	Investigation	Technology management	1
		Satisfaction with existing systems	1
		Technology readiness	1
Organizational	Governance	Top management support	8
	Financial benefits	(Technical) competencies	8
	Financial protections	Resources	8
		Organizational size	5
		Organizational structure	4
		Strategy	3
		Organizational readiness	3
		Culture	3
		Organizational innovativeness	2
		Interdisciplinary collaboration	1
		Perceived financial cost	1
		Organizational secrecy policies	1

Environmental		Knowledge and information	1
	Legal protections	Competition / industry pressure	8
		Governmental regulations	7
		Customer readiness	2
		Trust	2
		Industry requirements/ characteristics	2
		External partners / trading partners	2
		Perceived governmental pressure	1
		Perceived pressure from society	1
		Access to external expertise	1
		Public funding	1
		Customer and community support	1

Table 4: The success factors from Miller (2021) and Hamm & Klesel (2021)

As shown in Table 4, the literature review findings by Miller (2021) predominantly highlight technological factors, whilst the other literature review findings present 12 technological factors, 13 organizational factors, and 11 environmental factors; a more even distribution. A possible cause for the remarkable difference is the scope of the literature research. According to Miller (2021, p.379): “Researchers place the burden of responsibility for ethical decisions from AI systems on the system developers”. Hence the burden for moral decision-making by AI is of technological nature according to researchers, which would explain why technological success factors enhance the results of the technology and are therefore mentioned more frequently. Also, the two literature reviews hint towards the severe context-dependency of enablers for AI indicated by the following statement: “The requirements for the system, moral issues, and all aspects of the project are impacted by the context (country, industry sector, functional topic, and use case) of the algorithm” (Miller, 2021, p.386).

Comparing the previous results to the industry-specific study by Dora et al. (2021) (ref.4), which focuses on the CSFs of the implementation of AI in the Food Supply Chain (FSC), shows interesting results as well. The study uses Rough-SWARA, a method for capturing experts’ judgement and assigning a relative importance to variables, to capture the relative importance of CSFs in the FSC. The findings indicate that technology readiness, customer satisfaction, clear vision and strategy, competitor pressure, top management support, regulatory compliance, and information sharing among the supply chain are considered to be the most important factors. The full overview of results is presented in Table 5 below.

Type of success factor	Weight	Description of the success factors	Weight
Technological	0.458	Technology readiness	0.405
		Relative advantage/ perceived benefits	0.169
		Data complexity	0.051
		Compatible facilities	0.102
		Sufficient privacy and security	0.277
Organizational	0.084	Clear vision and strategy	0.387
		Top management support	0.290
		Change management	0.029
		Sufficient resources	0.169
		AI provider commitment	0.099

		Culture and environment	0.055
Environmental	0.308	Demand volatility	0.189
		Regulatory compliance	0.279
		Ethics in data collection	0.051
		Competitor pressure	0.389
		Institutional based trust	0.100
Human	0.163	Training for employees	0.185
		AI implementation team	0.098
		Job security for employees after implementation	0.050
		Information sharing among Supply Chain	0.278
		Customer satisfaction	0.400

Table 5: The success factors including the relative importance (from Dora et al. 2021, p.11)

Comparing the factors identified by Dora et al. (2021) (Table 5) to the success factors stated by Hamm & Klesel (2021) (Table 4, column 3) shows the high degree of similarity in factor description between the two. On the one hand, it seems that solely the factor *Information sharing among Supply Chain* is novel due to the FSC context. The high degree of similarity amongst the factors indicates a relation between general and industry-specific CSFs. On the other hand, the importance and interrelationship between the suggested CSFs might be different. Even though the degree of importance is not investigated within the literature review by Hamm & Klesel (2021), the measure that can be taken from the literature review is the number of times the success factor is mentioned in the literature; organizational factors are mentioned 48 times, technological factors 36 times, and environmental factors 28 times. Organizational factors being least important by experts in the FSC field (weight 0.084, Table 5) whilst being mentioned the most by academia (48 times, Table 4) is interesting and might imply the importance and interrelationship of the factors to be context dependent as well.

2.3.4 Enablers of chatbot implementation

Having synthesized literature on the enablers to AI implementation, a similar strategy is applied to find relevant literature on the enablers of chatbot implementation. By using a similar search term in the Scopus and Google scholar database, solely 3 Journal articles and 1 master's thesis seemed to focus on the CSFs for chatbot implementation in organizations. The main results are summarized in Table 6 shown below. From the key findings highlighted in the last column, a number of similarities between the general CSFs discussed before and the enablers of chatbot implementation can be seen (e.g., technology availability, top management support, change management, and defining project goals).

#	Reference	Study design	Key findings
1	Quiroz Martinez, M. A., Mayorga Plua, S. E., Gomez Rios, M. D., Leyva Vázquez, M. Y., & Plua Moran, D. H. (2021). Chatbot for Technical Support, Analysis of Critical Success Factors Using Fuzzy Cognitive Maps. In M. Botto-Tobar, S. Montes León, O. Camacho, D. Chávez, P. Torres-Carrión, & M. Zambrano Vizuete	Using Fuzzy Cognitive Maps (FCG), qualitative, descriptive research is applied to determine the CSFs for the technical support of a chatbot.	In hierarchical order: Database and response loading, definition scope chatbot, chatbot evaluations, user support, implementation of programs and processes, design of security and UI, project schedule, chatbot responses, chatbot algorithm, flow of chatbot.

	(Eds.), Applied Technologies (pp. 363–375). Springer International Publishing.		
2	Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do Chatbots fail? A Critical Success Factors Analysis. Proceedings of the 42nd International Conference on Information Systems (ICIS 2021), September.	Analysis of 103 chatbots, 20 expert interviews, and a literature review to derive 12 CSFs which are evaluated during a focus group.	Technology availability, user centric cases, chatbot promotion, chatbot design, chatbot progress, top management support, project resources, developmental strategy, chatbot developing team, usefulness, usability, trust.
3	Zhang, J. J. Y., Følstad, A., & Bjørkli, C. A. (2021). Organizational Factors Affecting Successful Implementation of Chatbots for Customer Service. Journal of Internet Commerce, 0(0), 1–35. https://doi.org/10.1080/15332861.2021.1966723	Analysis of six organizations that implemented chatbots for customer service. 14 interviews have been conducted with employees to determine organizational factors influencing the implementation.	Five factors: work and team organization, change management, competencies and competency acquisition, organizational resources, and defining performance measures.
4	Kousa, E. (2019). Exploring Success Factors in Chatbot Implementation Projects. https://www.theseus.fi/handle/10024/166422	Semi-structured interviews with 5 chatbot implementation experts to gain insights into the CSFs.	Defining clear project goals, involvement of stakeholders, forming coherent teams, support from leadership, agile project management, client involvement, testing, UX and conversation designers, continuous improvement.

Table 6: An overview of the results to enablers of chatbot implementation

Also, similar to the comparison between general CSFs and the CSFs for AI in the FSC, there are suggested CSFs that are solely applicable to chatbot implementation and can therefore be described as application- or context-specific. Examples of such factors suggested by the literature in Table 6 are: database and response loading, dialogue flow UX design, and chatbot evaluations.

2.3.5 Conclusion

Synthesizing the CSFs for AI and chatbot adoption shows on the one hand the similarity between general and industry-specific factors. This is noteworthy since complementing literature states the high context-dependency of the factors influencing the implementation (Bérubé et al., 2021; H. Chen et al., 2021; S. Kumar et al., 2021). On the other hand, due to the context-dependency of the technology, the interrelations, degree of importance, and change over time of these factors remain unknown. Overall, the literature review indicates that general CSFs can be used as a starting point for future research or analysis yet should carefully be considered and evaluated when applied to a specific context.

2.4 Prohibitors to Artificial Intelligence implementation

Besides the enablers of AI implementation, there are factors that can hinder the implementation of the technology. Throughout the following section, the prohibitors (often referred to as barriers) are explored. It is assumed that the BCs, enablers, and prohibitors are interrelated in a way and can therefore exhibit similar characteristics. In chapter 5, the interrelationship among the three sorts of factors will be explored more thoroughly.

2.4.1 Search strategy

To find valuable articles, the Google Scholar database is used with the search string: TITLE((Artificial Intelligence OR AI) AND (Barrier OR Barriers) AND implementation AND (Organization OR Organizations)). The Google Scholar database is used because of its wide range of articles and the relevance ranking algorithm within the search engine that helps to acquire an overview of the most significant articles (Beel & Gipp, 2009). The search string results in 227.000 results of which the first 9 have been reviewed due to their relevance to the search term. The 9 articles are presented in Table 7 including the study design, AI application, proposed types of barriers, and the quality appraisal. Thereafter, the same search string is used within the Scopus database. Eventually, this did not result in any additions to the list because of one of two reasons: either the articles did not match the inclusion criteria, or the articles were already added to the list (this was the case for references 2, 4, and 7).

#	Reference	Study design *Application*	Types of barriers	Quality appraisal: Include / exclude
1	Singh, R. P., Hom, G. L., Abramoff, M. D., Campbell, J. P., & Chiang, M. F. (2020). Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. Translational Vision Science and Technology, 9(2), 1–6. https://doi.org/10.1167/tvst.9.2.45	The study “aims to highlight the challenges and barriers to real-world AI adoption that impact the technology’s utility.” *Healthcare organizations.*	Social.	Exclude.
2	Bérubé, M., Giannelia, T., & Vial, G. (2021). Barriers to the implementation of AI in organizations: Findings from a Delphi study. Proceedings of the Annual Hawaii International Conference on System Sciences, 2020-Janua, 6702–6711. https://doi.org/10.24251/hicss.2021.805	Identifying the barriers to the implementation of AI in organizations based on a Delphi Study. *General application.*	Technological, Social, Business, Ethical, Legislative.	Include.
3	Paranjape, K., Schinkel, M., Hammer, R. D., Schouten, B., Nannan Panday, R. S., Elbers, P. W. G., … Nanayakkara, P. (2021). The Value of Artificial Intelligence in Laboratory Medicine. American Journal of Clinical Pathology, 155(6), 823–831. https://doi.org/10.1093/ajcp/aqaa170	Diagnose the anticipated challenges and solutions for the implementation of AI in Laboratory Medicine. *Laboratory medicine.*	Social.	Exclude.

4	Kumar, S., Raut, R. D., Queiroz, M. M., & Narkhede, B. E. (2021). Mapping the barriers of AI implementations in the public distribution system: The Indian experience. Technology in Society, 67(July), 101737. https://doi.org/10.1016/j.techsoc.2021.101737	Mapping the barriers of AI implementations in the PDS. *India's Public Distribution System (PDS).*	Technological, Social, Business, Ethical, Environmental, Legislative.	Include.
5	Kumar, S., Raut, R. D., Queiroz, M. M., & Narkhede, B. E. (2021). Mapping the barriers of AI implementations in the public distribution system: The Indian experience. Technology in Society, 67(July), 101737. https://doi.org/10.1016/j.techsoc.2021.101737	Barriers and pitfalls for AI in gastroenterology: Ethical and regulatory issues. *Gastroenterology.*	Technological, Social, Ethical, Legislative.	Include.
6	Renz, A., & Hilbig, R. (2020). Prerequisites for artificial intelligence in further education: identification of drivers, barriers, and business models of educational technology companies. International Journal of Educational Technology in Higher Education, 17(1). https://doi.org/10.1186/s41239-020-00193-3	"The drivers and barriers that currently affect data-based teaching and learning paths." *Educational technology (EdTech) companies.*	Social, Business, Environmental, Legislative.	Include.
7	Kumar, P., Bhamu, J., & Sangwan, K. S. (2021). Analysis of Barriers to Industry 4.0 adoption in Manufacturing Organizations: An ISM Approach. Procedia CIRP, 98, 85–90. https://doi.org/10.1016/j.procir.2021.01.010	Identifies potential barriers for the implementation of Industry 4.0 in manufacturing organizations. *Manufacturing organizations.*	Technological, Social, Business, Environmental.	Include.
8	Radhakrishnan, J., & Chattopadhyay, M. (2020). Determinants and Barriers of Artificial Intelligence Adoption – A Literature Review. IFIP Advances in Information and Communication Technology, 617(January 2021), 89–99. https://doi.org/10.1007/978-3-030-64849-7_9	Identifying the factors that facilitate- and hinder the adoption of AI. *Autonomous Vehicles, Big Data Analysis, Cognitive Engagement, Robots, Medicine.*	Technological, Social, Business, Ethical, Environmental, Legislative.	Include.
9	Burgess, A. (2018). The Executive Guide to Artificial Intelligence. The Executive Guide to Artificial Intelligence. https://doi.org/10.1007/978-3-319-63820-1	Explaining how AI technology will impact business. *General application.*	Technological, Social, Business, Ethical, Environmental, Legislative.	Include.

Table 7: The results of the first search string

2.4.2 Selection criteria

The selection criteria, similar to the selection criteria for the enablers of AI implementation, entails the generalizability of the results presented in the literature. Several of the articles that discuss the implementation barrier(s) of AI are focused on a specific field or application. As a result, the barrier(s) are field specific and not generalizable to different industries (S. Kumar et al., 2021). For example, references 1 and 3 (Paranjape et al., 2021; Singh et al., 2020) discuss barriers for AI adoption in healthcare focusing on techniques that affect social acceptance. Even though social acceptance is of importance for AI implementation, the study results do not contribute to the synthesis of generalizable barriers to implementation in organizations at this point and will therefore be excluded.

2.4.3 Synthesized literature

In Table 8, a synthesized list of the barriers identified by the included literature references 1 to 9 are presented. The barriers are subdivided over the three types of barriers based on the TOE framework. By dividing the barriers according to the same framework as the enablers in the previous section, it is more comprehensible to gain an overview of the differences and similarities between the two sorts of factors.

Type of barrier	Description	Mentioned in ref.#
Technological	Lack of quality data.	2, 8, 9, 7
	Low volume of available data.	2, 7
	IT infrastructure issues.	2, 4, 5, 7
	Data governance issues.	2, 5, 7, 8
	Security and confidentiality issues.	2, 5, 7, 8
	Complexity of the algorithm and coding.	4, 8
Organization	Insufficient availability of talent.	2, 4, 7
	Lack of skills for industrialization.	2
	Lack of understanding of the technological aspects of AI.	2, 4, 7, 9
	Resistance to change / negative attitude.	2, 4, 6, 7, 8, 9
	Change management issues / lack of understanding.	2, 4, 6, 7, 8, 9
	Lack of accountability and responsibility.	4, 5, 6, 8, 9
	Lack of transparency.	4, 5, 8, 9
	Perceived risk / concerns.	8
	Over-inflation of expectations (hype)	9
	Lack of understanding of business potential of AI.	2
	Lack of top management support.	2
	Job disruptions	7
	Lack of strategic vision.	2, 5, 6
Environment	Uncertain Return on Investment.	2, 7
	Lack of (public) economic policies.	4, 7
	Ethical issues.	2, 4, 5, 8
	Encoded bias.	5, 9
	Political issues.	4
	Lack of alignment with stakeholders.	4
	Lack of established framework for implementation.	4
	Lack of transparency in decision making.	4, 9
	Concerns and anxiety	6, 8, 9
	Immaturity of legal environment.	2, 4, 5, 9
	Data sovereignty.	6, 8
	Privacy concerns related to the training data.	6, 8, 9

Table 8: A synthesized list of the prohibitors described in the selected literature

As can be concluded from Table 8, the degree to which barriers are identified by the literature differs per type of barrier. The most acknowledged barriers are *resistance to change*, *change management issues*, and *lack of accountability*; all types of barriers categorized in the 'organization' type of the TOE. Also, technological barriers related to the volume, quality, and handling of data appear to be well-acknowledged by the literature.

2.4.4 Prohibitors to chatbot implementation

Exploring the Scopus and Google Scholar databases, no literature has been found on the prohibitors of chatbots in organizations. Only one article by Gudala et al. (2022) appears relevant, yet it explores the barriers to AI powered voice bots for older adults. In short, the authors' findings state that the main identified barriers include technology familiarity, costs, security, and privacy concerns. These prohibitors are more focused towards the adoption of the technology rather than the implementation of it in organizations and will therefore not be taken into account throughout the remainder of this research.

2.4.5 Conclusion

Interestingly, many similarities can be found in the enablers and prohibitors presented in the synthesized tables. For example, enablers described in the synthesized literature are: change management, ethics in data collection, and top management support. Whilst barriers described in the synthesized literature are: change management issues, ethical issues, and lack of top management support. The similarities suggest that the named factors are relevant to the success of AI implementation. As a result, if the factor is not managed properly it is suggested to be a barrier, whereas if the factor is managed well it is suggested to enable or enhance the success of the implementation. Also, similar to the conclusion on the enablers, there is no hierarchy or interrelation amongst the barriers suggested throughout the literature.

