# The Socratic Dialogue as a Method for Virtue Ethics in AI: A Case Study

**Maarten Wilders, Íñigo Martínez de Rituerto de Troya, Roel Dobbe**

Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands
maartenwilders@gmail.com, i.m.d.r.detroya@tudelft.nl, r.i.j.dobbe@tudelft.nl

## Abstract

This paper investigates how the Socratic Dialogue can cultivate moral virtues among AI practitioners, focusing on core virtues identified by Hagendorff as essential to ethical AI practice: justice, care, honesty, responsibility, practical wisdom (phronesis), and fortitude. Using a case study conducted at a financial bank's transaction monitoring department, we examine how structured ethical deliberation cultivates the dispositions needed to navigate moral complexity in AI ethics. Seven participants, including data scientists, legal specialists, and ethics experts, engaged in a facilitated Socratic Dialogue centered on an ethical dilemma involving AI-driven detection of financial fraud and terrorist activity. Through abductive analysis, we found that tensions emerged within key virtues, illustrating the complexities of ethical decision-making in AI systems. Through the collective nature of the dialogue, participants developed a more refined and context-sensitive understanding of key virtues, exploring what it means to be just, caring, honest, or responsible in practice by navigating within them. This process cultivated practical wisdom not as a solitary trait, but as a relational capacity fostered through shared reflection and moral reasoning. Additionally, the method strengthens fortitude, encouraging AI practitioners to voice ethical concerns despite situational pressures. While challenges remain, such as time investment, facilitation demands, and existing power imbalances, the Socratic Dialogue offers a promising foundation for virtue-oriented AI ethics that moves beyond compliance frameworks toward deeper moral engagement.

## 1 Introduction

As organizations develop, deploy, and use AI systems that increasingly shape decisions in business, law enforcement, healthcare, and daily life (Bankins et al. 2024; Qian, Siau, and Nah 2024), the need for normative reflection and guidance continues to grow. In response to ethical concerns about AI (Stahl et al. 2022), a prominent approach has been principles-based ethics (Jobin, Ienca, and Vayena 2019; Hagendorff 2020; Kazim and Koshiyama 2021), which guides AI development through high-level values such as fairness, transparency, and privacy. These principles, widely adopted by governments, organizations, and international bodies, provide an ethical foundation to ensure AI systems align with societal values.

However, despite their widespread acceptance, principles-based approaches have faced significant criticism. Munn (2023) argues that ethical principles, such as beneficence, autonomy, and justice, are often too vague, contested, and context-dependent. This ambiguity facilitates ethical "box-ticking", where organizations claim adherence to ethical standards without meaningful reflection or implementation. Moreover, these principles often exist in isolation, detached from broader company culture and lacking substantial ethical engagement, limiting their practical application. Ethical failures in AI are frequently linked to deeper systemic issues, such as corporate cultures that prioritize efficiency and profitability over ethical reflection and adjustment (Lauer 2021).

Beyond these structural concerns, empirical research further questions the effectiveness of principles-based ethics. Studies indicate that ethical knowledge alone does not necessarily translate into ethical decision-making. Even professional ethicists do not act more ethically than non-ethicists (Hagendorff 2022; Schwitzgebel 2009; Schwitzgebel and Rust 2014) and within AI development, ethical principles have been found to exert little influence on practitioners' decision-making processes (McNamara, Smith, and Murphy-Hill 2018). This suggests that principles alone are insufficient to ensure ethical AI development (Mittelstadt 2019). Hagendorff (2022) further emphasizes that ethical behavior cannot be secured by principles alone; rather, it requires cultivating moral character and embedding ethical practices within organizational structures. Ultimately, normative principles are ineffective unless actively acknowledged, internalized, and reinforced through cultural and organizational measures.

These critiques highlight the limitations of relying solely on abstract principles and point to the value of a complementary approach: virtue ethics. Rooted in Aristotelian philosophy (Kraut 2022), virtue ethics centers on cultivating moral character, including traits such as justice, honesty, courage, responsibility, and care, rather than merely following prescribed rules. It holds that ethical behavior arises from the dispositions and judgment of individuals, developed through practice and habituation over time. By shifting attention from static rules to character development, virtue ethics bridges the gap between ethical principles and prac-

tice, ensuring ethical considerations are not only conceptually sound but also deeply embedded in everyday decision-making.

Although the literature on virtue ethics in AI is expanding (Constantinescu et al. 2021; Hagendorff 2022; Hayes, Fitzpatrick, and Ferrández 2024; Raquib et al. 2022; Vallor 2016), much of it remains conceptual, focusing on theoretical justifications rather than practical applications. To make virtue ethics more actionable for AI practitioners, empirical research is needed to explore how virtues essential to ethical AI can be cultivated in practice. This paper addresses this gap by examining the Socratic Dialogue as a method for fostering virtue development within an AI practitioner community.

While the Socratic Dialogue has been studied as a method in its own right (Saran and Neisser 2004; Boers 2022) and in relation to education (Knezic et al. 2010), moral values in civil–military relations (van Baarle and Verweij 2008), and moral inquiry in HCI (Maaike Harbers et al. 2019), it has not yet been examined in the context of AI ethics.

