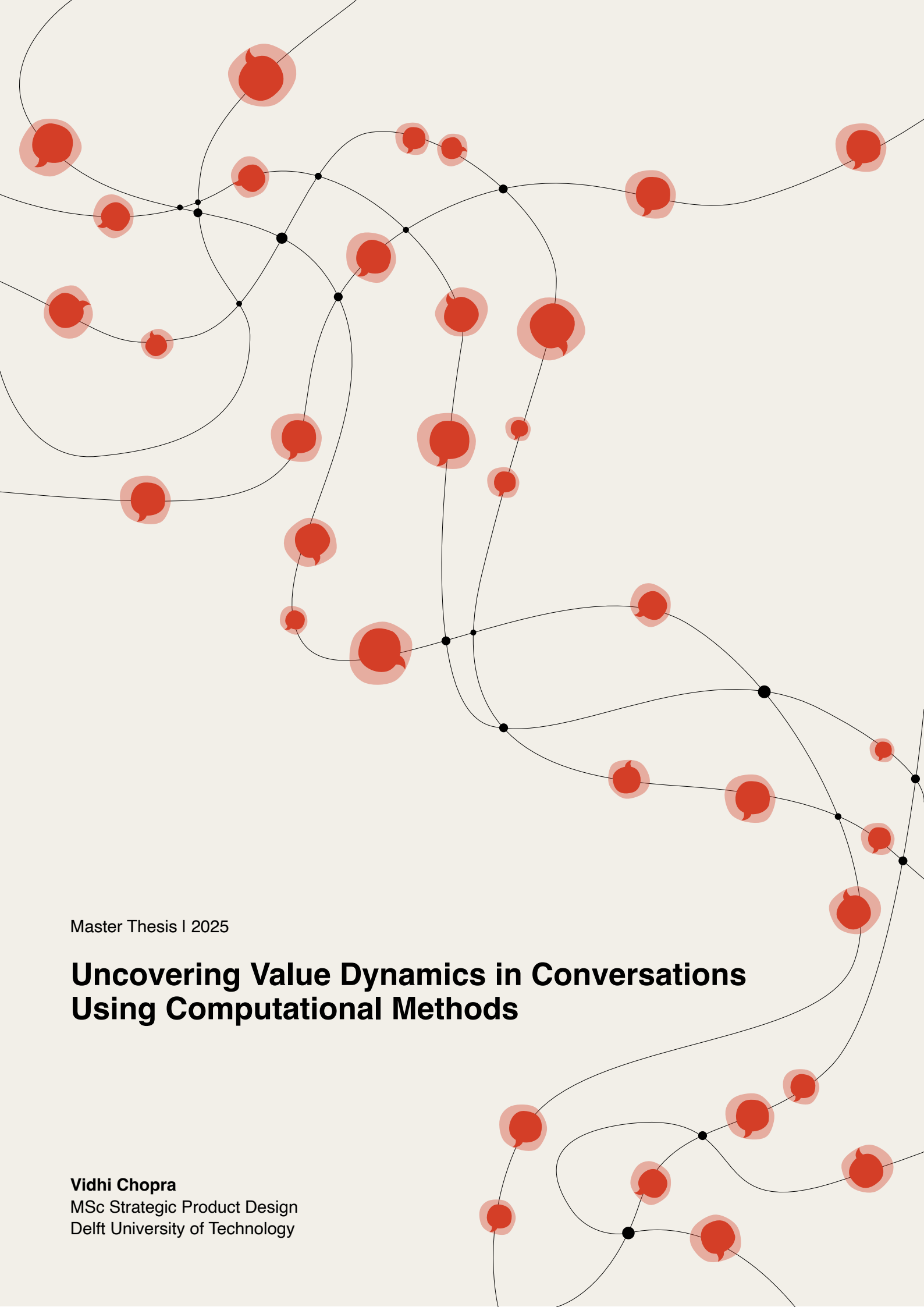


An abstract graphic featuring numerous red speech bubbles of varying sizes, each with a small tail pointing towards the bottom right. These bubbles are interconnected by a network of thin, black, curved lines that create a complex, web-like structure across the page. The background is a light, neutral color.

Master Thesis | 2025

Uncovering Value Dynamics in Conversations Using Computational Methods

Vidhi Chopra
MSc Strategic Product Design
Delft University of Technology

Uncovering Value Dynamics in Conversations Using Computational Methods

Master Thesis | Uncovering Value Dynamics in Conversations Using Computational Methods

Author

Vidhi Chopra

MSc Strategic Product Design
Faculty of Industrial Design Engineering
Delft University of Technology (TU Delft)

Supervisory Team

Chair, TU Delft | Dr. Marina Bos-de Vos
Mentor, TU Delft | Dr. Senthil Chandrasegaran
Department of Design, Organization and Strategy (DOS)

24th July | 2025

Acknowledgement

This thesis marks the end of a long journey. It wouldn't have been possible without the generosity, encouragement, and belief of the many people around me.

To my supervisors, Marina and Senthil — thank you for believing in me every step of the way. Your guidance helped shape not just this thesis but the way I now think about values and computational methods. Thank you for giving me the freedom, for creating space to explore the unknown, and for sharing the excitement and curiosity throughout the process. I'm especially grateful for your continued support beyond the scope of this project — for introducing me to Eric and Shreyan, whose thoughtful input and perspective added great value along the way.

This project would not have been possible without the trust of the people at INSEAD. Kirsten, Bibi, Arjan, and Ries — thank you for making the early phase of this research happen. Your encouragement gave this work a meaningful stage and helped it grow in directions I couldn't have imagined at the start.

To my parents and brother — thank you for standing by me and supporting me. For reminding me that this was just one moment in a much longer journey.

To Ishita and Anmol, for checking in from across countries and time zones.

To Alek, thank you for stepping in every time I fell short — for motivating me and for being there through the highs and lows. To Gayathri, Lina, Viktoria, Komal, and Sejal — thank you for walking this journey alongside me. To Heimin, for sharing the milestones and to Ben, for reminding me it's going to be fine in the end.

To Sayak and Shubham, for being pillars of support and to Prithvi, Jabez, and Anuj, for staying close and showing up.

A heartfelt thank you to the new friends I made through this research — Juwe, Renee, and Angela — thank you for hyping this project at every stage. To Amy, for being a great collaboration partner during the project.

To everyone who believed in me — especially in the quiet, meaningful ways, and especially when I didn't believe in myself — thank you. This thesis may have my name on it, but it was shaped by all of you.

Summary

This thesis explores how Natural Language Processing (NLP) can be used to uncover explicit and implicit value dynamics in conversations — especially in high-stakes, multi-stakeholder contexts where what is said and what is meant often diverge. The research takes a designed approach to a computational challenge, situated at the intersection of conversation as data and NLP as a method, to surface underlying value shifts that influence decision-making but often remain hidden.

Grounded in literature review across design, linguistics, and AI, the project frames the challenge around a conceptual gap — where individuals may profess certain values but express others in practice. This disconnect is termed the Value Expression Gap, which becomes both a framing concept and design target for the project. The approach follows a Research through Design (RtD) methodology, where iterative prototyping, deployment, observation, and refinement enable the method to evolve alongside insights.

In the prototyping phase, a computational tool was built using sentence embeddings and cosine similarity to assign values to spoken utterances. The method was then deployed at a global leadership conference hosted by INSEAD, where conversations were captured and analyzed for value patterns. This real-world testing not only confirmed feasibility but also revealed audience curiosity around value contradictions and alignment.

Insights from deployment prompted a grounding phase, where participatory workshops were observed to understand how people actually talk about values — including expressions of aspiration, discomfort, and uncertainty. These nuances were then used to refine the model with more diverse value utterances, layered meaning types, and a logic that allows for multiple value matches per statement.

Some insights like co-occurrence and tentativeness were included; others like silence, hedging, and conflict are proposed for future exploration.

The final phase focused on application, using semi-structured interviews with professionals across industries to assess real-world relevance. Seven key use cases emerged — two focused on external stakeholders (e.g. tailoring negotiation strategy, assessing stakeholder fit) and five focused on internal alignment (e.g. value-based leadership, cultural shifts, opportunity spotting). Interviewees emphasized the importance of transparency, organizational readiness, and ethical use for future adoption.

Key Contributions

- A New Lens on Dialogue: Offers a novel way to support reflection by revealing which values are emphasized, ignored, or shifting during conversation.
- The Value Expression Gap: Identifies the disconnect between professed and practiced values, contributing to both design theory and AI interpretability.
- Making Values Visible: Surfaces latent patterns, blind spots, and tensions that shape alignment and group dynamics.
- AI in Decision-Making: Positions NLP as a transparent, assistive lens — supporting sensemaking rather than prediction.
- Designers’ Role: Reframes designers as facilitators of ethical AI use, ensuring that value-driven tools remain grounded in human meaning-making.

Rather than presenting a final answer, this work opens up a new direction — one where designers and AI co-create reflective tools for dialogue and decision-making.

Acronyms and Glossary

AI	Artificial Intelligence
DL	Deep Learning
GPT	Generative Pretrained Transformer Model
LLM	Large Language Model
NLP	Natural Language Processing
ML	Machine Learning
RtD	Research through Design

Cosine Similarity	A mathematical method for measuring the similarity between two vectors (often representing text). It is used in natural language processing to assess how similar two pieces of text are in terms of their meaning.
Deliberation	Deliberation is a process of meaningful conversation and reasoning through which people reflect on diverse perspectives, often leading to new understanding, value shifts, or collective decision-making.
Example Utterance	A sample sentence crafted to show how a specific value might be expressed in conversation. These are used to “teach” the model what that value could sound like in real speech.
Token	A token is the smallest unit of text a language model reads — which could be a word, part of a word, or punctuation, depending on the tokenizer.
Transformer architectures	A deep learning model that uses self-attention to understand relationships between words in a sentence, enabling high-precision language understanding.
Utterance	A single spoken sentence or phrase from a conversation that is being analyzed by the model to detect underlying values.
Value Expression Gap	The mismatch between participants’ self-reported value priorities and the values actually expressed during conversation. It highlights the difference between stated intentions and observable dialogue.

Contents

Acknowledgement	5	
Summary	6	
Acronyms and Glossary	7	
1. Introduction	10	
1.1. Project Context	12	
1.2. Problem Framing	12	
1.3. Research Questions	13	
1.4. Approach Overview	14	
2. Methodology	16	
2.1. Research Techniques	18	
2.2. Theoretical Foundation	19	
2.3. Research through Design (RtD)	20	
2.4. Decision-Maker Interviews	24	
3. Theoretical Foundations	26	
3.1. Understanding Values	28	
3.2. Value in Design and Decision-Making	28	
3.3. Value Shifts and Complex Contexts	29	
3.4. Conversations as a Lens	29	
3.4.1. Value Engagement in Dialogue	30	
3.4.2. Why NLP? A Methodological Rationale	31	
3.5. Value Framework Used	33	
4. Developing the Prototype	36	
4.1. From Theory to Prototyping	38	
4.2. Context for Testing: The Forum	38	
4.3. Method Logic	39	
4.4. Assumptions and Starting Points	45	
4.5. Early Decisions for Advancement	46	
5. Deployment and Evaluation	50	
5.1. Pre-Survey for Value Prioritization	52	
5.2. Scoping Meaning from Elicited Values	52	
5.3. The Value Expression Gap	56	
5.4. Learnings to Next Steps	57	
6. Observing Value-in-Use	60	
6.1. Design of Exploratory Workshops	62	
6.2. How People Talk About Values	62	
6.3. From Insight to Design Focus	65	
7. Refining and Reflection	66	
7.1. Refining the Method	68	
7.1.1. Expanding the Value Dictionary	68	
7.1.2. Evaluating Value Elicitation Strategies	71	
7.1.3. Adding a Motivational Layer	77	
7.1.4. Adopting a Soft Assignment Approach	78	
7.2. Demonstrating Results	79	
7.3. Reflection on the Method	87	
8. Real-World Relevance	90	
8.1. From Development to Deployment	92	
8.2. Strategic Use Cases and Value Potential	95	
8.3. Integration, Adoption and Ethical Fit	102	
9. Discussion	106	
9.1. Reflection on Research Questions	108	
9.2. Limitations & Future Work	111	
9.3. Contributions	112	
Use of Artificial Intelligence	114	
References	115	
Appendix		

Chapter 01

Introduction

This chapter introduces the core premise of the thesis — that values shape decisions but are often left unspoken in collaborative settings. It presents the central research question and outlines the Research through Design (RtD) approach, positioning conversation as the lens and NLP as the method for uncovering value dynamics.

- 1.1. Project Context
- 1.2. Problem Framing
- 1.3. Research Questions
- 1.4. Approach Overview

1.1. Project Context

Design is rarely straightforward. At the heart of it lies decision-making — shaped by the values of those involved in and impacted by the process. These values shape how decisions are made and how strategies unfold. They influence what we notice, what we prioritize, and what we’re willing to trade off.

“Values are like fingerprints. Nobody’s are the same, but you leave them all over everything you do.”

- Elvis Presley

While values shape decisions, design is in turn expected to create value across multiple dimensions — for users, stakeholders, systems, and futures who may hold diverging priorities (Boradkar, 2010). But what constitutes ‘value’ is often not shared, and in collaborative settings, underlying ideals and motivations frequently remain unspoken. Collaborating actors bring personal and professional values to the table that must somehow be reconciled (Bergema, Kleinsmann, & Valkenburg, 2011).

Yet, in many design processes, values remain invisible. Actors either avoid articulating the values at play or focus too narrowly on familiar ones, overlooking others that may be just as important (Van Onselen & Valkenburg, 2015). For instance, within an organization, leadership might push for rapid scalability and competitive advantage, while on-the-ground teams prioritize sustainability or long-term social impact.

These conflicting values — especially when people themselves are unaware of their own values — can lead to internal misalignment, confusion in decision-making, or missed opportunities for coherence. As Bos-de Vos (2018) highlights, many such difficulties can be traced back to values that were never made explicit or openly discussed.

Kenter et al. (2016a) highlights how through conversation, individuals engage in a values-based exchange of perspectives—reflecting, challenging, and sometimes reconfiguring what they consider important. In multi-stakeholder design contexts, where priorities must be negotiated and meaning co-constructed, these dialogues often serve as the backbone of decision-making.

This research begins at such a crossroads—with a belief that conversations matter, and a question: Can we detect values as they appear in natural conversation, even when they aren’t directly named — or consciously known by the speaker?

The journey started with the exploratory possibility of computationally identifying values in spoken utterances. But as the process unfolded, it revealed much more and a new potential for using conversation analysis to support deeper reflection and more intentional decision-making.

This research does not aim to produce a definitive tool. Instead, it positions conversation as a critical lens, and computational methods—particularly natural language processing (NLP)—as a complementary means for surfacing what values are expressed and being implied. It builds on existing efforts of exploring ways to bring implicit values to the surface in co-creative settings.

As Schwartz (2016) reminds us, values may be implicit—but when made visible, they become stronger drivers of action. Analysing conversation may not solve every challenge, especially when people themselves are unaware of what guides their decisions. But it’s a powerful step toward mapping value dynamics as they unfold—and creating space for more reflective, inclusive, and value-aware decisions.

1.2. Problem Framing

As introduced earlier, this research starts from the premise that values are deeply embedded in collaborative decision-making discussions — but they are not always made explicit.

1.3. Research Questions

To explore whether computational methods can help surface values in conversation in meaningful ways, this research was guided by three sub-questions — illustrated in the Figure 01 below.

First, it focused on how NLP techniques could be prototyped to detect both explicit and implicit values. Then, it examined what kinds of insights these values could offer. Finally, it considered how such insights might support real-world decision-making.

Together, these threads help answer the core research question:

Can computational methods be used to uncover and analyze value dynamics in conversation in ways that are meaningful for decision makers?

While frameworks exist to help practitioners reflect on what they believe or prioritize, there is still limited understanding of how values are actually expressed and interpreted during the course of real conversations.

Conversations are where much of design and strategy work happens. They are spaces where trade-offs are debated, priorities emerge, and value judgments unfold — often without being directly named. But when values stay implicit, several problems can arise. Teams might move forward on a decision without realizing they hold conflicting priorities — for example, one group optimizing for speed, while another quietly prioritizes long-term sustainability. Or a stakeholder might feel unheard, not because they weren’t given time to speak, but because the value driving their concern — like inclusivity — was never articulated in a way others recognized.

These hidden dynamics can lead to misalignment, slowdowns, or decisions that appear aligned on the surface but carry unresolved tensions underneath. This creates an opportunity to explore how computational methods might offer new ways to surface those underlying value dynamics — not as replacements for human dialogue, but as tools to enhance awareness, prompt reflection, and improve alignment across complex teams.

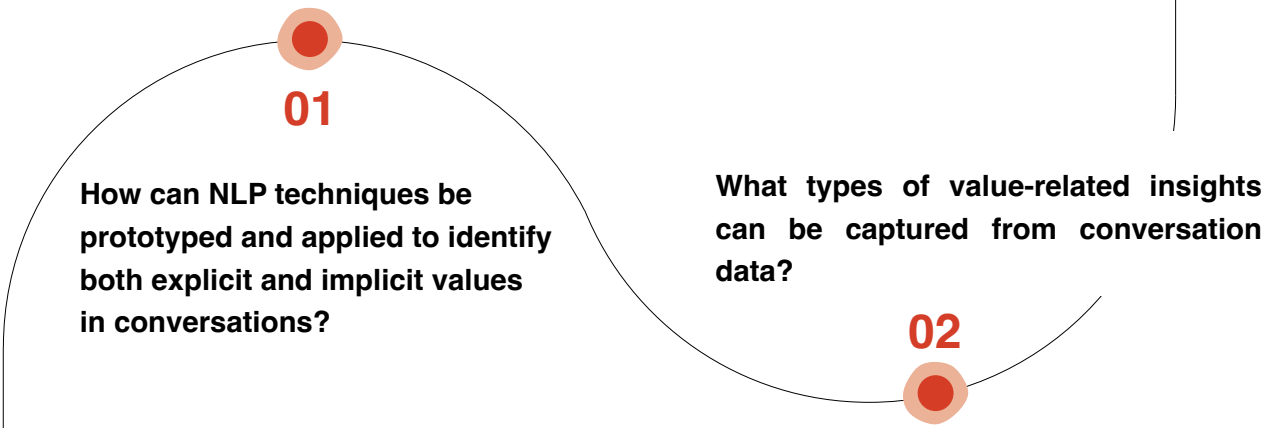


Figure 01: Showing sub-research questions that structure the logical flow to answer the main research question.

1.4. Approach Overview

This project follows a Research through Design (RtD) methodology — an approach where designing becomes a way of researching (Stappers and Giaccardi, 2014). Rather than starting with a fixed solution, the process unfolds through iterative making, deploying, observing, and refining. The prototype is not just a tool but a lens — used to both test technical possibilities and surface new insights about how values are expressed and understood in conversation.

The project unfolds across three overarching spaces (see Figure 02). It begins in the problem space, where the challenge is defined through literature and theoretical grounding. From there, it moves into the solution space, where a prototype is developed, deployed in a real-world context, grounded through observational workshops, and refined for greater accuracy and interpretability. Finally, the process transitions into an opportunity space, where the potential applications, impact, and future possibilities of the method are explored. This layered progression ensures that the research is not only responsive to the problem at hand, but also open to broader relevance and real-world use.

The overall journey is mapped through six interconnected stages (also see Figure 02):

Framing: The project begins by grounding the challenge in literature on values, design, and collaborative decision-making. This phase establishes why both explicit and implicit values in conversation matter — and what could be gained from surfacing them.

Prototyping: A low-fidelity prototype is built to explore the feasibility of computational value elicitation using NLP. This early version focuses on assigning values to utterances using a small dataset and basic logic.

Deploying: The prototype is tested in a real-world setting — a leadership conference — to observe not just its feasibility, but also its desirability by introducing it to an engaged and relevant audience.

Grounding: Informed by learnings from the deployment phase, it became important to better understand how

values are actually expressed in real conversations. This led to observing participatory workshops, where values were surfaced, discussed, and debated in natural, multi-stakeholder settings.

Refining: Insights from deployment and workshop observations fed into refining the method. The model was expanded with a broader dataset, incorporated more diverse expression styles, improved matching logic, and included an additional motivational layer to enhance both accuracy and interpretability.

Applying: In the final step, stakeholder interviews and reflective synthesis explored how such a method could be applied in organizational decision-making contexts. This phase — while assessing relevance and adoption potential — also revealed key considerations needed to bring this research closer to real-world implementation.

Each stage informed the next, forming an evolving process where technical and conceptual development progressed hand in hand. This iterative approach was essential to make sure the method is feasible but also desirable and viable.

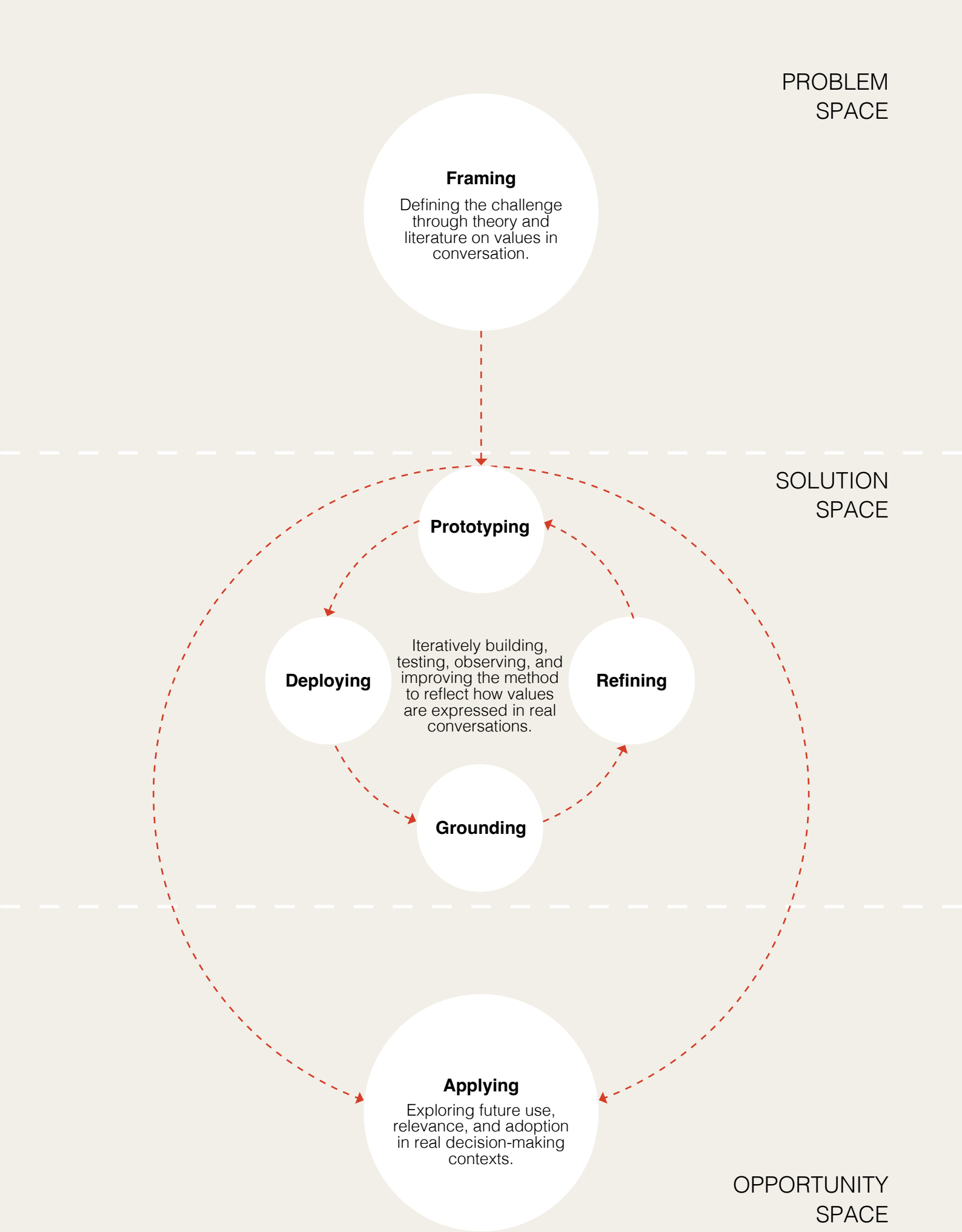


Figure 02: Showing the 6 interconnected stages of the project approach across problem, solution, and opportunity spaces.

Chapter 02

Methodology

This chapter outlines the research approach, grounded in a Research through Design (RtD) methodology that enabled iterative learning through making and testing. It introduces the three core research tracks — theoretical foundation, RtD cycles, and expert interviews — and explains how they collectively shaped the prototype and guided the inquiry.

- 2.1. **Research Techniques**
- 2.2. **Theoretical Foundation**
- 2.3. **Research through Design (RtD)**
- 2.4. **Decision-Maker Interviews**

2.1. Research Techniques

To build on the project approach outlined earlier, this section provides a clear overview of the research techniques used throughout the project — detailing how each method contributed to shaping the prototype and uncovering insights. While the project followed a Research through Design (RtD) methodology — where new questions and directions emerged through cycles of making and reflecting — the methods themselves were carefully chosen and sequenced to support different phases of the inquiry.

The research unfolded across three main technique clusters:

- a **theoretical foundation** rooted in literature and frameworks;
- an **RtD** cycle consisting of four phases — prototyping, deployment, workshop observation, and refinement;
- and **decision-maker interviews** to evaluate contextual relevance and potential application.

Together, these methods helped explore whether values could be computationally elicited from conversation, and how, in what form, and toward what ends. The Table 01 below concisely summarizes each technique’s purpose, source, and relation to the overall process.

Technique	Sub-Phase	Source	Purpose
Theoretical Foundation	Literature Review	Academic papers, frameworks and models	To frame the challenge space, clarify key concepts, and inform the design of the first prototype
Research through Design (RtD)	Prototyping	Self-coded prototype based on literature-derived value categories	To explore feasibility of value elicitation using NLP with a small-scale dataset
	INSEAD Alumni Forum Europe 2025 (Leadership Conference)	Pre-survey with Alumni, Conversations at dilemma sessions	To observe the prototype's performance in a real-world decision-making context and test desirability
	Workshop Observation	Workshops with TU Delft Dream Teams	To understand how values are expressed in natural dialogue
	Prototype Refinement	Based on results from conference and workshops	To improve model accuracy, diversify expression styles, and enhance interpretability
Decision-Maker Interviews	Semi-structured Interviews	Semi-structured interviews with five professionals across aviation, energy, consultancy, policy and AI	To explore practical relevance, adoption potential, and broader applications of the method

Table 01: Summarizing the techniques, sources, and purposes of each research phase within the overall methodological approach.

2.2. Theoretical Foundation

The literature review in this research was designed to strategically build a layered understanding of the opportunity space at the intersection of values and conversation. It begins with foundational readings on the nature of values — emphasizing their layered, contextual, and evolving nature. This foundation was critical for recognizing that values are rarely fixed, and are often left implicit in decision-making.

From there, literature on values in design and decision-making added an applied perspective: it showed that while values play a central role in shaping choices, they are frequently unspoken. This highlighted a gap — if values influence action but remain hidden, then surfacing them could support more aligned and intentional outcomes.

This conceptual base was further expanded by insights into value shifts in complex, co-creative environments. The literature revealed that values don't just sit still — they evolve through dialogue, negotiation, and interaction. This insight brought conversations into focus as a dynamic site for observing how values are revealed, adapted, or left unspoken in real time.

This made it possible to frame conversation as data — not just as a medium for decisions, but as a rich source of value expression. Literature on value engagement in dialogue reinforced this, positioning conversation as both a mirror and mechanism for value dynamics

Once this framing was in place, the methodological rationale for NLP became clear. If conversation can be treated as data, then NLP provides tools to explore patterns, detect implicit signals, and scale the analysis — without replacing human interpretation. It offers a way to explore subtle and evolving value cues computationally.

To support this method, a structured value framework was selected. It provided consistent categories for detection and interpretation, helping bridge conceptual insights and computational feasibility.

Together, this body of literature informed both the direction and methodology of the project — enabling it to sit at the intersection of conversation as data and NLP as an approach, where the research opportunity emerged (See Figure 03).

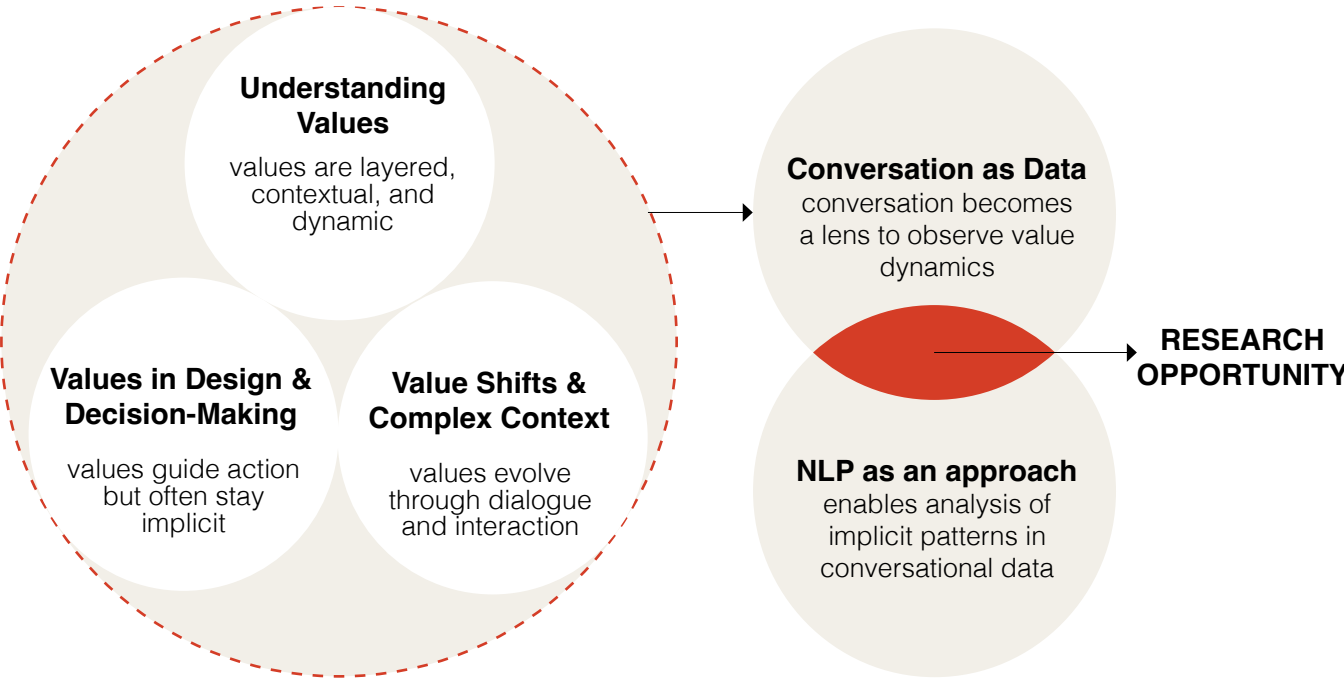


Figure 03: This body of literature informed both the direction and methodology of the project — enabling it to sit at the intersection of conversation as data and NLP as an approach, where the research opportunity emerged.