3 Social complexity

3.1 Introduction

The effect AI has on individuals' personal lives and society as a whole is becoming increasingly visible (Miller, 2021). As such, governments have started to outsource decision making on law enforcement, immigration, and social welfare to machines, whilst AI-based recommendation systems on social media create a filter bubble that results in a state of intellectual isolation for individuals and the polarization of society (Borgesius, 2016; H. Liu et al., 2019). These two examples serve to highlight the conflicting interests of the different actors affected by the technological advancements. As stated by Benbya et al. (2021, p.2): "While AI technologies offer many positive benefits to organizations, their introduction often creates significant unintended (or intended) consequences for individuals and organizations. Since the impact of AI implementation varies greatly among stakeholders, decisions to decouple stakeholders from the process of designing, implementing, and using AI systems often lead to the ultimate failure of the system". Therefore, the following chapter focuses on an extended actor analysis that will help to understand the multi-actor environment and the contrasting perspectives that these actors might have.

3.2 Actor analysis

To conduct an actor analysis, the first step is to make an inventory of all actors. This is done by constructing a causal diagram and asking the two questions of: *which actors can influence the important factors?* And *who has an interest in, or is affected by, the problem situation or the possible solutions?* The causal diagram shows the relationship between the most important variables that can affect the objective. The objective is set to be: the implementation of an AI solution in an organization. To determine the variables that are taken into account, the literature on Critical Success Factors for AI implementation by Dora et al., Hamm & Klesel, and Miller (2021; 2021; 2021), discussed in the section on enablers of AI implementation, is used as a starting point. For simplicity, similar success factors are combined into an umbrella variable. For example, algorithm complexity is used as an umbrella term for algorithm accuracy, consistency, and transparency. The final causal diagram is shown in Figure 6.

Subsequently, it can be determined which actors have an influence on the suggested factors. As can be deduced from the causal diagram, a distinction can be made between an organization's internal variables and the external variables. The internal variables (such as: organization's strategy, support, characteristics, and change management) are all influenced largely by the internal employees within the organization. To distinguish between internal employees, the four-tier pyramid of management levels is used. The pyramid distinguishes between: executives, senior managers, middle managers, and workers. Additionally, an internal department that can have a significant effect on the implementation of AI is the IT department of an organization. Therefore, these five actors will be considered as the main internal actors.

Besides the internal actors, a large number of external actors are present. Competitors that influence the market pressure, regulatory or institutional bodies that shape the legal environment, organizations that sell AI services, AI implementation consultants, and (in)direct consumers are such examples. Indirect consumers will be referred to as society, since these individuals do not choose to consume the service yet are affected by the impact it has. All in all, the actors can be divided into three overarching categories: the industry, the government, and the society.

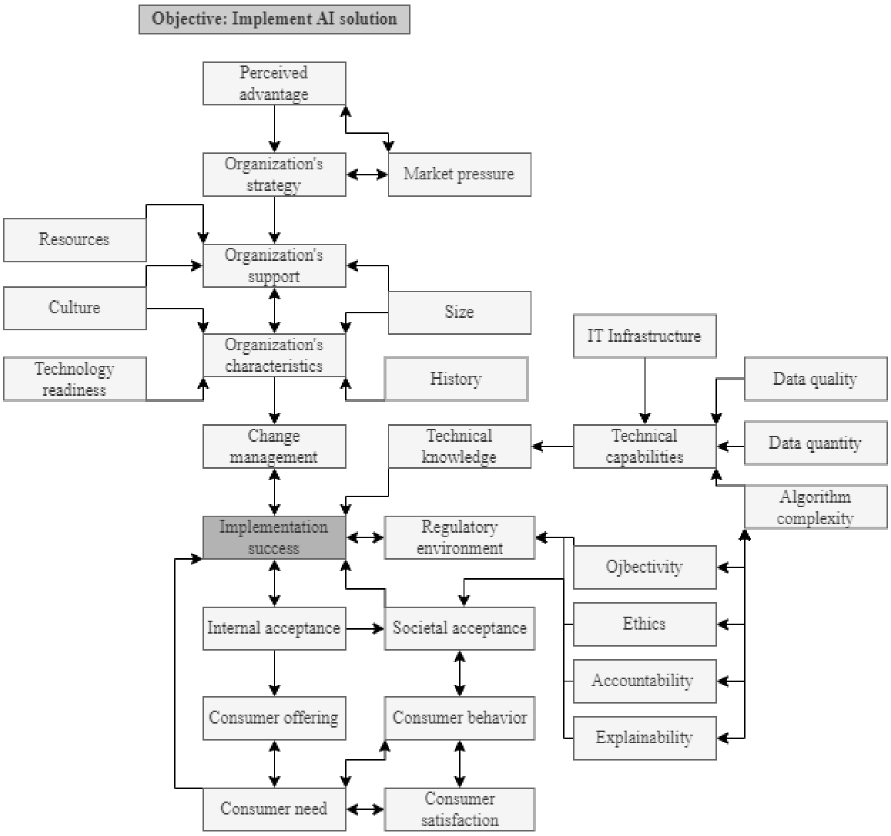

Figure 6: The causal diagram visualizing AI implementation

Having ascertained the involved actors, their distinct objective(s), interest(s), and criticality are determined. By determining these facets, it is possible to expose potential conflicts of interests or misalignment in needs and wishes. Throughout this research, an actor's objective is defined as "what actors want to achieve in a concrete situation", and interest (or fundamental objective) as "why do actors want to achieve these things?". Lastly, the criticality of an actor describes the ease with which the actor can be replaced. For the criticality score, 1 indicates it is easy to replace the actor whereas 5 indicates it is impossible. The objective remains to be the implementation of an AI-based solution. An overview of the established actors including their perspective is presented in Table 9.

Actor	Description	Objective	Interest	Criticality (1-5)
Executives (CEO)	The highest-ranking position in the	Strengthen the organization's position.	To increase profitability, achieve	4

	organization, often concerned with the organization's strategy, mission, and success.		set goals, or accomplish a mission.	
Senior manager	Focuses on certain aspects or departments of the company and takes responsibility for its performance.	Execute the organization's strategy and manage the departments accordingly.	To reach the company's achievements and fulfil his/her personal role.	3
Middle manager	Is the leading entity of a project team or department.	Execute the organization's strategy and manage their own department accordingly.	To reach the company's achievements and fulfil his/her personal role.	3
Workers	Are part of a team or department and do not have a managing role.	Carrying out the objectives set up by the higher management.	To satisfy the higher management. Potentially increase their work performance.	2
IT department	Provide the necessary IT means and knowledge for the organization to do business.	Providing support for the organization to conduct business (efficiently).	Working with new technologies and contributing to the organization's success.	4
Competitors	(In)direct competitors that operate in the same market.	Ensuring their own position or not falling behind the other competition.	Learning from best-practices.	4
Government	Regulating entity that sets boundaries regarding applicable rules and regulations.	Shape a fair playing field in which organizations can innovate and compete whilst protecting consumers and society.	A healthy playing field attracts organizations and stimulates a country's national wealth.	5
AI vendors	Provide AI-based solutions as a product or service.	Sell their services to organizations.	Their business model is based upon selling AI-based services.	2
AI implementation consultants	Provide advice on how to implement AI-based solutions.	Sell their services to organizations.	Their business model is based upon selling their knowledge on implementation processes.	3
Consumers	People that products or services provided by the organization.	Higher quality products or services.	Increase in personal life due to higher quality or lower costs.	4
Society	The people that do not buy the products or services directly yet can be affected by the technology.	There is no objective for this group since the society is affected, rather than interested.	Not being negatively influenced by the adoption of the technology.	5

Table 9: An overview of the identified actors

The actors, their objectives, their interests, and their criticality stated in Table 9 will be used in subsequent sections to determine the actors' interrelations and distribution of power. These factors will thereafter influence the stakeholder management approach discussed.

3.3 Formal relations and interdependencies

3.3.1 Legislation

To determine the boundaries of the playing field the actors are in, the formal rules and regulations that apply to the use of AI in the Netherlands are explored at first. In the Netherlands, Agentschap Telecom is the regulator and executive agent regarding the rules and regulations for IT. The agent states, there is no legislation for AI-applications in general at the moment (Agentschap Telecom, 2021). Nonetheless, "Clear, general, and robust legislation for regulation is needed to build and maintain trust. Trust is needed for the user acceptance and to capitalize on the great number of possible benefits of AI" (Agentschap Telecom, 2021, *translated from Dutch*). The agent also refers to the proposal made by the European Commission (EC) in April 2021. The proposal is based upon the whitepaper published by the EC and discusses the plan of action, guidelines, and suggested legislation for the use of AI in member states. The suggested legislation is discussed in the following section.

Ethical guidelines

The EC has announced a European strategy to AI in 2019 stating three ethical guidelines that the technology must comply with to be trustworthy. The three ethical guidelines are (Europa Decentraal, 2021):

1. The AI must be legal, thereby complying with all of the applicable rules and regulations;
2. The AI must be ethical, thereby complying with ethical and moral values;
3. The AI must be robust from a technical and a social viewpoint, since the system can (unintentionally) do harm, even though the initial intention was for good use.

The guidelines also include an assessment list of seven requirements that must be fulfilled. The seven requirements are: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, accountability (European Commission, 2019).

Typology of AI-applications

Also, in 2021, the EC has suggested another set of guidelines based upon a new AI-application typology since it deemed the previous three ethical guidelines insufficient for mitigating all of the potential risks. The typology distinguishes AI-applications on the base of potential risk; the higher the risk, the stricter the legislation. The degree of risk is determined based on the sector and the intended use, in particular from the viewpoint of protection of safety, consumer rights, and fundamental rights (European Commission, 2020). The EC distinguishes between four types of risk:

1. Unacceptable risk: AI-applications that pose an unacceptable risk to society and violate the country's constitution. An example is the Social Credit System used in Asia that scores individuals based on their behavior and acts accordingly.

2. High risk: AI-applications that pose a high risk and can potentially have a significant effect on stakeholders. Examples of such applications are the use of AI to grant visas for immigration or to assess one's financial situation. Ethics and societal effects are especially important for high-risk applications.
3. Medium risk: AI-applications that pose a medium risk to society can be applications such as chatbots or deepfake videos. For the medium risk applications, transparency, traceability, and explainability are important to maintain.
4. Low risk: AI-applications that pose a low risk to society entail applications such as spam filters or recommender systems. The effect that these systems have on stakeholders is limited yet possible (unintended) negative consequences must not be forgotten.

Regulatory framework

To stimulate the use and development of AI, the commission proposed a regulatory framework solely for the high-risk AI-applications. “The new regulatory framework for AI should be effective to achieve its objectives while not being excessively prescriptive so that it could create a disproportionate burden, especially for SMEs” (European Commission, 2020, p.17). As a result, the requirements that make up the regulatory framework are (European Commission, 2020):

1. Training data: requirements aimed at ensuring protection of personal data, complying with the General Data Protection Regulation (GDPR), and ensuring the training data is sufficiently broad to prevent discrimination on all relevant dimensions.
2. Data and record-keeping: documentation on the programming, training methodologies, and data must be kept allowing problematic actions to be traced back and verified.
3. Information to be provided: to ensure transparency to all stakeholders, clear information must be provided on the AI’s capabilities, limitations, and use cases.
4. Robustness and accuracy: outcomes should be reproduceable, robust, and accurate during all life cycle phases. Also careful consideration of the implications and effects of the system is required prior to the design.
5. Human oversight: involvement of human beings must be maintained to keep an oversight of the system performance and not undermine human autonomy. This could include the approval of human assessment before the output of the system becomes effective.
6. Specific requirements for certain particular AI applications (e.g., biometric identification): special attention must be paid to applications that gather personal data that might pose risks to fundamental rights.

3.3.2 Mapping relations

By mapping the formal relations, the interactions among stakeholders can be visualized. For the current context, a distinction is made between the organization, the market, and “outside actors”

(Figure 7). The organization entails all the internal employees as mentioned previously. The market entails the companies and consumers that operate in the same sector and are assumed to be affected by the same business challenges, regulatory context, and potential solutions. Lastly, the “outside actors” are the stakeholders that are positioned outside of the previous categories. These stakeholders do have an effect, and can be effected by, the organization that wishes to implement an AI-based solution.

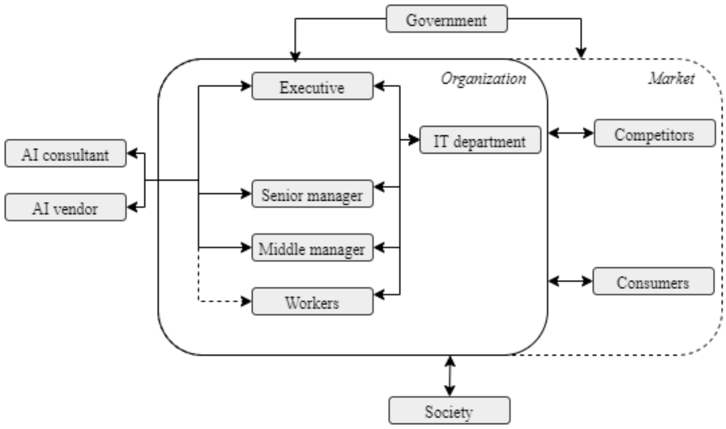

Figure 7: The formal relations between the actors

3.3.3 Actor positions

The final aspect that must be considered before the implications and stakeholder management can be discussed is the actor positions. To do so, a Power/Interest (PI) grid is used. The power refers to the impact an actor or organization can have on the future of the objective. The interest refers to the degree of incentive an actor has for reaching the objective. PI grids typically help to determine: which actors’ interests and power must be included, which coalitions of actors must be encouraged or discouraged, and the reasoning for an actor’s perspective (Bryson, 2004). Besides the power and interest, an actor’s attitude (positive or negative interest) towards the objective is not shown within the PI grid. It has been decided not to include this third aspect in Figure 8 since it is highly situation-dependent and therefore difficult to make a general statement on.

The PI quadrant is divided into four sub-categories; *players*, *subjects*, *crowd*, and *context setters*. *Players* describe the high-power and high-interest actors that must be managed closely. *Subjects* refer to the high-power and low-interest actors that might not have significant interest in the current objective yet can have a great impact and should therefore be kept satisfied. *Crowd* refers to the low-power and low-interest group of stakeholders. This group does not play a significant role in reaching the objective. Finally, *Context setters* are the low-power, high-interest actors that are affected by the objective, yet do not impact the future of it greatly. This group should be considered during the stakeholder management since their position can influence the acceptance of the objective. By using Figure 7: *The formal relations between the actors* in combination with the criticality and objective defined in Table 9: *An overview of the identified actors*, the following PI grid shown in Figure 8 is constructed.

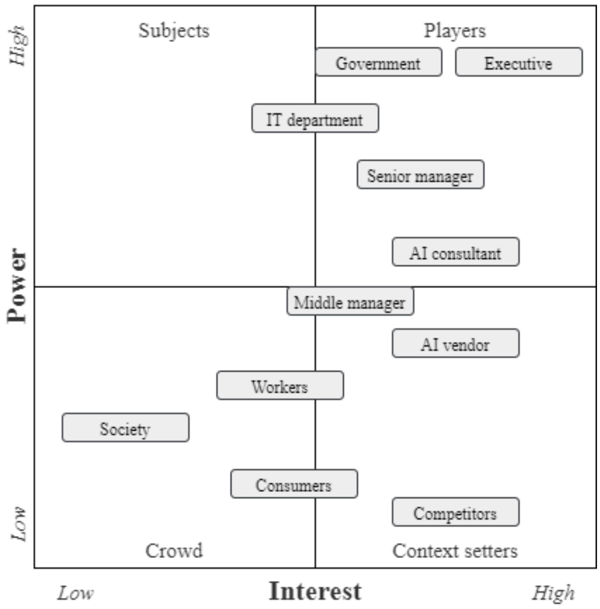

Figure 8: The PI grid showcasing each actor and their relative position

3.4 Implications and stakeholder management

By delineating the complexity of the actor situation illustrated by *Figure 7: The formal relations between the actors,* Figure 8: *The PI grid showcasing each actor and their relative position,* Table 9: *An overview of the identified actors*, and the current legislation, the implications and proposed stakeholder management approach are explored in the following section. This is done according to three trade-offs that are deduced from the previously presented materials. The three trade-offs that have been identified are: *the risks vs. the benefits, innovation vs. regulation, and the one vs. the many*.

3.4.1 Trade-offs

The risks vs. the benefits

The first trade-off addresses the far-reaching potential that AI offers and is clearly addressed in the opening-paragraph of the whitepaper by the European Commission (2020, p.1):

> *Artificial Intelligence is developing fast. It will change our lives by improving healthcare (e.g., making diagnosis more precise, enabling better prevention of diseases), increasing the efficiency of farming, contributing to climate change mitigation and adaptation, improving the efficiency of production systems through predictive maintenance, increasing the security of Europeans, and in many other ways that we can only begin to imagine. At the same time, Artificial Intelligence (AI) entails a number of potential risks, such as opaque decision-making, gender-based or other kinds of discrimination, intrusion in our private lives or being used for criminal purposes.*

The paragraph highlights the societal benefits that the technology can offer by helping to solve complex problems such as climate change, world hunger, and healthcare. Whilst simultaneously posing potential risks to minority groups, individuals, and illegal activities. The opportunities and risks for AI in society is extensively elaborated on throughout the paper by Floridi et al. (2018). The paper frames the discussion based upon four opportunities that AI offers centered on the fundamental points in human development: “*who we can become* (autonomous self-realization); *what we can do* (human agency); *what we can achieve* (individual and societal capabilities); and *how we can interact with each other and the world* (societal cohesion)” (Floridi et al., 2018, p.690). In each of the four opportunities, AI can be used, misused, and underused. An overview of the discussion is presented in Figure 9.