We presents a case study of a real-world application of this method at a financial bank, where AI practitioners responsible for financial transaction monitoring participated in a Socratic Dialogue to reflect on the ethical implications of AI in preventing financial fraud and terrorism. Through this inquiry-based practice, the study explores how Socratic Dialogue can help practitioners engage in shared reflection, express virtues in morally complex situations, and embed ethical considerations into professional practice. Guided by this focus, this study asks, *in what ways can participation in a Socratic Dialogue cultivate key moral virtues among AI practitioners?*

The structure of this paper is as follows. Section 2 introduces Hagendorff's virtue-based framework for AI ethics, which serves as the central analytical lens of this research. Section 3 presents the Socratic Dialogue as a method for ethical reflection, followed by Section 4, which outlines the methodological approach of this study. Section 5 integrates the Socratic Dialogue method with a real-world AI case on fraud and terrorism detection, showing how it cultivates virtues in practice. Section 6 provides a reflection of these findings. Section 7 discuss the broader relevance of our findings for AI ethics. Section 8 addresses potential limitations of applying the Socratic Dialogue in practice. Section 9 concludes with reflections on the role of the Socratic Dialogue in cultivating key moral virtues among AI practitioners. Section 10 reflects on the positionality of the first author.

## 2 A Virtue Ethics Approach to AI

The foundation for ethical AI lies in the cultivation of virtues, as outlined in Hagendorff's virtue-based framework (2022). Unlike deontology, which emphasizes universal principles guiding actions, virtue ethics focuses on the character and dispositions of individuals. This approach shifts attention from abstract rules to intrinsic motivations that align with ethical decision-making. Hagendorff identifies four essential virtues for AI practitioners: justice, honesty, responsibility, and care. When internalized, these virtues help ensure AI technologies are trustworthy and beneficial.

Each virtue connects directly to key ethical principles (Fjeld et al. 2020; Hagendorff 2020; Jobin, Ienca, and Vayena 2019). Justice aligns with fairness, encouraging systems that avoid discrimination and promote equity. Honesty supports transparency, motivating openness about technical limitations and research findings. Responsibility relates to accountability, guiding practitioners to take ownership of outcomes and address harm proactively. Care ties to AI safety, encouraging attention to harm prevention and stakeholder well-being. By embodying these virtues, practitioners act ethically without relying solely on external enforcement (Hagendorff 2022).

However, aligning with ethical principles is often difficult due to bounded ethicality (Hagendorff 2022): the cognitive, emotional, and situational constraints that hinder people from acting on their values. Ethical decision-making requires not only knowing what is right but also overcoming internal and external barriers. In AI, bounded ethicality underscores the need to cultivate virtues that help bridge the gap between awareness and action.

To address these challenges, Hagendorff introduces second-order AI virtues: practical wisdom and fortitude. Practical wisdom, or *phronesis*, is the Aristotelian "mother of all virtues," referring to the ability to reason and act virtuously in complex, real-world situations (Crisp 2014). Unlike theoretical knowledge (episteme) or technical skill (techne), phronesis is contextual, requiring intellectual reasoning, moral perception, and life experience. It enables practitioners to identify morally relevant features and balance competing values such as fairness, efficiency, and accountability, while choosing both the right ends and the right means.

Fortitude complements practical wisdom by fostering resilience against situational pressures that may compromise ethics (Hagendorff 2022). These include financial incentives, organizational norms, or conformity pressures that deprioritize ethical concerns (Lauer 2021). Fortitude empowers practitioners to resist these pressures by challenging unethical norms, questioning problematic directives, or advocating for responsible practices under time constraints. Together, practical wisdom and fortitude counteract the systemic and psychological barriers to ethical AI and reinforce the foundational virtues through thoughtful, context-sensitive moral decision-making.

## 3 The Socratic Dialogue: Background

The Socratic Dialogue, rooted in the method of questioning developed by Socrates (469–399 BCE), aims to refine moral judgment through dialogue rather than to assert definitive answers to moral dilemmas. Classical dialogues often led to *aporia*, the realization of one's own ignorance, but modern adaptations focus on developing ethical insight through structured inquiry.

In the 20th century, Nelson (1929) and later Heckmann (1981) formalized a method for examining moral questions through aiming for reasoned consensus. More recently,

Kessels, followed by Mostert and Boers, reoriented the dialogue toward situated judgment: the exploration of what a virtuous response should be in a specific context (Kessels 1997; Kessels, Boers, and Mostert 2009). Their approach values dissensus as a tool for ethical learning and cultivates practical wisdom rather than fixed moral truths.

In this contemporary form, a Socratic Dialogue is a structured dialogue in which participants explore a real moral dilemma drawn from experience, collectively investigating a "question of conscience" through five phases (see Results section). Central to this method is the judgment-free examination of a real-world experience shared by one participant, with others imagining themselves in that situation to explore emotional, cognitive, and action-based responses. Rather than aiming for consensus, the process facilitates a shared inquiry where divergent views enrich understanding (Boers 2022), making it particularly suited to contexts such as AI ethics, where challenges are contested and values may conflict. This study draws on Kessels' adaptation of the Socratic Dialogue, interpreted through Boers' lens as a philosophical practice for cultivating practical wisdom by engaging with concrete experiences and confronting moral complexity. The approach aims to support the development of virtues essential for ethical AI practice, including justice, honesty, care, responsibility, practical wisdom, and fortitude, which are further examined in the Results and Discussion sections.