2.3. Research through Design (RtD)

While the literature review helped define the research opportunity — at the intersection of conversation as data and NLP as an approach — it was now time to start exploring. This project adopted a Research through Design (RtD) approach to do just that.

RtD offered a way to explore a new and relatively uncharted domain: using computational methods to surface values from natural conversation. Given the layered, contextual, and often implicit nature of value expression, the topic could not be fully addressed through traditional research methods alone. RtD allowed the research to unfold iteratively — by designing, deploying, and refining a working method in real-world settings while generating insights throughout the process.

The following subsections outline the four key phases that structured this journey: Prototyping, Deployment, Observation, and Refinement — each building upon the last to evolve both the method and its purpose (See Figure 04).

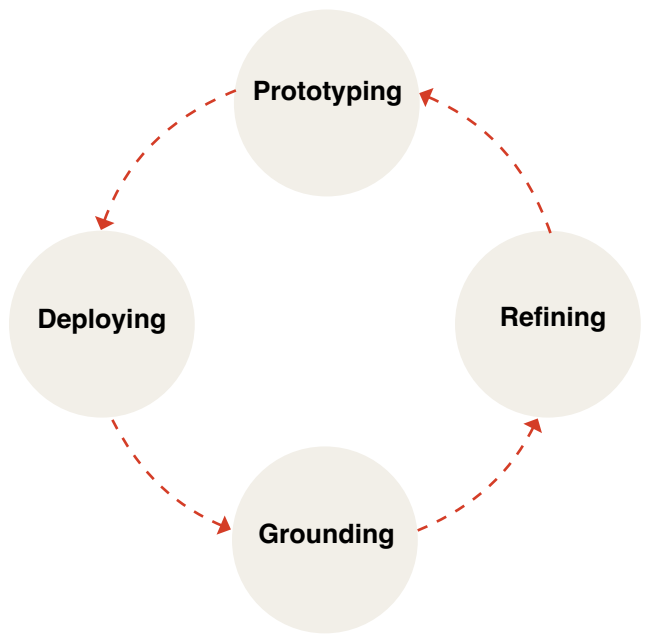


Figure 04: Illustrates the four iterative phases of the Research through Design journey — Prototyping, Deploying, Grounding, and Refining — each phase building upon the previous to evolve both the method and its purpose.

PROTOTYPING

To bring the value elicitation method to life, a stepwise prototyping process was followed — moving from environment setup to core logic, dataset creation, and refinement. As shown in Figure 05, the development began with defining the environment for prototyping, followed by establishing the method logic for value detection. A curated dataset of values and example utterances was then created to prompt the model, grounded in an existing theoretical framework. To enable early testing, several foundational assumptions were introduced, which ultimately led to early decisions like the pre-trained model selection based on observations during this phase.

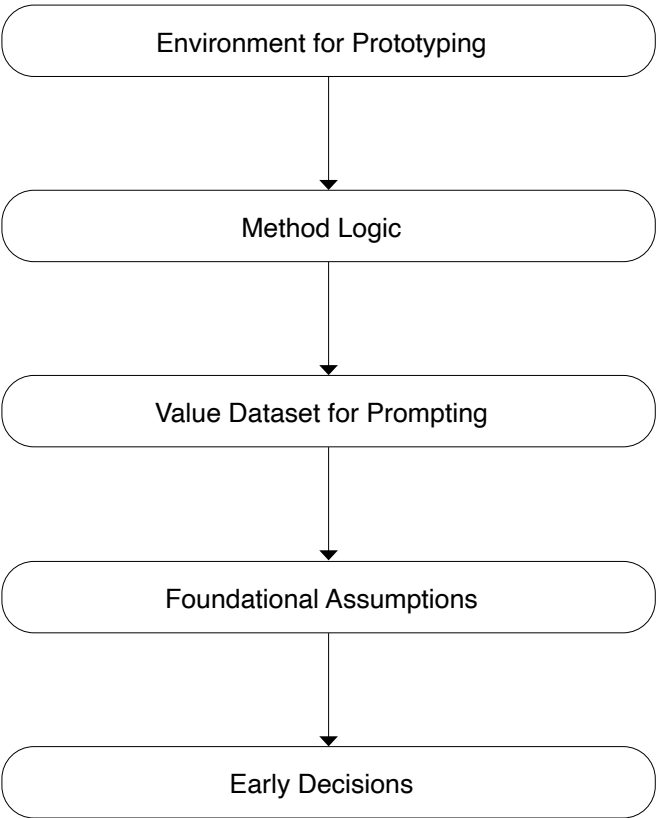


Figure 05: Depicts the stepwise process of method development during the prototyping phase — from setting up the environment and method logic to curating the value dataset, grounding assumptions, and making early design decisions.

Environment for Prototyping

The prototype was developed in Python within a fully local environment. This setup provided complete control over experimentation and ensured data privacy — an important consideration given the conversational nature of the input data. The local environment also supported rapid iteration during the early testing phase.

Method Logic

To computationally compare utterances with abstract value concepts, sentence embeddings were used in combination with cosine similarity. This approach is well-established in past research for its effectiveness in capturing semantic similarity across textual inputs (Reimers & Gurevych, 2019; Clark, 2015). In this method, both the utterances and example expressions for each value were encoded as vector embeddings, and their semantic proximity was measured to determine potential value matches.

Value Dataset for Prompting

A curated subset of 20 values was selected from a larger framework of 108 values (Bos-de Vos, 2020) to define the initial scope of the prototype. These values were chosen for their relevance to leadership, collaboration, and decision-making — aligning with the intended deployment context. To represent each value, four example utterances were created using OpenAI’s ChatGPT. This setup allowed controlled exploration of value detection using a manageable dataset during the early development phase.

Foundational Assumptions

To enable initial experimentation, the method relied on a few simplifying assumptions:

- Each utterance reflects one dominant value.
- Values are expressed clearly enough for NLP detection.
- High similarity scores indicate meaningful alignment between utterance and value expression.

These assumptions enabled the development of a baseline logic and supported early analysis.

Early Decisions and Model Selection

Based on iterative prototyping needs, the segmentation of utterances was adapted to focus on shorter units, better suited for capturing semantic nuance. The logic was expanded to register multiple value matches per utterance, increasing the interpretive range of the prototype. After evaluating different model options, the all-MiniLM-L6-v2 transformer was selected for its

balance of semantic performance and computational efficiency, supporting fast, scalable comparisons within the local development environment.

DEPLOYING

This phase focused on testing the prototype in a real-world decision-making context. The process began with a pre-survey to capture participants’ self-reported value priorities, followed by the deployment of the prototype at a leadership forum to computationally elicit values from actual conversations (See Figure 06). The two data sources — stated preferences and observed dialogue — were then compared, offering insights into alignment, divergence, and value patterns. These insights not only demonstrated the prototype’s interpretive potential but also highlighted its current limitations, informing the next steps in refinement.

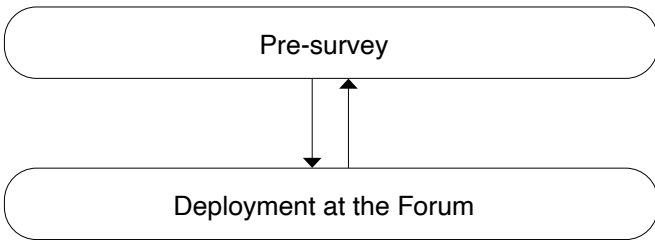


Figure 06: Shows the deployment process—capturing stated values via pre-survey and comparing them with values elicited from real conversations.

Context for Deployment

To evaluate the value elicitation method in a live setting, the prototype was deployed at the INSEAD Alumni Forum 2025, a leadership event designed to foster open dialogue on themes such as sustainability, innovation, governance, and leadership. The Forum brought together senior decision-makers from diverse sectors, making it a relevant context for testing how values emerge in real-time, multi-stakeholder conversations.

Pre-Survey for Value Prioritization

Before the deployment, a short pre-survey was distributed to 50 participants via Qualtrics. The aim was to establish a baseline understanding of self-reported value priorities in both general and context-specific decision scenarios. The value set used in the survey matched the prototype’s curated list of 20 values,

allowing for consistent comparison. The survey served not as a validation tool, but as a complementary lens for interpreting the elicited conversation data. Participation was anonymous and ethically approved; the full survey structure is provided in Appendix A.

Generating Deployment Insights

During the event, the prototype was used to analyze transcripts of spoken interactions. Three types of insights were generated from the output:

- Value Frequency: to observe dominant and missing values.
- Value Type Emphasis: to understand broader categories being prioritized.
- Value Co-occurrence: to detect patterns of values that appeared together.

These insights were synthesized and shared with the Forum organizers, who selected specific outputs to be visualized in a short closing ceremony video.

The Value Expression Gap

Post-deployment, the elicited values were compared to pre-survey results. This contrast revealed a Value Expression Gap — highlighting mismatches between what participants said they prioritized and what was actually expressed in conversation. This step helped frame the deployment as an inquiry into the difference between professed and practiced values, offering a new angle for interpretation and reflection.

Informing the Next Steps

These learnings informed both technical adjustments and the conceptual direction for the next phase — revealing that, before refining the method, it was essential to dig deeper into how people actually talk about values in practice.

GROUNDING

The final research phase focused on grounding the method in how values are expressed in practice. While the deployment phase highlighted technical and conceptual limitations, this phase deepened the understanding of value expression by observing it in live, collaborative settings.

Three exploratory workshops were conducted to simulate multi-stakeholder dialogue and surface values in action:

- **Two workshops** with Dream Team members from different project teams at TU Delft
- **One workshop** as part of a Climate Fresk session, aligned with the global Climate Fresk movement

The workshops were co-designed and facilitated in collaboration with another graduation student at TU Delft as part of a research initiative focused on creating “brave spaces” — environments where participants feel psychologically safe to express personal and professional values. Each session offered participants an opportunity to engage in open conversation, roleplay, and reflective exercises on values in team settings. The three-part structure included: identifying and exploring a personal value, responding to conflicting perspectives, and reframing value tensions. This phase informed two key directions (see Figure 07): different ways in which people naturally express values, and what additional interpretive layers could improve the method.

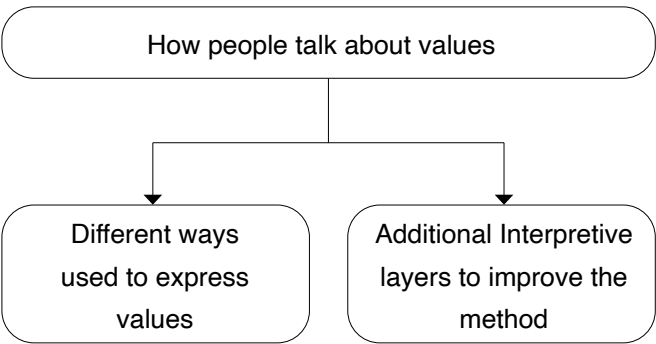


Figure 07: Captures two key takeaways from observing real conversations — the variety of ways people express values and the need for added interpretive layers to enhance method accuracy.

Patterns of Value Expression

Participants did not articulate values in neat, labeled terms. Instead, values surfaced in ways that were:

- Indirect and Layered
- Difficult to Name Explicitly
- Sensitive to Role and Context
- Charged with Emotion, Hesitation, or Soft Signals

These patterns revealed that value expression is nuanced, contextual, and often implicit — challenging the assumptions of earlier value detection logic.

Implications for Method Refinement

These insights directly informed the refinement phase by prompting an expansion of the value example dataset to include more narrative, emotional, and implicit expressions — making the method more responsive to how values are actually communicated.

They also pointed to future development opportunities beyond the scope of this thesis. Adding interpretive layers such as role-awareness, tension mapping, or uncertainty detection could help the method account for the complex, evolving nature of human value expression.

REFINEMENT

The final phase of RtD focused on refining the prototype to improve its semantic coverage, conversational relevance, and analytical robustness. The targeted enhancements were made based on prior observations from the deployment and workshop insights. Figure 08 below illustrates the flow of the refinement phase.

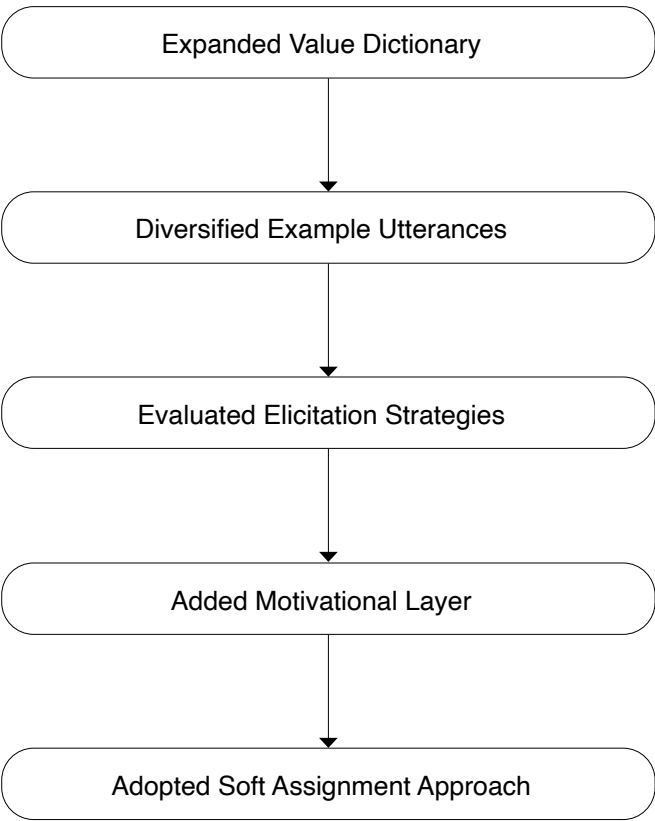


Figure 08: Outlines the sequential steps taken during the refinement phase — each aimed at enhancing the prototype's semantic, conversational, and analytical depth.

Expanded Value Dictionary

The full set of 108 values from the Bos-de Vos (2020) framework was incorporated to improve representational breadth. Each value entry was structured with attributes such as value type, motivational goal, and descriptive narrative.

Diversified Example Utterances

For each of the 108 values, eight example utterances were generated using ChatGPT — two per expression type (stories, expressive phrases, implicit, explicit), grounded in workshop findings. This aimed to mirror natural variations in how values are expressed in real dialogue.

Evaluated Elicitation Strategies

Three similarity-based scoring strategies were tested for assigning values to utterances: average vector, average similarity, and maximum similarity. The average similarity approach was selected for its balance of nuance, stability, and interpretability.

Added Motivational Layer

Each elicited value was tagged with a broader motivational goal (e.g., security, mastery, enjoyment), based on predefined mappings in the framework. This added a deeper interpretive layer to the output.

Adopted Soft Assignment Approach

The logic was updated to surface the top 5–10 most semantically aligned values per utterance, allowing for overlap and ambiguity. This replaced the earlier single-value assignment and enabled more open-ended analysis.

Together, these refinements strengthened the prototype's capacity to process varied conversational inputs and offer richer insight into the values being expressed.

2.4. Decision-Maker Interviews

To explore how the value elicitation method might extend beyond its development phase and into real-world decision-making contexts, five semi-structured online interviews (30–45 minutes each) were conducted with professionals across diverse sectors — including aviation, consulting, energy, AI, and digital governance (See Table 02). While each interviewee brought domain-specific expertise, all held strategic or leadership roles involving multi-stakeholder decision-making.

This diversity ensured a broad, non-sector-specific view of how the method might be received, challenged, or applied in practice. Each session began with a discussion on how values play a role in organizational decisions, followed by a prototype demonstration and open reflection. Thematic insights were derived from both recurring patterns and distinct perspectives on strategic relevance, integration, and ethical considerations. The interview guide is provided in Appendix B.

Interviewee	Company	Role	Length and Format
Interviewee 01	KLM (aviation industry)	Program Manager	45 min, online
Interviewee 02	the can do company (business innovation consultancy)	Director	30 min, online
Interviewee 03	Vattenfall (energy industry)	Global Procurement Head	45 min, online
Interviewee 04	Dehurdle (AI Coaching App)	Founder	30 min, online
Interviewee 05	Digital Governance Advisory (Digital Transformation)	Managing Partner	30 min, online

Table 02: Overview of decision-maker interviews, detailing sector, role, and session format to capture diverse, strategic perspectives on real-world relevance and application of the method.

Chapter 03

Theoretical Foundations

This chapter outlines the transition from theory to implementation, detailing the development of a low-fidelity prototype designed to computationally detect values in conversation. It introduces the technical setup, core logic using sentence embeddings and cosine similarity, and the early design decisions that shaped the model's interpretive capabilities.

- 3.1. Understanding Values
- 3.2. Value in Design and Decision-Making
- 3.3. Value Shifts and Complex Contexts
- 3.4. Conversations as a Lens
 - 3. 4.1.Value Engagement in Dialogue
 - 3. 4.2.Why NLP? A Methodological Rationale
- 3.5. Value Framework Used

3.1. Understanding Values

The term value is widely used in both everyday language and academic discourse, but it carries a variety of meanings that often depend on context. When discussing the concept of value, Friedman et al. (2013) highlight that the term might often be narrowly interpreted as economic worth, such as the price or utility of a product, but in reality, it encompasses much broader meanings.

One of the earliest academic attempts to define values more broadly comes from Kluckhohn (1951), who describes a value as ‘a conception, explicit or implicit, distinctive of an individual or characteristic of a group, of the desirable, which influences the selection from available modes, means, and ends of action’. This definition emphasizes that values are not just preferences or tastes, but deeply embedded conceptions that guide how individuals and groups make decisions and prioritize actions.

Despite this diversity in definitions, researchers broadly agree that values are not grounded in objective facts, but are shaped by what people care about, desire, and believe in (Friedman et al., 2013). To clarify this conceptual diversity, scholars have proposed a useful distinction: values can be approached from two core perspectives—values as guiding principles and values as qualities with worth (Martinsuo, Klakegg, & van Marrewijk, 2019; Bos-de Vos, 2020). The former refers to intangible beliefs and ideals, such as human or cultural values, that shape individual and collective behavior (Schwartz & Bilsky, 1987; Rokeach, 1973). The latter reflects perceived qualities assigned to objects, actions, or outcomes within specific contexts, such as economic, social, or ecological value (Boztepe, 2007; Boradkar, 2010; Den Ouden, 2012).

This dual perspective offers a strong foundation for understanding how values influence design and decision-making. At the same time, it is important to recognize that value is not a static concept—it is both situated and emergent, shaped by specific actors and conditions. Bos-de Vos (2020) emphasizes that designing for divergent values is a temporal and fragile process, continuously reshaped by

evolving perspectives and interactions. Similarly, De Wildt et al. (2021) highlight that the meaning and application of values can shift depending on organizational or societal contexts. Vargo et al. (2017) support this by framing value as multidimensional and co-constructed.

This understanding is central to this research. Since the project aims to computationally detect values in conversations, it must work with an understanding of value that is layered, contextual, and dynamic—able to account for not just what is said, but how meaning is constructed and shifts depending on use and interaction.

3.2. Value in Design and Decision-Making

Values are not passive concepts but active motivators of human behavior. Schwartz (2016) notes that the distinguishing feature of any value lies in the type of goal or motivation it expresses. As foundational psychological constructs, values guide how individuals set goals, make decisions, and interpret their surroundings (Moll et al., 2016; Rohan, 2000).

This motivational role extends into the realm of strategy and organizational decision-making. Rindova and Martins (2017), building on Weber’s notion of value-rational action, argue that values influence not just what decisions are made, but how strategists perceive opportunities, evaluate resources, design actions, and engage with stakeholders. They identify four strategic functions of values: as attentional structures, valuation lenses, design principles, and identity markers. Values can thus shape what is noticed, how it is judged, and how strategy is aligned and communicated—often providing stability and coherence in complex contexts.

Despite their strategic relevance, values are rarely made explicit in collaborative design settings. This lack of articulation may result in submerged conflicts, where actors assume shared goals while pursuing divergent value priorities (Van Onselen & Valkenburg, 2015). The design discipline has increasingly acknowledged the

importance of surfacing such latent values. Bos-de Vos (2020) emphasizes the need for teams to become more aware of and open about the values shaping their contributions, especially in co-creative environments. Similarly, Friedman et al. (2013) urge designers to consider who is affected by design choices, which values are implicated, and how to navigate trade-offs—e.g., between moral rights and aesthetic appeal.

Together, these studies underline the importance of uncovering, discussing, and prioritizing values within design and decision-making processes. Doing so not only improves alignment and collaboration but also ensures that the resulting strategies or products reflect a more thoughtful, ethical, and context-sensitive approach. This body of work provides a foundation for the present research by highlighting the need to make value-related dynamics more visible—especially as they unfold through everyday interactions such as conversation.

3.3. Value Shifts and Complex Contexts

While surfacing values is essential in design and decision-making, it is equally important to acknowledge that values are not fixed. In today’s complex and rapidly shifting environments, decision-makers face high degrees of uncertainty, interconnectedness, and change. Scarlett (2013) notes that under such conditions, managers and designers must navigate challenges related to information, communication, coordination, and action. In these fluid settings, values can be expected to adapt as people respond to new contexts, crises, collaborations, or competing priorities.

Therefore, social learning, as a mechanism that enables values to evolve at both individual and collective levels, becomes central to this research. Social learning is defined as a change in understanding that emerges through interaction and dialogue within a social context, such as a design team or stakeholder group (Reed et al., 2010). Kenter et al. (2016) describe it as a process of deliberation, where people reflect on their values, negotiate conflicting viewpoints, and revisit trade-offs through shared reasoning.

Through exposure to different perspectives, individuals and groups may revise what they consider important, and may shift how they define outcomes or success (Pahl-Wostl, 2006; Reed et al., 2010). Value change, then, is not simply a shift in preference—it is often a communicative and participatory process grounded in mutual understanding (Habermas, 1984). In such settings, value negotiation becomes a collective act, not just an internal shift.

Importantly, social learning does not only lead to value change—it can also enhance decision-making. Cundill and Rodela (2012) suggest that social learning improves decisions by increasing awareness of human–environment interactions and by strengthening the relational and problem-solving capacities of stakeholders. This is echoed by Brymer et al. (2018), who highlight the role of shared reflection in building collective agency and resilience in complex settings.

These processes are particularly likely to occur in collaborative, dialogic environments where people not only express views but co-construct knowledge (Steyaert et al., 2007; Reed et al., 2010). Recognizing the evolving nature of values is vital when developing tools or interventions that aim to analyze, support, or reflect upon value-based decision-making in real-world conversations.

3.4. Conversations as a Lens

In collaborative settings, conversations serve as more than exchanges of information—they are sites of reflection, negotiation, and meaning-making. Unlike one-way communication, interactions in dialogue provide richer opportunities for learning, perspective-taking, and even behavioral change (Beratan, 2007). This makes conversation a particularly powerful mechanism for uncovering how values are expressed, challenged, or reshaped during decision-making. The following two sections explore this premise further: first, by examining how individuals engage with values through dialogue, and second, by outlining how computational methods, specifically NLP, can help surface and analyze these value dynamics at scale.

3. 4.1. Value Engagement in Dialogue

To uncover the dynamic nature of values, it is first essential to understand how individuals engage with them, especially in decision-making and strategy contexts. Not all individuals are equally conscious of their values, nor do they uniformly rely on them to guide their choices. As Schwartz (2016) notes, value-based assessments often occur outside conscious awareness. Even when individuals articulate their values, they may only partially understand how those values inform their decisions. Values tend to operate within interrelated systems of beliefs, attitudes, and intentions (Feather, 1996; Rindova & Martins, 2017), making them difficult to isolate or analyze in a straightforward manner

Given this often-implicit nature of values, observing individuals' communication and interactions becomes a promising route to uncovering them. Interactions can reveal how people frame problems, negotiate meaning, and respond to trade-offs—insights not always accessible through direct questioning or surveys. Through conversation, individuals engage in a values-based exchange of perspectives, reflecting, challenging, and sometimes reconfiguring what they consider important (Kenter et al., 2016a).

Kenter et al. (2016a) provide a theoretical foundation for this approach through their Deliberative Value Formation (DVF) model (see Figure 09) . The model illustrates how individual values are not simply revealed but actively shaped through the process of group deliberation. It differentiates between transcendental values—broad guiding principles like justice or autonomy—and contextual values, which reflect what is deemed important in a specific decision-making situation. These are not directly transferred but mediated through evolving beliefs, norms, and social learning. For example, when a community deliberates over how to use a vacant lot, individuals may initially suggest different practical solutions. However, through dialogue, they begin to surface and align on transcendental values like well-being, sustainability, and equity. These values then guide the formation of contextual preferences, such as choosing a community garden over a parking lot — a decision that reflects newly shaped shared understanding.

The DVF model also highlights key outcomes of deliberation, such as a shift in value orientation toward shared goals and changes in trust dynamics. These outcomes form a strong foundation for applying this research approach.

Critically, the DVF model positions conversation as a site where values become “more explicit and contestable,” and where value change, conflict, or convergence can be observed. It highlights how values are not merely stated but actively constructed and reshaped through deliberation. This communicative process—of expressing, negotiating, and revising values—plays a vital role in shaping group dynamics, stakeholder alignment, and decision-making. Conversations thus become not only a reflection of individual value systems but also a space where those systems are enacted, challenged, and transformed.

This perspective is central to the present research, which analyzes conversations to trace how individuals and groups reason, justify, and express values in real time. By examining how values emerge and evolve in dynamic, multi-stakeholder dialogues, the study moves beyond static assumptions and grounds its application in theory. It uncovers both hidden tensions and common ground, offering critical insights for value-informed decision-making.

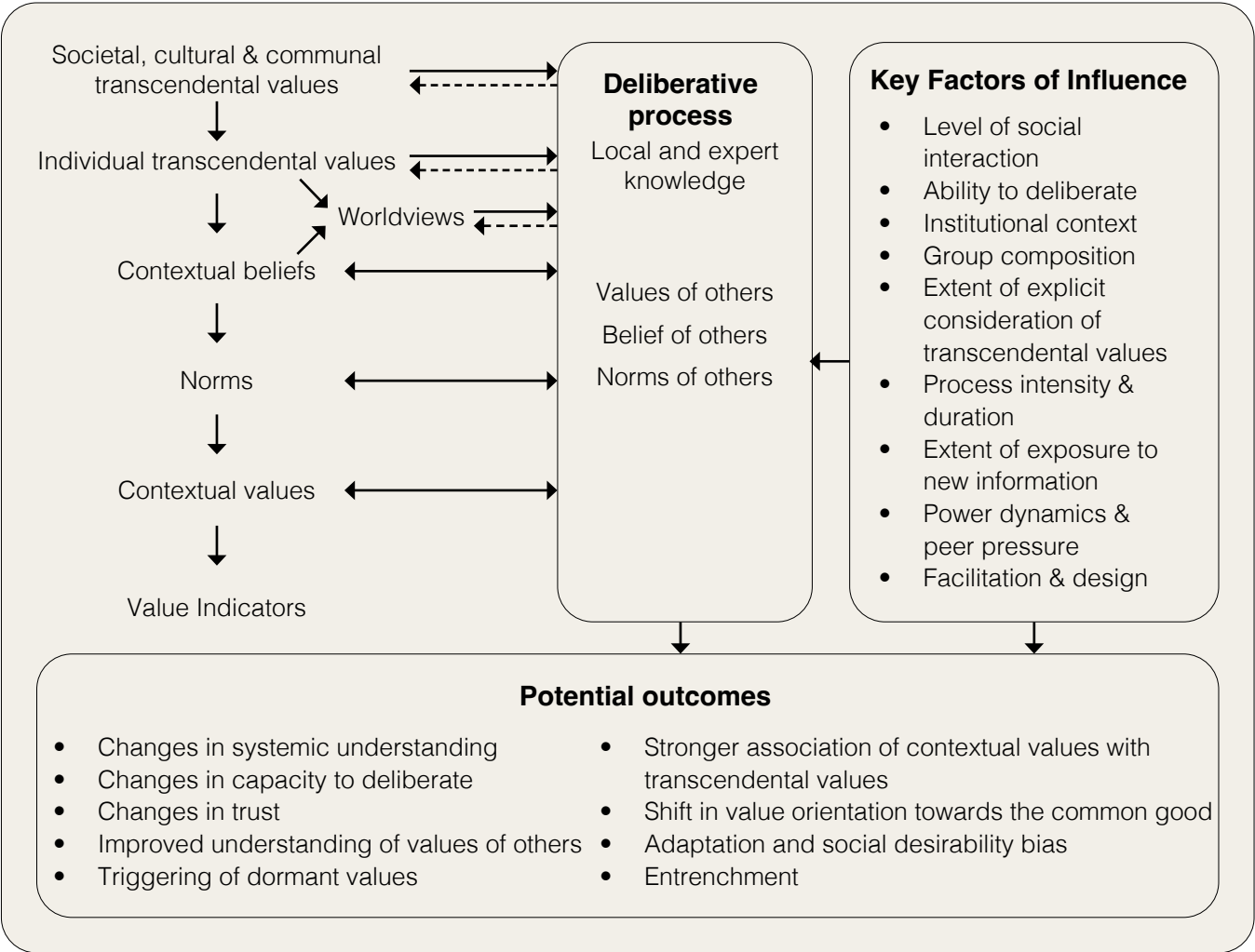


Figure 09: Adapted from Kenter et al. (2016a), this model illustrates how values are not just revealed but actively shaped through group deliberation—transforming broad transcendental values into contextual values via beliefs, norms, and social interaction. It positions conversation as a site for value change, alignment, and contestation.