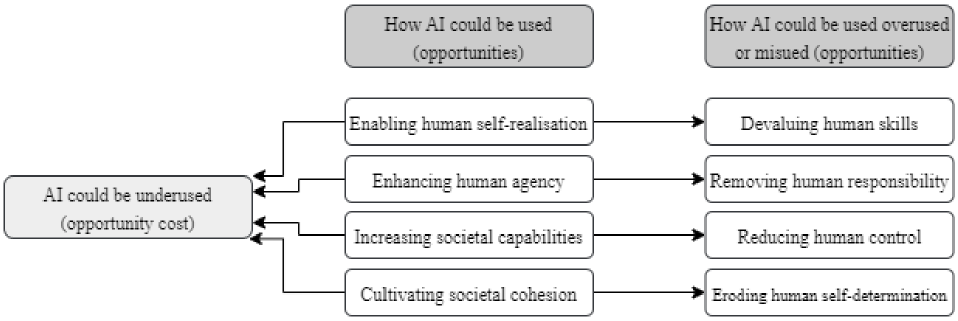

Figure 9: The four opportunities AI offers (from Floridi et al., 2018, p.691)

The authors state that unintended overuse or misuse caused by greed, harmful intent or misaligned incentive must be avoided. Whereas fear, misplaced concerns, or overreaction can lead to an underuse of the technology. The fine line between use, misuse, and underuse leads to the second trade-off; the societal vs. the business value.

Innovation vs. regulation

The third trade-off depicts a paradox that is familiar in different sectors and applications outside the field of AI as well yet is especially relevant to technological advances; the regulation paradox. “Technology symbolizes markets, enterprise, and growth, while regulation represents government, bureaucracy, and limits to growth” (Wiener, 2004, p.483). Often, regulation is therefore seen as a barrier to innovation since it burdens organizations with legal compliances and legislative thresholds (Aghion et al., 2021). As advancements in AI-applications require technological innovation, it can be suggested that regulating this process will slow down the innovation process. Nonetheless, the extensive societal impact that AI has might cause this situation to be different.

Apart from technological innovation, the diffusion and acceptance of technology contribute largely to the social adoption. Both aspects are assessed and described by the Diffusion of Innovation (DOI) and Technology Acceptance Model (TAM) respectively. The DOI theory indicates that the diffusion of a technology is dependent on innovation characteristics such as complexity, relative advantage, observability, compatibility, and trialability (H. Chen et al., 2021). As well as multiple sociological, organizational, psychological, and economic variables (Butler & Sellbom, 2002). The TAM estimates the intended use and behavior based upon the perceived usefulness and perceived ease of use (Davis & Venkatesh, 1996). Therefore, considering the current technological capabilities of AI, the possible causes of underuse suggested by Floridi et al. (2018), and the aspects considered by the DOI and TAM, it can be argued that social AI adoption might profit from a form of government regulation. The reason for this is because government regulation can stimulate social trust and (partly) mitigate risks of misuse (Miltgen & Smith, 2015). Fortunately, it appears as if the EC is knowledgeable of this trade-off which is highlighted by the statement on the regulatory framework for high-risk applications; "… while not being excessively prescriptive so that it could create a disproportionate burden, especially for SMEs*"* (European Commission, 2020, p.17).

Besides, regulations protect organizations also. By establishing regulations, boundaries are set to which organizations must comply in order to be not held accountable or liable if negative externalities occur. If the organizations operate within the set boundaries, they are free to innovate, implement, and make use of state-of-the-art technologies without being at risk. As a result, regulations can be argued to stimulate innovation as well.

The one vs. the many

The fourth and final trade-off is described as *the one vs. the many*. As can be seen in Figure 8: *The PI grid showcasing each actor and their relative position* and Table 9*: An overview of the identified actors*, the largest groups of actors in absolute numbers (workers, society, consumers), are not the decision makers that shape the implementation process of AI in organizations. Nonetheless, these "powerless" groups are the ones being affected by the implementation the most and are therefore key to the adoption, diffusion, and final success of the technology. Therefore, the inclusion of their perspective, their needs and wishes, and protection by design must be actively managed by the decision makers; all contributing to a multistakeholder approach. As concluded by Floridi et al. in their paper on Ethical AI for a Good Society (2018, p.701): "We also believe that creating a Good AI Society requires a multistakeholder approach, which is the most effective way to ensure that AI will serve the needs of society, by enabling developers, users and rule-makers to be on board and collaborating from the outset".

All three trade-off depict a delicate aspect related to the development of AI. Whereas one describes the potential risks and benefits, the other showcases to the asymmetric distribution of power for the design and development of these technologies. Overall, all three trade-off constitute to an overarching tension between *the social value vs. the business value of AI*. Today, most drivers for the development, adoption, and implementation remain to be mainly economic (Cubric, 2020). Also, very few businesses consider the human, organizational, and wider societal factors for the adoption of the technology even though AI is known to enlarge social divides and inequalities (Cubric, 2020; Hagerty & Rubinov, 2019). This is shown in diverging interests amongst stakeholders (Table 9: *An overview of the identified actors)* and creates the tension between the societal and the business value of AI. As a result, careful stakeholder management is required to design sustainable AI systems.

3.4.2 Stakeholder management

The delineation of the social institutions, human actors, and the technology portray the hybrid nature of the system at hand; a so-called socio-technical system (Kroes et al., 2006). The troublesome nature of such a system is the fact that the three components are integrated and cannot be treated separately. Also, “many actors within the socio-technical system are continuously changing (re-designing) the system” (Kroes et al., 2006 p.28). This causes the system to be dynamic and ever-changing. Due to this, it is important to engage the actors using a multistakeholder approach during the development, implementation, and maintenance of the system.

Floridi et al. (2018) describe four action points that constitute to this approach. The four action points are: to assess, to develop, to incentivize, and to support. The assessment captures the capacity of existing institutions and the presence of current liability foundations, specifically focusing on sustainability over time and stimulating the inclusion of dialogue between stakeholders. Also, it must be assessed to what extent AI can penetrate our lives based upon societal values and public opinion. Following, the development point addresses the need for a legal, controlling, and social framework. Sufficing to assess accountability and compliance, identify unwanted consequences, and enhance the explicability, respectively. Thirdly, (financial) incentive actions to actively stimulate the development for the social applications of AI, cross-disciplinary and cross-sectoral cooperation, and the inclusion of social and ethical considerations in projects must be maintained. Finally, the support action point describes the supportive education that can enhance executive board’s decision making, the development of self-regulatory codes of conduct, and more channels of educational curricula.

All in all, the aforementioned action points describe separate required developments for the future of AI yet are alike in the way that each actively includes the engagement of the involved actors (government, industry, and civil society).

3.5 Conclusion

The socio-technical environment characterized by continuous changes in the legal, social, and technical facet of the process gives rise to a delicate situation in which each decision must be carefully considered. The goal of this chapter was to explore the affected actors and lay the foundation for chapter 4, on the definition of “success” from different stakeholder perspectives. The key findings include a condensed overview of the current legislation, the relations amongst stakeholders, four identified trade-offs, and a suggestion on stakeholder management. The combination of these four findings contribute to the conclusion that a multi-actor governance is required to capitalize maximally on the potential of AI solutions. Therefore, the need for a thorough understanding of each stakeholder and their interests is critical for managing the diverging needs and wishes.

4 Different perspectives on success

4.1 Introduction

To be capable of doing research on the success of AI implementation in organizations, a definition of "success" must be established. As highlighted throughout the section on social complexity, there are many contrasting perspectives and (opposing) objectives related to the implementation of AI solutions. As a result, succeeding in achieving the objective of one actor might lead to the disappointment of another, affected, actor. "Thus, success and failure are difficult to define and measure since they mean different things to different people"(Graeme & Walter, 2008, p.733). Because of the difficulty in defining success and failure, projects are initiated regularly without having decided upon clear goals (Albert et al., 2017). Despite the fact that this decreases the likeliness for the project to succeed (Remenyi & Sherwood-Smith, 1991).

As a result, the sub-question for this research is formulated as: *When is an AI implementation deemed successful?* By providing different actors' perspectives on the definition of a successful implementation, the differences, the commonalities, and the potentially opposing views that these key figures might have can be exposed. Also, being able to either formulate a general definition of success, or expose the multifaceted aspects of the definition, will help to: assess the success of implementation projects, guarantee the inclusion of multiple stakeholders' perspectives during the implementation, and decrease the resistance to change by satisfying the wishes of affected stakeholders. Also, a study by Graeme & Walter (2008, p.733) suggests that "when success criteria are formally defined and then measured, IT project outcomes are improved, and project resources are better utilized". Hence by being able to formalize success requirements for AI implementation projects, the overall chance of succeeding can be enhanced.

To elaborate on the definition of success, two actor perspectives are investigated; the industry's perspective and the government's perspective. These two actor groups are explored because both are characterized as being the most powerful stakeholders that have a high interest in the implementation of the technology and are hard to replace (see Figure 8: *The PI grid showcasing each actor and their relative position* and Table 9: *An overview of the identified actors*). This is considered to be interesting since the variety in their fundamental criteria for success shines light on multiple facets, thereby enabling the research to cover a broad range of aspects.

4.2 Data collection methods

To gather relevant data, a combination of literature search and interviews is used as the data collection techniques. On the one hand, literature search is used to acquire insights on the perspective of the government. The government entails the departments and institutions that shape the legal landscape in the Netherlands. Because new Acts can be proposed by many actor groups, including political parties, citizens, and the media, it is considered to be unfeasible to expose their individual views. Therefore, an analysis of the legal environment and (grey) literature search is used to deduce the government's view.

On the other hand, interviews are used to gain insights on the perspective of industry practitioners because the industry predominantly shapes the use of the technology. Interviews are preferred over a survey or focus group because of multiple reasons. Firstly, no existing literature has been found on the topic from an industry perspective which makes this section exploratory. Because of the lack of literature, predefining survey statements would rapidly impose researcher's bias onto the statements. Secondly, whereas a focus group is meant to stimulate discussion amongst participants, interviews can be used to identify a participant's viewpoint on a particular subject thereby exposing the participant's subjectivity (Qu & Dumay, 2011). Lastly, by using individual semi-structured interviews, the rationale of each participant can be explored thoroughly without interruptions or influences by other participants. This allows the researcher to reveal an interviewee's criteria for success and the rationale for these criteria. Exposing the rationale is considered to be of added value since it can help to better understand the norms and values of the stakeholder group.

4.2.1 Interview protocol

The interview protocol that is used is presented in Appendix C: Interview protocol 1. The protocol is set-up according to the six stages described in Appendix B: Qualitative research interviews and follows a semi-structured approach. By conducting a semi-structured interview, comparison amongst interviewee's answers is possible whilst maintaining the possibility to adjust the structure and flow of the interview according to the interviewee's professional role and responses. The protocol consists of the following questions:

1. What is your role during AI related implementation projects?
    a. Could you give an example of the activities you do during a project?
    b. What are your responsibilities during these projects?
2. When would you consider an AI implementation project to be successful?
3. What would it take for a project to be defined as a failure?
4. How does your organization currently determine whether an implementation project has been successful?
    a. Is this a standard set of criteria or do these vary by project?
    b. At which stage of the project are the criteria normally defined?
    c. Have the initial criteria changed during any of the projects?
5. Has your organization experienced a failing project in the past?
6. Have you experienced any reoccurring problems during implementation projects?

Also, the purpose and the goal of each interview question is briefly explained in Table 10.

Question	The purpose	The goal
1	The question serves as an introducing question to kick start the conversation.	By answering the question, the relation between the technology and the interviewee is exposed. This will help to group the interviewee to a specific stakeholder group.
1.a	The probing question stimulates the interviewee to be explicit in his/her answer and draws out more complete narratives.	By knowing explicit tasks, the stakeholder grouping can be assured.
1.b	The follow-up question helps to map tasks to responsibilities.	By being aware of the responsibilities of the interviewee, the rationale for the success facets can be better understood.
2	The direct question draws out a direct response to the personal definition of success.	The direct response is the most important response to define "success" for different stakeholder perspectives.
3	The direct question draws out a direct response to the definition of project failure.	Knowing when a stakeholder considers a project to be a failure can be as valuable as knowing when it is a success.
4	The direct question draws out a direct response to the organization's success indicators.	The difference between personal and organizational success indicators is exposed. Also, empirical examples are gathered that determine success (such as proposed KPIs).
4.a	The follow-up question helps to elicit a more complete description of the success indicators	The degree to which organizations consider the generalizability of success indicators is revealed.
4.b	The follow-up question helps to elicit a more complete description of the success indicators	Helps to expose whether or not the success indicators are set at the start of a project. (Challenging the literature by Albert et al. (2017))
4.c	The follow-up question helps to elicit a more complete description of the success indicators	Helps to expose the possible presence of changes in objectives during a project and exposes a dynamic or static approach.
5	The direct question draws out a direct response to failing implementation projects in the past.	Prohibitors to AI implementation can be exposed.
6	The direct question draws out a direct response to the cause of problems during projects.	Difficulties of AI implementation can be exposed. Also critical facets that have been forgotten at the start of the project might be exposed.

Table 10: The purpose and goal of each interview question

4.3 The industry's perspective

4.3.1 Selection of interviewees

To gain an understanding of the perspective of industry practitioners, actor groups positioned as *players* (high-power, high-interest) and *context setters* (low-power, high-interest) (Figure 8: *The PI grid showcasing each actor and their relative position*) are considered to be key due to their high criticality (Table 9: *An overview of the identified actors*). Therefore, theoretical sampling is used to select the actor profiles of the interviewees. As shown in Table 11, the interviewee, the actor profile, and the supplier/client role is indicated. The actor profile labels the job description of the interviewee and is highlighted since it can support the rationale of the statements made. The supplier/client column indicates whether the interviewee is on the supplying or receiving end of the technology. Again, this is noted such that it can serve as a rationale for the relationship the interviewee has to the implementation.

Interviewee	Actor profile	Supplier/ Client
1	AI vendor	Supplier
2	AI consultant	Supplier
3	Developer	Supplier
4	CEO	Client
5	Project manager	Supplier

6	Developer	Supplier
7	Developer	Supplier
8	IT Department	Client

Table 11: An overview of the anonymized interviewees

In total, eight interviews have been conducted that took 20 minutes on average (excluding the introduction and debriefing). The interviewees range from technical actor profiles (developer and IT department) to client focused actor profiles (AI consultant and project manager) and differ in the sorts of AI application they work with, the sector they work in, and the organizations they work for. The high variability amongst interviewees is maintained to gather a general overview of the industry perspective on the facets that contribute to the success of AI.

4.3.2 Data analysis

To structure and systematically analyze the qualitative data gathered during the interviews, qualitative coding is used. Qualitative data can be grouped objectively into *codes* and *themes*, using a coding scheme. *"*Codes identify a feature of the data (semantic content or latent) that appears interesting to the analyst*"* and is the most elemental part of the raw data (Braun & Clarke, 2006, p.18). More broadly, analyzing reporting patterns or structures in the data can be marked as themes. By doing so, the qualitative data can be grouped per case into the same group whereafter it is more comprehensible to gain a structured, general overview. The data analysis consists of two iterations. For the first iteration, three codes and one theme are used as the coding scheme. For the second iteration, the theme from the first iteration is split into two codes to gain a deeper understanding of both actors' perspectives. The codes and themes that are used will be elaborated on in the introductions of both iterations.

The first iteration

The codes that are used during the first iteration of analysis are based upon the research by Graeme and Walter (2008) named *Success in IT Projects: A Matter of Definition?* The authors define three groups of success facets based upon interviews conducted with 72 senior managers in 36 companies. The three groups are: *project management success* (e.g., time, budget), *technical success* (e.g., capabilities, requirements), and *business success* (e.g., return on investment, delivery of benefits). In this iteration, a theme is added named *social success.* The theme is assigned to success facets related to social effects, ethical concerns, or human factors such as acceptance and satisfaction. The theme is added because the literature reviews on enablers and prohibitors in sections 2.3 and 2.4 show that the social factor is also suggested to contribute to the success of AI implementations.

Success\interviewee	1	2	3	4	5	6	7	8	∑
Project management	0	1	0	0	0	1	0	0	2
Technical	1	3	2	1	2	4	2	2	17
Business	5	1	0	4	1	1	1	2	14
Social	3	2	1	0	3	4	1	2	16

Table 12: The number of times each category of success facets is addressed by the interviewee

The results of the first round of analysis are displayed in Table 12. The table highlights the interviewee and the number of times the interviewee has addressed the relevant success facet as response to the second interview question. Examples of technical success facets stated during the interviews are: the speed of response, the degree of automation, the matching score, and the ability to execute desired functionalities by the new system. Stated business success facets include: increased effectiveness due to reduced costs or improved service, total use of the new system, and changes in Key Performance Indicators (KPIs). Finally, social success facets related to customer satisfaction, customer experience, and the adoption of the system have been mentioned repeatedly.

Regarding the results of the first iteration, multiple actor profiles appear to confirm expectations based on role and responsibilities. For example, developers (interviewees 3, 6, 7) state many technical and social success facets related to the system's capabilities and adoption yet rarely address the business facets. On the contrary, the AI vendor and CEO (interviewees 1,4) mention the (proposed) added business value more often as the relevant measure of success. Aggregating the results of all interviewees (shown in the last column, Table 12), suggests a fair distribution of importance for all of the facets except for project management. The limited importance of project management might be explained by one of the results presented by Graeme and Walter (2008), which shows it is possible to have project management success without business success and vice versa. Because the added business value of AI is often the driver for organizations to start the implementation, it might be suggested that it is therefore the most important success facet for industry players regardless of whether project deadlines or budget are met (Neuhüttler et al., 2020).

Another explanation for the unaddressed facets related to project management success can be the project methodology mentioned by interviewees 1, 3, 6, 7, and 8. All of these industry players mentioned the agile way of working used rather than the classic linear waterfall approach. By using an agile methodology, projects are divided into smaller, iterative periods that can be tested and revised quickly. Doing so allows the team to add technical features, adjust KPIs, and set new deadlines continuously. As an example, Interviewee 6 stated:

"If sometimes more money has to be made available because of delays, then it does not automatically mean that the project is a failure, it just means that there should be a shift in the timeframe a bit and more money has to be made available. And if the client is okay with that, then we add extra sprints, for example, an agile scrum, and that is done".

The citation expresses the dynamic goals related to project management success facets, which might explain the limited addressed presence of such facets during the interviews.

During the first iteration of analysis, the same codes and theme are used to identify the reasons for industry practitioners to define a project as a failure (based on question 3 of the interview protocol). This is considered to be insightful since it can expose the fundamental facets of an implementation from the industry perspective. The results are shown in Table 13.

Failure\interviewee	1	2	3	4	5	6	7	8	∑
Project management	0	0	0	0	0	1	0	0	1
Technical	1	1	3	2	1	2	1	0	11
Business	1	0	0	1	0	0	0	0	2
Social	3	1	0	1	1	0	0	2	8

Table 13: The number of times each category of cause of failure is addressed by the interviewee

The contrast in results between facets to define success (Table 12) and to define failure (Table 13) is most significant for the business facets. This is interesting and might be explained by a cause-and-effect relationship. Interviewees have most frequently defined the failure of an implementation due to unsupported technical requirements or capabilities. According to the responses, the unmet requirements can be caused by legal restrictions, low quality data, or algorithm complexity for example. Thereafter, user acceptance, change management, and customer satisfaction have been mentioned as social facets contributing to the failure of an implementation frequently. Interestingly, business failure by “lack of change after implementation” and “no measurable results” have only been mentioned three times throughout the interviews. Besides, considering the named facets mentioned by the interviewees, the enablers of AI implementation (section 2.3.3), and the prohibitors to AI implementation (section 2.4.3), many similarities can be seen.

Exploring the results of the second and third interview question suggests a hierarchy amongst the facets, of which the technical facet seems to be fundamental. Thereafter, the social facets are required to reach the proposed added business value and successfully complete the implementation. The proposed hierarchy and relation amongst success facets is presented in Figure 10 in the form of a strategy map. The implementation starts by assessing the fit of AI and the organization after which the agile project methodology is started. The dark grey ovals show the possible negative outcomes, based upon the causes of failure mentioned by the interviewees (interview question 3). The white ovals show the possible positive outcomes, based upon the success facets mentioned by the interviewees (interview question 2). In the end, the implementation is considered successful if the business can profit from the new system. This is presented as an “increase in business efficiency”. The proposed strategy map visually shows the proposed relations among the types of facets contributing to the success of AI implementation.

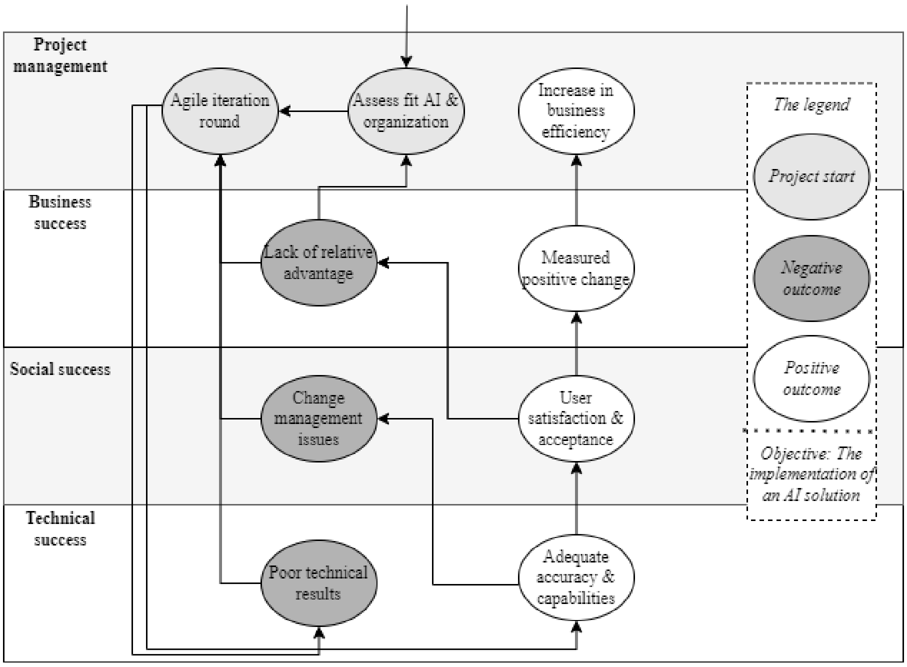

Figure 10: A visual strategy map of the chronological cause and effect among the success facets

The second iteration

For the second iteration, the "social" theme is divided into the codes*: internal social effect* and *external social effect*. The division is made to expose the scope of interest among industry players on the social effect an implementation can have. Internal social effect regards to topics related to change management, customer satisfaction, and user acceptance. External social effect regards more extensive effects such as ethics, bias, and effect on society. The results are presented in Table 14.

The aggregated results presented in the last column clearly show the industry players' focus is foremost on the social effects directly related to their implementation project. Again, this might be explained by a cause-and-effect relationship since internal social effects have a direct effect on the implementation and thereby potentially cause short term changes to the business success, whereas external social effects are less noticeable and might not cause negative externalities on the short term. Another explanation is related to the agile methodology deployed by the industry. If "simple" solutions are created at first, after which more advanced features are added iteratively without a final objective or end goal, it is cumbersome to assess the indirect long term social effects at the start of a project

Social effect\ Interviewee	1	2	3	4	5	6	7	8	∑
Internal	6	3	1	1	2	4	1	4	22
External	0	0	0	0	2	0	0	0	2

Table 14: The number of times internal and external effects are mentioned by the interviewees

Final remark

The final question stated in the interview protocol addresses reoccurring problems encountered by the interviewees. Surprisingly, human intervention or manual override (MO) of the automated system was mentioned multiple times as a forgotten capability at the start of a project. The interviewees described the high performance of AI automated systems in general, yet the underperformance of the system for edge cases. An example was given by interviewee 7:

"For example, during Christmas. During that time of the year we have a different assortment of products and consumer demands are different as well. Then, we want to make sure there is a functionality inside the solution that allows experienced employees to override the system's demand forecast".

4.4 The government's perspective

The government entails the departments and institutions that shape the legal landscape. Because, in the Netherlands, new Acts can be placed on the political agenda by many actor groups such as political parties, citizens, or even the media, it is considered to be non-viable to conduct interviews with each of these actor groups to expose their view on successful AI implementation. Therefore, it is chosen to derive the government's perspective on successful AI implementation based upon active legislation and grey literature regarding AI. As discussed in section 3.4.1, on the three trade-offs present in the socio-technical environment, the government benefits from a landscape in which the advantages of AI can be exploited yet society is protected at the same time. These two goals are used as the fundamental idea throughout the analysis of the government's perspective.

The current legislation

Following the legislation presented in section 3.3.1, the most important pillars for a successful AI implementation from a government perspective are deduced. In short, Agentschap Telecom (2021) strives for clear, general, and robust legislation to build and maintain trust, enhance user acceptance, and capitalize on the benefits of AI. The EC proposes three ethical guidelines, supported by seven requirements for AI. For high-risk applications, there is a regulatory framework proposed consisting of six additional requirements. A condensed visual overview is presented in Figure 11.

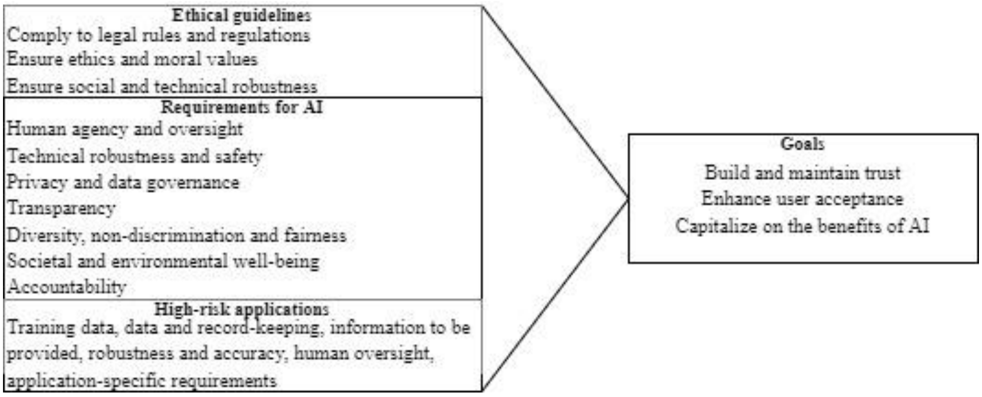

Figure 11: A visual overview of the current legislation (from section 3.3.1)

Deduced from the legislation presented in Figure 11, the government clearly balances between stimulating the development and innovation of AI whilst protecting society for (unintended) misuse. As one might expect, there are no statements made regarding project management success, whilst internal social success (user acceptance) and business success (capitalize on the benefits of AI) are addressed in some form. Meanwhile, emphasis is put on the external social success and technical capabilities. Specifically, whereas the industry appears to focus on the technical capabilities regarding system functionalities, the government seems to be more concerned about technical facets such as transparency, non-discrimination, and need for human oversight.

Grey literature

To either support or challenge the conclusions drawn from the current legislation, grey literature is explored. On the 8th of October 2019, the Dutch government published a Strategic Action Plan on AI stating the goals, future actions, and required resources. The plan consists of three main components each supported by sub-components (Rijksoverheid, 2019, p.6, *translated from Dutch*):

1. Exploiting the economic and social benefits;
    a. AI offers solutions to social challenges.
    b. AI is used for public services.
    c. Entrepreneurship involving AI is stimulated.
2. Creating the right terms and conditions;
    a. AI research in the Netherlands is of high quality and must set the bar for the rest of Europe.
    b. The Netherlands offers excellent education on AI.
    c. The Netherlands possesses quality data as input for AI solutions.
    d. The Netherlands is at the frontline of the use of IT services and AI solutions.
3. Strengthening the fundamentals;
    a. Public values and human rights must be protected.
    b. AI must be trustworthy and fair for all users.
    c. Markets must be open, competitive, and must protect consumers.
    d. The safety of citizens, organizations, and institutions must be maintained.

To reach the proposed goals and safeguard the three components, the government puts emphasis on the need for cooperation between public and private bodies, as well as on an international level. Also, an inclusive, human-centered approach must be taken to ensure a trustworthy and explainable AI that empowers society (Rijksoverheid, 2019).

All in all, the grey literature addresses similar facets as the ones discussed in the sections on social complexity and the current legislation. The final paragraph published by Rijksoverheid highlights the need for a multi-actor governance approach in which best practices are shared amongst practitioners. Also, to stimulate the use and development of AI whilst ensuring safety, the government hints towards a human-centered approach. Such approach includes the human-perspective in every design step to ensure it is actively present in the final design thereby safeguarding public values and human rights. Furthermore, the grey literature includes the exploitation of benefits as the first component in the Action Plan. This indicates the awareness of the Dutch government on the significant potential that AI can offer, which has been referred to as the facet *business success* before.

4.5 The differences in perspectives

To compare both perspectives on the facets that contribute to a successful AI implementation, a relative scale of importance is composed. The scale depicts the degree of emphasis placed upon a success facet derived by the researcher from the analysis performed previously. To this scale, 1 indicates the actor perspective places little to no emphasis on the facet, whereas 5 indicates the facet is considered to be critical to the success of AI implementation. The results are presented in Table 15.

Success \ actor perspective	The industry	The government
Project management	3	1
Business	5	5
Technical	4	4
Internal social	4	3
External social	1	5

Table 15: An overview of emphasis placed upon the deduced success facets

To elaborate on Table 15, both actor views seem to emphasize the business potential of AI as the most critical success facet. This suggests that to implement AI successfully, an increase in business efficiency is a necessary facet. Besides the business facet, differences in importance amongst facets emerge that appear to be in line with the actors' objectives and interests explored in section 3.2 (Table 9: *An overview of the identified actors*). The industry values the technical capabilities and internal social facets highly, as both contribute to reaching the intended business potential (see Figure 10: *A visual strategy map of the chronological cause and effect among the success facets*). At the same time, little attention is paid to external social effects that influence society to a greater extent. Meanwhile, the government's view on successful use of AI includes the need for transparency and ethical use of the technology to safeguard public values.

4.6 Conclusion

Having analyzed the two actors' views on their definition of a successful AI implementation, several conclusions can be drawn. Firstly, the facets for success and failure mentioned by the interviewees are in line with the synthesized literature on enablers and prohibitors. By being able to delve deeper into the rationale of the interviewees, a suggestion on the relationship between these facets is presented in the form of a strategy map. The strategy map can be used during future projects as guidance for setting temporary (agile) goals or requirements, as well as for the analysis of failing implementation projects to locate the cause of failure. Secondly, a difference in "successful AI implementation" for two actor perspectives is exposed. Whereas the industry's focus is on gaining a business advantage by being able to make use of new technological capabilities, the government's focus stretches further by including the external social effects and transparency of the technology also. All in all, the two different views do not seem to be opposing but seem to put emphasis on different aspects. Therefore, formulating a single definition of "a successful implementation" is hard. Instead, it is suggested to define the applicable success facets of an implementation by evaluating the socio-technical environment (as executed in section 3) and determining the key success facets based on the success typology used throughout this analysis (project management, business, technical, internal social, and external social). As a result, implementation-specific success facets can be formulated that will help to include a multi-actor perspective on an AI project.

5 The boundary conditions for chatbot implementation

5.1 Introduction

As discussed in section 2.3 on the enablers of AI implementation, there are many factors suggested by literature that enhance the implementation of the technology to an extent thereby potentially increasing the project's success. After, in section 2.4, the suggested prohibitors to AI implementation have been synthesized. This section zooms in on the boundary conditions (BCs) for chatbot implementation in organizations. BCs are different from enablers and prohibitors because BCs are conditions that *must* be satisfied. This means that these conditions must be fulfilled before the implementation can be successful, whereas enablers and prohibitors are suggested to improve or hinder the implementation, respectively. The BCs are important to be aware of because they are critical factors for a successful implementation and can serve as a tool for future implementations. The tool can be used as guidance throughout future projects to ensure all social and technical factors are addressed, as well as to identify unaddressed factors in failing implementations.

But, because no literature has been found on the BCs for chatbot implementations, this section takes an exploratory and application-based approach to formulate these. By using an application-based approach, this section aims to avoid the context-dependency, which is highly present for the suggested enablers and prohibitors, and establish BCs that are generalizable to other chatbot implementations in different contexts as well.

To analyze the BCs for chatbot implementation, an iterative approach is taken using the definition of "success" from an industry perspective obtained in the previous chapter. Therefore, the intended goal for chatbot implementation is set to be: adding business value by increasing business efficiency. For the first iteration, a novel conceptual framework is introduced that guides the analysis process to establish a first draft of BCs. Thereafter, during the second iteration, duplicate BCs are discarded and the remaining BCs are divided according to the TOE framework. The second iteration of BCs is thereafter tested with two empirical cases discussed in Section 5.4.3. In the end, the cases are used to confirm, gain a deeper understanding of, and relate the proposed BCs.

5.2 The approach

To perform the first iteration of the analysis, a conceptual framework for the implementation of chatbots is used. Because no such framework has been found in the available literature, a novel one is created. The new framework is a combination of three existing frameworks, of which two have been discussed in previous chapters.

- The general chatbot architecture by Adamopoulou & Moussiades (2020) is used to capture the technological requirements of the implementation, previously shown in Figure 2: *A general chatbot architecture (from Adamopoulou & Moussiades, 2020, p.*380)

- To expose the social interactions and highlight the multistakeholder dimension related to the implementation, an applied version of Figure 7: *The formal relations between the actors* is used.
- To include and categorize different forms of institutions, legislation, and social norms, the four-layer model by Williamson (2000) is used.

Combining the three frameworks leads to the inclusion of the technological, social, and legal aspects of the implementation. The final framework is presented in Figure 12.