3. 4.2. Why NLP? A Methodological Rationale

The communicative nature of value expression suggests that conversations offer a meaningful unit of analysis for understanding what people value. Unlike static texts, conversations are co-constructed in real time and shaped by mutual responsiveness, turn-taking, and contextual framing (Yeomans et al., 2023).

Natural Language Processing (NLP)—an interdisciplinary field at the intersection of linguistics, computer science, and artificial intelligence—offers tools to analyze such conversations at scale. By transforming unstructured language into structured behavioral data, NLP makes it possible to extract patterns, themes, and latent signals from large volumes of text (Hirschberg & Manning, 2015; Jurafsky & Martin, 2017).

While human analysis remains valuable for deep contextual interpretation, it can be inconsistent, time-consuming, and difficult to scale (Yeomans et al., 2023). NLP, in contrast, offers consistent formatting, cost efficiency, and the ability to identify implicit values that may not be explicitly named—an essential capability when analyzing complex, context-dependent constructs like values (De Wildt et al., 2021). This is possible because some NLP models are trained to recognize underlying patterns and contextual associations between words. Additionally, systems like TalkTraces (Chandrasegaran et al., 2019) and other research (Aseniero et al, 2020; Chen et al., 2025) have also demonstrated NLP’s ability to support real-time capture and visualization of conversations, addressing key limitations of manual review.

The approach in this study adopts a hybrid mindset—using NLP to structure and detect patterns, while keeping space for human interpretation for further refinement and ethical reflection. The rationale for choosing NLP for the purpose of this research is shown in the Table 03 below.

Criteria	NLP Approach	Human Coding	Sources
Scalability	Processes large datasets rapidly and cost-effectively	Time- and labor-intensive. Effort scales linearly	Yeomans et al., 2023; De Wildt et al., 2021
Implicit Nature Handling	Detects values expressed implicitly through context	Risk of missing subtle expressions or implicit meanings	De Wildt et al., 2021
Real-Time Application	Enables live capture and analysis of discussions	Post-discussion analysis only	Chandrasegaran et al., 2019
Contextual Nuance	Interprets meaning beyond keywords using embeddings and sentence structure	May miss contradictions or nuanced shifts	Yeomans et al., 2023
Consistency & Bias Control	Offers consistent outputs with auditable model logic	Subject to cognitive and social biases	Yeomans et al., 2023
Cost & Efficiency	Low-cost once implemented; ideal for rapid iteration	High-cost in repeated or large-scale studies	Chandrasegaran et al., 2019
Support for Human Insights	Structures data for deeper interpretation	Strong in nuance, weak in pattern detection at scale	Yeomans et al., 2023

Table 03: This table compares NLP and human coding across key criteria relevant to value elicitation—such as scalability, cost-efficiency, contextual nuance, and real-time application.

3.5. Value Framework Used

As previously discussed, the term value encompasses a variety of meanings and interpretations across disciplines. This conceptual richness, while valuable, also presents challenges, particularly when one aims to systematically detect or engage with values in real-world contexts such as design or decision-making. Existing theories of values and approaches for integrating diverse, and sometimes conflicting, values remain scattered across academic fields, leaving designers with little structure or shared vocabulary for what they are encountering (Bos-de Vos, 2020). To address this, the present study adopts an existing framework (see Table 04) developed by Bos-de Vos (2020), known as the Design for Divergent Values framework that provides a coherent lens for interpreting values in the context of collaborative and strategic conversations.

In line with the duality outlined earlier, between values as guiding ideals and value as perceived worth, several scholars have described this conceptual split as ‘values’ in the plural (ideals) versus ‘value’ in the singular (worth) (Boradkar, 2010; Martinsuo, Klakegg, & van Marrewijk, 2019). Rather than searching for a single definition of value, this study embraces both perspectives, following the approach proposed by Bos-de Vos (2020), who encourages viewing value through multiple lenses to capture its full complexity.

The framework introduces three degrees of value-specificity, offering a structured way to analyze how values function at different levels:

- Overarching value dimensions – broad categories such as human, ecological, or economic values;
- Underlying motivational goals – the reasons or drivers behind why a value matters in a specific context;
- Concrete value examples – specific, situational expressions of values, often reflected in everyday decisions or trade-offs.

By using this layered framework, the study is able to identify and classify values expressed in conversations at different levels of abstraction. It provides both the flexibility to accommodate different perspectives and the structure needed for computational interpretation.

Importantly, this framework is not intended to be exhaustive or definitive. As Bos-de Vos (2020) notes, it should be seen as a starting point - a first stepping stone towards developing tools and methods that can support designers in navigating divergent values in complex, collaborative settings. In the context of this research, it serves as a shared foundation for surfacing values from conversations in a way that is consistent, scalable, and reflective of real-world complexity.

		Type of Value	Motivational Goal	Value Examples
Value as Guiding Principles Influence strategic decisions related to value co-creation and value capture (Rindova & Martins, 2017) Value as Qualities with Worth	OF INDIVIDUALS	Human Values (e.g. Schwartz & Bilsky, 1987)	Enjoyment	pleasure, self-indulgement, gratification, sensuous enjoyment, happiness at work,...
			Security	physical safety, psychological/ mental health, integrity,...
			Achievement	achievement, competence, success,...
			Self-Direction	autonomy, self-sufficiency, independence, intellectualism,...
			Restrictive-conformity	conformity to social expectations,...
			Prosocial	altruism (e.g. acting in best interests society/client), benevolence, kindness, love,...
			Social power	dominance, status, influence, social control, power, leadership, authority,...
			Maturity (cannot be actively attained)	wisdom, tolerance, faith in one's convictions, deep emotional relationships, appreciation for the beauty of creation,...
	OF GROUPS	Cultural Values (e.g. Schwartz 2006)	Autonomy	Intellectual autonomy: broadmindedness, curiosity, creativity, ... Affective autonomy: pleasure, exciting life, varied life, ...
		e.g.: - teams - organizations - economic sectors - nations	Embeddedness	social order, respect for tradition, security, obedience, wisdom,...
			Egalitarianism	equality, social justice, responsibility, help, honesty,...
			Hierarchy	social power, authority, humility, wealth,...
			Harmony	world at peace, unity with nature, protecting the environment,...
			Mastery	ambition, success, daring, competence,...
	FOR PEOPLE	Use Value (e.g. Bocken et al., 2013; Ravasi et al., 2012; Ekstrom, 2011)	Utility	functionality, convenience, usability, efficiency, durability, time management, accessibility, appropriateness, compatibility,...
			Well-being & development	health, comfort, safety, growth, knowledge development,...
			Symbolic meaning	expression of identity, signal of social status, prestige, stature, ... historic value, brand value, political value, aesthetic value, ...
			Emotional meaning	fun/joy, pleasure, appreciation,...
		Social Value (e.g. Boradkar, 2010; Den Ouden, 2011)	Social propriety	human health, safety, security, justice, privacy,...
			Social wealth	minimize/no labor exploitation, fair living wages, maximize opportunity for workers, efficiency,...
		Economic Value (e.g. Bowman & Ambrosini, 2000)	Money	income, profit, wealth, affordability, rents, economic sustainability,...
			Other economic value	reputation, competitive advantage, innovation, commercial relationship,...
	FOR PLANET	Ecological / Environmental Value (e.g. Bocken et al., 2013)	Preservation of the planet	emission regulations/ reduction, product safety, re-use of existing material, sustainability, long lasting neighborhood,...

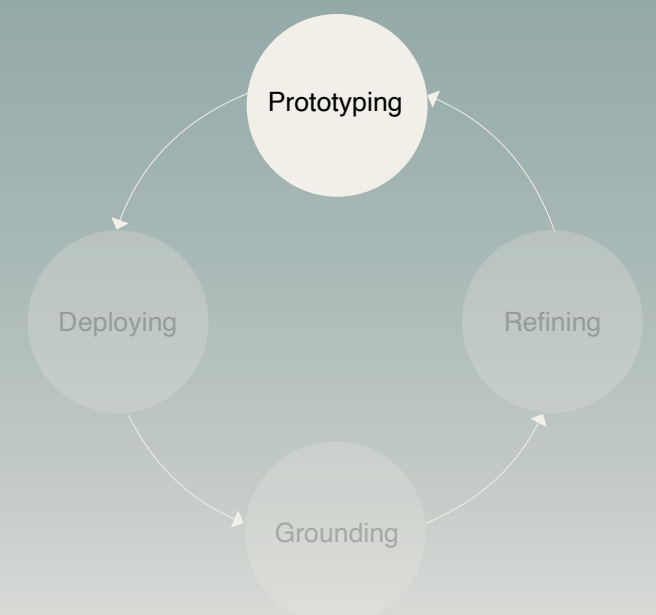
Table 04: Framework by Bos-de Vos (2020), adopted for this research to identify and classify values at multiple levels.

Chapter 04

Developing the Prototype

This chapter outlines the transition from theory to implementation, detailing the development of a low-fidelity prototype designed to computationally detect values in conversation. It introduces the technical setup, core logic using sentence embeddings and cosine similarity, and the early design decisions that shaped the model's interpretive capabilities.

- 4.1. From Theory to Prototyping
- 4.2. Context for Testing: The Forum
- 4.3. Method Logic
- 4.4. Assumptions and Starting Points
- 4.5. Early Decisions for Advancement



4.1. From Theory to Prototyping

Building on the theoretical grounding from earlier chapters, this phase shifts focus—from understanding why values matter in conversation to exploring how to computationally surface them. Rather than aiming to build a polished product, the goal was to test feasibility through a low-fidelity prototype that could act as a research probe for detecting values in real-time dialogue.

Since earlier sections have already established conversation as a meaningful lens for observing values, and NLP as a fitting technique for detecting latent patterns at scale, this section explains the rationale behind the specific technologies chosen for the prototype.

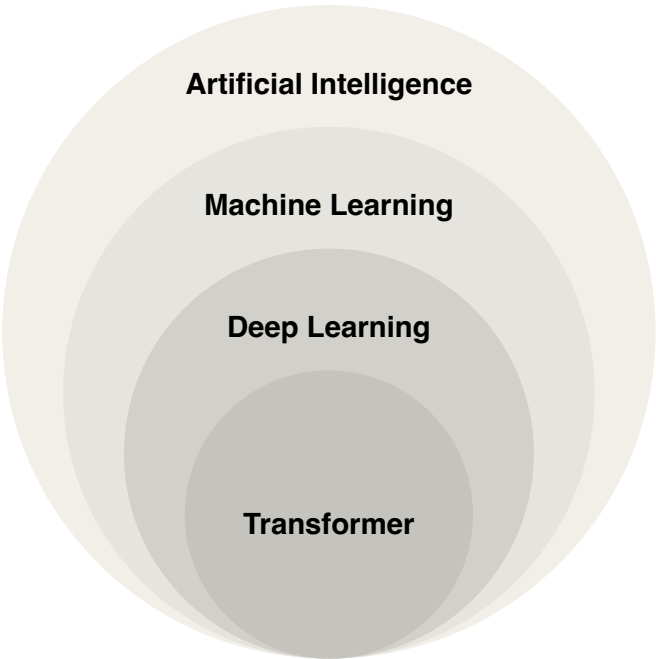


Figure 10: Layered structure of AI technologies adapted from Caelen & Blete (2024), illustrating how Transformers—used in this project—are a specialized architecture within DL, which itself is a subset of ML, all under the broader domain of AI.

To set the stage, let’s briefly unpack the layered structure of AI technologies used in this project (see Figure 10). Caelen and Blete (2024) visualize this as a nested system: at the top is Artificial Intelligence (AI), the broad field focused on enabling computers to mimic human behavior. Within that sits Machine Learning (ML), where systems learn from data without explicit programming. A layer deeper, Deep Learning (DL) leverages neural networks to detect complex patterns in large datasets.

At the core of DL, Transformer architectures now lead the way in language modeling—able to capture sentence structure, context, and meaning with high precision.

Rather than working broadly across all these layers, this research takes a more focused approach. It works intentionally with the lower layers—transformers, deep learning, and machine learning—to retain clarity, transparency, and alignment with the conceptual foundation laid out earlier. It uses a pre-trained transformer model that fits both the theoretical and practical needs of the research. On top of this, it applies machine learning techniques to build a logic for comparing language patterns—allowing the prototype to detect potential value expressions in conversation.

This setup strikes a balance between feasibility and depth. Instead of training a complex model from scratch, the approach uses a reliable, off-the-shelf transformer to ensure speed and consistency—freeing the research to focus on testing whether value dynamics can be meaningfully explored through this layered architecture. The next step, then, was to translate this stack into a functional prototype logic capable of surfacing values in real-world dialogue.

4.2. Context for Testing: The Forum

At the start of the project, an opportunity emerged to engage with a real-world setting that would go on to shape the prototype’s development: a high-level leadership conference bringing together senior decision-makers from diverse sectors. Participants came together to reflect on pressing societal themes—sustainability, innovation, governance, and leadership—through facilitated, open conversation.

The event was designed to foster deep dialogue. Each theme encouraged participants to challenge perspectives, build on each other’s ideas, and reflect collaboratively. This made the setting not only rich in value-laden discourse, but also an ideal reference point for exploring how values surface in real-time, multi-stakeholder conversations.

4.3. Method Logic

Rather than testing with a fully formed method, the project used this context to scope the problem and choices. For instance, the early prototype didn’t attempt to work with the entire list of 108 values from the value framework (Table 04). Instead, a curated list of 20 values was selected based on their relevance to the conference themes and leadership language (Illustrated in Fig 11). These values became the input for the model, and the guiding lens for how value reasoning would be analyzed.

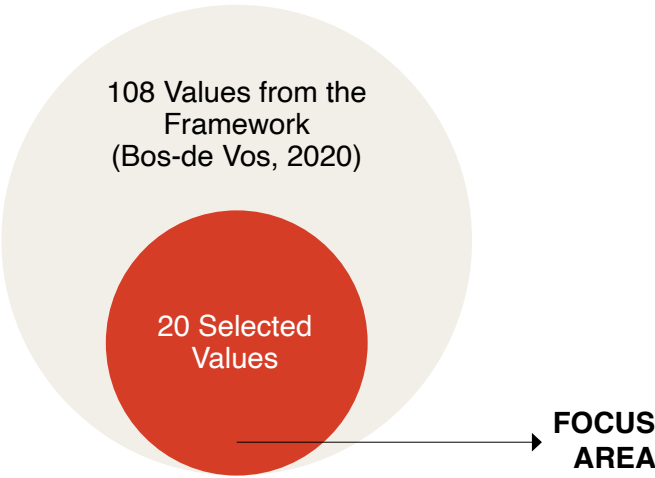


Figure 11: Illustrates how a subset of 20 values was selected from the full framework of 108 values (Bos-de Vos, 2020). This focused selection became the basis for the early prototype and shaped how value reasoning was explored in the project.

By grounding the early development in this environment, the prototype was able to focus on feasibility in a high-stakes setting. It wasn’t about real-time output or evaluation, but about seeing whether value dynamics—both implicit and explicit—could be computationally surfaced in ways that made sense within the flow of natural conversation.

This contextual grounding played a critical role. It influenced not just what the prototype looked for, but how it reasoned about language, how values were defined, and how insights might be fed back to participants in meaningful ways. In this sense, the leadership conference didn’t just test the prototype—it helped design it.

With the testing context in place, the next step was to translate the idea into action—by operationalizing the value elicitation method into a working prototype. The goal was to create a structured, repeatable system that could analyze spoken dialogue and surface the values embedded within it.

To do this, conceptual ideas about values had to be transformed into something a machine could interpret. That meant building a logic: a system that could process utterances, compare them meaningfully, and match them to specific values based on how they were expressed.

This section walks through that system. It begins by outlining the value framework selected as the foundation of the method. From there, it explains the key design decisions—how values were categorized, how example expressions were selected, and how similarity scores were used to identify alignment between utterances and values.

Together, these components shaped the prototype’s core functionality—and determined the kind of insights it could generate from natural, real-world conversation.

Value Framework Adoption

This project adopted the value framework developed by Bos-de Vos (2020), which treats values as both guiding principles and qualities of worth. The framework is structured across three layers: six overarching value types, 23 motivational goals, and 108 specific value examples. This layered structure offered both the conceptual clarity and the practical granularity needed to explore values in design and organizational contexts.

To make the method computationally viable at the sentence level, the focus shifted from abstract value types to specific value examples. While high-level categories helped structure the thinking, they weren’t expressive enough to capture the richness of how values appear in real speech. So instead of modeling categories, the prototype worked directly with the value examples.

Throughout the report, these “value examples” are referred to simply as “values” to keep things clear and consistent with how users and participants naturally talk about them.

For the initial prototype, 20 values were selected from the full set of 108 (see Table 05). This selection was based on their relevance to the Leadership Conference context and the need to keep the model lightweight and interpretable.

Type of Value	Value Examples
Human Values	Gratification, Independence, Leadership
Cultural Values	Creativity, Equality, Responsibility, Social Power, Success
Use Values	Usability, Accessibility
Social Values	Human Health, Privacy, Security, Fair Living Wages, Efficiency
Economic Values	Profit, Affordability, Innovation, Competitive Advantage
Ecological / Environmental Value	Sustainability

Table 05: Lists the 20 values selected, along with their respective value type from the full framework of 108 for the initial prototype. These were chosen for their relevance to the Leadership Conference and to ensure interpretability and have control over the data during the early phases.

To operationalize these values in the prototype, each one was paired with four example utterances that illustrated how it might realistically appear in conversation (see Table 06 for examples). These examples were generated using OpenAI’s ChatGPT, with prompts designed to reflect both explicit and implicit expressions of each value.

Using a large language model (LLM) like ChatGPT allowed the project to simulate the varied and natural ways people express abstract concepts like values. This approach helped the machine better make connection between real-world utterances with the values they conveyed. Further details and supporting literature for this generation process are provided in Appendix D.

Type of Value	Value Examples	Example Utterance
Human Values	Independence	“I value being able to make my own decisions.” “Having the freedom to act on my ideas is non-negotiable.” “I work best when I’m trusted to find my own way.” “Autonomy gives me the clarity and confidence to lead.”
Cultural Values	Autonomy	“In our culture, we encourage open-ended thinking and exploration.” “We believe innovation thrives when creativity is nurtured collectively.” “This team values originality as a driver for progress.” “Our organization promotes experimentation as part of its identity.”

Table 06: Illustrates with examples of how values were operationalized in the prototype by pairing each selected value with four example utterances. These utterances, generated using ChatGPT, simulate both explicit and implicit ways in which people naturally express values in conversation.

Breaking Down the Tech

Before outlining how the prototype works, this section takes a brief detour — a learning pit stop to unpack the key technical terms. By clarifying the foundational concepts, the rest of the chapter becomes easier to follow.



SENTENCE EMBEDDING

turning sentences into points in high-dimensional spaces

Imagine every sentence said gets turned into a point on a giant invisible map. The closer two points are, the more similar the sentences.

The ones in the example on the left would land close together on the map because they carry a similar meaning. This transformation is called a sentence embedding.

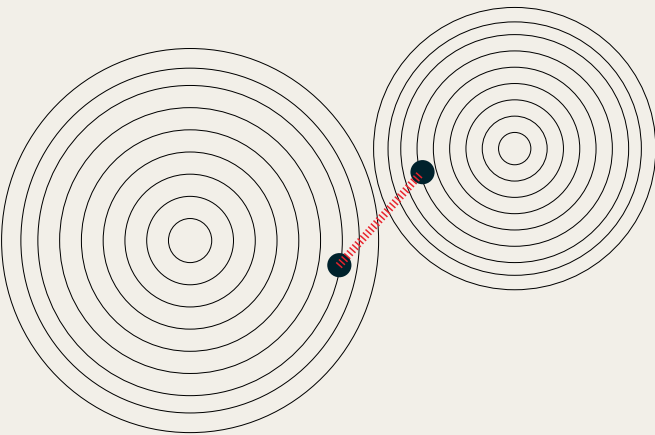


TRANSFORMERS

context-savvy language models

Transformers are AI models trained on large language datasets to understand meaning in context. They’ve proven to be good at picking up tone, intent, and subtle cues.

A sentence transformer takes a full sentence and turns it into a point— called an embedding — that captures the meaning of that sentence.



COSINE SIMILARITY

measuring how close meanings are

Cosine similarity tells us how similar two sentences are based on their position on the meaning map. If the similarity score is close to 1, the meanings are very similar. If it’s closer to 0, they’re less related.

In the illustration on the left, the length of the line shows how far apart the meanings are — a bigger distance means less similarity. (See Appendix J for Detailed Explanation)

Prototype Logic

To take the value elicitation method forward, a computational workflow was developed using sentence embeddings — an approach well-suited for measuring semantic similarity between varying textual inputs (Reimers & Gurevych, 2019). According to literature, representing sentences as embeddings and comparing them using cosine similarity is a common (Clark, 2015; Chandrasegaran et al., 2019), efficient (Reimers & Gurevych, 2019) and effective approach for capturing the nuanced semantic relationships within multi-turn dialogues (Liu et al., 2023).

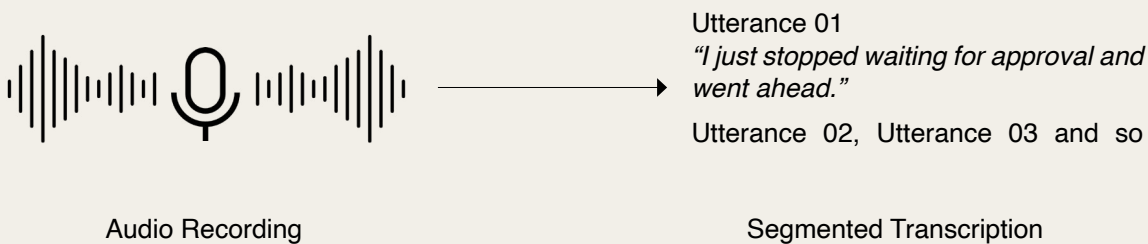
Let’s look at the flow of the prototype with an example below:

The value with the highest similarity score — based on comparison with the averaged embedding for each value — was assigned to the utterance. In the prototype, no minimum similarity threshold (i.e., a cutoff for similarity score below which no value would be assigned) was applied; the aim was to observe patterns and test feasibility over precision.

The Figure 12 illustrates the full pipeline — from spoken conversation to computed similarity scores — culminating in a heatmap that makes the relationship between values and utterances visible. This representation helped reveal patterns in value expression and served as a foundation for deeper analysis.

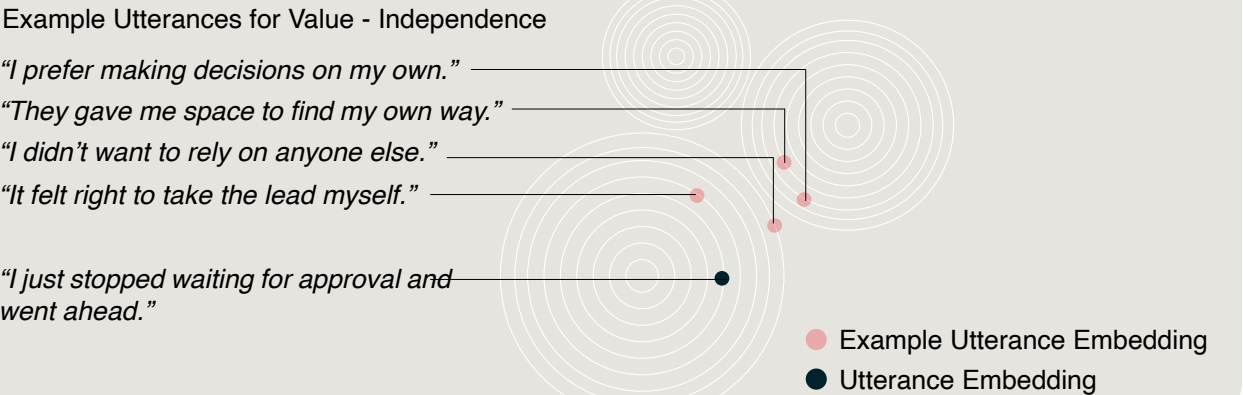
Step 01: Transcription and Segmentation

Audio recordings from conversation transcribed into text and then split into individual utterances.



Step 02: Sentence Embedding Generation

Each utterance, including example utterances for each value, are then converted into a sentence embedding using a pre-trained transformer model. These embeddings numerically represent the semantic meaning of the sentence.



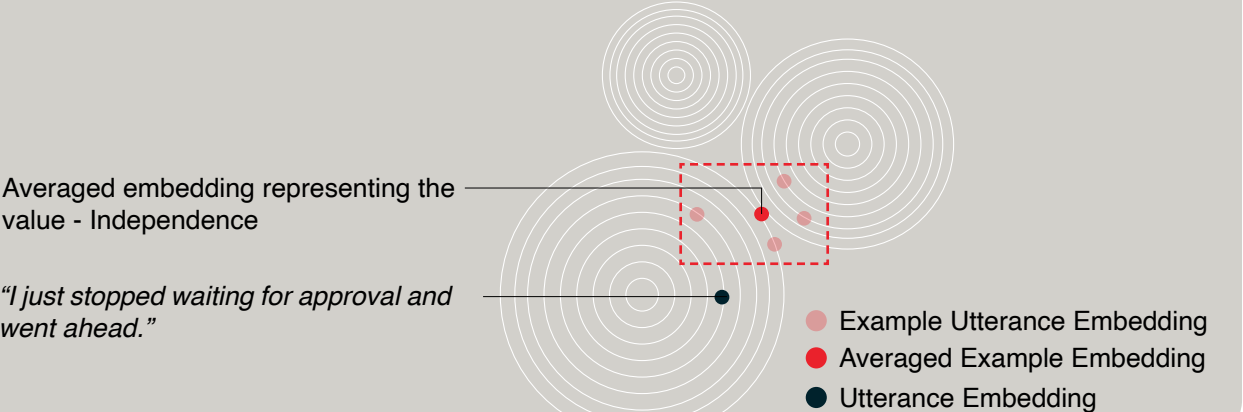
Step 04: Similarity Comparison

Using cosine similarity (see Appendix J), the prototype measures how semantically close each utterance is to the averaged embeddings of each value.



Step 03: Value Embedding Averaging

Each predefined value is associated with four example utterances. The model computes the average embedding of these examples to represent that value.



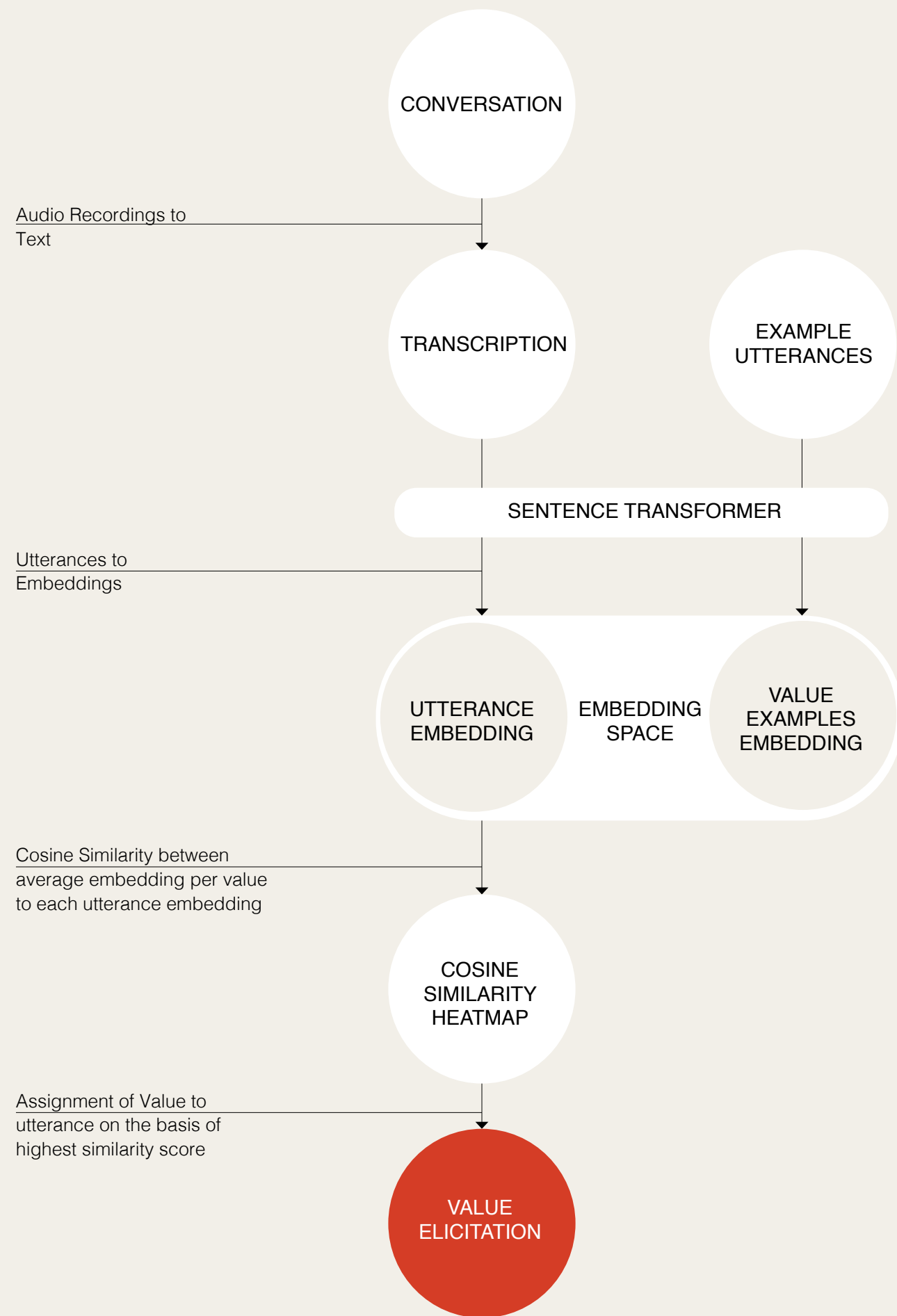


Figure 12 shows the full pipeline of the prototype: starting with recorded conversation, followed by transcription, sentence embedding of both utterances and example values, calculation of cosine similarity, and final value elicitation based on the highest similarity score.