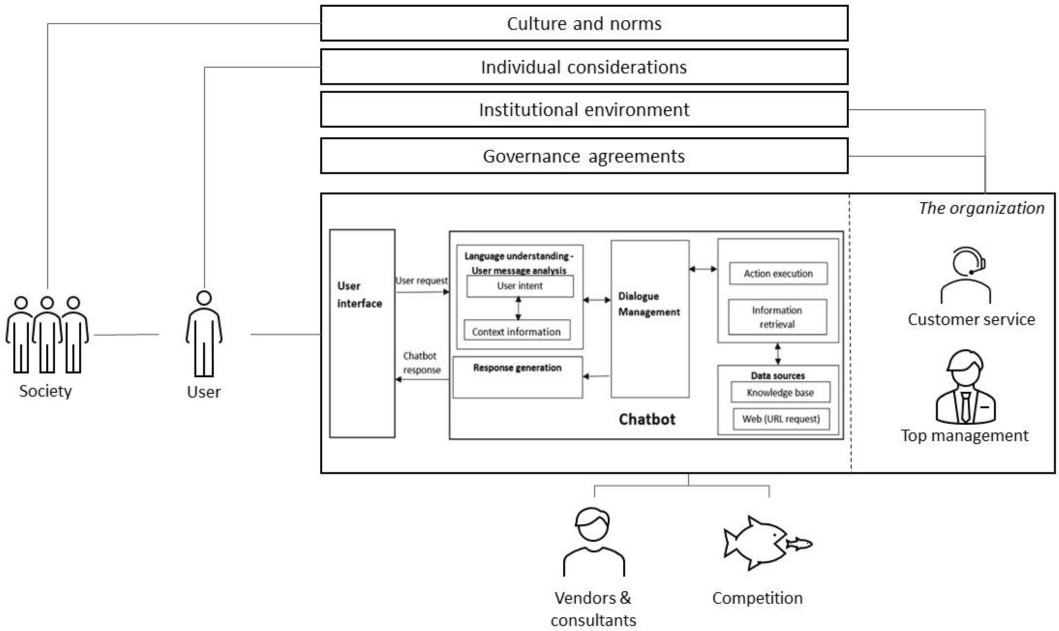

Figure 12: The conceptual framework used for the analysis of BCs

With the use of the conceptual framework, previously addressed design choices, trade-offs, and success factors are assessed on their urgency. Assessing the urgency is done by questioning: *can the implementation be completed without the condition? Can the implementation be successful without the condition?* or *can the implementation achieve its intended business value without the condition?*

5.3 The systematic analysis of the boundary conditions

For the first iteration of analysis, the previously evaluated sections and their respective factors are assessed on the urgency to a successful chatbot implementation. Table 16 shows the results of the first iteration, which is the first tentative version of the boundary conditions needed to successfully implement a chatbot within an organization.

Boundary condition
There must be high quality data available.
The chatbot must have an understandable user interface (language, visual, sound).
The chatbot must comply with local data privacy and confidentiality regulations.
The chatbot must be data secure.
The chatbot must have a natural usage control.
The chatbot must have legal compliance.
There must be tool availability.
The chatbot must be scalable.

There be customer readiness.
There must be top management support.
There must be available resources.
There must be chatbot provider commitment.
There must be high customer satisfaction.
There must be an internal understanding of the technology.
There must be a secure and reliable IT infrastructure.
The use must be ethical.
There must be data exchange possible among data sources.
The organization must decide on the chatbot type.
The organization must decide on the knowledge domain.
The organization must decide on response generation.
The organization must decide on the bot's capabilities (file-system permission).
The organization must execute regular chatbot evaluations.
There must be a perceived advantage (internal and external).
The chatbot must fit the organization's strategy.
There must be technological readiness inside the organization.
There must be internal acceptance inside the organization.
There must be social acceptance outside the organization.
There must be explainability of the responses.
There must be high algorithm accuracy.
The chatbot must be in line with local cultural norms and values.
The chatbot must be providing advantages to the user.
The social effects must be discussed and mitigated if necessary.
There must be internal incentive.
The chatbot must not exhibit bias.
There must be perceived usefulness by the user.
There must be a high ease of use for the user.
User requirements must be included during the development.
Human override must be possible by the user and/or the customer service.
Internal social effects must be evaluated.
External social effects must be evaluated.
Added business value must be evaluated.
Project goals must be set.
There must be alignment amongst stakeholders inside the organization.

Table 16: The first iteration of the boundary conditions

The first iteration results in 43 BCs. Even though duplicate conditions have been rejected during the analysis, there are BCs that overlap to an extent. Therefore, the second iteration will consist of: determining an umbrella condition for overlapping BCs and separating the BCs according to the TOE framework. This is done by first grouping the 43 BCs to one of the three types of BCs and then determining which conditions overlap. Finally, 35 BCs remain that are presented in Table 17.

Type of boundary condition	Boundary condition
Technology	High quality data must be available.
	Input and output data must be secure.
	There must be tool availability.
	The chatbot must be scalable.
	There must be a secure and reliable IT infrastructure.
	Data exchange amongst data sources must be possible.
	The right chatbot type and capabilities must be determined.
	The response generation must be explainable.
	The responses must be accurate.
	Human override must be possible by the user.
Organization	There must be top management support.
	There must be available resources.
	There must be an internal understanding of the technology.

Environment	The chatbot must fit the organization’s strategy. There must be technological readiness inside the organization. There must be internal acceptance inside the organization. There must be internal incentive to implement the chatbot. Internal social effects must be evaluated and mitigated if necessary. Added business value must be evaluated. Project goals must be set. There must be alignment amongst stakeholders inside the organization. The UI (User Interface) must be understandable and natural to use (language, visuals, sound). There be customer readiness. There must be chatbot provider commitment. There must be high customer satisfaction. The use must be ethical. The organization must execute regular chatbot evaluations. There must be a perceived advantage (internal and external). The chatbot must be in line with local cultural norms and values. The chatbot must be providing advantages to the user. There must be social acceptance outside the organization. There must be perceived usefulness by the user. There must be a high ease of use for the user. The chatbot must not exhibit bias. Must comply with local regulations (data privacy, confidentiality, security).

Table 17: The second iteration of the boundary conditions

5.4 Empirical testing of boundary conditions

The resulting BCs are plotted against two empirical cases to investigate whether these are supported by empirical evidence. By doing so, insights are obtained regarding the BCs’ validity, empirical presence, possible relations, and effects. The following sections describe the interview protocol that is used and the case selection. Thereafter, the cases are explored.

5.4.1 Interview protocol

The interview protocol that is used is presented in Appendix D: Interview protocol 2. The protocol is set-up according to the six stages described in Appendix B: Qualitative research interviews. By having the interviewee(s) determine if the BC is applicable to the case before the interview has started brings two advantages. First, the interviewee has had time to think about the relation between the BCs and the case, which is expected to enhance the discussion that follows. Second, the interviewee specificizes for each BC if it is applicable to the case thereby completing one of the three goals for the case study; testing the empirical presence. During the interview, the interviewee’s interpretation of the BC’s meaning is discussed also. As a result, a more refined description of the BC is acquired. The more refined description of the BC is thereafter discussed on satisfaction to the case.

5.4.2 Case selection

The two case studies are finalized AI chatbot implementations facilitated by the company CM.com. The company provides organizations with the tools and resources to set up their own Conversational AI Cloud. Examples of their long-term clients are DHL, ING, and Takeaway. For the case selection, theoretical sampling is used to explore two cases that differ in terms of success and context. By having two cases in which one has been more successful than the other, an understanding of the effects of the BCs is wished to be exposed. This is done by evaluating the differences in satisfied BCs and the success of the case. Also, by sampling two cases that have been implemented in a different context (sector and organization), the goal is to enhance the generalizability of results. Moreover, relevant case information is obtained by conducting interviews with the use of the interview protocol described above. The interview is conducted with two persons per case to gain an objective overview of the implementation process. Also, a total of four interviews have been conducted with solely two people at CM.com. By this means, both employees have been interviewed twice; once for each case.

5.4.3 Case study descriptions

Case study A concerns a chatbot implementation for an organization that provides an overview of insurances, energy contracts, and other financial products on their website. The motivation for the implementation was to support the customer service department because high workloads were experienced by the employees due to limited staff capacity. Therefore, the company wished to implement a closed knowledge domain chatbot with intent classification to answer FAQs and facilitate the first contact with the website's users. The goal was to have the chatbot go live as soon as possible. In the end, it took approximately four months to do so. Due to incorrect expectation management, this was longer than the company had hoped for. Nonetheless, both interviewees have described this case as being successful.

Case study B concerns a chatbot implementation for an organization that loans mortgages. The incentive for the company to implement a chatbot was to stimulate contact reduction for the customer service department. As interviewee 2 stated: "The company's organization decided in a top-down manner that a chatbot had to be used since competitors deployed similar techniques". At the start, no specific time schedule had been set for the implementation regarding goals, chatbot capabilities, or budget. The project has been ongoing for almost two years now and is described as less successful (or "challenging") by both interviewees. A summary of both cases is shown in Table 18.

Characteristic \ case study	A	B
Sector	Comparison website for insurances, energy contracts, and financial products.	Website to request loans and mortgages.
Project time	4 months.	2 years.
Chatbot motivation	Support the customer service department.	Support the customer service department.
Chatbot requirements	Answer FAQs, closed knowledge domain, intent classification, redirect to correct follow-up live-chat.	
Classified as	Successful.	Less successful.
Implementation consultant	CM.com	CM.com

Table 18: A descriptive summary of cases A and B

Type of boundary condition	#	Boundary condition	Description	Present in case A: ☑/☒/~	Present in case B: ☑/☒/~
Technology	1	The right chatbot type and capabilities must be determined.	Decisions regarding the chatbot typology, the chatbot design, the necessary chatbot capabilities, and the division between responsibilities of customer service and the chatbot must be clearly defined.	☑	☒
	2	High quality data must be available.	High-quality data must be available as input for the chatbot to train the algorithm and achieve an accurate model, used for example in intent classification. High-quality data must be available as output for the chatbot to provide the user with useful information for his/her request.	☑	☑
	3	Input and output data must be secure.	Confidentiality, security, and data privacy must be ensured for the input and output data of the chatbot. This includes masking of personal user information and is a prerequisite to complying with the GDPR.	☑	☑
	4	The response generation must be accurate.	The user must receive a response that is applicable to the question.	☑	☒
	5	There must be a secure and reliable IT infrastructure.	The physical IT infrastructure or cloud computing system that is needed to support the chatbot must be secure and reliable.	☑	☑
	6	The chatbot must be scalable.	It must be possible to upscale the chatbot to multi-channel, multi-users, and multi-tasking. Multi-channel implies the chatbot is accessible on different devices. A design choice must be made between an omnichannel bot (one that can track users across the multiple channels) and a multi-channel bot (one that cannot track the user across the channels yet supports the standard conversation). Examples of multi-tasks are: retrieving personal user information, adjusting personal user information, and lead generation.	☑	☑
	7	There must be tool availability.	(Open source) tools and packages to support and implement chatbots. For example, problem-specific customized solutions for specific problems. Often in the form of a Content Manager System (CMS), front-end builder, or support portal that can be provided by an implementation consultant.	☑	☑
	8	Human override must be possible.	Human override by the organization must be possible to change a chatbot's response to a question. Also, human override must be possible for a user to bypass the bot and contact customer service (by means of a live chat or audio conversation).	☑	☑
	9	Data exchange amongst data sources must be possible.	Read and write actions by the chatbot can happen in different databases or data lakes. Therefore, data exchange amongst these sources must be possible. This can be hindered by legacy systems or differences in the data structures.	~	~
	10	The response generation must be explainable.	Explainability and transparency of the trained algorithm are key. It must be explainable why a specific question is answered by a specific answer. If intent classification is not accurate, random answers can be provided to normal questions. An example in which this has failed is Microsoft's chatbot in 2016 (see Neff & Nagy (2016)).	☑	~
Organization	11	There must be internal incentive for the implementation.	Internal stakeholders must be incentivized to stimulate the implementation of chatbot technology.	☑	☑
	12	There must be available resources.	There must be a sufficient budget, time, people, etc.	☑	☒
	13	There must be technological readiness inside the organization.	There must be data, an IT infrastructure, a website, and employees that are capable of facilitating the implementation of the technology.	☑	☑

	14	There must be internal understanding of the technology.	Internal employees must understand the underlying concepts of the technology to better optimize the bot, respond to the changes in workflow, and capitalize on the benefits for the organization.	☑	☒
	15	There must be internal acceptance inside the organization.	Internal stakeholders must be willing to accept the implementation of chatbot technology and the changes it will bring.	☑	☑
	16	Added business value must be evaluated.	It must be evaluated what the proposed added business value of the chatbot will be to the organization. For example, will costs be reduced? Will new leads be generated? Or will the customer experience be improved?	☑	☒
	17	Project goals must be set.	Project goals such as the budget, the time schedule, and the chatbot's capabilities must be set.	☑	☒
	18	There must be alignment amongst stakeholders inside the organization.	Internal stakeholders must be aligned regarding the added business value, the expectations, the estimated benefits and costs, and the expected effects.	☒	☒
	19	The chatbot must fit the organization's strategy.	The use of chatbot technology and the implications it has for customer service, sales, and users must be in-line with the long-term plans of the organization.	☑	☒
	20	There must be top management support in the organization.	The top management should support the decision to start the implementation of chatbot technology. This allows for changing company structure and workflows.	☑	☑
	21	Internal social effects must be evaluated and mitigated if necessary.	It must be evaluated how the chatbot affects the organization's internal environment. For example, will the implementation result in a reduction of jobs in the customer service department? Or will the implementation change the way in which customer service and Sales communicate and work?	~	~
Environment	22	The UI must be understandable and natural to use.	The users must be able to use the chatbot's interface by considering factors such as the spoken language, the use of visuals, the use of sounds, content design, and the layout of the chatbot.	☑	☑
	23	The chatbot must comply with local regulations.	Regulations regarding saving data, providing answers based on user data, confidentiality on data sharing, and user tracking must be complied with.	☑	☑
	24	There must be customer readiness.	Customers must be able to use the chatbot efficiently and effectively and accept the changes it brings.	☑	☒
	25	There must be high customer satisfaction.	Customers must be happy with the use of the chatbot and experience an increase in customer experience or a higher level of customer service.	☑	☒
	26	There must be high ease of use for the user.	Examples that enhance the ease of use include a clear UI, quick responses provided by the bot, and the user does not need to repeat information during a conversation.	☑	☑
	27	The chatbot must be providing advantages to the user.	The user must experience advantages such as faster customer service, constant availability of the service, more easily digestible responses, or more personal help from the chatbot.	☑	☒
	28	The chatbot must not exhibit bias.	The chatbot cannot make a distinction between users in generating answers or understanding requests.	☑	☑
	29	The chatbot must be in line with local cultural norms and values.	Standards local society lives by and shared expectations amongst behavior must be ensured. This relates to the language used in the response generation and the information presented.	☑	☑

30	The organization must execute regular chatbot evaluations.	The responses and effects the chatbot offers must be regularly evaluated. For evaluation techniques, see section 2.2.5 chatbot evaluation.	☑	☑
31	The chatbot’s use must be ethical.	The use of the chatbot must be transparent to the user, the chatbot must not exhibit signs of racism, sexism, or abuse within the communication.	☑	☑
32	There must be social acceptance outside the organization.	Users and non-users of the chatbot must accept the deployment of the technology. This relates to sensitive contexts or public bodies that can come across social resistance when implementing such technology or negative chatbot experiences that are shared online.	☑	☒
33	There must be chatbot provider commitment.	If there is a chatbot provider, the provider must be committed for a longer period to maintain a steady chatbot performance.	☑	☑

Table 19 The case study results showing the BC’s description and empirical presence

☑ = The BC is satisfied, ☒ = the BC is not satisfied, ~ = The BC is not applicable.

5.4.4 Case study results

During all four interviews, the second iteration of BCs has been discussed with the interviewees. During the discussions, the interviewees' interpretation of the BC and its applicability to the cases have been the subjects of conversation. By debating over the implied meaning of the BCs, a more refined description has been obtained. As a result, the BCs *there must be perceived usefulness by the user* and *the chatbot must be providing advantages to the user* have been merged because of the overlap in description. Also, the BC *there must be perceived advantage (internal and external)* has been discarded since it is already described by the BCs *added business value must be evaluated* (for the internal part) and *there the chatbot must be providing advantages to the user* (for the external part). The third and final iteration of BCs including the description and empirical presence is presented in Table 19. The case study results are elaborated on in the following sections.

Case study A results

For the first case study, only *alignment amongst stakeholders inside the organization* (BC 18) is considered to be an unsatisfied BC. This is because the most prominent problem during this project was the expectation management amongst the organization's departments and the implementation consultant, resulting in an underestimation of the required workload. As stated by interviewee 1: "*Even though the project was a success, it did take longer than we expected due to two assumption errors*". At first, the internal knowledge base that was used by the customer service department to answer customer's questions manually was expected to be suitable as the knowledge base for the chatbot. However, due to unstructured data and large amounts of text used for the manual response generation, the knowledge base was not usable. This took time to adjust which caused delay to the original time schedule. The second misalignment concerned the use of intent classification. The organization wished to use intent classification to enhance the chatbot's performance and improve the customer experience. However, a more important project goal was to deploy the chatbot as quickly as possible. Since intent classification requires time and effort to be optimized, deployment of an accurate chatbot took longer than anticipated by some stakeholders. As a result, the discussed BC is unsatisfied even though the case is described as successful.

Case study B results

For the second case study, 12 BCs are considered to be unsatisfied. According to both interviewees, the implementation started out "challenging" because the organization did not know what capabilities the chatbot must possess, what project goals must be met, and in what way the chatbot would support the customer. This results in not satisfying BCs 1, 16, and 17. The organization solely specified the top management wanted "the best chatbot of the Netherlands" and that the decision was taken in a top-down manner; which caused some departments to disagree with the decision (BC 18). Thereafter, only a single person within the organization was dedicated to the implementation of the bot. Because the single employee was not able to understand the required underlying technological concepts, response generation was not optimized thereby affecting the chatbot's overall performance (BCs 4, 12, 14). Due to a poor performing chatbot, no advantages were provided to the user and the overall user satisfaction was low (BCs 25, 27). Finally, the environment in which the organization operates makes for a challenging chatbot implementation. Because mortgages can have an extended impact on a customer's life, there is less willingness to be assisted by a bot (BCs 24, 32). Also, questions that are asked by the customers are often case-specific rather than general (FAQs), this makes it hard for a chatbot to provide valuable answers to user requests (BC 19).

Non-applicable BCs

As can be seen in Table 19, there are two BCs (BCs 9 and 21) that are not applicable to any of the two cases. The first unapplicable BC, *data exchange amongst data sources must be possible,* is not present because both chatbots make use of a central data source to provide the data for the chatbot's responses. Also, because both projects concerned a chatbot that is solely able to answer questions rather than adjust or retrieve personal information as well, access to a single database was adequate. Secondly, *the internal social effects must be evaluated and mitigated if necessary* is marked as unapplicable because the incentive for both projects was to support the customer service department. By being incentivized to support the customer service department rather than cutting employees from the department, no negative internal social effects have been found regarding the two internal social effects. Therefore, this BC was also concluded to be unapplicable to both cases.

The two non-applicable BCs lead to the discussion on the generalizability of the proposed BCs. As confirmed by both interviewees, BCs 9 and 21 are valid if a different chatbot type is deployed or if the social effects are more extensive. Therefore, the BCs are not discarded from the proposed list. On the contrary, BCs 18, 19, and 20 are typical success factors suggested in the literature and might therefore not be considered BCs by some. As a result, the final iteration of BCs displayed in Table 19 is arranged according to the expected generalizability per type of BC; the more generally applicable BCs are stated above the more case-specific BCs within the TOE.

General remarks

Next to the satisfaction of BCs to the two cases, there are numerous general remarks that stood out during the interviews. Firstly, there is no qualitative data being collected to assess the user experience or to measure user satisfaction. The difficulty in obtaining such data was also mentioned in section 2.2.5 by Vijayaraghavan et al. (2020) as the drawback for using human assessment as chatbot evaluation technique. The only ways in which user experience is measured is by having a one-click survey at the end of a chatbot conversation to rate the experience, or by looking at the number of users that had to be directed to the customer service live-chat after being assisted by the chatbot. This is remarkable because the use of a chatbot is claimed to provide advantages to the user also (Følstad et al., 2018; Nguyen, 2019). Not measuring the satisfaction levels could lead to the use and maintenance of an ineffective and inefficient chatbot that does not reach its intended goal.

Secondly, BC 28 states *the chatbot must not exhibit bias* to prevent discrimination and unfair treatment of specific user groups. During the interviews, interviewee 2 spoke about the chatbot's capability to provide different answers to different user groups based on the device that was being used to support the HCI (e.g. laptop, smartphone, or tablet). The interviewee stated that the generated response was in essence the same, yet could differ in the presented layout or extensiveness. After a discussion, it has been decided that the BC is satisfied because the bot does not exclude user groups and the core answer is the same. Nonetheless, it should not be forgotten how well the chatbot is able to define its user and manipulate its response accordingly. This is highly related to the redistribution of power mentioned in section 2.2.6 on the ethics related to chatbot implementation.

Thirdly, as mentioned by interviewee 1: "*After deploying a chatbot, there is always some delay in seeing positive results. This is because the organization's user must get used to the chatbots, and we need to get used to the questions asked to provide the correct answers*". This is an interesting statement because the interviewee clearly addresses the intent classification and response generation that must be optimized causing the delay in measured business advantaged, and the change from HHI to HCI to which the user must adjust to.

5.5 Conclusion

Having completed three iterations of analysis, a list of 33 BCs for the successful implementation of a chatbot is proposed. The proposed BCs have been tested by assessing the applicability to two cases to determine the BCs' validity, empirical presence, and possible relations. As a result, 31 out of 33 BCs have been applicable to either one of the cases thereby providing evidence to support their empirical presence. On the contrary, the two non-applicable BCs imply a context-dependency of the proposed BCs as well. As a result, the context of a chatbot implementation could influence the extensiveness and criticality of the proposed factors. This implication can be supported by a discussion on the need for BCs 18, 19, and 20; which some might categorize as success factors rather than BCs. Therefore, a suggested hierarchy amongst the BCs is introduced based upon their general applicability to a chatbot. Finally, it is concluded that the proposed BCs can serve as a starting point to an implementation yet are not fully inclusive or fully exclusive. Besides, the results of case study B insinuate a relation between the conditions in which one BC can be highly affected by the (un)fulfillment of another BC. All in all, the 33 BCs presented in Table 19 have been discussed with two industry experts which resulted in a more refined description of the conditions and the validation of the empirical presence of most.

6 Discussion

6.1 Key findings

To the main research question

Thirty-three BCs for chatbot implementation are identified by analyzing theoretical background, the socio-technical environment, and the diverging perspectives on the facets related to implementation success. After, a multiple case-study methodology is used to verify the BCs' validity, empirical presence, and (possible) interrelations. Thereby, the main research question: *How do the socio-technical boundary conditions affect the success of chatbot implementation in organizations?* can be answered.

The proposed BCs have a significant effect on the implementation success of chatbots in organizations. This outcome is based upon the number of satisfied BCs per case, the success of each case as indicated by the interviewees, and the project time needed to complete the implementation. Also, the findings suggest two important contextual factors that influence the answer to the main research question, which are the number of applicable BCs and the definition of success.

First, by testing the validity and presence of the BCs for case studies A and B, a context-dependency amongst the conditions is suggested. This implies that, depending on the type of chatbot, not every BC is applicable and must therefore be fulfilled to successfully implement the bot. This implication is supported by the theoretical background on the enablers and prohibitors of AI, which are found to be context-dependent as well. As a result, the final iteration of BCs shown in Table 19 is arranged according to the expected generalizability per TOE category; the more generally applicable BCs are stated above the more case-specific BCs.

Second, the significant effect of the BCs is assessed from the industry's perspective on success. Thus, the intended goal of the chatbot implementation is to increase business efficiency and thereby create added business value (see Figure 10: *A visual strategy map of the chronological cause and effect among the success facets*). If the perspective on success is different, for example by taking the government's perspective and emphasizing the transparency of the bot or its social effects, the effects of the proposed BCs would not be the same.

To RQ1: What are the factors affecting AI implementation?

To determine the affecting factors, a literature search on the suggested enablers and prohibitors to AI implementation is performed in combination with a more in-depth exploration of AI for engagement (chatbots). The findings consist of two key conclusions. First, all three sections conclude that the applicable factors are context-dependent. For that reason, the interrelations, the degree of importance, and the change over time amongst these are barely discussed in a general sense. Second, even though the factors are context-dependent, many similarities can be found between general and industry-specific factors proposed in the literature. This causes the general factors to be a good starting point for defining the industry-specific or case-specific factors.

To RQ2: When is AI implementation deemed successful from different actor perspectives?

To investigate the most prominent actors and their positions, an extensive analysis of the socio-technical landscape is performed. Thereafter, the industry's perspective and the government's perspective are used to answer the research question. By conducting eight interviews with industry practitioners, the industry's perspective on "successful AI implementation" is mapped based on five success facets (project management, business, technical, social internal, and social external). During the same interviews, the causes of failure are investigated. This results in a suggested hierarchy and interrelation between the success facets and is presented in the form of a strategy map (Figure 10: *A visual strategy map of the chronological cause and effect among the success facets*).

For the government's perspective, the current regulatory environment in the Netherlands is used to deduce the hierarchy amongst success facets. By comparing the two results, both views appear to be highly interested in the business potential that AI solutions offer. Yet, whereas the industry is more concerned with the organizational acceptance of the technology (social internal facet), the government actively aims to protect its society (social external facet). This conclusion is like two of the trade-offs identified in section 3.4.1, which are *the risks vs. the benefits* and *the social value vs. the business value*.

To RQ3: What are the socio-technical boundary conditions for chatbot implementation?

The final research question aims to synthesize a list of socio-technical boundary conditions for chatbot implementation, test these, and observe if there is a relationship between them. By developing a novel framework for chatbot implementation (Figure 12: *The conceptual framework used for the analysis of BCs*), a list of BCs is proposed based on the information gathered throughout the chapters prior to Chapter 5. After, the proposed BCs are tested on two cases that differ in terms of success. As such, a difference in the number of fulfilled BCs is present between the successful case (case A, fulfilling 31/33 BCs) and the less successful case (case B, fulfilling 18/33 BCs). By discussing the proposed BCs with the interviewees, all 33 BCs are validated and described more elaborately. Also, case B suggests a possible relation between the BCs. It seems that if certain BCs are not fulfilled, the fulfillment of other BCs is affected. But further research is required to make a statement on a specific hierarchy. Finally, like the context-dependency mentioned in RQ1, a context-dependency of BCs is suggested by the two non-applicable BCs. These two non-applicable BCs have been validated and confirmed by the interviewees yet were not applicable to the two cases at hand.

6.2 Contributions

The contributions of this research consist of three parts. At first, the analysis of the industry's perspective on a "successful AI implementation" has led to a suggested hierarchy amongst success facets, presented in the form of a strategy map. The strategy map can be used as a high-level framework for future AI implementations to set (temporary) goals and demystify the progress being made. This can be useful, especially for organizations working in an agile manner, because it allows beginning every iteration from the same starting point. Also, the strategy map can be used for failing projects to identify the point(s) of failure.

Second, by comparing the industry's perspective and the government's perspective on "success", differences in emphasis on success facets is exposed. By exposing these differences, it will be easier to facilitate a multistakeholder approach throughout the whole AI project life cycle in future projects, potentially contributing to the design of more sustainable AI systems.

Third, 33 BCs for the implementation of a chatbot in an organization are suggested. The suggested BCs can be used as guidelines for the implementation of future chatbots. Satisfying the BCs during the design and implementation of the chatbot is suggested to enhance the overall success and can be used to ensure that the technical and social implications are addressed. Also, future research could investigate whether fulfilling the BCs decreases project time and project costs and/or increases business value and customer satisfaction, which are some of the expected effects.

6.3 Reliability of the results

The reliability of results addresses whether the data collection techniques and analytical procedures would reproduce consistent findings if repeated in another setting or by another researcher. The two errors through which unreliable data can be acquired are error and bias, both of which can be caused by the researcher and the participant. Participant error is about factors that can alter the responses, whereas the bias is about factors that can lead to false responses. As for the data collected throughout this research, neither of these two risks appear to be high. This is because all the interviews have been conducted in private, took on average 20 minutes (for the first eight interviews), and did not address any sensitive topics about the interviewees' organizations or their relationship with the organization.

Moreover, researcher error regards factors that alter the researcher's interpretation. To decrease the risk of having such errors, no more than two interviews have been conducted on the same day, the color coding of the qualitative data was not performed on the same day as conducting the interview, and the color coding was revised twice on completeness and correctness. Also, the risk of implying researcher bias has been mitigated actively by using the current literature as starting point, using interview protocols, and discussing the results with (external) supervisors and experts working on chatbot implementations daily. Unfortunately, one should be aware of possible unintended researcher bias. The sections in which this could have had the most significant effect include the analysis of the government's view on "success" (section 4.4), the comparison between the two perspectives on success (section 4.5), and the analysis of the BCs (section 5.3). The possible effect of the unintended research bias might lead to a discussion on the conclusions being drawn. Nonetheless, all three of these sections entail an elaborate reasoning for the conclusions. Therefore, one should be able to identify the researcher's reasoning of choices and conclusions.

6.4 Validity of the results

The validity of the results consists of three parts: the construct validity, the internal validity, and the external validity. The construct validity elaborates on the reasoning for the applied research strategy and methodologies. For this research, RQ2 and RQ3 are considered to conduct exploratory research since no relevant literature has been found addressing these topics. Therefore, using qualitative research methods (interviews and case studies) seems applicable (Stebbins, 2001). However, exploratory research normally implies using inductive reasoning to develop a theory; which was only the case for sub-research question 2. For question 3, the related available literature gathered throughout sub-research question 1 was used to test the proposed theory (BCs) in a deductive manner. Because the proposed theory is based upon well-established literature, the deductive testing of theory seems reasonable, yet uncommon.

The internal validity addresses the assumption of the causal relationship between two variables, which are the proposed BCs and the success of a chatbot implementation throughout this research. In short, the internal validity is defined as the degree to which the results represent the truth. Throughout this research, the internal validity was increased by basing the proposed BCs on available literature and analyzing two real world chatbot implementations that have been executed in different contexts. Also, even though the sample size of the case studies was limited, the analysis of both cases has been executed with the help of two experts from the field. Doing so has enhanced the practical implications and increases the likelihood of the analysis to be truthful.

The external validity is about whether the study's research findings can be generalized to other settings. For the findings related to the hierarchy amongst success facets, these may be generalizable but this cannot be stated with certainty. The qualitative data gathered that led to these findings is based upon eight interviews with industry practitioners employed in the Netherlands. Because theoretical sampling was used to acquire a heterogenous mix of interviewees, the aggregated results are expected to be generalizable to other settings. However, because this has not been tested by conducting the same interview protocol with practitioners in different contexts, this cannot be stated with certainty. Second, for the findings related to the differences in "success" from two actor perspectives, the generalizability of results is expected to be limited. Reason for this is because the government's actor perspective is mainly based upon the Dutch legislation, and therefore may not be generalizable to different countries. Conversely, Dutch legislation is influenced by legislation set up by the European Commission. Therefore, for other European countries the differences might not be that significant, which enhances the generalizability of results. Third, for the findings related to the BCs, the goal was to establish BCs that are generalizable by taking an application-based approach and using theoretical sampling to analyze two cases set in a different contexts. But because the two cases have been supported by the same implementation consultant (CM.com), it cannot be stated with certainty that the goal was reached. Therefore, the BCs must be tested on different cases to be able to make a grounded claim on the generalizability.

6.5 Limitations

There are several limitations to this research that are acknowledged and elaborated on in the following section. Being aware of the limitations should help to interpret the findings correctly.

Firstly, the research has been conducted in the Netherlands and further research is needed to determine if the findings are valid outside of the Netherlands. Research stages that might be affected by the research location include the social analysis (chapter 3), the interview results (chapter 4), and the effect of the BCs on the success of chatbot implementations (chapter 5). External factors such as local culture, norms and values of inhabitants, or the degree of development in a country could have an influence on the results obtained.

Secondly, the industry perspective for the definition of "success" in chapter 4 is based upon eight interviews, which is a limited amount. Normally, the number of qualitative research interviews is ought to be sufficient when information saturation is reached. Even though the interview responses do include similarities regarding success factors and causes of failure, there is no guarantee that information saturation has been reached. Therefore, conducting the same interview protocol with a more extensive group of interviewees would strengthen the presented data. Also, the qualitative interview data is analyzed in a quantitative manner using color coding. The assumptions made based on the quantitative analysis are thereafter generalized without being certain if this is applicable to a wider group of industry practitioners.

Thirdly, the research was performed by a single researcher, which could have led to unintended researcher's bias. To avoid researcher bias, all the decisions made during the research have been elaborated on by means of a rationale and supporting (academic) evidence. Nonetheless, the sections in which the researcher's bias could be present include the analysis of the government's view on "success" (section 4.4), the comparison between the two perspectives on success (section 4.5), and the analysis of the BCs (section 5.3).

Fourthly, the empirical testing of the BCs was performed by means of interviewing two employees employed by the same organization. Therefore, future research could investigate if the validity, presence, relations, and effects are like other chatbot implementations in different organizations.

6.6 Future research

Based on the findings and limitations, there are multiple aspects for future research suggested. First, ranking methods such as the q-methodology or the best-worst method can be used to gain a deeper understanding of the industry's perspective on "successful AI implementation". Doing so could strengthen, or challenge, the strategy map and provide more insights into what is deemed important by the industry. Also, conducting additional interviews with other industry practitioners or investigating different actor perspectives could be valuable to more elaborately answer RQ2.

Second, the suggested BCs can be tested on cases set in different organizations or contexts to challenge their validity and generalizability. Also, testing the BCs on new cases can provide more insights into the relationship between them. This was already one of the objectives for this research (RQ3c) yet was not fully answered. Moreover, by testing the BCs on different applications, the external social effects might be considered as well. For example, by taking the government's perspective on "success" and determining how the social effects relate to the suggested BCs.