4.4. Assumptions and Starting Points

The prototype was developed using Python in a local JupyterLab environment and iterated with regular testing against conversational transcripts and feedback loops integrated into its logic. The goal at this stage was to structure a computational approach that could process human conversation and return value-related insights in a form that could be explored and iterated upon.

Because the project aimed to explore how values manifest in real-time discussions, a deliberate decision was made to focus exclusively on verbal content as the core data source. While nonverbal signals—such as tone, gesture, or facial expressions—can influence how messages are perceived, verbal content plays a disproportionately central role in how meaning is constructed and interpreted (Yeomans et al., 2023).

Similarly, Lapakko (1997) emphasizes that although nonverbal signals may enhance or nuance a message, the verbal layer remains the primary carrier of informational content. This provided a clear rationale for using conversational transcripts as the foundation for this prototype.

To make early development tractable, a set of simplifying assumptions were made about how values appear in language and how computational methods might capture them. These assumptions were not meant to be exhaustive or final, but rather to serve as working hypotheses that could scaffold the method's design and offer initial insights. Table 07 below outlines the key assumptions, what they enabled us to analyze, and the kinds of insights they opened up during this early stage of development.

Initial Assumptions	What It Allows Us to Analyze	Potential Insights
Each utterance reflects one dominant value	Count value frequency across utterances	Identify most emphasized values in the conversation
Values are expressed clearly enough for NLP detection	Track values that are absent or weakly expressed	Spot overlooked or underrepresented value
High similarity scores reflect meaningful alignment of	Use similarity score to detect co-occurrence of values	Observe value clusters (e.g. sustainability + affordability)

Table 07 outlines the key assumptions behind the value elicitation method, what each assumption enabled the analysis to capture, and the types of insights this approach made possible in the prototype's early stage.

These assumptions formed the backbone of the method's early logic—shaping not only how the prototype was built but also which types of value-related patterns were expected to emerge. While the method would later be refined to challenge or expand on these assumptions, they offered a necessary starting point for learning how computational detection of values could unfold.

4.5. Early Decisions for Advancement

Before deploying the prototype at the INSEAD Alumni Forum 2025, a series of foundational decisions were made to progress its technical maturity and conceptual clarity. These decisions were shaped by a combination of relevant theory and early, small-scale observations using the prototype. While not exhaustive or definitive, these choices reflected what made the most sense at the time to move forward. Rather than claiming to offer the perfect solution, they served as working assumptions and informed the direction of development. Together, they laid the groundwork for the value elicitation approach explored in this report.

01

Transformer Model Selection

A key early decision in developing the value elicitation prototype was selecting an appropriate sentence transformer model to embed utterances and value examples into a shared embedding space. Based on comparative model studies (Colangelo et al., 2025), the model all-MiniLM-L6-v2 was selected for its balance between semantic quality and computational efficiency—both of which were crucial for testing the method in near real-time contexts like live conversations or post-session summaries.

Two primary considerations supported this choice (Colangelo et al., 2025):

- **Semantic Robustness:** The model generates meaningful sentence-level embeddings that allow the detection of nuanced conceptual similarities between everyday speech and abstract value expressions.
- **Computational Efficiency:** MiniLM-L6-v2 significantly outperforms larger models like Mpnet in speed. Since this method was envisioned as a potential real-time tool, efficiency was a practical necessity. The model’s ability to return results quickly without sacrificing interpretability supported its use in dynamic settings, such as summarizing a live panel discussion.

According to Colangelo et al. (2025), while Mpnet offered slightly higher average similarity scores (0.71 vs. MiniLM’s 0.66), the tradeoff in speed and responsiveness made MiniLM-L6-v2 a more viable foundation for this phase of development. Future versions of the method may experiment with other models if precision becomes a higher priority.

02

Utterance Segmentation

Initially, utterances were structured based on speaker turns — that is, one full statement made by a single speaker before someone else responded. However, during early usage of the prototype, it became evident that this approach often resulted in segments that blended multiple ideas or reasoning threads. This led to semantic flattening, where dominant values overshadowed subtler yet relevant ones.

Take the following early utterance:

“We’ve been working on making our solutions more scalable and accessible, especially for communities that can’t afford high-tech interventions. It’s a constant balance — trying to stay ahead with new ideas while also thinking about the long-term impact we leave behind.”

Top value rankings with similarity score:

Innovation: 0.601
Efficiency: 0.496
Affordability: 0.465
Success: 0.463
Creativity: 0.458

While technically within the 256-token limit supported by MiniLM-L6-v2 (where a token represents a word or part of a word used in language processing, see Appendix K), the utterance reflects values of innovation, efficiency, affordability, success, and creativity. The embedding, however, is heavily weighted toward innovation, which—while relevant—only captures the speaker’s intent on a broad level.

To address this, the segmentation strategy was adapted. Rather than relying on full speaker turns, the prototype adopted the default segmentation provided by the transcription tool, which splits input based on natural pauses. This results in smaller, more focused units of meaning:

“We’ve been working on making our solutions more scalable and accessible, especially for communities that can’t afford high-tech interventions.

Top value rankings with similarity score:

Efficiency: 0.523
Affordability: 0.514
Innovation: 0.503
Accessibility: 0.472
Success: 0.443

“It’s a constant balance — trying to stay ahead with new ideas while also thinking about the long-term impact we leave behind.”

Top value rankings with similarity score:

Innovation: 0.594
Leadership: 0.490
Sustainability: 0.484
Creativity: 0.450
Social Power : 0.394

These refined segments allow the prototype to distribute value salience more appropriately. Instead of flattening multiple expressions into a single dominant category, the segmentation brings out subtler layers of meaning—revealing values like accessibility, leadership, and sustainability that might have been overlooked otherwise. By isolating focused expressions of thought, the prototype more accurately reflects the complex interplay of values present in real conversations.

While all-MiniLM-L6-v2 can technically process up to 256 tokens, shorter input segments tend to yield sharper and more targeted embeddings. This is because focused segments help the model concentrate on a single idea, reducing semantic noise that often arises from longer, multi-thematic utterances. As a result, the similarity scores produced are more distinct and interpretable. Early observations confirmed that shorter, well-defined

inputs consistently led to more accurate and intuitive value rankings, enhancing the precision of the elicitation process.

For the sake of convenience and feasibility at this stage, the segmentation produced by the transcription tool (based on natural pauses) was adopted. However, this approach has limitations. Future studies could explore more deliberate segmentation strategies—such as splitting after full stops or using custom pre-processing rules—especially in light of the current limitations of transcription tools. Such refinements may further improve the precision and interpretability of value attribution in complex dialogues.

03

Reflecting Co-occurring Values

During early testing of the prototype, it became evident that many utterances naturally conveyed more than one value. Our initial assumption was that each utterance would express one dominant value (See Table 07 from before). Accordingly, the model was first configured to assign only the top-scoring value to each utterance. However, this approach quickly proved reductive. It masked the layered reasoning embedded in natural speech, where speakers often express blended or competing priorities within a single statement.

To better reflect this nuance, the prototype was modified to record not only the highest, but also the second-highest scoring value. This adjustment was particularly useful during the INSEAD Alumni Forum 2025 deployment, where the goal was to observe patterns and elicit insight without overwhelming the audience with complexity. Displaying the top two values for each utterance struck a pragmatic balance between interpretability and nuance, allowing participants to recognize value combinations without diluting the clarity of individual results.

Take, for example:

“It’s a constant balance — trying to stay ahead with new ideas while also thinking about the long-term impact we leave behind.”

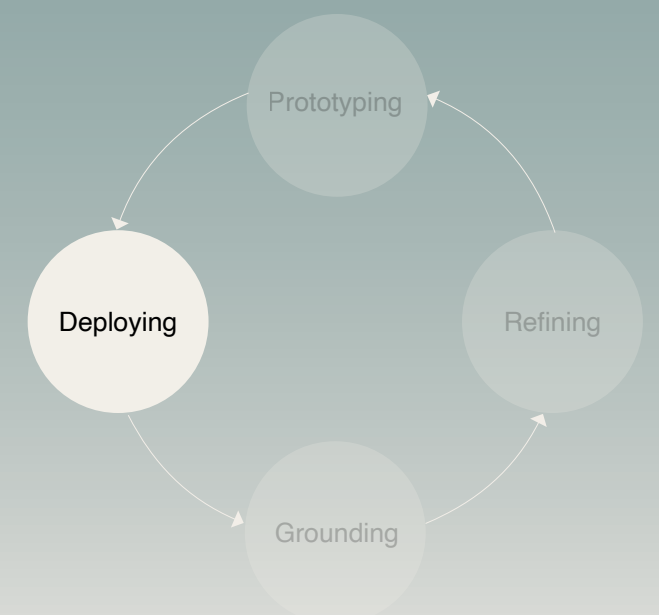
Rather than simply assigning Innovation, the second-highest value Leadership added an important dimension, highlighting the layered intent behind the statement. This realization later inspired a methodological shift: rather than hard-assigning one or two values per utterance, we began treating value scores as soft assignments—allowing for the recognition of multiple, overlapping values where relevant. This created the foundation for a deeper analysis of value co-occurrence patterns in conversation.

These early observations and design decisions were instrumental in evolving the prototype from a conceptual probe into a more grounded and interpretable method. They revealed that choices about segmentation, model selection, and value assignment strategies carry significant implications for how value dynamics surfaced from conversation. While not claiming precision, this phase laid the technical and conceptual foundation for further deployment—shaping a method capable of engaging with the nuanced and layered nature of value expression in collaborative dialogue. With this foundation in place, the next step was to explore how the prototype would perform in a real-world setting and what meaning it might hold for those engaging with its outputs.

Chapter 05 Deployment and Evaluation

This chapter covers the prototype's first real-world test at the INSEAD Alumni Forum, where a pre-survey and live transcripts were used to compare self-reported and expressed values. The resulting insights revealed a "Value Expression Gap," uncovering mismatches that raised critical questions about model limitations and the nuances of spoken value expression.

- 5.1. Pre-Survey for Value Prioritization
- 5.2. Scoping Meaning from Elicited Values
- 5.3. The Value Expression Gap
- 5.4. Learnings to Next Steps



5.1. Pre-Survey for Value Prioritization

Before deploying the prototype at the INSEAD Alumni Forum, a short pre-survey was conducted to establish a baseline understanding of participants’ value priorities. This step was not intended to validate the prototype, but to contrast what individuals say they prioritize with what actually emerges in conversation—framing the deployment as an exploration of value expression in practice. By capturing self-reported preferences beforehand, the survey added a layer of contextual interpretation to the insights generated from the conversation analysis.

The survey was hosted on Qualtrics and shared with 50 conference participants. It took approximately two minutes to complete and asked individuals to reflect on their professional values—both in general and in specific decision-making contexts. This created a point of comparison with the values that would be computationally elicited from transcripts during the Forum.

The **first part** of the survey asked participants to rank six broad value types—human, cultural, ecological, economic, social, and use values—based on their overall professional priorities (See Figure 13). This helped surface a high-level perception of which value domains participants considered most important.

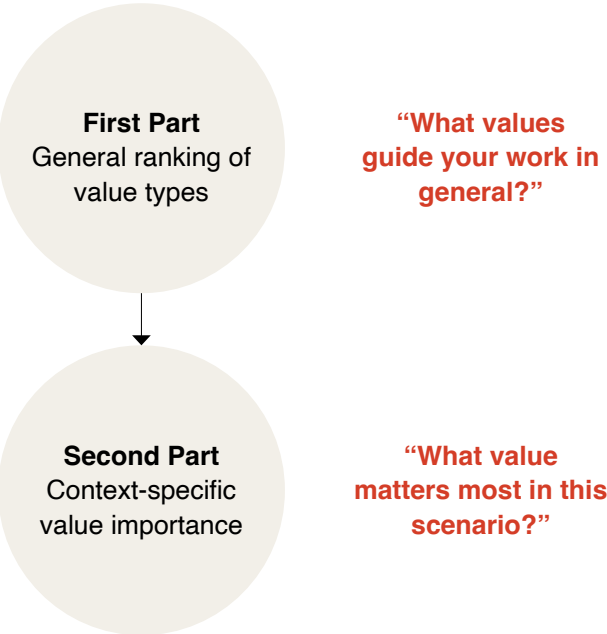


Figure 13 illustrates the two-part survey design: a general ranking of value types followed by a context-specific ranking to reveal shifts in value priorities.

The **second part** introduced decision scenarios inspired by different conference tracks (See Figure 13). For each scenario, participants were asked to rank the importance of several specific values within that context. This approach acknowledged that value preferences are often context-dependent, and aimed to surface potential trade-offs or shifts in how values are prioritized across domains.

To avoid overloading participants, a curated subset of 20 values was offered for each question—these were the same values used in the prototype (See Table 05). Keeping the value set consistent ensured a fair comparison between what participants selected in the survey and what the model detected in conversations.

The survey results served as a complementary lens for interpreting conversation data—highlighting where values aligned, diverged, or remained unspoken.

Participation was fully voluntary and anonymous, with no personal data collected. All responses were stored securely and used solely for research purposes.

A full version of the pre-survey, including question structure and value options, can be found in Appendix A.

5.2. Scoping Meaning from Elicited Values

After the prototype was able to assign values to spoken utterances, attention turned to what types of meaningful insights this output could reveal. The aim was to explore the kinds of value patterns that became visible—and how these might support reflection, comparison, or awareness among participants.

Three types of insights emerged from this phase:

- Dominant and Missing Values
- Prioritization by Value Type
- Value Co-Occurrence Patterns

Dominant and Missing Values

The first insight came from tallying the frequency of each value mentioned across the sessions. This simple aggregation revealed which values dominated the conversation — and, equally importantly, which ones didn’t.

This chart in Figure 14 shows how often each of the 20 selected values was elicited from the transcripts. Values like sustainability and accessibility appeared frequently, while others such as privacy or efficiency were rarely mentioned.

This distribution helps surface underlying priorities or blind spots within the discussion, inviting participants to reflect on which values are driving the conversation—and which may be unintentionally overlooked.

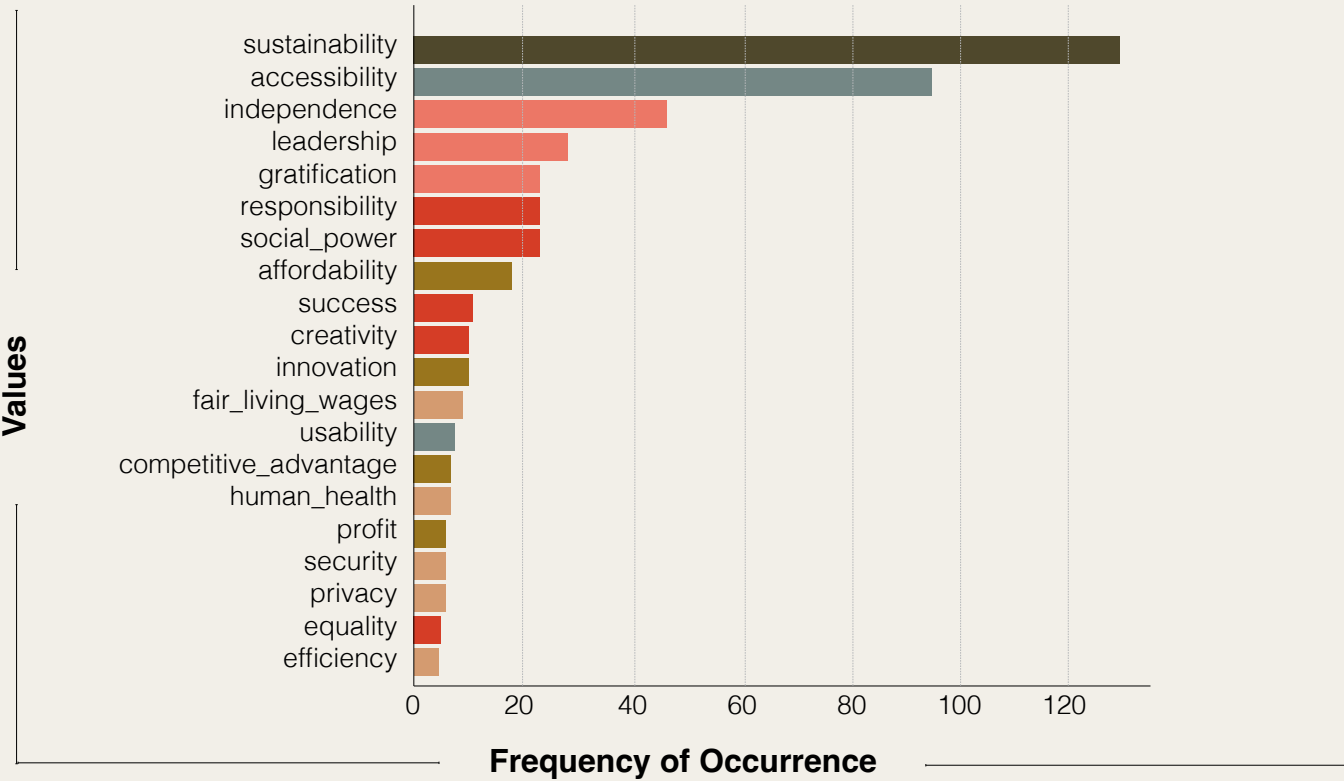


Figure 14 visualizes the frequency with which each of the 20 selected values appeared in the transcripts, revealing both dominant and rarely mentioned values across the conversations.

Value Type Legend

- Economic
- Human
- Cultural
- Environmental / Ecological
- Use
- Social

Prioritization by Value Type

To gain a higher-level view of value dynamics, the prototype included an additional layer of logic: each value was tagged with its corresponding value type (Human, Social, Cultural, Use, Economic and Ecological). This allowed the system to track not just individual value mentions, but the broader categories of meaning being emphasized.

The graph in Figure 15 shows how value types were distributed across the dialogue. For instance, economic values like Efficiency and Profit were dominate, while social values such as Privacy or Security were less represented.

By surfacing which categories of values receive the most attention—and which are underemphasized—this view supports critical reflection on the types of concerns that are shaping the conversation, and whether this distribution aligns with the group’s intended goals or ethical considerations.

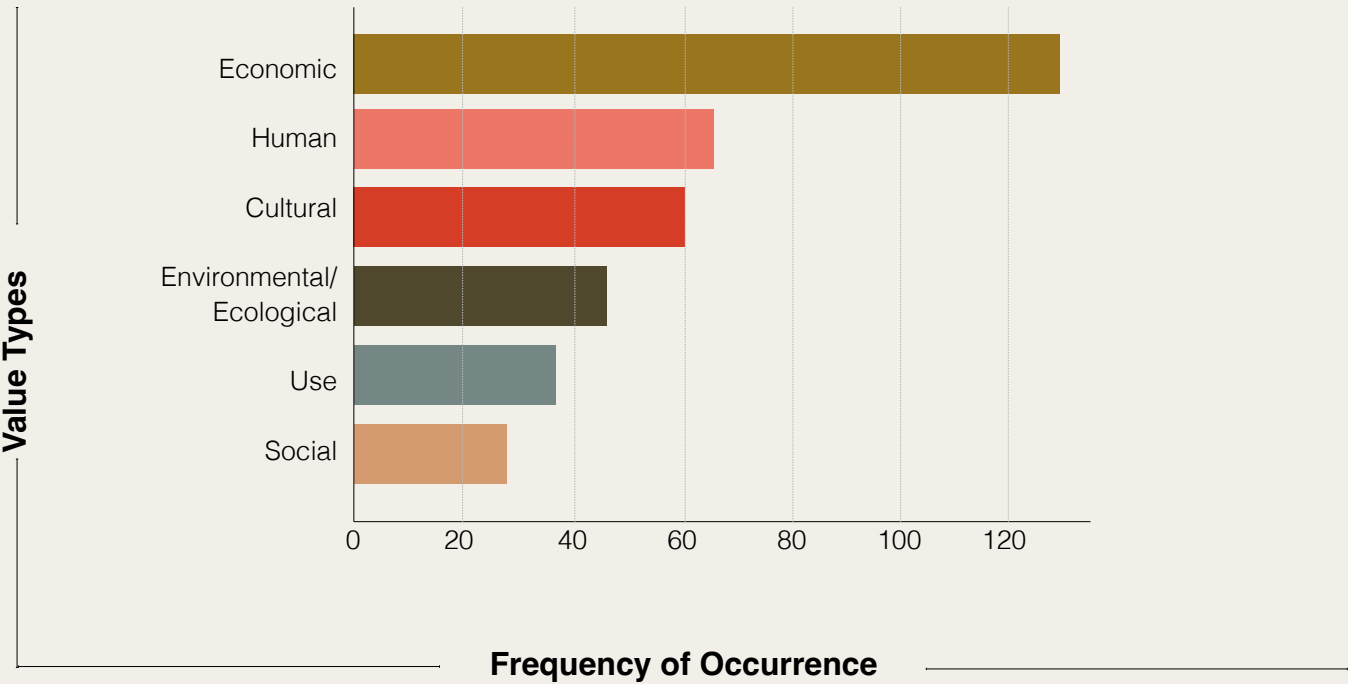


Figure 15 displays the total frequency of each value type mentioned across conversations, offering a higher-level view of which categories of values received the most emphasis.

Value Co-Occurrence Patterns

Lastly, because the prototype recorded not just the top value but also the second-highest scoring value for each utterance, it became possible to detect patterns of value co-occurrence.

This insight opens new doors for future analysis: rather than treating values as isolated signals, they can be seen as interacting elements within a broader ecosystem of priorities, trade-offs, tensions, and values that may need to be prioritized together. For example, Figure 16 shows that sustainability frequently co-occurred with profit and competitive advantage—21 and 18 times respectively—suggesting an underlying interconnectedness between these values that warrants further attention and investigation.

Identifying which values tend to appear together can help uncover implicit connections or underlying tensions in the dialogue—prompting deeper reflection on how values are negotiated, bundled, or balanced in complex decision-making contexts.

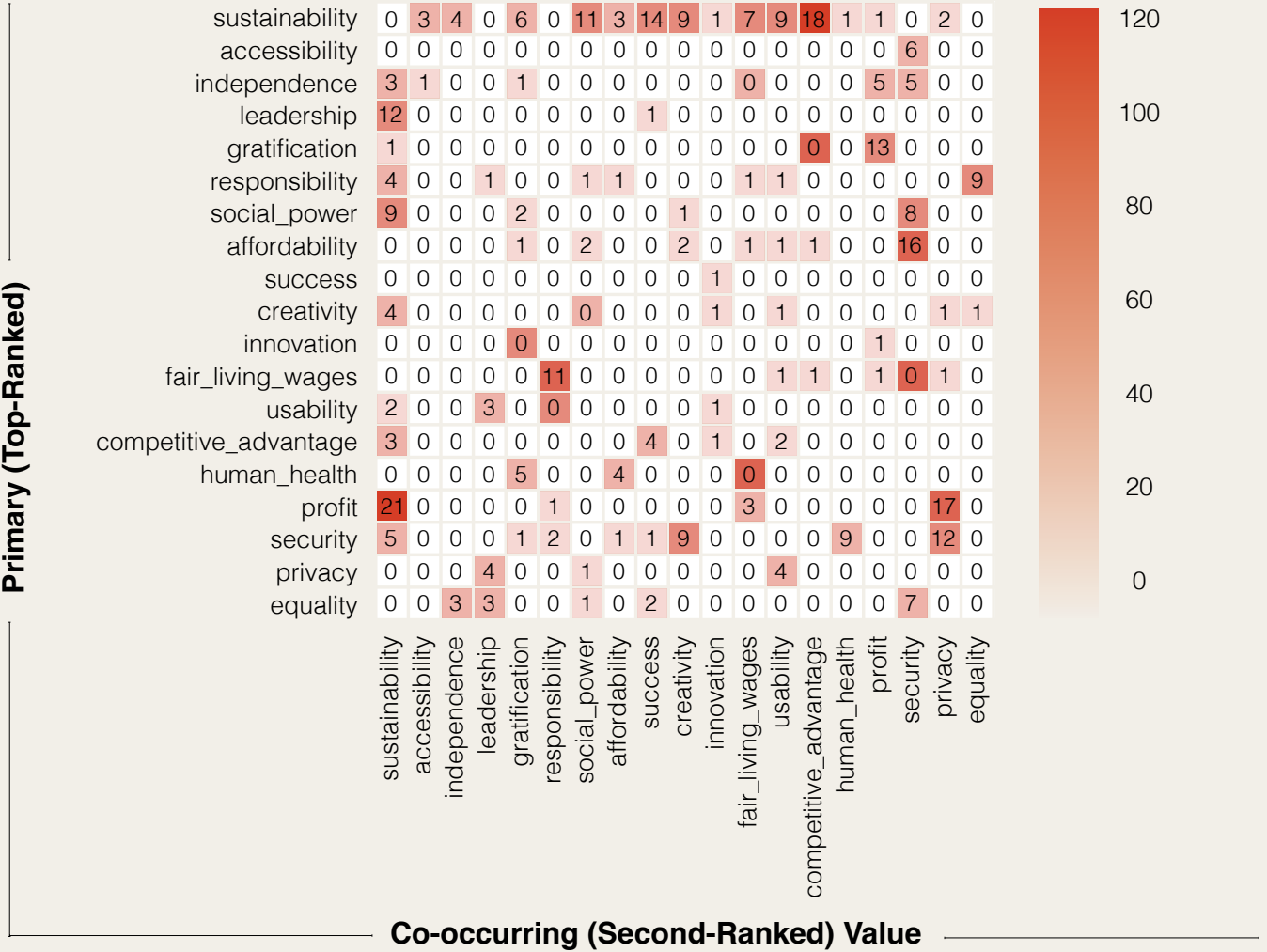


Figure 16 presents a heatmap showing how often values were mentioned together as primary and secondary rankings, revealing patterns of co-occurrence that highlight potential value clusters or trade-offs in conversation.

These early insights revealed the potential power of the prototype—not just to identify which values were present, but also to surface what was left unsaid, and how values intertwined in dialogue. Based on this, we decided to present three kinds of insights during the deployment: the most frequently expressed values, the ones that went missing despite being ranked high in the pre-survey, and the value pairs that frequently appeared together in conversation. These visualizations were intended to prompt reflection and spark awareness among participants, even at this early stage of prototyping.

5.3. The Value Expression Gap

This process of deployment of the value elicitation method revealed a deeper tension—one that pointed to a silent gap between what people say they value and what actually gets expressed in the room.

As we compared the self-prioritized value types from the pre-survey with the values surfaced in actual conversations, a clear mismatch began to emerge.

From what we say we value, to what we actually bring into the conversation.

For instance, social values, ranked as a top priority in the survey, surfaced far less frequently in discussions (See Figure 17). Conversely, ecological values, though not always highly prioritized by individuals, were more prominently voiced.

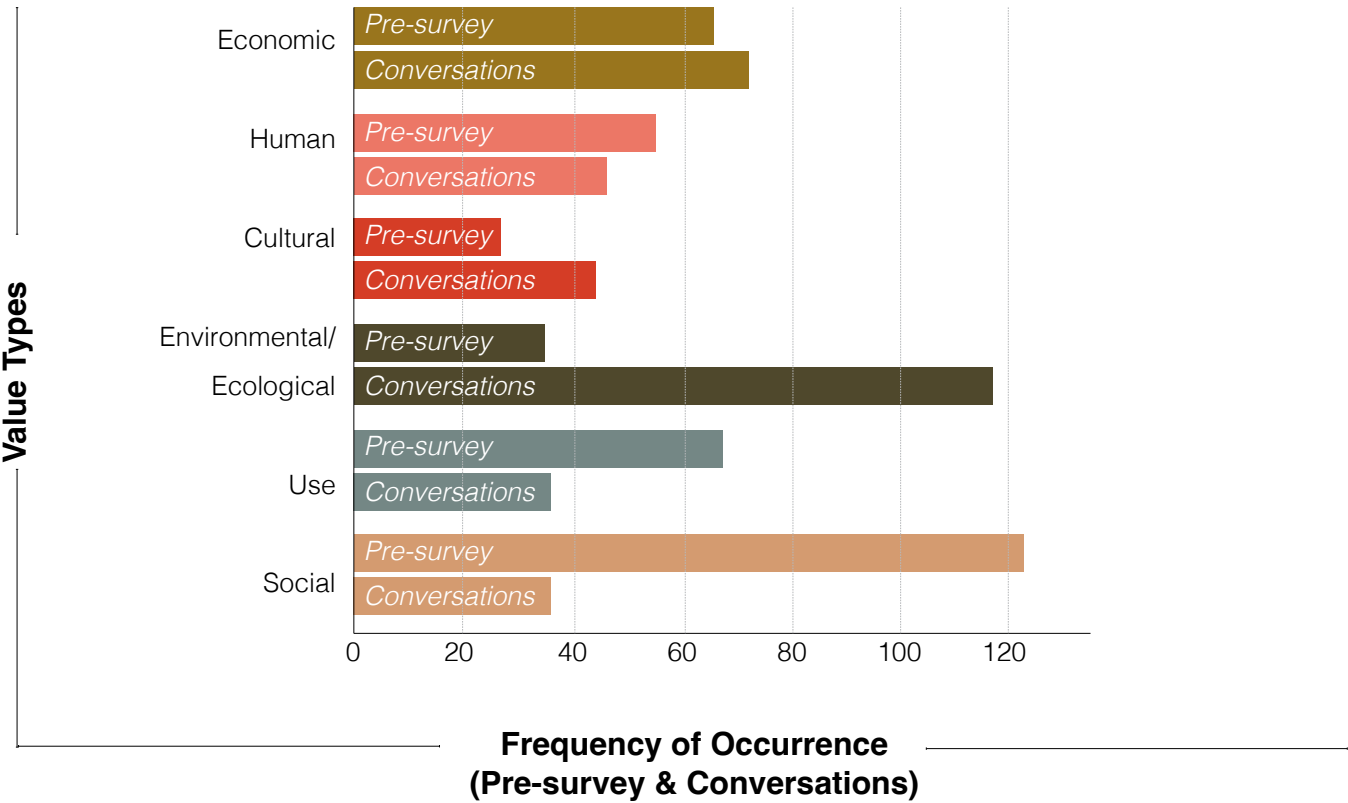


Figure 17 visualizes the value expression gap by comparing the frequency of value types ranked in the pre-survey with how often those same types were expressed in actual conversations.

This contrast led to a defining insight of the research: **The Value Expression Gap** — the difference between professed values and the values actually revealed through everyday conversations and decisions.

This idea echoes—but is distinct from—the concept of the Value–Action Gap, which deals with the space between what people believe and how they behave in behavioral and sustainability studies (Essiz et al., 2022). What was observed suggests that another gap may come even earlier in the process: a gap between internal priorities and expressed contributions. And if certain values remain unspoken in key decision-making moments, it raises the question—could that silence influence what ultimately gets acted upon?

The purpose of highlighting this gap is not to diagnose or explain its causes, but to bring it into view. Rather than offering an answer, this insight aims to open reflection—inviting individuals and teams to consider:

- Which values they bring into decision-making spaces,
- Which values go unspoken, and
- What this might mean for the outcomes they co-create.

While future research can further investigate the social, cultural, or psychological reasons behind this phenomenon, the intention at this stage was simply to acknowledge its presence—because seeing the gap is often the first step toward engaging with it.

5.4. Learnings to Next Step

The first deployment of the prototype culminated not in a formal evaluation, but in a public moment of reflection. During the Forum’s closing ceremony, a short video visualizing the surfaced values was shown to all participants. It highlighted dominant values, overlooked ones, frequently co-occurring pairs, and — most strikingly — the gap between participants’ pre-survey priorities and the values they expressed during the event. This final showcase wasn’t just a summary; it was a test of the prototype’s ability to spark awareness, prompt reflection, and support meaningful storytelling.

What Emerged Through the Process

Along the way — from early conversations about the experiment to the selection of what would be shown — several key observations were made:

Ethical considerations arose early, particularly around the idea of surfacing values from high-stakes conversations. Since the Forum was a closed conference involving decision-makers, concerns were raised about the implications of analyzing such discussions. However, trust was maintained through prior consent from speakers and the prototype’s local implementation — all processing occurred offline, with no data stored or shared beyond the experiment.

Collaborative insight selection took place when the prototype’s outputs were reviewed by the organizing team. From a wide range of detected patterns, the most resonant were:

- Dominant and missing values: which showed what received attention and what was left unspoken.
- Value co-occurrences: revealing how certain values appeared in tandem, hinting at underlying tensions or bundled priorities.
- The value expression gap: showing mismatches between stated priorities (pre-survey) and actual dialogue, which became the most discussed and memorable takeaway.

What These Reflections Revealed

These moments of use, acceptance, and reflection pointed to several important insights:

- **Proof of potential:** The prototype worked — not only in a technical sense, but in its ability to provoke discussion, curiosity, and awareness.
- **Real-world resonance:** Acceptance by the organizers and inclusion in the public program demonstrated its relevance and adaptability in real settings.
- **The “aha” moment of the value expression gap:** This gap offered a mirror to participants, helping them reflect on the dissonance between what they say they value and what they actually emphasize in conversation — revealing blind spots and prompting self-inquiry.

Toward a Refined Prototype

While promising, the prototype also showed clear limitations that shaped the direction of future development:

- **Limited Value Set Constrained Expressiveness:** The model operated on a selected set of 20 values tailored to the Forum’s theme. While this ensured contextual relevance, it restricted the expressiveness of the model, sometimes failing to capture values that were present but not among the prompted set.
- **Over-Simplified Assignment Logic:** The system was configured to assign only the top one or two values per utterance. While this improved readability, it often failed to capture cases where more than two values were simultaneously present — especially in complex or layered expressions.
- **Need for Deeper Understanding:** The curiosity sparked by the prototype also revealed a desire to understand the why behind the values expressed. While the model could surface what values were present, it lacked the depth to interpret underlying motivations or contextual tensions.

These limitations, along with the earlier observations from the field, were further synthesized through a structured SWOT analysis (See Appendix E) conducted after the Forum. The analysis helped bring clarity to what was working, what needed improvement, and where the greatest opportunities for refinement lay. By mapping the strengths (e.g., ability to provoke reflection), weaknesses (e.g., limited value granularity), opportunities (e.g., deeper stakeholder engagement), and threats (e.g., ethical concerns in sensitive contexts), the SWOT served as both a reflective and strategic tool.

A Shift Toward Accuracy and Interpretive

Rather than broadening the prototype’s functionality or scaling prematurely, a strategic pivot was made: to go deeper, not wider. While the first deployment revealed exciting possibilities — from uncovering co-occurring values to identifying overlooked patterns — it also became clear that a strong foundation was essential before building further. Ensuring accuracy and interpretability at this stage would provide the necessary grounding for future research and more advanced applications.

The following refinements were prioritized:

- **Expand and diversify value utterance examples,** to better reflect how values are actually spoken about — both implicitly and explicitly.
- **Implement soft assignment logic,** enabling the model to capture multiple values per utterance when layered priorities emerge.
- **Lay groundwork for future layers,** such as motivational intensity, value conflict, or sentiment — to support deeper analysis and richer insights down the line.

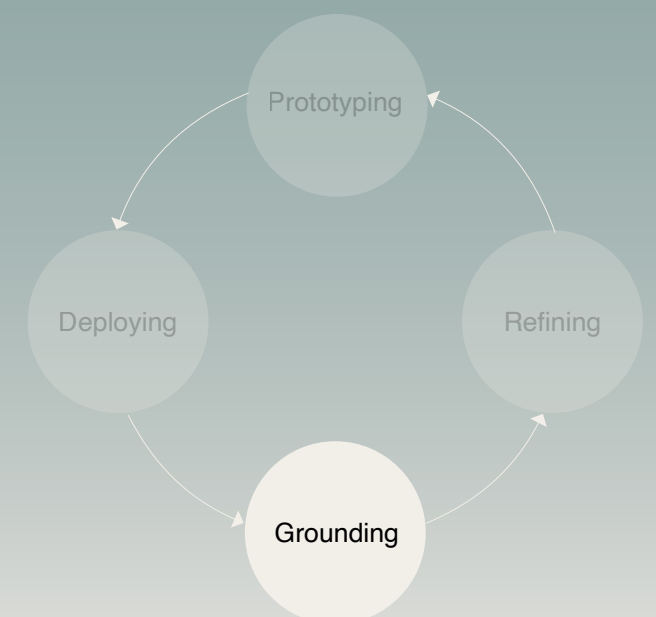
It was also recognized that improving accuracy wasn’t just about refining code. It meant better understanding how values show up in natural language. As a result, the next step was to observe real-world workshops focused on values in practice, in order to surface linguistic nuances and everyday expressions that may have been overlooked during early development.

Chapter 06

Observing Value-in-Use

This chapter grounds the method in real-world conversations by analyzing how values surface during three collaborative workshops. The findings emphasized that values are often indirect, emotional, and context-dependent — insights that shaped how the prototype was refined to better reflect how people naturally speak about what matters to them.

- 6.1. Design of Exploratory Workshops
- 6.2. How People Talk About Values
- 6.3. From Insight to Design Focus



6.1. Design of Exploratory Workshops

To better understand how people naturally articulate values in conversation, a series of exploratory workshops were designed and conducted. These sessions aimed to observe how values surface, shift, and sometimes clash during group discussions and collaborative reflection.

The workshops were co-designed and facilitated in collaboration with another graduation student at TU Delft as part of a research initiative focused on creating “brave spaces” — environments where participants feel psychologically safe and confident to express personal and professional values. Participants were university students, representing diverse roles across student-led, project-based teams — including designers, engineers, and project managers — to simulate realistic collaborative settings.

Three workshops were conducted:

- **Two workshops** with Dream Team members from different teams at TU Delft
- **One workshop** as part of a Climate Fresk session, aligned with the global Climate Fresk movement

Each session included 3–4 participants and followed a consistent three-part structure:

- **Deep Dive into Values:** Participants identified and reflected on a personal core value, with the help of values tokens (see appendix F), relevant to team collaboration, then explored its deeper motivations and limits through paired dialogue and group reflection.
- **Clash of Perspectives:** Using provocative statements and role-switching, participants silently responded to conflicting viewpoints, surfacing implicit tensions and examining values from multiple perspectives.
- **Flip the Tension:** Participants selected a key tension and reframed it by exploring what might be lost without it. This opened a discussion on how to productively navigate competing values rather than resolving them.

Throughout the workshops, structured observation notes were made, paying close attention to the language, patterns, and framing participants used when expressing values. The next section will be discussing these observations in detail.

6.2. How People Talk About Values

The workshops revealed that people rarely talk about values in neat, labeled terms. Instead, values tend to surface in indirect, fragmented, or emotionally charged ways — embedded in stories, expressed through tone, or hinted at through metaphor and action. Even when participants were offered a predefined list of values, many struggled to articulate what they stood for without multiple rounds of reflection. This underscored an important insight: value expression in natural conversation is rarely straightforward. It is layered, contextual, and often intuitive — making it both rich in meaning and challenging to detect. This section unpack key patterns observed during the workshops and what they imply for the development of a value elicitation method.

A. Indirect and Layered Expressions

One striking observation from the exploratory workshops was how rarely values were expressed through direct or literal language. Instead, participants articulated what mattered to them through a wide range of conversational forms — many of which would fall outside the reach of traditional value detection methods.

These expressions can be grouped into four ways:

- **Stories, narratives, or lived experiences**
Participants often shared personal anecdotes or concrete situations to express what they valued, even without naming the value explicitly. For example, a participant remarked:
“I don’t want to build the car alone. In the end, if everyone’s not happy, you have to do it alone.”
- Workshop Participant

While the word “collaboration” or “team happiness” was never used directly, the narrative conveyed a deep commitment to collective well-being and shared ownership.

- **Implicit expressions via actions, emotions, or design choices**

In some cases, values surfaced through emotional framing, decisions, or subtle justifications of action. For instance, one participant explained:

“You help them get help, so your group can still focus. That’s still your main value.”
- Workshop Participant

This framing implies a sense of responsibility or balance between individual support and group performance — even though no value was overtly named.

- **Expressive phrases (slang, jargon, metaphor)**
Some values were hinted at through metaphorical language or commonly used phrases. For example:

“It’s like a circle of death.”
- Workshop Participant

This metaphor reflected the pressure and emotional toll of inefficient team dynamics, pointing toward values like mental well-being or organizational clarity without explicitly labeling them.

- **Explicit naming of values (core or aspirational)**
While less frequent, there were also moments where participants did name values directly — often when prompted or after moments of reflection. Examples included:

“Fairness is really important to me,”
- Workshop Participant
“I guess I’m always aiming for independence.”
- Workshop Participant

Implications

These findings make clear that if a prototype is to detect values with any degree of realism or nuance, it cannot rely solely on direct mentions or pre-defined keywords. Value expression in conversation is layered, emotional, and often implicit. As such, the next step focused on enriching the value example dataset to include a broader

variety of expression types — from lived narratives and emotional cues to metaphors and indirect phrasing. This became the primary focus for refinement at this stage of the project.

B. Values Were Hard to Put into Words

Another recurring observation was how difficult it was for participants to name the values they were referencing — even when those values clearly shaped their thinking or actions. Despite having a list of value tokens available during the session, many participants circled around an idea or feeling for several minutes before settling on a word that felt right. In some cases, they never found one at all.

One participant, for example, described repeated efforts to support a struggling teammate but couldn’t immediately identify what value was driving that behavior. Only after reflection did they begin to associate it with care or responsibility. Another participant shared a detailed story about wanting to be taken seriously in group discussions, but hesitated to call it respect until others helped name it.

This difficulty wasn’t due to a lack of thoughtfulness — rather, it pointed to the complexity of values as lived experiences. Articulating them often required a mix of introspection, social prompting, and contextual framing.

Implications

This reinforced the core premise behind the prototype: that computational tools might support value articulation by surfacing patterns and signals that are hard to recognize — even for the speaker themselves. The challenge of naming values doesn’t mean they aren’t present; it means that a method capable of detecting latent or intuitive value expression can offer real support in making the implicit more visible.

C. Values Shift with Role and Context

Value expression was not only shaped by personal beliefs but also by the roles participants adopted during the workshop. When participants were asked to speak from the perspective of different stakeholders — such as a project manager, teammate, or external partner — the values they emphasized often shifted accordingly.

For example, in one session, a participant spoke from the role of a team lead under tight deadlines and focused heavily on efficiency and accountability. Yet, when speaking as a peer reflecting on group dynamics, the same person emphasized well-being and inclusivity. This shift didn't mean one set of values was more authentic than the other — it showed that value expression is situational, influenced by responsibilities, expectations, and interpersonal dynamics.

In another instance, participants who initially expressed strong opinions softened or reframed their views after role-switching exercises. These shifts revealed not only the presence of multiple value perspectives but also how empathy and context can bring latent values to the surface.

Implications

These observations highlight that value expression is fluid rather than fixed. Any detection method must remain sensitive to this variability — understanding that people speak from shifting positions, and that what they express is often shaped by the context they are placed in. While this dynamic quality is not yet modeled in the current prototype, it points to the importance of designing datasets and systems that reflect multiple framings of the same value.

D. Emotions, Hesitations, and Soft Signals

Not all value expressions were clear or confident. In many cases, participants spoke with hesitation, uncertainty, or emotional weight — especially when navigating tensions or sharing personal experiences. These soft signals, while subtle, carried meaningful information about what participants valued and how firmly those values were held.

Some participants used emotionally charged language — not to name a value directly, but to express how certain situations made them feel. One participant, when reflecting on productivity pressure, remarked: “That’s the frustrating part.” Another, navigating group responsibility, shared: “I feel helpless.” These statements conveyed an underlying tension between productivity and well-being without naming either.

Hesitations also revealed something important. Phrases like “I’m not sure, but I guess...” or “Maybe it’s just how I was raised” signaled that participants might still be in the process of forming or reevaluating their values. These moments weren’t noise — they were evidence of values in transition.

Implications

Emotional tone, uncertainty, and hesitation should not be dismissed as irrelevant or unclear. Instead, they can be treated as valid expressions of evolving values — especially in collaborative settings where people are negotiating priorities in real time. A prototype designed to detect values must be attuned to these soft cues, recognizing that value articulation is not always confident, consistent, or complete.

While a wide range of observations emerged from the workshops, this research chose to focus on a select few that could be directly addressed within the scope of the current prototype. These were considered most critical for improving the model’s immediate accuracy and relevance. On the other hand, the remaining insights point to promising directions for expanding the depth and sensitivity of the value elicitation method over time.

6.3. From Insight to Design Focus

These insights confirmed that improving the prototype’s ability to detect values would require more than algorithmic tuning — it would require grounding the model in how people actually talk.

To move forward, the refinement process focused on one clear priority: enriching the value example dataset. The original dataset had been developed using straightforward example phrases. While useful for initial testing, it lacked the expressive diversity observed in real conversations.

By training the prototype on this expanded range of expressions that were observed, the goal is to improve its ability to surface values that reflect not only what people say, but how they say it — including narrative expressions, emotional or metaphorical phrasing, implicit cues and explicit naming.

Other insights from the workshops — such as the influence of role and context, or the need to capture value tensions and hesitations — were recognized as essential for future development. These will inform the next staged beyond this thesis, where additional layers such as role-awareness, tension mapping, or confidence detection can be designed to extend the model’s interpretive depth.

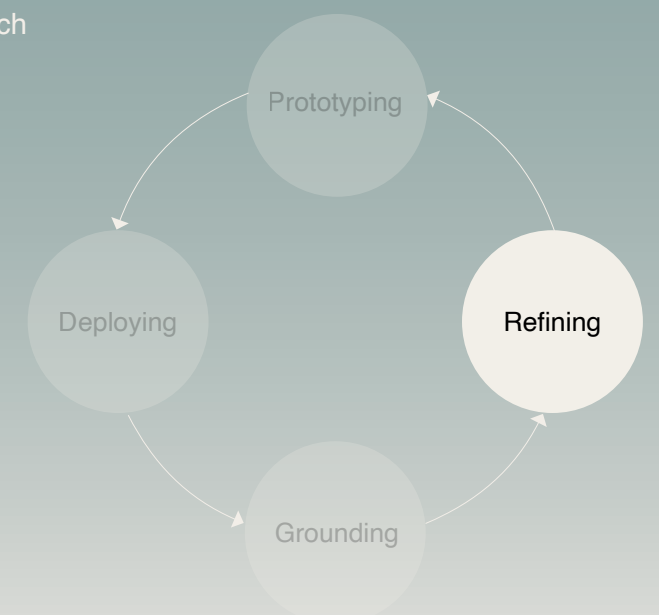
This shift marked a turning point: from simply identifying values, to exploring the nuanced ways in which they emerge in real-world dialogue — making the prototype not just more accurate, but more attuned to the complexity of human expression.

Chapter 07

Refining and Reflection

This chapter outlines the major refinements made to enhance the prototype's accuracy and interpretability. These included expanding to 108 values, diversifying utterance examples, testing scoring strategies, introducing motivational goals, and shifting to a soft assignment method — all aimed at better capturing the nuance and ambiguity of real-world value expression.

- 7.1. **Refining the Method**
 - 7. 1.1. Expanding the Value Dictionary
 - 7. 1.2. Evaluating Value Elicitation Strategies
 - 7. 1.3. Adding a Motivational Layer
 - 7. 1.4. Adopting a Soft Assignment Approach
- 7.2. **Demonstrating Results**
- 7.3. **Reflection on the Method**



7.1. Refining the Method

Building on learnings from early testing, deployment, and workshop observations, this section focusses on improving the prototype’s accuracy and interpretive depth. This phase aimed to refine the underlying method by enriching the value example dataset, testing more robust strategies for value assignment, and introducing an additional motivational layer — all designed to better reflect the nuanced ways in which people express values in conversation.

A key shift during this phase was also the move from hard assignment (one or two fixed values per utterance) to a soft assignment approach, allowing the model to capture multiple overlapping value signals when appropriate, rather than reducing rich expressions to a single label and open the model to further exploration.

7. 1.1. Expanding the Value Dictionary

In the early prototype, only 20 selected values were used, which constrained the accuracy and interpretive depth of the model. Many important value signals went undetected simply because they were not represented in the prompting dataset.

To overcome this limitation and build a stronger foundation, the full set of 108 values from the project’s underlying framework was incorporated into the model (See figure 18). While this set is still not exhaustive, it provides a comprehensive starting point — one that significantly expands the model’s semantic coverage and improves its ability to align with the varied and nuanced ways values appear in real conversations.

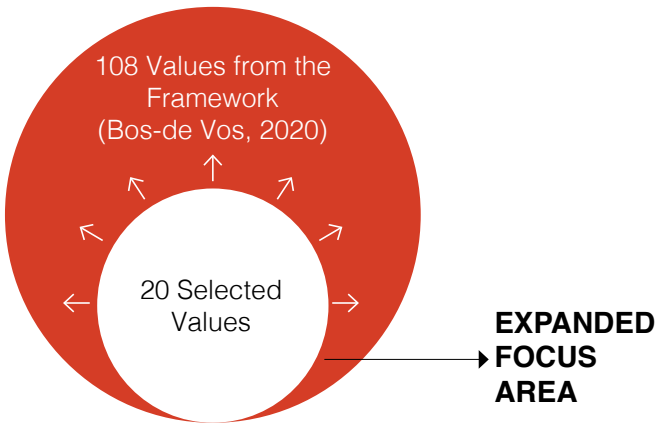


Figure 18 illustrates how the model expanded its scope by moving from 20 selected values to the full set of 108 values defined in the Bos-de Vos (2020) framework.

Building the Value Dictionary

To create a consistent and interpretable foundation for value elicitation, a comprehensive value dictionary was developed. Each value was documented and built on the existing framework (Bos-de Vos, 2020) using a structured format consisting of:

Value Perspective
Identifies if value is a Guiding Principle (Individual/Group) or Quality with Worth (People/Environment).
Type of Value
Broad overarching dimension of value (e.g., Human, Cultural, Economic, Use, Social, Ecological).
Motivational Goal
Specific motivational goal from the adopted framework.
Value
Concrete value explicitly mentioned in the framework.
Description
Narrative derived from the framework and constructed with the help of ChatGPT. (See Appendix D for generation details)

This layered structure served multiple purposes. First, it ensured clarity for both readers and users of the prototype by offering a common reference point for what a particular value means when surfaced by the model.

Second, it helped disambiguate values that share labels but differ in meaning depending on context — for example, the word pleasure could represent a human, cultural, or use-oriented value. To address this, each value was tagged with its value type (e.g., pleasure_human, pleasure_cultural) to avoid confusion both in analysis and communication.

Third, and perhaps most importantly, the dictionary laid the groundwork for prompting LLMs to generate realistic utterance examples. The value descriptions formed the basis for prompting data, making sure that generated sentences aligned conceptually with the intended meaning of each value.

Given the volume of values (108), the descriptions for the dictionary were created using OpenAI’s ChatGPT for efficiency. This allowed for rapid prototyping while maintaining reasonable coherence. However, it is acknowledged that future versions should reference more literature-based or domain-specific sources to enhance validity and trustworthiness. (See Table 08 for example of the value description and Appendix H for the full dictionary.)

Value	Description
Pleasure_human	Pleasure is a human value rooted in the motivational goal of enjoyment. It guides individuals to seek experiences that produce immediate positive feelings, often through activities that are emotionally or physically satisfying.
Pleasure_cultural	Pleasure as a cultural value reflects a group’s emphasis on individual enjoyment and emotional satisfaction. It supports norms that encourage personal expression and well-being.
Social Order	Social order is a cultural value that emphasizes structure, cohesion, and regulation within a group. It guides shared expectations that maintain stability and cooperation.

Table 08 presents sample descriptions from the value dictionary, which was used to help language models generate contextually accurate utterances during prototyping.

Example to show Improvement:

Utterance:

“We implemented clear roles and schedules to avoid confusion.”

Top Values Revealed by the earlier version of the prototype:	Top Values Revealed by the latest version of the prototype:
social power (0.4150)	time management (0.547)
equality (0.351)	autonomy (0.4719)
accessibility (0.332)	efficiency_social (0.4557)

The earlier model couldn’t detect time management, autonomy, or efficiency_social simply because they weren’t among the original 20 values — highlighting how limited coverage constrains insight.

Generating Example Utterances

While the ideal scenario would involve generating a large and diverse dataset of utterances for each value, this phase aimed to maintain a manageable scale in order to observe how small changes impacted the model’s performance.

Therefore, it was decided to generate eight example utterances per value, allowing for controlled experimentation while ensuring adequate variation.

These examples were grounded in the expression types identified during earlier workshop observations (See Figure 19):

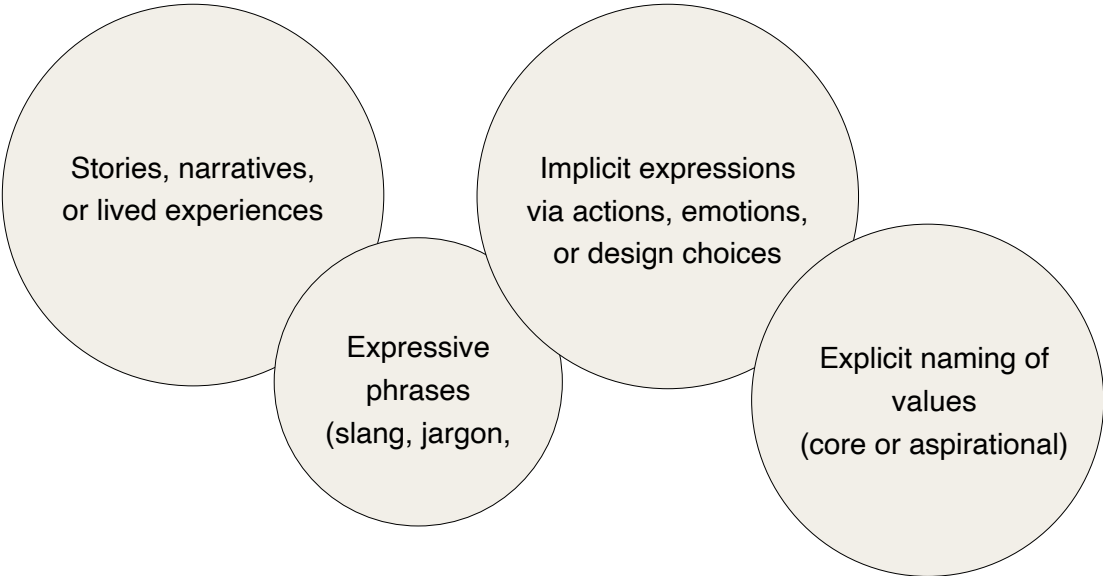


Figure 19 illustrates four types of value expression identified through earlier workshop observations: stories or lived experiences, implicit expressions through actions or emotions, expressive phrases (e.g., slang or jargon), and the explicit naming of values.

For each value, two utterances were generated per expression type, resulting in a total of eight utterances per value (See Table 09 for an example). These were created using OpenAI’s ChatGPT, with prompts that combined the value’s definition from the dictionary and specific instructions to simulate how that value might realistically be expressed in spoken dialogue. (Refer to Appendix D for the generation prompts.)

This curated set of utterances aimed to improve the model’s ability to detect both explicit and implicit value expressions, while reflecting the conversational richness and ambiguity observed in real-world speech.

By expanding the value dictionary and diversifying the example utterances, the prototype became better equipped to recognize values in naturalistic conversations — not just textbook phrases. This shift laid the groundwork for more realistic and context-aware value detection, allowing the method to move beyond surface-level interpretation and capture the nuances embedded in everyday speech.

Value: Pleasure

- 1. Stories, Narratives, or Lived Experiences
 - “Last weekend, I spent hours cooking with my partner—it wasn’t about the food, just the joy of being in the moment.”,
 - “Every time I visit that beach, I feel a kind of deep, simple happiness I can’t get anywhere else.”,
- 2. Expressive Phrases (Slang, Jargon, Metaphor)
 - “This is my guilty pleasure—don’t judge!”,
 - “That’s pure bliss in a bottle.”,
- 3. Implicit via Design Decisions, Actions, Emotions
 - “We added haptic feedback to make interactions feel more delightful.”,
 - “She lingered on the onboarding screen because the animation made her smile.”,
- 4. Explicit Naming (Core/Aspirational)
 - “Pleasure is a core value in how we craft user experiences.”,
 - “I try to design moments that spark immediate joy and pleasure.”

Table 09: Examples of eight generated utterances for the value “Pleasure,” covering four distinct expression types shown in Figure 19.

Example to show Improvement:
Utterance:

“That was straight fire after all that hustle”

Top Values Revealed by the *earlier* version of the prototype:

responsibility (0.095)
competitive_advantage (0.086)
success (0.072)

Top Values Revealed by the *latest* version of the prototype:

gratification (0.125)
responsibility (0.095)
competitive_advantage (0.086)

Once the dataset included slang-based expressions — like “The grind was real, but that win hit different” for gratification — the latest model could correctly pick up the value from casual phrasing. To see the testing process and examples used for slangs and stories, refer to Appendix I.

7. 1.2. Evaluating Value Elicitation Strategies

During the conference phase, the prototype used a single value assignment strategy: averaging the embeddings of all example utterances for a given value to create one ‘summary meaning.’ The model then compared a conversation utterance to this average vector to determine the closest value. While effective for a limited set of consistent examples, this approach began to fall short as the dataset evolved.

Once the value dictionary was expanded and enriched with a diverse range of expression types — from stories and metaphors to slang and implicit cues — it became evident that a single averaged embedding could no longer represent the richness of a value. Different examples of the same value often varied so widely in tone and structure that their average lost semantic sharpness. In other words, the summary meaning began to blur the distinctions that made value detection meaningful.

To address this, the project explored two additional strategies, resulting in a total of three options for value assignment. This section evaluates each approach — averaging embeddings, averaging similarity scores, and using the maximum similarity — to identify which method best balances accuracy, nuance, and robustness when dealing with naturalistic conversation.

Exploring Strategies

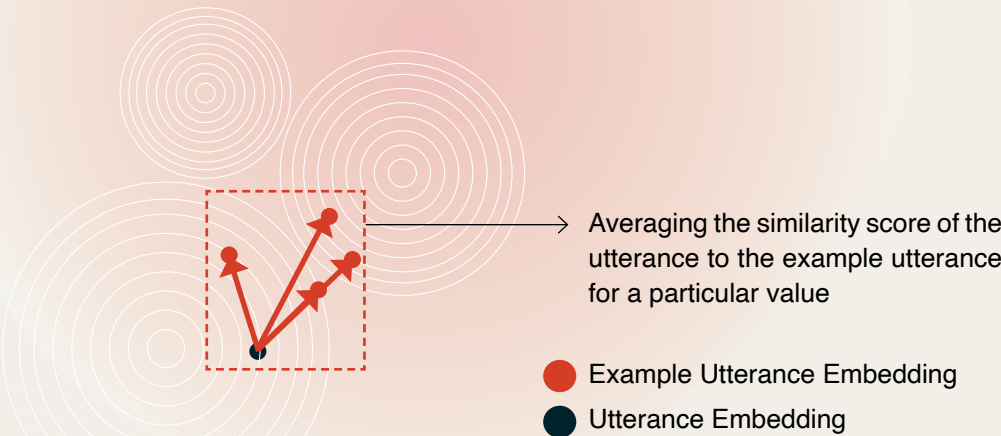
To identify the most accurate and interpretable way of surfacing values from conversation, three different strategies were tested. Each method used the same input: a sentence embedding of an utterance and a set of example embeddings for each value in the dictionary. However, they differed in how they calculated the “closeness” or relevance of a value to the utterance.

Option 02: AVERAGE SIMILARITY

Compare to each example, then take the average.

Instead of averaging the value examples first, each sentence was compared to every example individually, and the resulting similarity scores were then averaged.

This approach is more sensitive to nuance, as it captures the range of ways in which a value might be expressed.

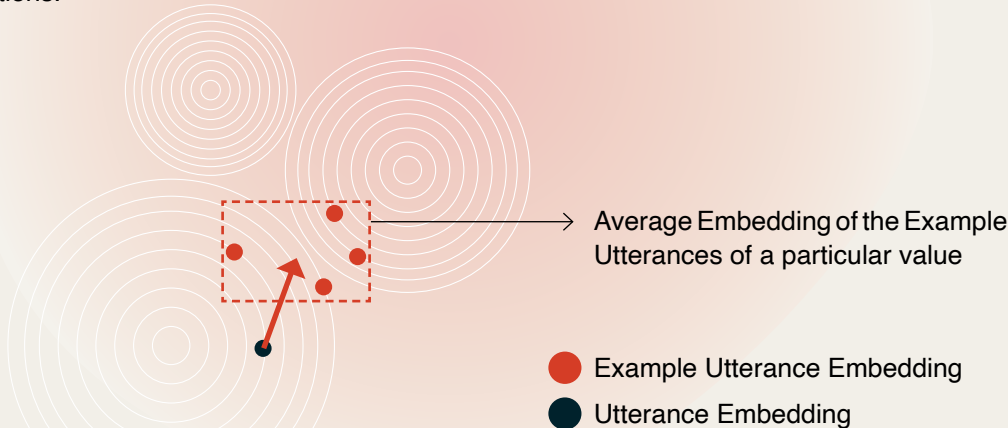


Option 01: AVERAGE VECTOR

Blend all meanings into one, then compare.