Thirdly, research can be conducted to determine if an organization and a chatbot are always a good fit. As suggested by the context in which the chatbot for case study B was deployed, there might be organizations that will not profit from implementing a chatbot due to the environment they operate in. Therefore, it will be valuable to investigate if there are BCs related to the fit of a chatbot and an organization.

6.7 Reflection

Having discussed the outcomes and limitations of the research, future research can profit from reflecting on several of the research decisions made. First, the choice to use the TOE framework to categorize the proposed enablers, prohibitors, and BCs. At the start of the research, the reasons to use this framework were because it is widely applied to describe technological adoption, it is used in literature reviews on the prohibitors of AI adoption, and it distinguishes between social and technological factors which is useful for analyzing socio-technical problems. Even though the framework has been valuable throughout the research by offering a similar overview to distinguish between the influential factors, the use of the PESTEL (Political, Economic, Social, Legal, Environmental, Technological) analysis, formerly known as PEST analysis, may have been better (Sammut-Bonnici et al., 2015). The PESTEL analysis has two advantages compared to the TOE framework. First, by making a distinction between six categories instead of three, a more detailed overview is created. Second, the six categories appear to be similar to the five success facets (project management, business, technical, social internal, and social external) used throughout SQ2 to an extent. Therefore, if the same framework was used to answer all three research questions, the comparison of results would be simplified.

Second, the theoretical background starts by exploring the enablers and prohibitors of AI adoption rather than focusing on chatbots immediately. Doing so comes with an advantage and a disadvantage. The advantage of considering the whole field of AI at first is that it allows for inspiration and ideation whilst mitigating the risk of acquiring a tunnel vision early on. The process of diverging and converging the research scope is often used within problem-solving situations and was therefore applied for this research also. The disadvantage of doing this is the risk of losing direction and (potentially) drowning in unrelated literature. This can cause the research to drift away from the main research question and into irrelevant sub-topics. Resulting in a loss of time and a blurred research objective.

6.8 Conclusion

Due to advances in AI technology, the number of organizations willing to implement AI-based solutions is extremely high. But, to design and deploy sustainable AI systems, the social implications caused by the technological system must be evaluated. Because of the “chatbot tsunami”, a term that is used to describe the enormous rise in chatbots facilitated by organizations offering the Software as a Service, this research focused on the effects of the socio-technical BCs on the success of chatbot implementation in organizations.

As a result, the findings show that satisfying the suggested BCs has a positive effect on the success of the chatbot implementation. This practically implies that ensuring the satisfaction of the proposed BCs during future chatbot implementations enhances the success of the implementation. As a result, project time and project costs of an implementation may be decreased whilst the added business value and customer satisfaction may be increased. But, these effects must be confirmed by future research before grounded claims can be made. For future research, it is therefore suggested to investigate the generalizability of the boundary conditions to different contexts. To do so, other organizations in different sectors must be analyzed. Also, a similar research approach is suggested to investigate the socio-technical BCs to different AI applications such as recommendation systems and healthcare applications to enhance the social acceptance of such applications as well.

References

Abdi, H. (2007). The Method of Least Squares. *Encyclopedia of Measurement and Statistics 1*, 530–532.

Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. *IFIP International Federation for Information Processing*, *May*, 373–183. https://doi.org/10.1007/978-3-030-49186-4

Agentschap Telecom. (2021). *Regelgeving Kunstmatige Intelligentie (AI)*. https://www.agentschaptelecom.nl/onderwerpen/kunstmatige-intelligentie/regelgeving-kunstmatige-intelligentie-ai#:~:text=Regulering van AI in Nederland,die al wel gereguleerd zijn.

Aghion, P., Bergeaud, A., & van Reenen, J. (2021). The Impact of Regulation on Innovation. *NBER Working Paper No. 28381*, 1–76.

Ahmad, O. F., Stoyanov, D., & Lovat, L. B. (2020). Barriers and pitfalls for artificial intelligence in gastroenterology: Ethical and regulatory issues. *Techniques and Innovations in Gastrointestinal Endoscopy*, *22*(2), 80–84. https://doi.org/10.1016/j.tgie.2019.150636

Albert, M., Balve, P., & Spang, K. (2017). Evaluation of project success: a structured literature review. *International Journal of Managing Projects in Business*, *10*(4), 796–821. https://doi.org/10.1108/IJMPB-01-2017-0004

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, *1*, 238–247. https://doi.org/10.3115/v1/p14-1023

Beel, J., & Gipp, B. (2009). Academic Search Engines, Google Scholar, Ranking Algorithm. *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, *1*(July), 203–203. www.docear.org

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Artificial Intelligence in Organizations: Implications for Information Systems Research. *Journal of the Association for Information Systems*, *22(2), xxx*, 25. https://doi.org/10.17705/1jais.00491

Bérubé, M., Giannelia, T., & Vial, G. (2021). Barriers to the implementation of AI in organizations: Findings from a Delphi study. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-Janua*, 6702–6711. https://doi.org/10.24251/hicss.2021.805

Borgesius, F. J. Z. (2016). *Should we worry about filter bubbles ? 5*(1), 1–16. https://doi.org/10.14763/2016.1.401

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brock, J. K. U., & von Wangenheim, F. (2019). Demystifying Ai: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review*, *61*(4), 110–134. https://doi.org/10.1177/1536504219865226

Bryson, J. M. (2004). What to do when stakeholders matter: Stakeholder Identificatixon and analysis techniques. *Public Management Review*, *6*(1), 21–53. https://doi.org/10.1080/14719030410001675722

Butler, D. L., & Sellbom, M. (2002). Barriers to Adopting Technology for Teaching and Learning. *Educause Quarterly*, *25*(2), 22–28. https://www.learntechlib.org/p/92849

Castro, F., Tedesco, P., Alves, H., Quintino, J. P., Steffen, J., Oliveira, F., Soares, R., Santos, A. L. M., & da Silva, F. Q. B. (2018). Developing a Corporate Chatbot for a Customer Engagement Program: A Roadmap. In D.-S. Huang, V. Bevilacqua, P. Premaratne, & P. Gupta (Eds.), *Intelligent Computing Theories and Application* (pp. 400–412). Springer International Publishing.

Chen, H., Li, L., & Chen, Y. (2021). Explore success factors that impact artificial intelligence adoption on telecom industry in China. *Journal of Management Analytics*, *8*(1), 36–68. https://doi.org/10.1080/23270012.2020.1852895

Chen, Q., Zhuo, Z., & Wang, W. (2019). BERT for Joint Intent Classification and Slot Filling. *DAMO Academy, Alibaba Group*, 1–4.

Chowdhary, K. R. (2020). *Fundamentals of artificial intelligence* (1st ed.). Springer New Delhi. https://doi.org/https://doi.org/10.1007/978-81-322-3972-7

Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the Bot : Anthropomorphism and Anger in Customer – Chatbot Interactions. *Journal of Marketing*, *86*(1), 132–148. https://doi.org/10.1177/00222429211045687

Cubric, M. (2020). Drivers , barriers and social considerations for AI adoption in business and management : A tertiary study. *Technology in Society*, *62*(December 2019), 101257. https://doi.org/10.1016/j.techsoc.2020.101257

Dahlin, E. (2021). Mind the gap! On the future of AI research. *Humanities and Social Sciences Communications*, *8*(1), 21–24. https://doi.org/10.1057/s41599-021-00750-9

Davenport, T. H. (2018). *The AI advantage: How to put the artificial intelligence revolution to work*. MIT Press.

Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. *International Journal of Human Computer Studies*, *45*(1), 19–45. https://doi.org/10.1006/ijhc.1996.0040

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, *40*(4), 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x

Dora, M., Kumar, A., Mangla, S. K., Pant, A., & Kamal, M. M. (2021). Critical success factors influencing artificial intelligence adoption in food supply chains. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2021.1959665

Europa Decentraal. (2021). *Artificial Intelligence*. https://europadecentraal.nl/onderwerp/digitale-overheid/data-en-technologie/artificial-intelligence/

European Commission. (2019). *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS - Building Trust in Human-Centric Artificial Intelligence*. *1*, 1–23.

European Commission. (2020). *White Paper on Artificial Intelligence - A European approach to excellence and trust*. *1*, 1–27. https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_fr.pdf

Fleming, N. (2018). How artificial intelligence is changing drug discovery. *Nature*, *557*(7707), S55–S57. https://doi.org/10.1038/d41586-018-05267-x

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Følstad, A., Brandtzaeg, P., Feltwell, T., Law, E., Tscheligi, M., & Luger, E. (2018). SIG: Chatbots for Social Good. *Conference on Human Factors in Computing Systems*, 1–4. https://doi.org/10.1145/3170427.3185372

Fontana, A., & Frey, J. H. (1994). Interviewing: The Arts of Science. *Handbook of Qualitative Research*, *i*, 361–376.

Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). Faster Is Not Always Better : Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction the 26 th European Conference on Information Systems (ECIS), Portsmouth , United Kingdom , June 23-

28 . Please note : Copyright is owned by th. *Proceedings of the 26th European Conference in Information Systems (ECIS), June*.
Graeme, T., & Walter, F. (2008). Success in IT projects : A matter of definition? *International Journal of Project Management*, *26*, 733–742. https://doi.org/10.1016/j.ijproman.2008.06.003
Graves, A., Mohamed, A., & Hinton, G. (2013). Speech Recognition with Deep Recurrent Neural Networks. *International Conference on Acoustics, Speech, and Signal Processing*, *3*, 5. https://doi.org/10.48550
Grudin, J., & Jacques, R. (2019). Chatbots, Humbots, and the Quest for Artificial General Intelligence. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3290605.3300439
Gudala, M., Ross, M. E. T., Mogalla, S., Lyons, M., Ramaswamy, P., & Roberts, K. (2022). Benefits of, Barriers to, and Needs for an Artificial Intelligence--Powered Medication Information Voice Chatbot for Older Adults: Interview Study With Geriatrics Experts. *JMIR Aging*, *5*(2), e32169. https://doi.org/10.2196/32169
Gupta, A., Hathwar, D., & Vijayakumar, A. (2020). Introduction to AI Chatbots. *International Journal of Engineering Research & Technology (IJERT)*, *9*(07), 255–258.
Hagerty, A., & Rubinov, I. (2019). Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence. *CoRR*, *abs/1907.0*. http://arxiv.org/abs/1907.07892
Hamm, P., & Klesel, M. (2021). Success factors for the adoption of artificial intelligence in organizations: A literature review. *27th Annual Americas Conference on Information Systems, AMCIS 2021*, *August*.
Hartman, F., & Ashrafi, R. A. (2002). Project Management in the Information Systems and Information Technologies Industries. *Project Management Journal*, *33*(3), 5–15. https://doi.org/10.1177/875697280203300303
Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245–250. https://doi.org/https://doi.org/10.1016/j.chb.2015.02.026
Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *Computer*, *29*(3), 31–44. https://doi.org/10.1109/2.485891
Jeffrey, P., Richard, S., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
Kostadinov, S. (2019). *Understanding Encoder-Decoder Sequence to Sequence Model*. Towards Data Science. https://towardsdatascience.com/understanding-encoder-decoder-sequence-to-sequence-model-679e04af4346
Kroes, P., Franssen, M., Van De Poel, I., & Ottens, M. (2006). Treating socio-technical systems as engineering systems: Some conceptual problems. *Systems Research and Behavioral Science*, *23*(6), 803–814. https://doi.org/10.1002/sres.703
Kumar, R., & Ali, M. M. (2020). A Review on Chatbot Design and Implementation Techniques. *International Research Journal of Engineering and Technology (IRJET)*, *07*(02), 2791–2800.
Kumar, S., Raut, R. D., Queiroz, M. M., & Narkhede, B. E. (2021). Mapping the barriers of AI implementations in the public distribution system: The Indian experience. *Technology in Society*, *67*(July), 101737. https://doi.org/10.1016/j.techsoc.2021.101737
Kvale, S. (1994). InterViews: An introduction to qualitative research interviewing. In *InterViews: An introduction to qualitative research interviewing.* Sage Publications, Inc.
Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lei, C., Shaohan, H., Furu, W., Chuanqi, T., Chaoqun, D., & Ming, Z. (2017). SuperAgent : A Customer Service Chatbot for E-commerce Websites. *Association for Computational Linguistics*, 97–102.

Link, B. (2013). Considering the Company's Characteristics in Choosing between SaaS vs. On-Premise-ERPs. *Wirtschaftsinformatik Proceedings*, 261–277. http://aisel.aisnet.org/wi2013/17

Liu, H., Lin, C., & Chen, Y. (2019). *Beyond State v Loomis : artificial intelligence , government algorithmization and accountability*. 1–20. https://doi.org/10.1093/ijlit/eaz001

Lokman, A. S., & Ameedeen, M. A. (2019). *Modern Chatbot Systems : A Technical Review* (Issue 3). Springer International Publishing. https://doi.org/10.1007/978-3-030-02683-7

McCarthy, J. (2007). WHAT IS ARTIFICIAL INTELLIGENCE? *Computer Science Department Stanford University*, 15.

Mellon, C. A. (1990). *Naturalistic Inquiry for Library Science: Methods and Applications for Research, Evaluation, and Teaching* (N. Greenwood Press, New York (ed.)). Greenwood Press, New York, NY.

Miller, G. J. (2021). Artificial Intelligence Project Success Factors: Moral Decision-Making with Algorithms. *Proceedings of the 16th Conference on Computer Science and Intelligence Systems, FedCSIS 2021*, *25*, 379–390. https://doi.org/10.15439/2021F26

Miltgen, C. L., & Smith, H. J. (2015). Exploring information privacy regulation, risks, trust, and behavior. *Information and Management*, *52*(6), 741–759. https://doi.org/10.1016/j.im.2015.06.006

Murtarelli, G., Gregory, A., & Romenti, S. (2021). A conversation-based perspective for shaping ethical human–machine interactions: The particular challenge of chatbots. *Journal of Business Research*, *129*(March 2019), 927–935. https://doi.org/10.1016/j.jbusres.2020.09.018

Neff, G., & Nagy, P. (2016). Talking to Bots : Symbiotic Agency and the Case of Tay. *International Journal of Communication*, *10*, 4915–4931. https://doi.org/1932–8036/20160005

Neuhüttler, J., Kett, H., Frings, S., Falkner, J., Ganz, W., & Urmetzer, F. (2020). Artificial Intelligence as Driver for Business Model Innovation in Smart Service Systems. In J. Spohrer & C. Leitner (Eds.), *Advances in the Human Side of Service Engineering* (pp. 212–219). Springer International Publishing.

Nguyen, T. (2019). *Potential effects of chatbot technology on customer support: A case study*. https://aaltodoc.aalto.fi/handle/123456789/38921

Nimdzi Insights, & Edge, P. (2019). *Artificial intelligence, localization, winners, losers, heroes, spectators, and you*.

Oh, K. J., Lee, D., Ko, B., & Choi, H. J. (2017). A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. *Proceedings - 18th IEEE International Conference on Mobile Data Management, MDM 2017*, 371–376. https://doi.org/10.1109/MDM.2017.64

Palasundram, K., Sharef, N. M., Nasharuddin, N., Kasmiran, K., & Azman, A. (2019). Sequence to Sequence Model Performance for Education Chatbot. *International Journal of Emerging Technologies in Learning (IJET)*, *14*(24), 56–68. https://www.learntechlib.org/p/217029

Paranjape, K., Schinkel, M., Hammer, R. D., Schouten, B., Nannan Panday, R. S., Elbers, P. W. G., Kramer, M. H. H., & Nanayakkara, P. (2021). The Value of Artificial Intelligence in Laboratory Medicine. *American Journal of Clinical Pathology*, *155*(6), 823–831. https://doi.org/10.1093/ajcp/aqaa170

Pinto, J. K., & Slevin, D. P. (1988). Critical success factors across the project life cycle. *Project Management Journal*, *19*(3), 67–75.

Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting and Management*, *8*(3), 238–264. https://doi.org/10.1108/11766091111162070

Rabionet, S. E. (2011). How I learned to design and conduct semi-structured interviews: An ongoing and continuous journey. *Qualitative Report*, *16*(2), 563–566. https://doi.org/10.46743/2160-3715/2011.1070

Radhakrishnan, J., & Chattopadhyay, M. (2020). Determinants and Barriers of Artificial Intelligence Adoption – A Literature Review. *IFIP Advances in Information and Communication Technology*, *617*(January 2021), 89–99. https://doi.org/10.1007/978-3-030-64849-7_9

Rapp, A., Curti, L., & Boldi, A. (2021). International Journal of Human - Computer Studies The human side of human-chatbot interaction : A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human - Computer Studies*, *151*(September 2020), 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

Refaeilzadeh, P., Tang, L., & Liu, H. (2016). Cross-Validation. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 1–7). Springer New York. https://doi.org/10.1007/978-1-4899-7993-3_565-2

Reim, W., Åström, J., & Eriksson, O. (2020). Implementation of Artificial Intelligence (AI): A Roadmap for Business Model Innovation. *Ai*, *1*(2), 180–191. https://doi.org/10.3390/ai1020011

Remenyi, D., & Sherwood-Smith, M. (1991). Maximize information systems value by continuous participative evaluation. *Log Information Management*, *12*, 14–31.