All example utterances for a given value — for example, the different ways people might express responsibility — were combined to create a single “summary meaning” using an averaged vector. The similarity of each sentence to this averaged meaning was then measured.

This method tends to generalize when the example expressions vary widely, such as across slang, stories, or formal definitions.

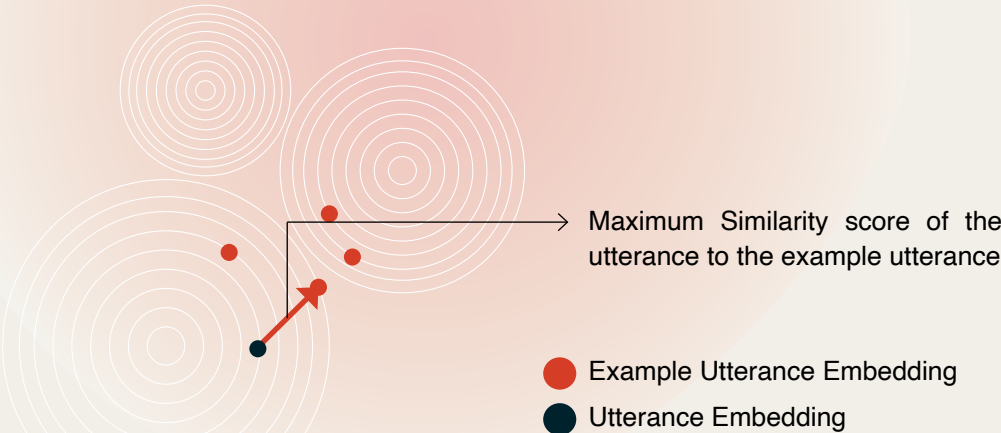


Option 03: MAXIMUM SIMILARITY

Find the closest match and go with that.

In this approach, each sentence was compared to every example, and the single highest similarity score was selected as the match.

While this method often identifies a strong match, it can also misfire — particularly when one example is an outlier or only loosely related to the sentence.



To evaluate which strategy could most effectively support nuanced value elicitation, all three methods were implemented and tested — from averaging embedding vectors to analyzing individual similarities. Each approach offered a distinct balance between generalization, specificity, and interpretability. To understand their real-world behavior, it was necessary to observe how these strategies performed when applied to actual conversation.

Example case:
Utterance:

“Honestly, I’m less worried about the numbers and more about whether people trust it. That’s what makes or breaks adoption in teams like this.”

Example Case: How the Strategies Compare

While the methods were tested across a broader set of utterances, one illustrative example is shared here to show how each approach handles nuance, context, and meaning. The following utterance is used to compare the performance of the three strategies:

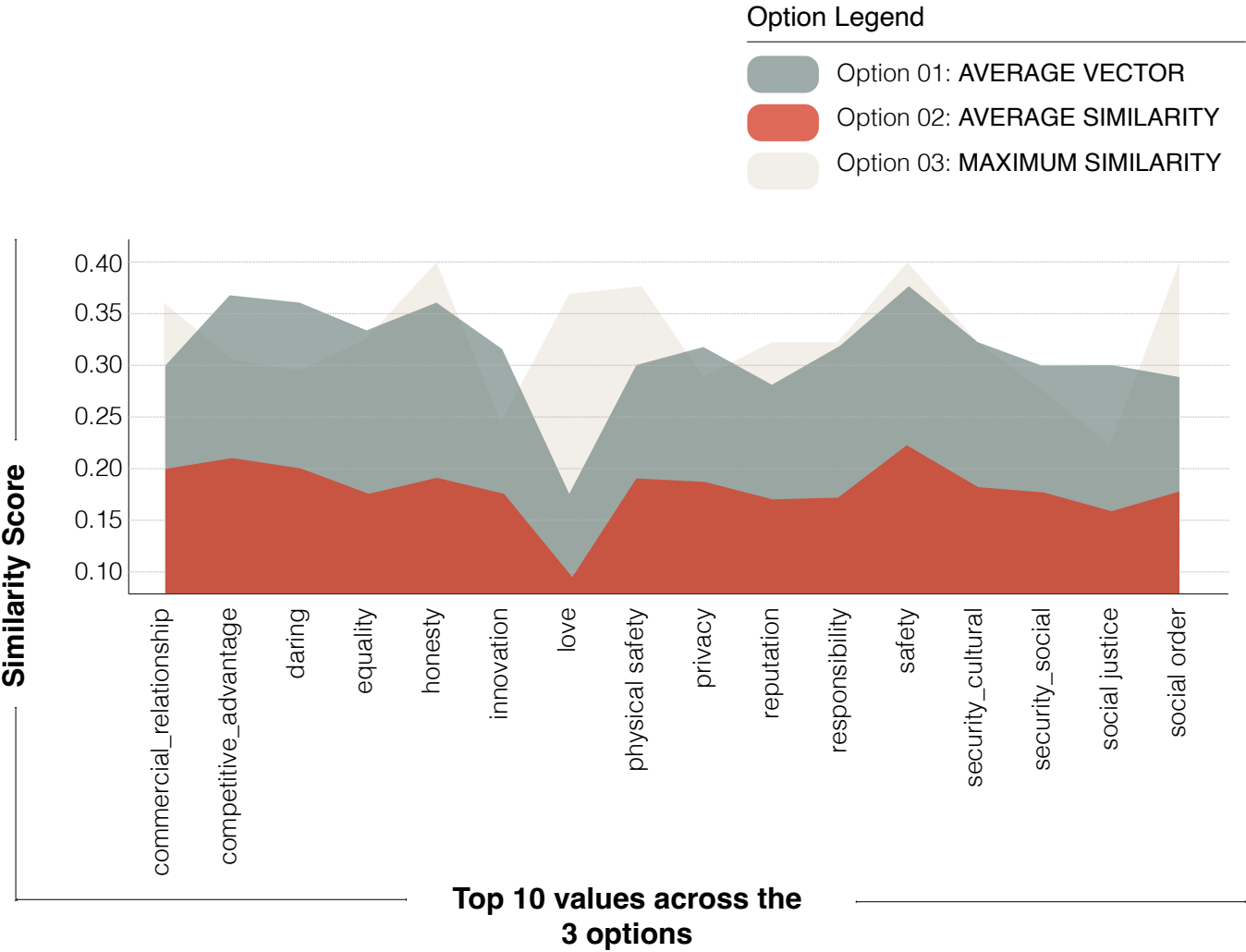


Figure 20: Comparison of three similarity scoring strategies applied to a single utterance, highlighting how each method surfaces different scores for the top 10 values—revealing how distinct approaches (average vector, average similarity, and maximum similarity) influence which values are prioritized and how strongly.

Option 01	Option 02	Option 03
AVERAGE VECTOR	AVERAGE SIMILARITY	MAXIMUM SIMILARITY
safety 0.3749	safety 0.2187	social_order 0.3996
competitive_advantage 0.3662	competitive_advantage 0.2072	honesty 0.3995
daring 0.3588	daring 0.2011	safety 0.3932
honesty 0.356	commercial_relationship 0.1975	physical_safety 0.3747
equality 0.3326	privacy 0.1874	love 0.3648
security_cultural 0.3189	physical_safety 0.1872	commercial_relationship 0.3555
responsibility 0.3188	honesty 0.1864	equality 0.3267
privacy 0.3135	security_cultural 0.1818	responsibility 0.3212
innovation 0.3133	security_social 0.1773	reputation 0.3207
social_justice 0.3018	social_order 0.1767	security_cultural 0.3179

Table 10: Top 10 value scores for a single utterance across three scoring methods—highlighting how each approach (average vector, average similarity, and maximum similarity) produces different rankings and reveals distinct interpretations of meaning.

Observations:

As seen in the graph comparing the top 10 values across the 3 options (Figure 20), there is substantial overlap between the top values revealed by Option 01 (Average Vector) and Option 02 (Average Similarity), while Option 03 (Maximum Similarity) produces a more distinct — and at times, unexpected — set of results.

- Option 01 tends to favor more generic, high-frequency values with relatively high confidence scores. However, its reliance on a single averaged representation (“summary meaning”) means it sometimes misses nuanced or context-specific values, flattening layered intent.
- Option 02 offers a balance — it shares many results with Option 01 (see Figure 20), but also captures more subtle values such as commercial_relationship. Notably, if the top 10 values were

based on higher similarity score across all options (see Table 10), commercial_relationship would not appear in Option 01. However, it is clearly surfaced in both Option 02 and Option 03, suggesting these strategies are more attuned to context-specific signals. While Option 02 yields lower confidence scores, the increased nuance and contextual relevance make it a more robust approach overall.

- Option 03 displays the highest confidence scores (see Figure 20) and surfaces values like social_order and responsibility, which align with the utterance’s concern about trust and group dynamics. However, it also introduces outliers — for example, physical_safety and love — that may not be contextually appropriate. These likely emerged from edge-case matches in the value example set (see Appendix L for details).

These observations laid the foundation for selecting the most appropriate value elicitation strategy. To support the decision further, the comparison was visualised using these observations and the classic learning trade-off between underfitting, robust generalization, and overfitting. (see Appendix M and Table 11)

Together, these comparisons made the trade-off clear: while Option 01 was too simplified to capture conversational nuance, Option 03 offered precision at the cost of stability — often surfacing false positives with high confidence. Option 02 struck the right balance. It was robust enough to handle varied expressions without overreacting to linguistic quirks, making it the most reliable and interpretable choice for value elicitation in real-world conversations.

	Option 01	Option 02	Option 03
	AVERAGE VECTOR	AVERAGE SIMILARITY	MAXIMUM SIMILARITY
Method	Average embedding of all examples per value	Average similarity with each example per value	Highest similarity with one example
Strengths	Stable and simple	Captures nuance across varied expressions	Picks up rare or subtle signals others may miss
Limitations	Overgeneralizes — flattens diverse meanings	Slightly lower scores, but interpretable	Overreacts — prone to outliers and false positives
Risk Profile	Underfit — too simplified to catch complexity	Balanced — handles diversity without overfitting	Overfit — sensitive to noise and individual quirks
Best Used When	Dataset is small and value expressions are uniform	Values are expressed in diverse, conversational ways	Controlled settings or when top-match accuracy is prioritized

Table 11: Summary comparison of the three similarity strategies, outlining their method, strengths, limitations, risk profile, and recommended use cases to inform selection of the most appropriate approach for value elicitation.

7. 1.3. Adding a Motivational Layer

To deepen the interpretive power of the prototype, an additional motivational goal layer was introduced. This layer builds on the framework adopted in the literature review (see Section X), where each value is associated with a broader motivational goal — such as enjoyment, well-being and development, mastery, security, or economic interest.

In practical terms, once a value is elicited by the prototype, it is also tagged with its corresponding motivational goal. This tagging was achieved by feeding the model the defined mappings between values and motivational goals, as outlined in the framework before from literature (See Table 12 for example and entire mapping in Table 04). As a result, users can now not only see what values are being expressed, but also gain insight into why — revealing the underlying drivers and thematic intentions behind conversations.

Motivational Goal	Value
Enjoyment	pleasure, self-indulgement, gratification, sensuous enjoyment, happiness at work,...
Security	physical safety, psychological/ mental health, integrity,...
Achievement	achievement, competence, success,...
Self-Direction	autonomy, self-sufficiency, independence, intellectualism,...
Restrictive-conformity	conformity to social expectations,...
Prosocial	altruism (e.g. acting in best interests society/client), benevolence, kindness, love,...
Social power	dominance, status, influence, social control, power, leadership, authority,...
Maturity	wisdom, tolerance, faith in one's convictions, deep emotional relationships, appreciation for the beauty of creation,...

Table 12: Illustrates example of how each value is linked to a broader motivational goal (e.g., enjoyment, security, achievement), enabling the prototype to reveal not just what values are present in a conversation but also the deeper drivers behind them.

Example: Motivational Breakdown of an Utterance

“Honestly, I’m less worried about the numbers and more about whether people trust it. That’s what makes or breaks adoption in teams like this.”

Top elicited values and their motivational goals:

Value	Motivational Goal
Safety	Well-being and development
Competitive_advantage	Other economic values
Daring	Mastery
Commercial_relationship	Other economic values
Privacy	Social proprietary

From this, it becomes evident that the speaker’s concern centers not just on isolated values but on a multi-motivational stance: prioritizing well-being and trust, while also safeguarding economic viability, individual courage, and social boundaries.

This added layer helps shift the analysis from what was said to what drives what was said — making the output more actionable for reflection, decision-making, or value alignment discussions in teams.

At this stage, the motivational goal is not detected directly from the utterance but assigned based on the value-to-goal mappings defined in the framework (see Table 04). While this offers useful interpretive cues, it assumes a fixed relationship between each value and its corresponding goal — a limitation that has the potential to be explored further.

7. 1.4. Adopting a Soft Assignment Approach

As observed throughout earlier chapters, a single utterance often reflects more than one value — sometimes layered priorities, subtle tensions, or co-existing goals. In real-world conversations, values are rarely clean-cut; they appear in clusters, overlaps, or as implicit undercurrents.

This became especially evident when analyzing the similarity scores across values for the same sentence. In many cases, multiple values scored closely, suggesting that narrowing down to a single “top value” would result in a loss of nuance.

Given that this research prioritizes exploration over prescription, a soft assignment approach was adopted. Surfacing top 5-10 values allowed to interpret which values feel most salient and why, offering space for reflection and nuance. Instead of enforcing a singular interpretation, this approach supports open-ended analysis — more aligned with the exploratory nature of this project.

Take, for instance, the earlier utterance again:

“Honestly, I’m less worried about the numbers and more about whether people trust it. That’s what makes or breaks adoption in teams like this.”

Value	Similarity Score
safety	0.2187
competitive_advantage	0.2072
daring	0.2011
commercial_relationship	0.1975
privacy	0.1874
physical_safety	0.1872
honesty	0.1864
security_cultural	0.1818
security_social	0.1773
social_order	0.1767

Table 13: Shows the top 10 value similarity scores for a single utterance under the soft assignment approach, illustrating how multiple values can score closely and co-exist — enabling richer, layered interpretations rather than forcing a single dominant value.

Moreover, this method opens new possibilities for understanding value interconnectedness. As shown in Table 13, for the utterance “Honestly, I’m less worried about the numbers and more about whether people trust it...”, values such as competitive advantage and daring appear with closely matched scores. If explored further, this could suggest that daring may play a role in achieving competitive advantage within such a context. While still speculative, soft assignment enables these interpretive layers to emerge — nuances that would likely be flattened or lost under a hard assignment approach.

With these refinements — including an expanded value dictionary, enriched example dataset, nuanced elicitation strategy, motivational layering, and soft assignment logic — the prototype is now better equipped to reflect the complexity and subtlety of value expression in real conversations. The next chapter demonstrates this refined method in use by analyzing a single conversation

7.2. Demonstrating Results

To show what the refined prototype can now surface, this section presents a single synthetic yet realistic conversation — ten utterances from a team discussion on deploying a new tool (the prompt used to generate this conversation can be found in Appendix N). The aim is to illustrate what the prototype can reveal in practice: how values emerge, overlap, and connect to deeper motivational drivers. This example serves two purposes — it invites the reader to reflect on the values surfaced by the prototype within the exchange, and it provides a grounded basis for the next subsection, where the output of the latest version is interpreted in context.

The Conversation

The excerpt on the right captures a short exchange between three team members. Their dialogue explores concerns around usability, trust, rollout strategy, and long-term impact — a context rich in value expression but sparse in explicit labeling.



The visual was created with the help of OpenAI's Sora.

We've been prototyping the tool for three weeks now. I think the biggest concern is not performance, but whether it's actually changing how people make decisions. We can't just deliver something shiny.

Utterance 01

Right. I've been talking to ops — if it makes them slow down, even a little, they'll ditch it. But if we can streamline their work without them noticing, it'll stick.

Utterance 02

That's why I pushed for that ambient feedback system. People shouldn't need a manual to figure out how it fits into their day.

Utterance 03

Agreed. Also, the pilot team mentioned something interesting — they felt like the tool helped them pause before rushing into decisions. That's not in the metrics, but... maybe it should be?

Utterance 04

Honestly, I'm less worried about the numbers and more about whether people trust it. That's what makes or breaks adoption in teams like this.

Utterance 05

I mean, trust is earned, right? We can't just tell them it's better — we need to show that it learns, adapts, and doesn't just throw generic insights at them.

Utterance 06

Exactly. And the rollout needs to be mindful — I don't want another case of "innovation theatre." If it doesn't reflect how they already work, we'll be back to square one.

Utterance 07

Can we also keep an eye on the footprint? The last iteration was heavier than expected, server-side. If we're aiming for long-term scale, we need to factor in efficiency early.

Utterance 08

Noted. I've been exploring edge processing anyway — could reduce dependency on external calls and actually speed things up.

Utterance 09

Good. Let's not forget — the goal isn't to build tech for tech's sake, but to actually support better decisions, at scale.

Utterance 10

What happens when the model listens?

Each utterance in the conversation was analyzed using the refined value elicitation method. Rather than assigning a single label, the model evaluates each utterance against all 108 values in the system and surfaces a distribution of matches, reflecting varying degrees of alignment.

The heatmap in Figure 21 below visualizes this distribution. Each row represents an utterance, and each column corresponds to a value in the system. Darker shades indicate stronger alignment. As the heatmap shows, most utterances relate to multiple values — not just one — highlighting the layered and overlapping nature of how people speak about what matters to them.

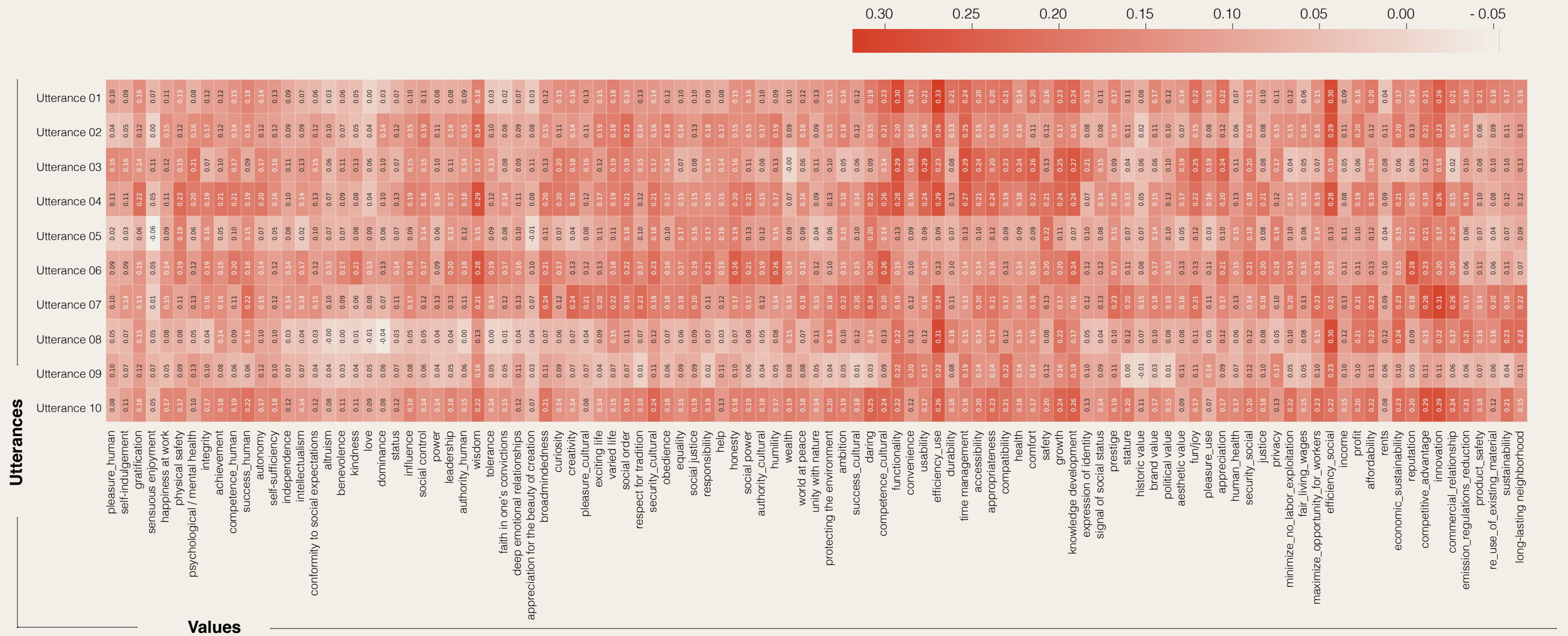


Figure 21: Heatmap showing similarity scores between 10 utterances and all 108 values in the system, revealing how each utterance aligns with multiple values to different degrees — highlighting the model’s ability to capture the layered, overlapping nature of value expression in conversation.

	Value	Value Type	Motivational Goal
Utterance 01	efficiency_use functionality efficiency_social	Use Value Use Value Social Value	utility utility social_wealth
Utterance 02	efficiency_social efficiency_use time management	Social Value Use Value Use Value	social_wealth utility utility
Utterance 03	usability time management functionality	Use Value Use Value Use Value	utility utility utility
Utterance 04	wisdom efficiency_use efficiency_social	Cultural Value Use Value Social Value	embeddedness utility social_wealth
Utterance 05	safety competitive_advantage daring	Social Value Economic Value Cultural Value	social_propriety other_economic_value mastery
Utterance 06	wisdom reputation competence_cultural	Cultural Value Economic Value Cultural Value	embeddedness other_economic_value mastery
Utterance 07	innovation competitive_advantage commercial_relationship	Economic Value Economic Value Economic Value	other_economic_value other_economic_value other_economic_value
Utterance 08	efficiency_use efficiency_social economic_sustainability	Use Value Social Value Economic Value	utility social_wealth money
Utterance 09	efficiency_social compatibility efficiency_use	Social Value Use Value Use Value	social_wealth utility utility
Utterance 10	innovation competitive_advantage efficiency_use	Economic Value Economic Value Use Value	other_economic_value other_economic_value utility

Table 14: Displays the top 3 values detected for each utterance, along with their corresponding value types and motivational goals. While it highlights key value signals, the note clarifies that many other values also scored similarly, meaning the table offers a simplified snapshot rather than a full picture of value expression.

We’ve been prototyping the tool for three weeks now. I think the biggest concern is not performance, but whether it’s actually changing how people make decisions. We can’t just deliver something shiny.

Utterance 01

Right. I’ve been talking to ops — if it makes them slow down, even a little, they’ll ditch it. But if we can streamline their work without them noticing, it’ll stick.

Utterance 02

That’s why I pushed for that ambient feedback system. People shouldn’t need a manual to figure out how it fits into their day.

Utterance 03

Agreed. Also, the pilot team mentioned something interesting — they felt like the tool helped them pause before rushing into decisions. That’s not in the metrics, but... maybe it should be?

Utterance 04

Honestly, I’m less worried about the numbers and more about whether people trust it. That’s what makes or breaks adoption in teams like this.

Utterance 05

I mean, trust is earned, right? We can’t just tell them it’s better — we need to show that it learns, adapts, and doesn’t just throw generic insights at them.

Utterance 06

Exactly. And the rollout needs to be mindful — I don’t want another case of “innovation theatre.” If it doesn’t reflect how they already work, we’ll be back to square one.

Utterance 07

Can we also keep an eye on the footprint? The last iteration was heavier than expected, server-side. If we’re aiming for long-term scale, we need to factor in efficiency early.

Utterance 08

Noted. I’ve been exploring edge processing anyway — could reduce dependency on external calls and actually speed things up.

Utterance 09

Good. Let’s not forget — the goal isn’t to build tech for tech’s sake, but to actually support better decisions, at scale.

Utterance 10

What the Prototype Can Reveal?

To make the results more interpretable, the table X on the previous page presents the top three value matches for each utterance, along with their corresponding motivational goals and value types. This offers a focused snapshot of the model’s layered output.

While limiting the view to the top three values does not capture the full richness of the model’s scoring — since many values have similar similarity scores — it serves as a helpful starting point to illustrate what the prototype can detect. A broader view is available in the heatmap shown earlier (see Figure 21), which reveals that each utterance often aligns with a wider network of values, many of which overlap or reinforce each other in subtle ways.

This kind of value mapping opens up multiple analytical possibilities for exploring a conversation in greater depth. The prototype’s layered output (see Table 14) makes it possible to:

Identify the dominant values shaping each utterance

For example, efficiency_use, competitive_advantage, and innovation appear multiple times across the conversation, highlighting a recurring focus on economic and operational impact. These dominant themes suggest that practical value delivery and strategic positioning are core concerns throughout the dialogue.

Understand the motivational framing behind those values

Many of the top values, such as efficiency_use and sustainability, are associated with motivational goals like utility, money, and social_wealth. This indicates that the conversation is not just about performance, but about scalable, trusted solutions that serve both organizational and societal needs.

Observe the types of values being prioritized

The table reveals a strong emphasis on use values and economic values, with occasional appearances of cultural and social values. This shows that while the team is primarily focused on implementation and impact, they also acknowledge the role of principles like trust, reputation, and wisdom.

Notice what is not being said

Despite discussing adoption and decision-making, explicit human values like gratification or independence are largely absent. This suggests that the emotional or individual perspective may be underrepresented in the conversation — a potential blind spot worth exploring in real-world reflection.

Track shifts and interplay over time

In earlier utterances, values like usability and functionality dominate, tied to immediate user experience. As the discussion progresses, efficiency, sustainability, and long-term thinking begin to surface — signaling a shift from short-term usability to broader strategic concerns. This progression shows how value dynamics evolve naturally in collaborative settings.

Together, these layers offer more than a surface-level reading. They allow for richer, more reflective engagement with everyday conversations — enabling researchers, designers, or teams to explore value dynamics in ways that support decision-making, alignment, and organizational learning.

In the next section, these insights are used to reflect on this specific conversation — not only to validate the prototype’s performance but to demonstrate its interpretive value in practice.

7.3. Reflection on the Method

This prototype does not mark the end of development — it marks the beginning. While the goal of this project was to explore how values can be surfaced from conversation using a NLP approach, the results so far are exploratory. What has been built is a foundation — and through the process of designing, deploying, and analyzing the prototype, a number of key learnings emerged that can help shape what comes next.

This section reflects specifically on the prototype results and technical implementation. A broader discussion of the full project is presented in the final chapter. Figure 22 below visualizes the four key themes along which the current prototype can be expanded in future work: making the system operational, deepening analytical power, scaling its scope, and expanding its conceptual foundations.

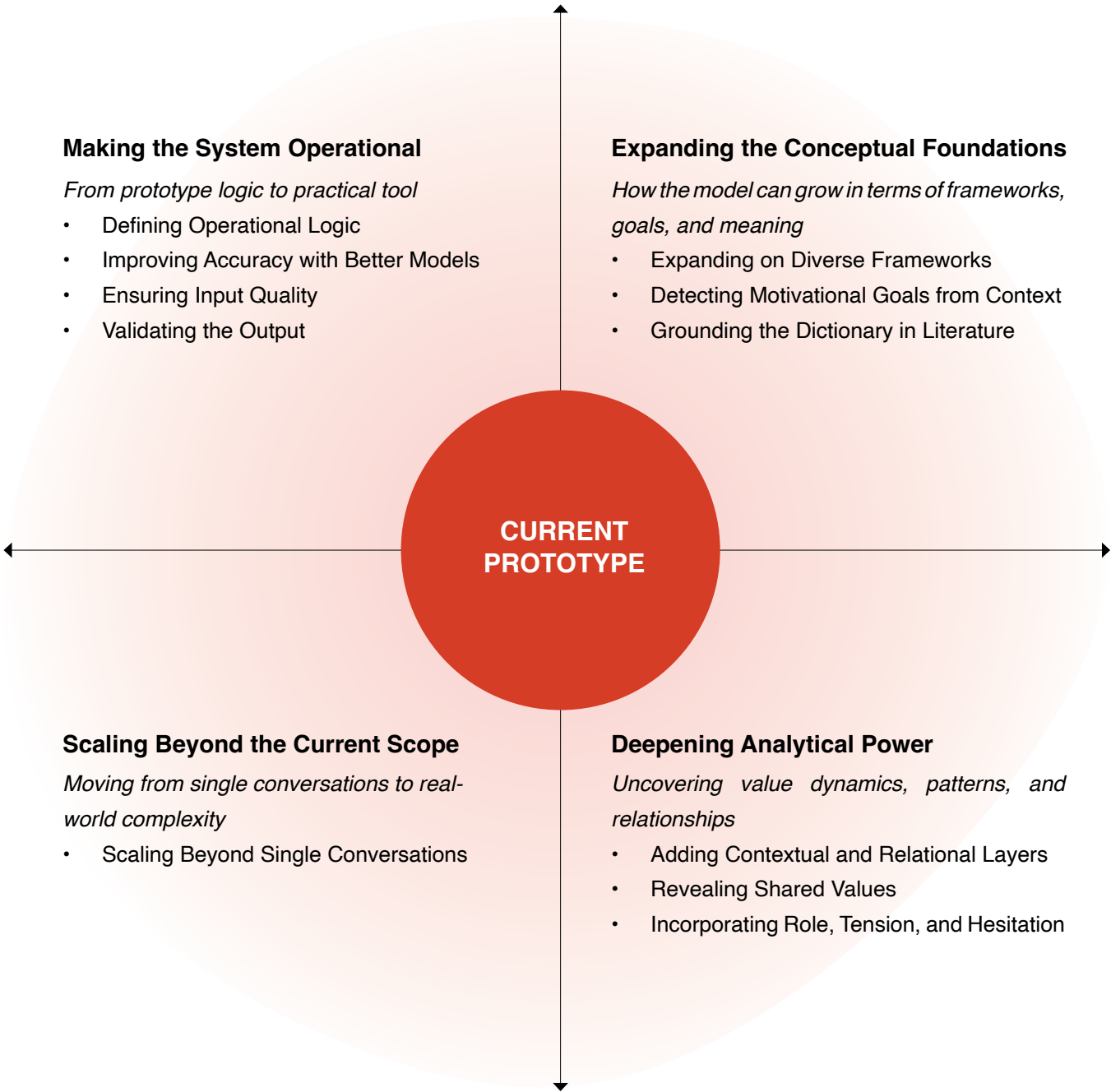


Figure 22: Four directions for improving the prototype
This diagram shows the future potential of the current prototype by outlining four areas for development: making it operational, expanding conceptual frameworks, deepening analytical power, and scaling to broader real-world settings.

Expanding the Conceptual Foundations

The prototype was built using one predefined value framework and a fixed set of motivational goals. While this helped demonstrate the method, future versions can expand its conceptual depth in several ways:

- **Integrating Diverse Frameworks**
Many disciplines offer their own value systems — from ethics to design to sustainability. These can be integrated to make the model more inclusive and adaptable. Technically, this is feasible, but conceptually, care is needed to align overlapping definitions and prevent redundancy.
- **Detecting Motivational Goals Dynamically**
In the current prototype, motivational goals are assigned based on the detected value (e.g., efficiency to economic interest). In practice, the same value can be motivated by different goals. Future versions could infer motivational framing directly from context or phrasing, rather than static linkage.
- **Grounding the Dictionary in Literature**
The current value dictionary was generated using LLMs. This made rapid prototyping possible, but future refinements should be grounded in real conversations and existing literature. Doing so would increase accuracy and contextual relevance.

Making the System Operational

As a prototype, the system is still exploratory. To evolve into a usable tool, it must address core questions around design logic, accuracy, and reliability:

- **Defining Operational Logic**
The current model assigns values using soft similarity scoring without a threshold. Any functional tool would require clearly defined parameters for:
 - Minimum similarity thresholds
 - Number of values shown per utterance
 - When to use soft vs. hard assignment
- **Improving Accuracy with Better Models**
The sentence transformer model was chosen on the basis of current requirements. Future iterations could use better models, self-trained models and value-rich data.
- **Ensuring Input Quality**
The system's accuracy depends heavily on the quality of transcripts. Noisy audio or poor transcription tools can lead to misleading results. Future use would require strategies for ensuring or correcting input quality.
- **Validating the Output**
So far, the interpretation of results has been based solely on reflective analysis by the author. No formal validation — such as expert annotation or participant feedback — has been conducted. A key next step is to design a structured validation process to assess the model's performance and reliability in real-world contexts.

Deepening Analytical Power

Beyond assigning values, the prototype can be developed to uncover how values behave in conversation — when they emerge, how they interact, and what roles they play.

- **Adding Contextual and Relational Layers**
Techniques such as clustering or hierarchical mapping can help show:
 - Which values co-occur and when
 - How utterances are thematically linked
 - What roles values play at different points in the conversation
- **Revealing Shared Values**
Future development could identify shared values by tracking the ones expressed by multiple speakers.
- **Incorporating Role, Tension, and Hesitation**
Observations from workshop settings showed that values shift based on speaker role, and that tension and uncertainty are often expressed emotionally or implicitly. These soft cues could be modeled using new layers — such as role-awareness, hesitation detection, and value tension mapping — to make the tool more responsive to nuance.

Scaling Beyond the Current Scope

The current method analyzes single conversations. Future development could support more dynamic, multi-session, and multimodal communication.

- **Analyzing Conversations Over Time**
Extending the model to analyze how values change across meetings, email threads or chats could support reflection at the team or organizational level. This would allow for deeper insight and reflection.

Chapter 08

Real-World Relevance

This chapter investigates how the method could be applied in organizational contexts through interviews with professionals from five industries. It identifies seven strategic use cases while also surfacing critical adoption conditions and ethical concerns — highlighting both the promise and the responsibility of using value elicitation in practice.

- 8.1. **From Development to Deployment**
- 8.2. **Strategic Use Cases and Value Potential**
- 8.3. **Integration, Adoption and Ethical Fit**

Please Note:

Throughout this chapter, the term tool is occasionally used to describe the prototype, although it is more accurately a code-based method developed to explore value elicitation in conversations. While it has not yet reached the level of a fully developed application, the term is used to reflect its potential for integration and real-world relevance. The reflections presented here are grounded in the assumption that this method could eventually be embedded into an interactive tool—one that individuals or teams might use to generate value-based insights.

8.1. From Development to Deployment

As the development of the value elicitation method matured, the focus gradually shifted from testing and refining its technical foundation to exploring its applicability in real-world settings.

During earlier phases, several potential application domains had emerged — ranging from organizational decision-making and academic research to civic participation. These early signals surfaced through exploratory workshops, ongoing discussions, and the Leadership Forum, where the method was first publicly deployed.

To better frame its future potential, the following possibilities are organized by key stakeholder domains that are increasingly expected to collaborate in value-driven decision-making (See Figure 23) :

Academic and Educational Contexts

- **Academic Research**
The method offers researchers a novel way to analyze stakeholder reasoning, detect implicit values, and study group dynamics in participatory or collaborative settings.
- **Design and Strategic Education**
In educational programs focused on design, strategy, or organizational behavior, the code can serve as a hands-on tool. It enables students to explore how NLP can be used to uncover conversational patterns and relate them to established value frameworks.

Business and Organizational Contexts

- **Organizational Decision-Making**
The method holds promise for supporting leaders in surfacing the value dynamics underlying their decisions. Early traction from the Leadership Forum highlighted its relevance for strategic contexts involving multiple stakeholders and competing trade-offs.
- **Team Collaboration and Student Groups**
In collaborative environments like TU Delft’s Dream Teams, like witnessed during workshops, the method could help reveal both individual and shared values, fostering alignment and mutual understanding within teams.

Civic and Governmental Contexts

- **Municipal and Political Settings**
The method could support more inclusive public dialogue and participatory governance. Applied in municipal settings or public consultations, it aligns with value-sensitive design principles by uncovering value tensions, overlooked priorities, and implicit assumptions in civic discourse.

Cross-Sector Dialogues and Events

- **Conferences and Events**
As demonstrated at the Leadership Forum, the method can help synthesize discussions by surfacing dominant or missing values—making events more reflective and insight-driven.
- **Joint Value Creation Sessions**
When integrated into stakeholder workshops or co-creation sessions, the method can act as a reflective prompt. It supports participants in identifying tensions, aligning goals, and enriching dialogue with computational insight.

Despite this range of possibilities, the scope for this phase of the research was intentionally narrowed to focus on organizational decision-makers. This choice was driven by the method’s natural alignment with decision-making scenarios and the momentum gained from its initial reception among professionals at the leadership forum.

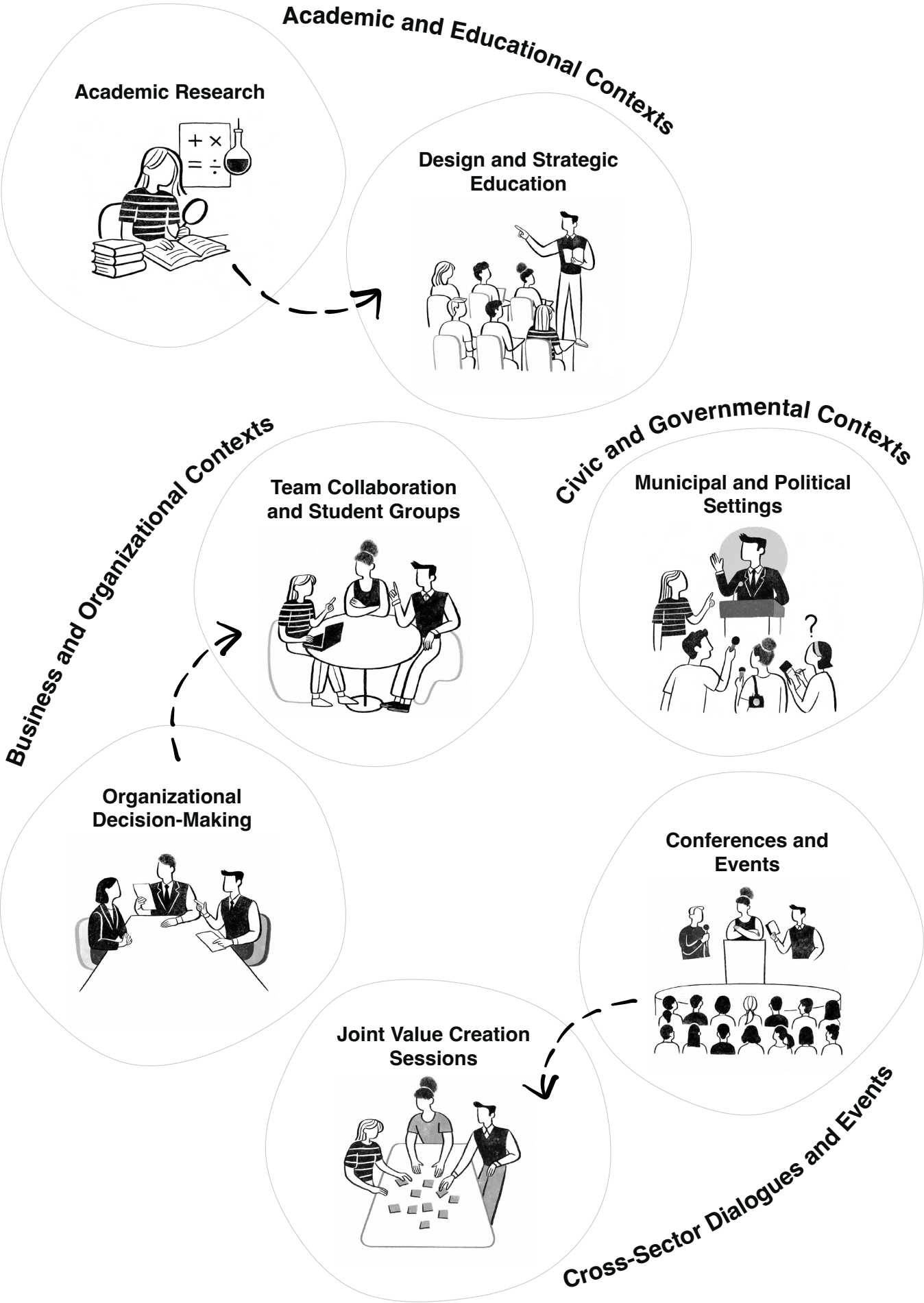


Figure 23: Application map showing contexts where the prototype can be applied (education, business, government, cross-sector) (The visuals were created by OpenAI's Sora)

Introduction

To explore how the value elicitation method might extend beyond its development phase and fit into real-world decision-making environments, five semi-structured interviews were conducted with professionals across diverse industries. The aim was to understand how such a tool might be perceived, used, or challenged in organizational settings—moving from proof of concept toward potential integration.

Interviewees were chosen based on their strategic roles and decision-making responsibilities. They included:

- the program manager at a leading aviation organisation (KLM),
- the director of a business innovation consultancy (the can do company),
- the global procurement head in energy sector organisation (Vattenfall),
- the founder of an AI coaching application (Dehurdle),
- and the managing partner at a digital strategy advisory (Digital Governance Advisory).

This selection ensured a spread of perspectives across sectors while maintaining the focus on individuals shaping strategic decisions.

Methodology

Each conversation lasted 30–45 minutes and was conducted online, following a consistent structure tailored to the interviewee’s domain. The session began with a broad discussion on how decisions involving multiple stakeholders are currently made in their context and how values feature in that process. This was followed by a demonstration of the value elicitation prototype using selected visuals and examples (see Appendix C for slides shown).

Participants were then invited to react — sharing their impressions, concerns, and ideas regarding the method’s strategic utility, ethical implications, integration possibilities, and potential use cases. Interviews were recorded (with permission), transcribed, and thematically coded. Themes were derived both from recurring discussion points (e.g., reflection, value misalignment) and from specific, actionable needs expressed during the sessions (e.g., “What would make me adopt this?”). The full interview guide is included in Appendix B.

Key Findings

The interviews revealed two key strands of insight. On one hand, participants identified a range of potential use cases for the value elicitation method — spanning reflection, team alignment, and strategic decision support. These possibilities centered on how the method could generate value in different organizational and collaborative contexts. On the other hand, the conversations also surfaced critical reflections on integration, adoption, and ethical fit — underscoring the need for transparency, contextual sensitivity, and safeguards in real-world use.

The following sections focus on these strands starting with the first strand: the strategic use cases and value potential perceived by interviewees across industries. The subsequent section will examine their concerns and recommendations around adoption conditions and ethical responsibility.

8.2. Strategic Use Cases and Value Potential

After discussing their current relationship with values in decision-making, each interviewee was introduced to the foundational logic of the method and shown how it surfaces values from spoken conversation using computational similarity. This demonstration included a few early insights from the real-world deployment at the INSEAD Leadership Forum — such as identifying which values were most frequently mentioned, which were rarely discussed, and which co-occurred, revealing possible tensions or synergies in group dialogue.

Of particular interest to many was the value expression gap — the disconnect between the values people claim to hold (professed values) and the values that actually emerge in their language during real conversations. Rather than serving as a validation exercise, the purpose of sharing these insights was to prompt reflection and explore potential relevance. Interviewees were not asked to verify or judge the accuracy of the results but to consider its strategic usefulness within their own contexts.

From Insight to Action

“So, the real value is in what is after this. These insights are just information.”

- Interviewee, Director (the can do company)

The most immediate and consistent reaction to these examples was an appreciation for the underlying data — but also a clear desire for what comes after. Decision-makers were not satisfied with surfacing values alone. They wanted to know: What does this mean for me? and more importantly, What do I do with it?

“You need next steps with every section. I am talking about action steps.”

- Interviewee, Program Manager (KLM)

This reaction marked an important pivot in how the value of the tool was understood. The goal wasn’t just to create awareness — it was to enable intentional action, strategic alignment, or even personal reflection.

Thinking in Terms of Results

While the interviews were meant to surface what kind of insight might be valuable, most participants spoke directly about the kind of result or transformation they wanted to see. For them, value elicitation was not an end in itself, but a lever toward broader impact — within teams, across organizations, or even at the leadership level.

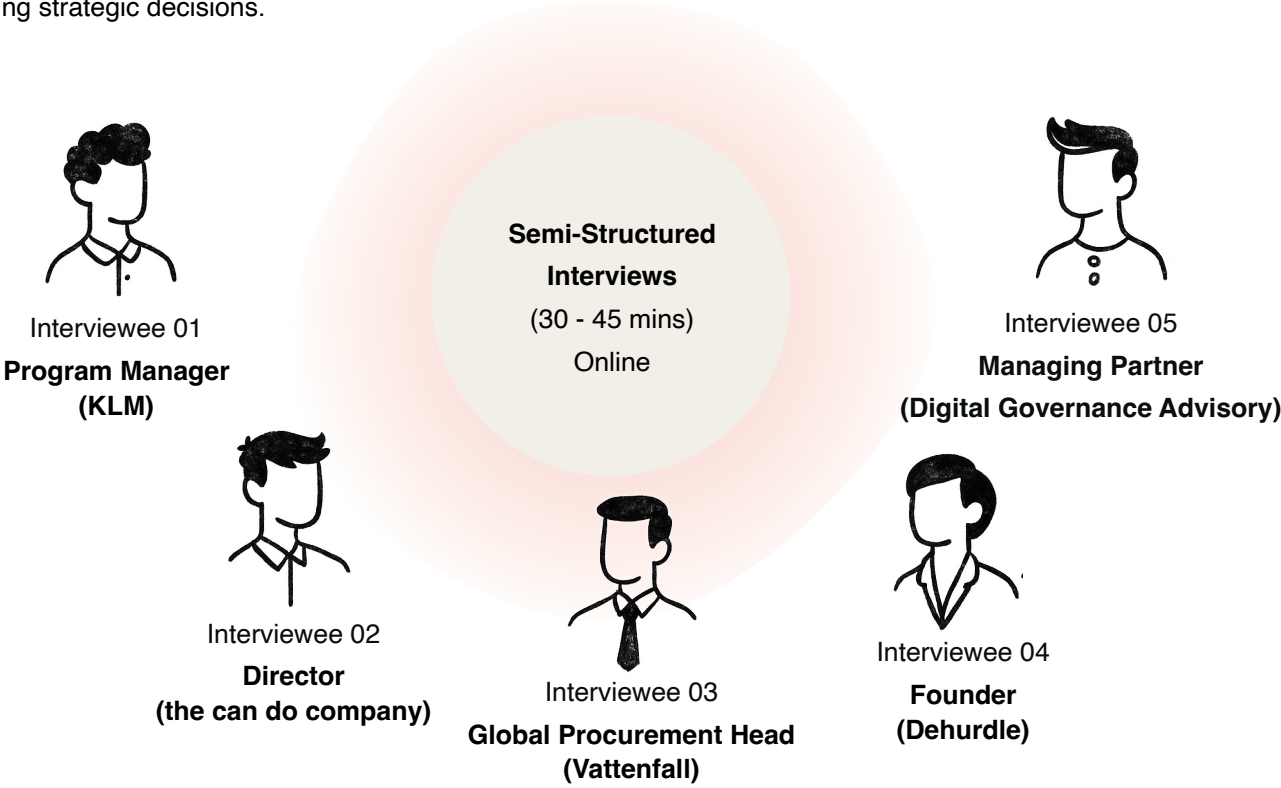


Figure 24: Overview of the semi-structured interviews conducted with five professionals across strategic roles and sectors, illustrating the diverse perspectives gathered to explore the method’s applicability in real-world decision-making contexts. (The visuals were created by OpenAI’s Sora)

“For me, the most fantastic thing we could achieve with this tool is that we actually become consistent with the values that we have and become a force for good as an organization.”

- Interviewee, Global Procurement Head (Vattenfall)

“I will come to use this input to improve myself, because it seems that I’m thinking in one direction, but I’m saying another direction.”

- Interviewee, Program Manager (KLM)

This led to a reframing: Instead of categorizing insights first, I began identifying result or impact themes, and then worked backwards to define what kind of insight, action, and data would be required to achieve that result—keeping the value elicitation method as the foundational layer that informs and enables this process.

Mapping Result-Oriented Use Cases

Across the five interviews, seven core impact-oriented use cases emerged — each outlining a potential way the value elicitation method could support reflection, alignment, or decision-making in real-world contexts. While there may be additional possibilities, the following use cases are grounded in this study’s interview data:

- Becoming more consistent with organizational values
- Reflecting individually on whether personal leadership communication expresses intended values
- Finding opportunities through value misalignment across teams or departments
- Tracking cultural shifts before and after interventions
- Designing teams based on compatible value motivations
- Assessing collaboration fit through shared or divergent values
- Adapting negotiations and pitches to match stakeholder values

Each was seen not only as a possible insight but as a directional lever for action — where understanding value dynamics could enable strategic shifts or interventions.

Two Critical Conditions for Real-World Use

Before shaping these use cases into actionable concepts, two important themes emerged across interviews — each pointing to a structural change needed in how the method works:

Longitudinal Value Tracking

Interviewees emphasized that to detect cultural shifts or monitor consistency, the method must analyze patterns across time — from recurring meetings to email threads. One-off insights were seen as insufficient for deeper organizational learning.

Integration of Custom Organizational Values

Several use cases — such as identifying value expression gaps or assessing team fit — rely on comparing expressed values with a predefined set of organizational or team values. This makes it essential to allow for custom value integration within the model.

These themes shaped how the method’s functionality and inputs would need to evolve to meet strategic needs — expanding from single-session analysis to context-aware, system-integrated reflection tools.

Translating Use Cases into Design Challenges

To bridge the gap between expressed needs and practical application, each interview-derived use case was re-examined as a design challenge. Starting from the strategic result that participants envisioned, I worked backward to identify:

- the **type of insight** needed to inform that result,
- the **action step** it could support,
- the **implications** this has for how the current value elicitation method would need to evolve, and
- the **ideal data inputs** required to make it operational.

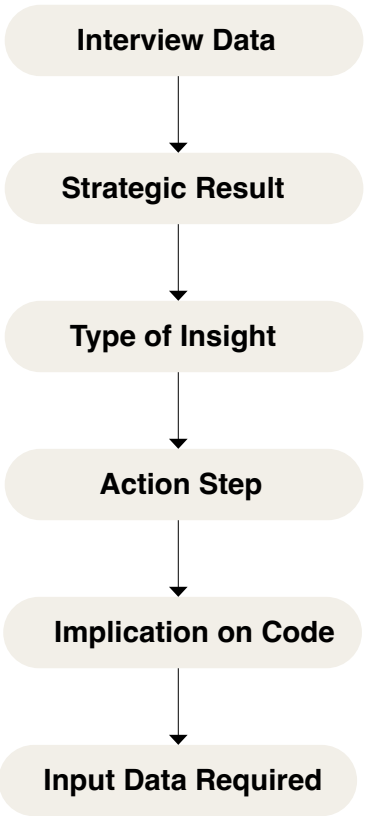


Figure 25: Translating interview-derived use cases into design challenges by tracing each strategic result backward through the type of insight required, the supported action, implications on the method’s evolution, and the ideal input data to operationalize it.

The table 15 summarizes this translation from interviews to implementation — showing how real-world aspirations demand specific forms of analysis, system behavior, and data readiness. It offers a structured starting point for guiding further design and research. Future work could further validate or expand on these action steps through literature and practical trials.

Reflection on the Use Cases

While the strategic use cases discussed in this chapter predominantly emerged within internal organizational contexts (See Table 15), this emphasis likely reflects both the interviewees’ professional roles and how the method was framed in the interview — as a support system for decision-making within teams and departments. In many cases, this led participants to consider applications within their own sphere of control first, particularly for fostering alignment, improving communication, and tracking change over time.

However, some interviewees did begin to extrapolate its relevance to external-facing contexts, such as understanding stakeholder values or pitching ideas more effectively. More notably, a few participants speculated about applications in the political domain — where value expression gaps are often stark between professed ideals and actual discourse. As one put it:

“They say in the political party, like we’re progressive and we think of the people... but in their discussions, how far do they really say what they say?”

- Interviewee, Program Manager (KLM)

Further, they also imagined a future where such a tool could be used to analyze political speeches for consistency and motivation. While these external applications were not the primary focus of this research, they point toward promising directions for future exploration — particularly in contexts such as politics, where value expression gaps are both visible and consequential. Therefore, the political domain, in particular, presents a compelling case for future research, where analyzing the alignment between professed values and public discourse could support transparency, accountability, and more values-conscious leadership.

Quotation	Strategic Result	Key Insight	Action Step	Implication on Code	Data
Internal					
“...the most fantastic thing we could achieve with this is that we become consistent with the values that we have and become a force for good as an organization...” - Global Procurement Head (Vattenfall)	Align Internal Culture with Professed Values	Dominant Values ; Mismatch between stated values and operational behavior	Compare internal communications with published value statements; adjust internal strategy or messaging accordingly	Value elicitation of dominant values + comparison with professed values to flag misalignment	Longitudinal meeting data, organisational professed values
“...I will come to use this input to improve myself because it seems that I'm thinking one direction, but I'm saying another direction...” - Program Manager (KLM)	Improve Leadership Communication	Dominant Values and Missing Values in personal communication	Review how leadership messaging reflects values; make updates to how key messages are framed	Value elicitation from leadership communication to surface dominant and missing values	Script of talk/presentation, personal reflection data
“...this can help you where you missed out on picking some golden Nuggets...” - Founder (Dehurdle) “...you could run their profiles, chats and emails to see the values and get insights that could help solve a problem....” - Director (the can do company)	Find Opportunities Through Misalignments	Shared vs. Divergent Values	Identify value gaps across departments and address them through joint planning or clarification	Value elicitation across teams to flag discrepancies and surface complementary values	Conversations/emails across teams
“it's very tough to measure the impact of team interventions...I'm curious, what if you do it a year later and see if it changes...kind of quantifies the implicit changes” - Director (the can do company)	Track Cultural Shifts Over Time	Shifting Values Over Time	Use before/after comparison to evaluate success of an intervention and inform next steps	Value elicitation over time to track longitudinal changes in value expression	Pre/post-intervention conversations over time
“...It can help you comprehend better understand things holistically. ...understand other person's point of view...” - Managing Partner (Digital Governance Advisory)	Build High-Trust, Values-Aligned Teams	Shared Values; Value-driven Motivation	Use value profiles when forming teams or assigning roles to improve fit	Value elicitation at individual level to build group profiles and detect compatibility	Conversations + individual profiles
External					
“...good reflection tool to see if the values of the company that I sit together with if they match with our organisational values...” - Global Procurement Head (Vattenfall)	Assess Collaboration or Stakeholder Fit	Stakeholder's Shared and Divergent Values	Compare your organization's core values with those of potential collaborators	Value elicitation with stakeholder profiles for value alignment matching flagging discrepancies	Conversations + stakeholder profiles
“...for negotiation, I would be crazy to not use it because it saves me a lot of money...” - Global Procurement Head (Vattenfall)	Tailor Stakeholder Pitches and Negotiation Strategies	Stakeholder's Value-Driven Motivations	Adjust pitch to match the values and concerns that matter most to the stakeholder	Value elicitation to identify motivational cues from conversations and inform message framing	Stakeholder conversations + stakeholder profiles

Table 15: Summary of interview-derived use cases translated into design implications — tracing each strategic outcome back to the needed insight, action step, code adaptation, and ideal input data.

Strategic Use Cases

How the Method Could Support Reflection, Alignment, and Decision-Making Across Organizational Settings

The visual was created with the help of OpenAI's Sora.

While not exhaustive, the use cases shown here reflect the directions that emerged most clearly across stakeholder interviews. Additional applications may surface as the method is further developed and deployed.

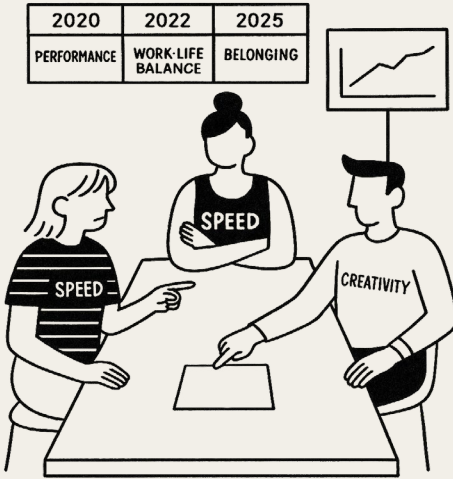
Improve Leadership Communication

Your words shape the culture — even when they miss the mark.



Track Cultural Shifts Over Time

Values evolve — are you keeping track?



Assess Collaboration or Stakeholder Fit

The wrong values match costs more than a missed deal.



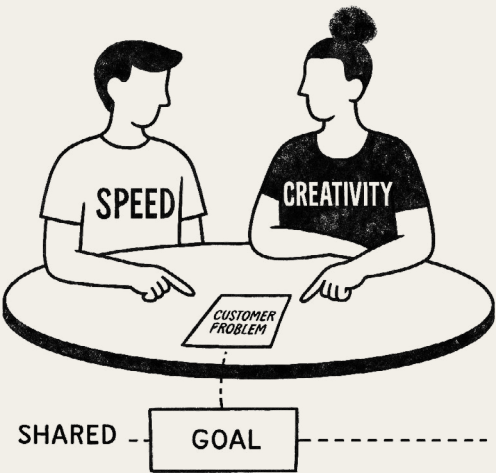
Align Internal Culture with Professed Values

What we say we believe isn't always how we behave.



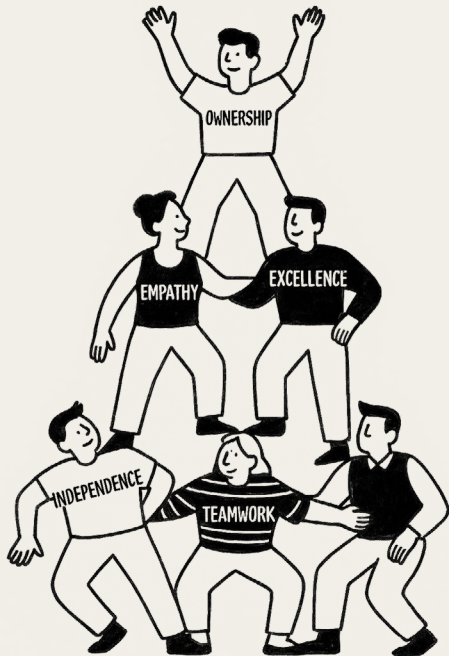
Find Opportunities Through Misalignments

Different values can still build the same bridge.



Build High-Trust, Values-Aligned Teams

Aligned values make stronger teams — misaligned ones crumble fast.



Tailor Stakeholder Pitches and Negotiation Strategies

You can't sell meaning if you don't know what matters.



8.3. Integration, Adoption, and Ethical Fit

As the interviews explored potential use cases of the value elicitation method, several other deeper themes emerged around its practical and ethical deployment. These themes highlight not just technical or functional needs, but the social, emotional, and organizational conditions under which such a method could be accepted, trusted, and meaningfully used.

Acceptance and Organizational Buy-In

Organizational Readiness

Interviewees emphasized that for this method to be meaningfully integrated, people within the organization must first believe in its value and be ready for the kind of feedback it offers. Surfacing values — especially revealing discrepancies between what is said and what is done — was described as inherently provocative. It requires a culture that sees feedback not as criticism, but as an opportunity.

“I’m a very nice guy and then at the conversation you see like I’m a jerk.”

- Interviewee, Program Manager (KLM)

“How willing and how well equipped people are to accept such fundamental feedback about who they are and what they do?”

- Interviewee, Global Procurement Head (Vattenfall)

Personality and Leadership Openness

Because of this, several interviewees noted that readiness is often linked to personality and leadership culture. For example, confident or curious leaders — those genuinely open to listening, learning, and understanding their teams — were seen as more likely early adopters.

“I think it requires certain personalities — confident personalities — to start using this and getting some awareness.”

- Interviewee, Managing Partner (Digital Governance Advisory)

“You want to try out new things because it’s such a great learning opportunity, right?”

- Interviewee, Global Procurement Head (Vattenfall)

Framing of the Intervention

The tone and framing of the intervention also emerged as crucial. Participants expressed that value-based insights should not feel like a diagnostic of what’s wrong — instead, they should highlight possibilities and new dimensions of understanding.