Rijksoverheid. (2019). *Strategisch Actieplan voor Artificiële Intelligentie*. https://www.rijksoverheid.nl/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie

Rodney Turner, J., & Müller, R. (2005). The Project Manager's Leadership Style As a Success Factor on Projects: A Literature Review. *Project Management Journal*, *1997*, 49–61.

Rong, X. (2014). *word2vec Parameter Learning Explained*. 1–21. http://arxiv.org/abs/1411.2738

Sammut-Bonnici, T., Galea, D., Mcgee, J., & Cooper, C. (2015). PEST analysis. *Wiley Encyclopedia of Management*.

Schensul, S. L., Schensul, J. J., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires*. Walnut Creek, Calif: AltaMira Press.

Serban, I. V, Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., Kim, T., Pieper, M., Chandar, S., Ke, N. R., Rajeshwar, S., Brebisson, A. De, Sotelo, J. M. R., Suhubdy, D., Michalski, V., Nguyen, A., Pineau, J., & Bengio, Y. (2017). A Deep Reinforcement Learning Chatbot. *ArXiv Preprint ArXiv:1709.02349*, 1–40.

Shewalkar, A., Nyavanandi, D., & Ludwig, S. A. (2019). *PERFORMANCE EVALUATION OF DEEP NEURAL NETWORKS APPLIED TO SPEECH RECOGNITION : RNN , LSTM AND GRU*. *9*(4), 235–245.

Singh, R. P., Hom, G. L., Abramoff, M. D., Campbell, J. P., & Chiang, M. F. (2020). Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. *Translational Vision Science and Technology*, *9*(2), 1–6. https://doi.org/10.1167/tvst.9.2.45

Stebbins, R. A. (2001). *Exploratory research in the social sciences* (40th ed.). SAGE Publications.

Tarhini, A., Ammar, H., Tarhini, T., & Masa'deh, R. (2015). Analysis of the Critical Success Factors for Enterprise Resource Planning Implementation from Stakeholders' Perspective: A Systematic Review. *International Business Research*, *8*(4). https://doi.org/10.5539/ibr.v8n4p25

Thorat, S. A., & Jadhav, V. (2020). A Review on Implementation Issues of Rule-based Chatbot Systems. *SSRN Electronic Journal*, *Icicc*, 1–6. https://doi.org/10.2139/ssrn.3567047

Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. K. (1990). *Processes of Technological Innovation*. Lexington books.

Vijayaraghavan, V., Cooper, J. B., & Rian Leevinson, R. L. (2020). Algorithm Inspection for Chatbot Performance Evaluation. *Procedia Computer Science*, *171*(2019), 2267–2274.

https://doi.org/10.1016/j.procs.2020.04.245
Wiener, J. B. (2004). The regulation of technology, and the technology of regulation. *Technology in Society*, *26*(2–3), 483–500. https://doi.org/10.1016/j.techsoc.2004.01.033
Williamson, O. E. (2000). The New Institutional Economics: Taking Stock, Looking Ahead. *Journal of Economic Literature*, *38*(3), 595–613. http://www.jstor.org/stable/2565421
Yang, Y., Liu, Y., Lv, X., Ai, J., & Li, Y. (2022). Anthropomorphism and customers' willingness to use artificial intelligence service agents. *Journal of Hospitality Marketing and Management*, *31*(1), 1–23. https://doi.org/10.1080/19368623.2021.1926037
Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, *46*(1), 53–93. https://doi.org/10.1162/COLI_a_00368

Appendix

Appendix A: Deep Learning introduction

Deep Learning is a subfield of ML based on artificial Neural Networks (NN). Whereas in ML feature extraction must be executed manually, this is not the case for DL. Feature extraction involves identifying and reducing the number of key variables needed to describe the data. In ML, careful engineering and domain expertise is needed for this task (Lecun et al., 2015). Since this is not the case for DL, complex patterns in large (unstructured) data sets which indicate how the machine should change its algorithm parameters are identified automatically. Being able to do so has brought breakthroughs in areas such as image, video, and speech recognition (Lecun et al., 2015).

The NN used in DL are inspired by biological NN present inside the human brain and consists of three types of layers (Jain et al., 1996). The three sorts of layers are: the input layer, the hidden layer(s), and the output layer. A schematic representation of a typical NN is shown in Figure 13 below.

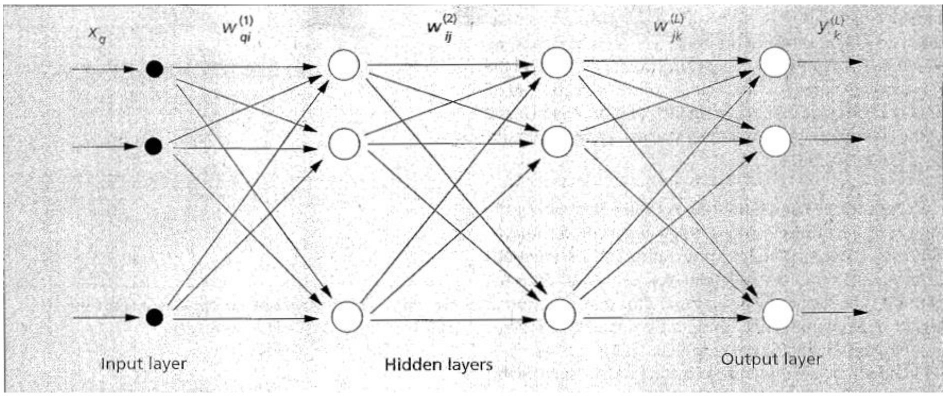

Figure 13: A schematic representation of a Neural Network (from Jain et al., 1996, p.38)

The learning algorithm is meant to approximate the weights assigned to the edges between the nodes. After the weights have been approximated, new inputs can be assigned a predicted outcome to classify these. To approximate the weights, the backpropagation learning algorithm is widely applied (Jain et al., 1996). The algorithm uses the partial gradient of the loss function with respect to the weight of each edge by the chain rule. By starting from the final hidden layer and iterating backwards, the weights are updated. These updates are computed after each training iteration to approximate the final values. For a more elaborate explanation on artificial NN and the applied learning algorithms, please refer to Jain et al. (1996).

Besides a normal NN, there are Convolutional NNs and (CNN) and Recurrent NNs (RNN). CNNs are used to process input data in the form of multiple arrays, such as images. This type will not be further elaborated on since it is not as applicable to this thesis. On the contrary, "RNNs process an input sequence one element at a time, maintaining in their hidden units a 'state vector' that implicitly contains information about the history of all the past elements of the sequence" (Lecun et al., 2015, p.441).

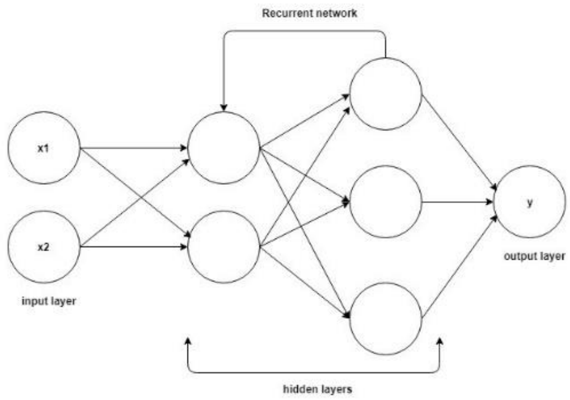

Figure 14: A schematic representation of a Recurrent Neural Network

Figure 14 shows a schematic representation of an RNN. The feedback loop that is incorporated inside the RNN is the key difference compared to a regular NN. The feedback loop allows the hidden layer(s) to not only consider the current input, but also the previous input(s). This is extremely helpful for use cases such as language modelling, machine translation, and speech recognition since it allows the machine to consider the context of the whole input sequence rather than that of a single input. Unfortunately, because of the backpropagated gradients, the weights inside the RNNs tend to explode or vanish over time, referred to as the Vanishing Gradient problem (Lecun et al., 2015). Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) have therefore been developed to overcome this problem.

Appendix B: Qualitative research interviews

Conducting research interviews is a powerful and commonly used method to collect qualitative data in exploratory studies (DiCicco-Bloom & Crabtree, 2006; Qu & Dumay, 2011). To enhance the quality of the collected data and decrease the risk of the researcher's subjectivity, the process of conducting research interviews consists of six stages as described by Rabionet (2011).

Stage 1: Selecting the type of interview.

The degree of structure is a common characteristic to distinguish between types of interviews (Fontana & Frey, 1994). The three types of interviews defined by this characteristic are: structured, unstructured, and semi-structured (Qu & Dumay, 2011). A brief overview of the three types is shown in Table 20. The table shows a description of the interview type and advantage(s).

	Description	Advantage
Structured	The interviewer asks the interviewee a series of pre-established questions, allowing for a limited number of responses. All interviewees are asked the same questions in the same order.	Minimizing researcher bias and increasing generalizability of the findings (Qu & Dumay, 2011). Data can easily be categorized and qualified as quantitative data. Can be used for hypothesis testing (DiCicco-Bloom & Crabtree, 2006).
Unstructured	There are no questions prepared beforehand by the researcher. The interview follows a more informal free-flowing conversation. All interviewees can be asked different questions.	New insights can be gathered that follow from a more personal reasoning. Each interview is structured to a personalized approach applicable to the interviewee.
Semi-structured	Organized around a set of open-ended questions that suffice to stimulate a conversation between the interviewer and interviewee.	Allow for the comparison of answers amongst candidates, whilst providing the opportunity to delve deeper into specific topics.

Table 20: An overview of the interview typology showing the description and advantages

Stage 2: Establishing ethical guidelines.

The second stage entails the establishment of ethical guidelines necessary whilst conducting interviews. Even though the subject of this research is not specifically on emotional or sensitive experiences, issues on confidentiality, consent, and purpose must be considered. DiCicco-Bloom & Crabtree (2006, p.319) state four ethical issues related to the interview process:

1. Reducing the risk of unanticipated harm;
2. Protecting the interviewee's information;
3. Effectively informing interviewees about the nature of the study;
4. Reducing the risk of exploitation

These four ethical issues can be mitigated by clearly setting the context at the start of the interview, providing a debriefing at the end of the interview, and anonymizing the obtained results. All in all, it is of utmost importance that the interviewer is honest, transparent, and thorough to the interviewee(s).

Stage 3: Crafting the interview protocol.

The interview protocol is made up of three stages: the introduction, the interview questions, and the debriefing afterwards. The introduction must set the context, explain the purpose of the interview, include statements of confidentiality, consent, and ask whether the interviewee has any remaining questions (Qu & Dumay, 2011; Rabionet, 2011). The introduction is an important stage since it serves to build trust among the interviewee and interviewer and explain the purpose of the interview. This will stimulate the interviewee to speak freely and truthfully about the subject (Mellon, 1990).

As described by Kvale (1994) there is a broad range of interview questions that can be used to stimulate the flow of the interview. For example, introducing questions help to start the conversation and suffice to stimulate the collection of contextual information, direct questions serve to obtain direct information about the subject, and probing questions are used to extract a more detailed explanation. By carefully setting up the interview questions, the flow of the interviewee's story can be ensured, the relationship between the interviewer and interviewee is preserved, and researcher's bias is avoided (Schensul et al., 1999). Because of the many variables related to the interview protocol, constructing one is an iterative process and "… often results in altering questions as the investigators learn more about the subject" (DiCicco-Bloom & Crabtree, 2006, p.316).

Finally, the debriefing afterwards must explain the participants on the consecutive steps and data handling afterwards. This can include: verification of transcribed interviews, publication of results, and anonymization.

Stage 4: Conducting and recording the interview.

Stage 4 describes the technical means that are used to conduct the interviews. Relevant aspects include the way of recording or note taking, transcription software, and the data analysis software (DiCicco-Bloom & Crabtree, 2006). Throughout this research, the most often applied means to conduct the interviews is Microsoft Teams (https://www.microsoft.com/nl-nl/microsoft-teams/group-chat-software/). Also, the same application is used to record the interview. The recording contains video and audio and is saved as an mp4 file. After the interview has been conducted, the recording is transcribed with the use of Otter (https://otter.ai). Thereafter, the transcribed interview is inspected by the researcher and sent to the interviewee for confirmation. Finally, the transcribed interview is analyzed with the help of Atlas.ti (https://atlasti.com/).

Stage 5: Analyzing and summarizing the interview.

As stated in the previous paragraph, the analysis of the qualitative data is done with the help of ATLAS.ti (https://atlasti.com/). The software supports the researcher in analyzing hidden complex phenomena in unstructured data by providing tools to (color) code, annotate, and locate findings. After having located relevant sections of information, similar sections are grouped together such that the process of deriving conclusion is facilitated.

Appendix C: Interview protocol 1

AI is broadly being used to automate processes (marketing and robotics), stimulate engagement (chatbots and voice assistants), provide new insights for decision making (big data analysis), and is even being used to enhance innovation process for new products (drug-discovery). Because of the wide range of applications, many organizations wish to implement AI-based solutions. Nonetheless, an important factor throughout any project is the question of: when is the project considered to be successful?

Since AI-based solutions can impact individuals and society to a great extent, the answer to this question is assumed to be highly dependent on one's relationship to the technology. Therefore, I would like to conduct this semi-structured interview that is set up to explore your perspective as a stakeholder and your point of view towards a successful implementation.

Before we start the interview, I have two remaining questions: do you agree with me recording the interview? And do you have any remaining questions yourself? Please remember that there is no right or wrong answer, I am solely interested in the differences among stakeholder perspectives.

1. What is your role during AI related implementation projects?
   a. Could you give an example of the activities you do during a project?
   b. What are your responsibilities during these projects?
2. When would you consider an AI implementation project to be successful?
3. What would it take for a project to be defined as a failure?
4. How does your organization currently determine whether an implementation project has been successful?
   a. Is this a standard set of criteria or do these vary by project?
   b. At which stage of the project are the criteria normally defined?
   c. Have the initial criteria changed during any of the projects?
5. Has your organization experienced a failing project in the past?
6. Have you experienced any reoccurring problems during implementation projects?

Thank you for your participation in my research. I will start transcribing the interview we have had today after this meeting. Once the transcription is finished, I will send it to you via e-mail for confirmation. Also, I will anonymize this interview such that your name, or any personal information, will be protected.

Appendix D: Interview protocol 2

During this research, a number of boundary conditions for the successful implementation of chatbots in organizations have been established. Boundary conditions are defined as: “the socio-technical constraints that must be satisfied to successfully complete an implementation”. Throughout this interview, I would like to discuss the aspects that the boundary conditions cover regarding the case at hand. The goal of this interview is to determine whether the established list of boundary conditions is applicable to the case. Therefore, I would like to hear your thoughts on: if the boundary condition were applicable to the case, how it was applicable, and to what extent it has had an impact on the success of the case.

Preparation before the interview

I would like to ask you to indicate if the boundary condition is applicable to the case before the interview. This will help to stimulate the discussion during the interview, and it will clearly show the number of applicable boundary conditions.

During the interview

The interview will consist of three 10-minute rounds in which we will discuss the three different types of boundary conditions (Technology, Organization, Environment). During each discussion round we will speak about:

- Why the boundary conditions are (not) applicable;
- If applicable, how the boundary conditions are present in the case by means of examples;
- If applicable, how the boundary conditions have had an influence on the outcome of the case;
- If there is a relation among the stated boundary conditions;
- If there is a hierarchy among the stated boundary conditions;

In the end, we will briefly discuss the effects of the chatbot implementation for the case at hand. We will discuss how the implementation has influenced the organization and its consumers as well as if anything would be done differently if the implementation were repeated. Also, I will ask you if the list of BCs does not address an important aspect related to chatbot implementation.

The list of boundary conditions is presented on the next page…

Type of boundary condition	Boundary condition	Applicable: R/S
Technology	High quality data must be available	
	Input and output data must be secure	
	There must be tool availability	
	The chatbot must be scalable	
	There must be a secure and reliable IT infrastructure	
	Data exchange amongst data sources must be possible	
	The right chatbot type and capabilities must be determined	
	The response generation must be explainable	
	The responses must be accurate	
	Human override must be possible by the user	
Organization	There must be top management support	
	There must be available resources	
	There must be internal understanding of the technology	
	The chatbot must fit the organization's strategy	
	There must be technological readiness inside the organization	
	There must be internal acceptance inside the organization	
	There must be internal incentive for the implementation	
	Internal social effects must be evaluated and mitigated if necessary	
	Added business value must be evaluated	
	Project goals must be set	
	There must be alignment amongst stakeholders inside the organization	
Environment	The UI must be understandable and natural to use (language, visuals, sound)	
	There be customer readiness	
	There must be chatbot provider commitment	
	There must be high customer satisfaction	
	The chatbot's use must be ethical	
	The organization must execute regular chatbot evaluation	
	There must be a perceived advantage (internal and external)	
	The chatbot must be in line with local cultural norms and values	
	The chatbot must be providing advantages to the user	
	There must be social acceptance outside the organization	
	There must be perceived usefulness by the user	
	There must be high ease of use for the user	
	The chatbot must not exhibit bias	
	The chatbot complies to local regulations (data privacy, confidentiality, security)	