“People can add more value or dimension to it, not thinking in limitations or risk involved, but also opportunities.”

- Interviewee, Program Manager (KLM)

Strategic Awareness

Buy-in was also tied to organizational awareness at the leadership level. If leadership understands the potential of this approach — not only for surfacing values, but for actively using them to build alignment or strategy — adoption is more likely.

“The leadership of the organization should be aware of the possibilities and the contribution of such tooling — and believe in it.”

- Interviewee, Managing Partner (Digital Governance Advisory)

Measurable Outcome

Lastly, it was clear that insights alone are not enough — organizations want to see measurable results. Several interviewees noted that for the method to be adopted seriously, it would need to demonstrate clear, trackable impact beyond qualitative reflections. Whether improving negotiation outcomes, strengthening collaboration, or accelerating alignment, stakeholders seek evidence of effectiveness — ideally tied to specific metrics.

“If I can quantify that using your tool actually boosted my negotiation results with X percent...”

- Interviewee, Global Procurement Head (Vattenfall)

“I’m rather focused on the results.”

- Interviewee, Managing Partner (Digital Governance Advisory)

This feedback suggests that for organizational buy-in, the tool must not only offer reflective or strategic value — it should also help prove that surfacing and working with values leads to better decisions, improved performance, or stronger relationships. Quantification doesn’t need to be perfect, but even directional metrics could help translate insights into legitimacy.

Transparency of the Use

A recurring theme in the interviews was the importance of clarity, honesty, and framing when introducing a value detection method in organizational contexts. The tool should never be seen as a mechanism for surveillance or judgment, but as a supportive aid for shared understanding.

“Let’s be completely honest about what we’re doing. Let’s also make it very clear that this is not some kind of assessment or a court case. We’re not going to judge people for the gap that we see between those values.”

- Interviewee, Global Procurement Head (Vattenfall)

Participants stressed that the tool should be seen as a reflective aid, not an evaluative mechanism. It should support open dialogue — not judge or label people. This was especially important in contexts like negotiations, where some saw potential while also raised ethical concerns about acting on insights without the other party’s awareness. The need for clear boundaries and transparent use was emphasized.

Timing and Context of Intervention

Participants emphasized that the value elicitation method would be most effective when used selectively, aligned with the purpose and dynamics of the meeting. Several interviewees saw high potential during quarterly reviews, team coaching sessions, or role assignment meetings — where values are directly relevant to alignment, decision-making, or group formation. Others pointed to progress meetings or negotiation settings, where real-time or reflective value prompts could guide discussions back to shared intentions.

“So I think there is some sort of a subset of meetings where this could be really of value.”

- Interviewee, Managing Partner (Digital Governance Advisory)

Some preferred mid-meeting reflection over real-time feedback, to avoid disruption while still enriching the conversation. However, participants also flagged that not all meetings warrant value analysis.

“It’s not really valuable to talk about values in every meeting... sometimes they’re so implicit, you don’t need to talk about them.”

- Interviewee, Managing Partner (Digital Governance Advisory)

The overall takeaway was that context matters: the method should be deployed intentionally, based on meeting purpose, timing, and readiness of participants.

Reflection vs. Real-Time Nudging

When asked about the preference between using the tool for reflection or real-time intervention, most interviewees favored using the tool for retrospective reflection. Mostly, the inclination was towards using it after meetings or projects — as a way to uncover patterns, spark dialogue, and drive self or team-level learning. Reflection was seen as a low-pressure, high-value moment to step back and evaluate whether values were truly present in the conversation.

“I can imagine doing this with my management team as a reflection.”

- Interviewee, Global Procurement Head (Vattenfall)

There was also interest in real-time applications, particularly in high-impact settings like alignment meetings, negotiations, or team-building sessions. However, real-time use was viewed as more context-dependent and requiring greater organizational maturity and psychological safety.

“Looking backwards is always easier... Giving live feedback and adjusting your plan as such requires huge maturity.”

- Interviewee, Program Manager (KLM)

Several interviewees expressed hesitation about real-time confrontation, suggesting that people may resist being held accountable in the moment — especially if it disrupts momentum or triggers defensiveness.

In essence, reflection was the preferred starting point — useful for deepening awareness, reinforcing values, and setting the stage for more timely interventions in the future, once trust and familiarity with the tool have been established.

Ethical Considerations

While the value elicitation method showed potential, interviewees expressed concerns around its ethical deployment.

A recurring theme was the need to protect the confidentiality of conversations, especially in team meetings where sensitive or emotional undercurrents might emerge. Participants stressed that not all values are meant to be shared or surfaced publicly — and that applying the method without explicit consent could result in overreach.

“What if it starts singling out individuals and their personal values which they didn’t want to share?”

- Interviewee, Global Procurement Head (Vattenfall)

“They don’t want to talk about the values because it’s not serving their professional goals”

- Interviewee, Managing Partner (Digital Governance Advisory)

Another concern was the potential for manipulation. In contexts like negotiation or stakeholder engagement, using inferred values to tailor communication without the other party’s awareness raised ethical questions. While such strategies could enhance persuasion, several interviewees emphasized that doing so without transparency risks undermining trust. This was not only framed as manipulative, but also as a breach of relational integrity — where one party holds an informational advantage about the other’s motivations without their knowledge or consent.

Ultimately, the tool’s credibility and impact rely not just on what it detects — but on how respectfully it is integrated into real-world contexts.

Chapter 09

Discussion

This chapter synthesizes the project’s impact, affirming that computational methods can meaningfully uncover value dynamics in conversation. It reflects on the tool’s desirability, feasibility, and viability, while acknowledging key limitations and charting future directions. The chapter concludes by highlighting conceptual and practical contributions — from the “Value Expression Gap” to repositioning AI as a reflective tool for ethical decision-making.

- 9.1. Reflection on Research Questions
- 9.2. Limitations and Future Work
- 9.3. Contributions

9.1. Reflection on Research Questions

This final reflection revisits the research questions that shaped the project, now considered in light of the method’s development, deployment, and real-world exploration. The section begins by addressing the sub-questions, each linked to specific phases of the report, before returning to the main research question. Rather than offering definitive conclusions, the answers reflect the exploratory nature of the work—grounded in what was done and observed, while pointing toward what is possible. In doing so, the reflection connects technical feasibility with practical viability and desirability, setting the stage for further development.

Sub-Research Question 1
How can natural language processing techniques be prototyped and applied to identify both explicit and implicit values in multi-stakeholder conversations?

This question is addressed through the iterative development of the prototype across Chapters 4 to 7. The foundational method logic (Section 4.3) used sentence embeddings and cosine similarity to detect value proximity. Early versions lacked depth due to a limited set of value examples, prompting refinements that expanded the value dataset and introduced implicit expressions such as slangs, metaphors, and emotional tones. Additional improvements — including a soft assignment approach, a motivational goal layer, and the evaluation of strategic elicitation strategies — made the output more nuanced and open to interpretation. The results of the latest prototype can be seen in Section 7.2, demonstrating its ability to detect both explicit and implicit expressions of values in conversation.

Sub-Research Question 2
What types of value-related insights can be captured from conversation data?

This question was explored across all phases of the research. Theoretically, the premise was that uncovering values in conversation offers a richer understanding of decision-making dynamics. During the Research through Design phase, the evolving prototype revealed several types of value-related insights, including dominant values, missing values, motivational goals,

of value types, soft assignment profiles, and the value expression gap (See section 7.2). These insights demonstrated that the method could go beyond keyword extraction and begin surfacing more complex relational patterns across values.

In the real-world relevance phase, the use cases derived from stakeholder interviews pushed this further. The table in Section X outlines the kinds of organizational outcomes stakeholders wanted to achieve—such as aligning culture with professed values, assessing stakeholder fit, or building value-aligned teams—and how each implied a specific type of value insight, required action, and system capability.

While these insights represent current capabilities, the potential for further development remains. With richer data inputs such as longitudinal conversations, speaker profiles, or additional layers, the method could begin uncovering even more complex value phenomena—such as identifying personal vs. shared values, mapping value interdependencies, or tracing shifts in alignment over time.

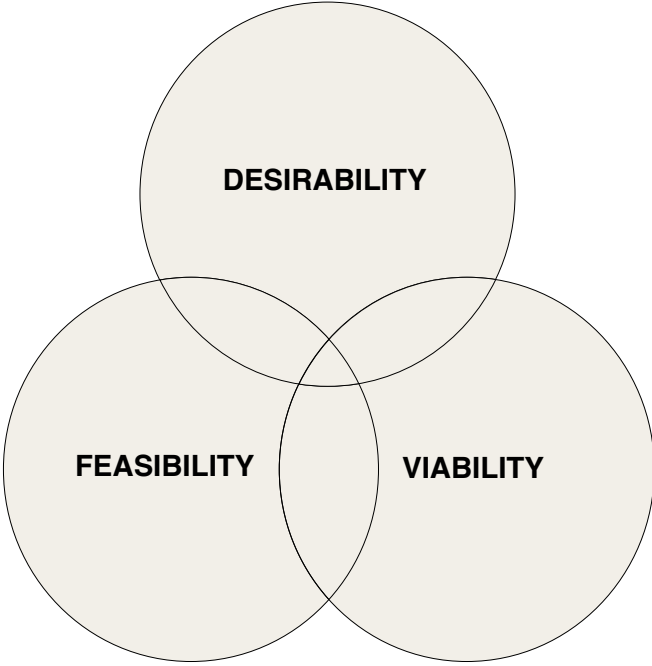
Sub-Research Question 3
How can the relevance of computational value analysis be understood and applied in real-world decision-making contexts?

While the value of uncovering implicit and explicit values in conversations was framed theoretically, Chapter 8 focused on how this potential was interpreted by decision-makers in real-world contexts. Interviews conducted with professionals across sectors revealed that the method was broadly understood as a reflective tool—capable of surfacing value tensions, prompting self-awareness, and guiding alignment within teams and organizations. However, the relevance of such a tool was not seen as self-evident. It depended heavily on how it was framed, the level of transparency involved, and whether its insights could translate into meaningful action.

The use cases identified ranged from internal applications like strengthening team alignment and improving leadership communication, to external ones such as tailoring stakeholder strategies and assessing collaboration fit (See table X). But beyond these, interviewees surfaced a range of enabling conditions—emphasizing that the tool’s impact would hinge on organizational readiness, ethical safeguards, and the ability to demonstrate tangible outcomes (see section 8.3). As these themes unfolded, it became clear that understanding and application are not just about functionality—they require trust, cultural fit, and a clear sense of purpose.

Main Research Question
Can computational methods be used to uncover and analyze value dynamics in conversation in ways that are meaningful for decision-makers?

The sub-questions already demonstrate that this is possible: computational methods can surface explicit and implicit values, reveal new forms of insight, and spark real-world relevance. However, to assess whether such a method holds long-term promise, this section evaluates whether it is desirable, viable, and feasible — providing a broader lens on its future and use potential.



Desirability
Do people want it?

Desirability was evidenced through multiple layers of engagement across the research journey. In literature, the tool filled a recognized gap — the need to articulate values in ways that are traceable and actionable. In practice, both forum participants and interviewees voiced a strong interest in using the method for personal reflection, strategic alignment, and communication clarity.

“...I will come to use this input to improve myself because it seems that I’m thinking one direction, but I’m saying another direction....”

- Interviewee, Program Manager (KLM)

“...for negotiation, I would be crazy to not use it, because it saves me a lot of money...”

- Interviewee, Global Procurement Head (Vattenfall)

But desirability went beyond curiosity or novelty. Decision-makers themselves articulated concrete, relevant use cases — from preparing negotiations and aligning teams to tracking value shifts over time. Their feedback revealed that the method doesn’t just raise awareness; it can clarify purpose, guide strategy, and reveal blind spots that are otherwise easy to miss.

By showing people what might be missing, the method created space for more intentional, values-aligned dialogue. And while the prototype still requires refinement, its underlying logic — and the impact it can have — was not only understood, but actively imagined by its potential users.

Feasibility

Can it be built and delivered?

Through prototyping and iterative deployment, the method was shown to work. It could detect values, identify their relative prominence, and visualize value dynamics in ways that prompted reflection and discussion. The development journey itself — from initial sentence embedding logic to additional layers like soft value assignment — showed that such a method can be built with accessible tools and refined in a short timeframe.

“You need next steps with every section. I am talking about action steps.”

- Interviewee, Program Manager (KLM)

However, feasibility isn’t just about working code. Several decision-makers raised concerns about integration: if the tool is to be truly useful in organizations, it must connect with existing workflows and technologies. And once it does, expectations rise — users want not just insight, but direction.

“The people we do business with will be the same for three years...”

- Interviewee, Global Procurement Head (Vattenfall)

This points to the gap between a working prototype and a functional product. The current method opens up possibility — but to sustain use, it must evolve to fit into long-term strategic processes, support repeat interactions, and offer concrete, context-aware recommendations.

So while feasibility is proven in principle, future iterations will need to balance adaptability, usability, and guidance — especially if the method is to move beyond reflective workshops into everyday decision-making.

Viability

Can it work long-term in the real world?

Throughout the interviews, decision-makers voiced enthusiasm about using the method in real-world contexts: leadership conversations, cultural reflections, even high-stakes negotiations. Its potential to support these moments was widely acknowledged.

“So I think there is some sort of a subset of meetings where this could be really of value.”

- Interviewee, Managing Partner (Digital Governance Advisory)

Yet with interest came caution. For the method to become viable at scale, it must move from possibility to proof. That means showing measurable benefits — not just perceived insight. Participants highlighted that without concrete outcomes, adoption would be difficult to justify.

“If I can quantify that using your tool actually boosted my negotiation results with X percent...”

- Interviewee, Global Procurement Head (Vattenfall)

This insight reflects a broader challenge: while the prototype sparked curiosity and recognition, stakeholders expect more from viable tools. They want repeatable results, integration with existing systems, and clear pathways from insight to action.

What this research revealed is that the method could be viable — but only through continued iteration and validation. If refined to meet organizational needs and supported by evidence of its effectiveness, it holds real potential to become a trusted part of value-driven decision-making.

9.2. Limitations and Future Work

This project set out to explore how computational methods might uncover and interpret value dynamics in conversation. While the prototype and deployment surfaced valuable insights, the following limitations shaped the outcomes — and inform potential directions for future research.

Contextual Breadth and Generalizability

The current findings are derived from a limited sample — a few conversations across industries and three exploratory workshops. As such, the results are not yet generalizable across teams, industries, or cultural contexts. This opens space for future research to test the method in high-stakes, real-time decision-making scenarios (e.g., policymaking, corporate strategy, NGO coalitions) and in more diverse settings. While the literature review included insights from multiple fields, the project did not scope down to specific decision-making domains (e.g., product development, public policy, or organizational design). Future iterations could benefit from clearly defining where and for whom this method creates the most impact.

Expression Gaps and Interpretive Subjectivity

While the concept of a “value expression gap” emerged as powerful, its underlying causes remain unclear. Factors like power dynamics, confidence, or cultural habits may play a role — but these were not explored. Future work could investigate to examine how social environments impact value expression. In addition, even though the prototype surfaces signals, the interpretation of value cues still depends heavily on human judgment. This subjectivity limits standardization — but it also reveals an opportunity for tools that support reflective, not prescriptive, dialogue.

Conversational and Data Modalities

The current method analyzes verbal conversation only, excluding tone, silence, or non-verbal expressions — dimensions that often carry another nuanced layer. It also doesn’t yet accommodate communication formats like email threads, messaging platforms, or asynchronous dialogue. Future exploration could investigate how to incorporate multimodal data to capture these additional layers of meaning.

Technical Trade-offs and Model Constraints

Some design choices were made for feasibility of this thesis research — for instance, using pause-based segmentation or staying within a local pipeline to ensure privacy. While this worked well for testing, these decisions may limit long-term scalability, especially if the tool is to be embedded in organizational platforms. Future research could evaluate the trade-offs between ethical deployment and technical integration.

Note: Several technical limitations and improvements (e.g., model choices, soft assignment approach) are already discussed in Section 7.3.

Behavioral Impact and Feedback Loops

At this stage, the method generates insights — but does not yet track whether these insights influence decisions or behaviors. Future research could explore how exposing participants to their own value patterns impacts alignment, collaboration, or performance. Does awareness change action? Or provoke resistance? Understanding this could inform how feedback is framed and acted upon in organizational settings.

Methodological Decisions and Theoretical Gaps

Some strategic decisions — like the logic for identifying action steps — were made based on what seemed appropriate and not grounded theoretical frameworks. Future work can connect these pathways to established design or behavior change theories to build stronger foundations for applied use.

9.3. Contributions

This research offers multiple contributions — conceptual, methodological, and practical — to the growing space where language, values, and decision-making intersect. While the method remains in early stages, its development and deployment surfaced new ways of thinking about value expression, supported reflection, and revealed promising opportunities for further application.

A New Lens on Values in Dialogue

This research contributes a novel way to surface and interpret values in conversations, building on the literature’s recognition that values often remain implicit in decision-making contexts (Bos-de Vos, 2020; Kenter et al., 2016a). While scholars have emphasized the importance of making value dynamics visible in design and strategy (Friedman et al., 2013; Rindova & Martins, 2017), most approaches rely on self-reporting or facilitation. This method moves beyond that — detecting values as they emerge through dialogue, not just as individuals name them. The tool acts as an assistant — not replacing dialogue, but enhancing it by surfacing what’s left unsaid.

By making value signals traceable, contestable, and analyzable, the prototype helps uncover hidden tensions, blind spots, or alignments — advancing the idea of conversation as a site for social learning and value formation (Brymer et al., 2018). It enables detection not just of what values are said, but opens door for future potential of how they interact — through co-occurrence, contradiction, or silence.

The Impact of Making Values Visible

Literature in values research highlights that individuals are not always consciously aware of the values guiding their decisions (Schwartz, 2016), and even when expressed, values are often embedded within layered belief systems and social contexts (Feather, 1996; Rindova & Martins, 2017). The Deliberative Value Formation (DVF) model (Kenter et al., 2016a) further emphasizes that values are not simply revealed but actively constructed through conversation, often becoming “more explicit and contestable” through dialogue.

This research contributes to that discourse by demonstrating how computational tools can make implicit value signals more visible. The method surfaces what is left unsaid — the emotional cues, implicit priorities, or recurring themes — offering individuals and teams a clearer view of the value dynamics shaping their interactions. In doing so, it not only supports reflection but also reveals misalignments, blind spots, or evolving tensions that might otherwise be overlooked in everyday conversation. This research also investigates potential use cases for decision-makers, where such a tool could create impact. These use cases illustrate how making values visible can drive reflection, alignment, and strategic action across diverse decision-making contexts.

The Value Expression Gap: A Conceptual Contribution

This thesis introduces the Value Expression Gap — the disconnect between the values people say they hold and the values that actually surface in their everyday speech. This builds on the understanding that values are often unconscious or situational (Schwartz, 2016; Rindova & Martins, 2017), and that people may not fully grasp how values influence their decisions (Feather, 1996). By identifying this gap, the research contributes a new conceptual lens to both design research and value-sensitive AI — one focused not on judgment, but on reflection and awareness.

Role of AI in Decision-Making

Unlike many AI systems designed for prediction or automation, this method repositions AI as a reflective companion — supporting human sensemaking rather than replacing it. In soft, qualitative domains such as stakeholder engagement or organizational reflection, the tool acts as a co-pilot: surfacing patterns, highlighting value co-occurrences or contradictions, and pointing toward meaningful questions. This challenges dominant narratives of AI as a neutral or purely rational optimizer and instead embraces a human-centered, interpretive approach to computational analysis.

At the same time, the tool was intentionally developed to be transparent and explainable. Every value assignment

can be traced back to concrete examples and similarity scores, offering interpretability that is often absent in black-box AI models. This supports critical engagement rather than blind trust — ensuring that users remain in control of meaning-making. By doing so, the project contributes to ongoing efforts in ethical, transparent AI, while opening new directions for the use of machine learning in reflective, values-driven contexts.

Designers’ Role and Collaborative Dynamics

While the method emerged from a design research context, its implications for the role of designers merit reflection. Designers sometimes act as facilitators of dialogue, interpreters of stakeholder needs, and stewards of ethical practice. This method does not replace those functions but invites reconsideration of how designers engage with value articulation in increasingly data-driven environments.

Designers could play a critical role in shaping how such a tool is used and further developed — making it not only usable and interpretable, but also ethically integrated. Rather than limiting designers, this shift opens new collaborative possibilities.

Use of Artificial Intelligence

At the core of this research lies a question about how artificial intelligence can help us reflect more consciously on the values embedded in our conversations. To explore that, AI didn't just remain a subject of inquiry — it also became a hands-on collaborator.

The value elicitation prototype was built using pre-trained sentence-transformer models, specifically MiniLM-L6-v2, to capture the semantic meaning of spoken utterances. Cosine similarity was then used to compare each utterance against a curated set of example expressions, allowing the prototype to assign the most likely values being communicated.

Crucially, the method was designed to avoid the black box problem often associated with AI. Every value assigned by the system could be traced back to a specific sentence similarity score and example utterance, making the output transparent, explainable, and open to human interpretation. Rather than replacing reflection, the AI acts more like a prompt — surfacing signals that invite deeper thinking, not prescriptive conclusions.

Beyond the prototype, AI tools like ChatGPT and Sora played a supporting role throughout the research process. They were used to:

- Refine language, check grammar, and unblock writing flow
- Generate the initial value dataset of 108 values with descriptions (based on a published framework) (see Appendix D)
- Produce hundreds of example utterances per value, across diverse expression types (e.g., explicit, metaphorical, emotional) (see Appendix D)
- Help create custom illustrations that brought abstract concepts to life in the report and presentations

In short, while this thesis critically examines the role of AI in supporting value-aware dialogue, it also showcases how AI — when thoughtfully applied — can become a practical collaborator in both design and research.

References

Aseniero, B. A., Constantinides, M., Joglekar, S., Zhou, K., & Quercia, D. (2020, October). MeetCues: Supporting online meetings experience. In 2020 IEEE Visualization Conference (VIS) (pp. 236-240). IEEE.

Beratan, K. K. 2007. A cognition-based view of decision processes in complex social-ecological systems. *Ecology and Society* 12 (1):27. <http://dx.doi.org/10.5751/ES-02103-120127>

Bergema, K., Kleinsmann, M., & Valkenburg, R. (2011). Exploring collaboration in a networked innovation project in industry. DS 68-3: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 3: Design Organisation and Management, Lyngby/Copenhagen, Denmark, 15-19 August 2011.

Bernthal, W. F. (1962). Value perspectives in management decisions. *Academy of Management Journal*, 5(3), 190-196.

Bos-de Vos, M. (2018). Open for business: Project-specific value capture strategies of architectural firms. Retrieved from <https://journals.open.tudelft.nl/index.php/abe/article/view/2399>

Bos-de Vos, M. (2020). A framework for designing for divergent values. *Proceedings of DRS*, 1. <https://doi.org/10.21606/drs.2020.374>

Boradkar, P. (2010). Valued Possessions: The Worth of Things. In *Designing Things: A Critical Introduction to the Culture of Objects* (pp. 45-74). Oxford: Berg.

Boztepe, S. (2007). User value: Competing theories and models. *International Journal of Design*, 1(2).

Brymer, A. L. B., Wulfhorst, J. D., & Brunson, M. W. (2018). Analyzing Stakeholders' Workshop Dialogue for Evidence of Social Learning. *Ecology and Society*, 23(1). <https://doi.org/10.5751/es-09959-230142>

Caelen, O., & Blete, M. A. (2024). *Developing Apps with GPT-4 and ChatGPT*. O'Reilly Media, Inc.

Chen, X., Yap, N., Lu, X., Gunal, A., & Wang, X. (2025). MeetMap: Real-Time Collaborative Dialogue Mapping with LLMs in Online Meetings. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1–35. <https://doi.org/10.1145/3711030>

Clark, S. (2015). Vector Space Models of Lexical Meaning. *The Handbook of Contemporary Semantic Theory*, 493–522. <https://doi.org/10.1002/9781118882139.ch16>

Colangelo, M. T., Meleti, M., Guizzardi, S., Calciolari, E., & Galli, C. (2025). A comparative analysis of sentence transformer models for automated journal recommendation using PubMed metadata. *Big Data and Cognitive Computing*, 9(3), 67. <https://doi.org/10.3390/bdcc9030067>

De Wildt, T. E., Van De Poel, I. R., & Chappin, E. J. L. (2021). Tracing Long-term value change in (Energy) Technologies: Opportunities of probabilistic topic models using large data sets. *Science Technology & Human Values*, 47(3), 429–458. <https://doi.org/10.1177/01622439211054439>

Den Ouden, E. (2012). *Innovation Design: Creating Value for People, Organizations and Society*. London: Springer.

Essiz, O., Yurteri, S., Mandrik, C., & Senyuz, A. (2022). Exploring the Value-Action gap in green consumption: roles of risk aversion, subjective knowledge, and gender differences. *Journal of Global Marketing*, 36(1), 67–92. <https://doi.org/10.1080/08911762.2022.2116376>

Friedman, B., Kahn, P. H., Borning, A., & Hultgren, A. (2013). Value sensitive design and information systems. In *Philosophy of engineering and technology* (pp. 55–95). https://doi.org/10.1007/978-94-007-7844-3_4

Habermas, J. 1984. *The theory of communicative action*, Vol. I. Beacon, Boston, Massachusetts, USA.

Kenter, J. O., Reed, M. S., & Fazey, I. (2016). The Deliberative Value Formation model. *Ecosystem Services*, 21, 194–207. <https://doi.org/10.1016/j.ecoser.2016.09.015>

Kluckhohn, C. (1951) Values and Value-Orientations in the Theory of Action: An Exploration in Definition and Classification. In: Parsons, T. and Shils, E., Eds., *Toward a General Theory of Action*, Harvard University Press, Cambridge, 388-433. <http://dx.doi.org/10.4159/harvard.9780674863507.c8>

Lapakko, D. (1997). Three cheers for language: A closer examination of a widely cited study of nonverbal communication. *Communication Education*, 46(1), 63–67.

Liu, C., Wang, R., Liu, J., Sun, J., Huang, F., & Si, L. (2021). DialogueCSE: Dialogue-based Contrastive Learning of Sentence embeddings. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. <https://doi.org/10.18653/v1/2021.emnlp-main.185>

Martinsuo, M., Klakegg, O.-J., & van Marrewijk, A. (2019). Introduction: delivering value in projects and project-based business. *International Journal of Project Management*.

Meincke, L., Mollick, E. R., & Terwiesch, C. (2024). Prompting Diverse Ideas: Increasing AI Idea Variance. *arXiv preprint arXiv: 2402.01727*.

Moll J, Zahn R, de Oliveira-Souza R (2016) The neural underpinnings of moral values. Brosch T, Sander D, eds. *Handbook of Value: Perspectives from Economics, Neuroscience, Philosophy, Psychology and Sociology* (Oxford University Press, Oxford, UK), 119–127.

Pahl-Wostl, C. (2006). The importance of social learning in restoring the multifunctionality of rivers and floodplains. *Ecology and Society*, 11(1), 10. DOI: 10.5751/ES-01545-110110

Reed, M. S., Evely, A. C., Cundill, G., Fazey, I., Glass, J., Laing, A., Newig, J., Parrish, B., Prell, C., Raymond, C., & Stringer, L. C. (2010). What is Social Learning? *Ecology and Society*, 15(4). <https://doi.org/10.5751/es-03564-1504r01>

Reimers, N. and Gurevych, I. (2019) Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Pro-cessing (EMNLP-IJCNLP)*, Hong Kong, November 2019, 3982-3992. <https://doi.org/10.18653/v1/D19-1410>

Rohan MJ (2000) A rose by any name? The values construct. *Personality Soc. Psych. Rev.* 4(3):255–277.

Rindova, V. P., & Martins, L. L. (2017). From Values to value: value rationality and the creation of great strategies. *Strategy Science*, 3(1), 323–334. <https://doi.org/10.1287/stsc.2017.0038>

Rokeach, M. (1973). *The Nature of Human Values*. New York: The Free Press.

Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. 2019. TalkTraces: Real-Time Capture and Visualization of Verbal Content in Meetings. In *Proceedings of CHI Conference on Human Factors in Computing Systems (CHI '19)*, May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA 14 Pages. <https://doi.org/10.1145/3290605.3300807>

Scarlett, L. 2013. Collaborative adaptive management: challenges and opportunities. *Ecology and Society* 18(3):26. <http://dx.doi.org/10.5751/ES-05762-180326>

Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, 53(3), 550-562.

Schwartz SH (2016) Basic individual values: Sources and consequences. Brosch T, Sander D, eds. *Handbook of Value: Perspectives from Economics, Neuroscience, Philosophy, Psychology and Sociology* (Oxford University Press, Oxford, UK), 63–84.

Stappers, P. and Giaccardi, E. (2014, January 1). Research through Design. *Interaction Design Foundation - IxDF*. <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design>

Steyaert, P., M. Barzman, J. P. Billaud, H. Brives, B. Hubert, G. Ollivier, and B. Roche. 2007. The role of knowledge and research in facilitating social learning among stakeholders in natural resources management in the French Atlantic coastal wetlands. *Environmental Science and Policy* 10:537-550. <http://dx.doi.org/10.1016/j.envsci.2007.01.012>

Tai, R. H., Bentley, L. R., Xia, X., Sitt, J. M., Fankhauser, S. C., Chicas-Mosier, A. M., & Monteith, B. G. (2024). An examination of the use of large language models to aid analysis of textual data. *International Journal of Qualitative Methods*, 23. <https://doi.org/10.1177/16094069241231168>

Van Onselen, L., & Valkenburg, R. (2015). Personal values as a catalyst for meaningful innovations: Supporting young designers in collaborative practice. Proceedings of the 20th International Conference on Engineering Design (ICED15), 27-30 July, Milan, Italy.

Vargo, S. L., Akaka, M. A., & Vaughan, C. M. (2017). Conceptualizing value: a service-ecosystem view. *Journal of Creating Value*, 3(2), 117-124.

Yeomans, M., Boland, F. K., Collins, H. K., Abi-Esber, N., & Brooks, A. W. (2023). A practical guide to conversation research: how to study what people say to each other. *Advances in Methods and Practices in Psychological Science*, 6(4). <https://doi.org/10.1177/25152459231183919>

Rohan MJ (2000) A rose by any name? The values construct. *Personality Soc. Psych. Rev.* 4(3):255–277.