At the heart of this process lies practical wisdom (phronesis), conceptualized as a tridimensional capacity of inquiring, judging, and acting (Rego et al. 2025). Rather than forming a fixed sequence, these dimensions represent mutually reinforcing aspects of moral perception and decision-making, drawing mainly on Thomas Aquinas (Regan 2005), Aristotle (Crisp 2014), and Naughton (2017).

**Inquiring** refers to studying and reflecting on the moral complexity involved in each particular situation. This requires the ability to perceive ethically relevant tensions in context. In the Socratic Dialogue, this ability is developed through the examination of a participant's lived experience, rather than abstract theory. Using *exemplary validity* (Boers 2022, p. 104), participants reconstruct a specific moral dilemma, exploring what was felt, thought, and done, reflecting Aristotle's view that ethical knowledge arises from experience.

**Judging** involves forming a reasoned judgment about what is morally appropriate. This happens through collective deliberation, where participants test their moral intuitions against those of others, reflect on underlying values and assumptions, and engage with disagreement to refine their views.

**Acting** entails the disposition to act in accordance with ethical reflection. Although dialogue itself does not prescribe behavior, it cultivates habits of reflection, attention, and self-awareness that support ethical action.

These dimensions frame the dialogical process presented in the Results section, where they are illustrated through a real-world use case in AI development.

# 4 Methodology

This study investigates how Socratic Dialogue can cultivate moral virtues among AI practitioners. To answer this question, the methodology is structured around two interconnected components. The first component focuses on the design of the Socratic Dialogue, analyzing how its structured format facilitates ethical reflection, builds moral character, and develops practical wisdom and fortitude. The second component centers on the content of the dialogue itself, particularly how participants engaged with an ethical dilemma on AI-driven detection of fraud and terrorism and developed more refined, context-sensitive understandings of key virtues (justice, care, honesty, responsibility). By combining an analysis of process and outcome, this study provides an integrated account of how moral virtues emerge, evolve, and become actionable in complex AI settings.

**Case Study** The study employs a case study approach to examine virtue development in real-world AI practice. The case was conducted in the Netherlands at a financial bank within its financial transaction monitoring department. This site was selected due to the ethically sensitive nature of AI use in fraud and terrorism detection, where practitioners regularly confront value-laden trade-offs involving fairness, security, privacy, and efficiency.

Seven professionals participated in the dialogue: four data scientists, two legal specialists, and one compliance and ethics expert. All participants were employees of the bank who were or had been involved in the development, deployment, and/or use of AI systems in the department. They were recruited through internal communication channels and participated voluntarily without receiving compensation.

The session was led by an external professional Socratic Dialogue facilitator, independent of the research team, who was compensated for their services. The dialogue lasted approximately 2.5 hours and focused on reconstructing a concrete moral dilemma encountered in the participants' work.

All participants provided informed consent prior to the session. The collected data was anonymized, and the bank reviewed all materials to ensure confidentiality. The study posed minimal risk to participants.

**Data Collection** The dialogue was documented through detailed minute-taking by a dedicated note-taker. Audio recording was deliberately omitted to avoid inhibiting participants, based on the facilitator's prior experience. Given that the case study serves primarily to illustrate concepts and processes, rather than to extract empirical insights from verbatim content, this approach was deemed appropriate. The minutes aimed to capture the flow and substance of the discussion as faithfully as possible, including direct quotes where feasible and summarized responses elsewhere. This ensured a comprehensive yet unobtrusive record of the session.

**Thematic Development and Maxim formation** The minutes of the dialogue were analyzed through thematic analysis. First, participant contributions, such as a judgment, statement, or reflection were coded with descriptive

labels (Johnny Saldaña 2015), for example, *role of empathy in criminality assessment* or *importance of system transparency*. The coding process was iterative and aimed to produce a clear, manageable set of codes.

Based on these codes, contributions were grouped into broader themes which reflected a range of perspectives, including opposing views, to capture how participants understood and approached the issues. Finally, the perspectives were condensed into maxims: concise, rule-like statements that summarized participant viewpoints while preserving their essential meaning (Boers 2022, p. 105). This process aligns with the Socratic Dialogue practice of distilling ethical reflections into generalizable principles, enabling a structured interpretation of the dialogue's key insights. While the conversion to maxims required a degree of generalization, care was taken to base these on participants' explicit statements, avoiding interpretations beyond what was directly expressed. By staying closely connected to the dialogue's content, the analysis aimed to respect the authenticity of participants' perspectives while allowing for meaningful synthesis. The dialogue was conducted in Dutch, and later translated to English prior to thematic coding. The overview of the process of going from dialogue to maxims is shown in Figure 1. Notation used in the figure is as follows: § denotes the index of a dialogue contribution, §§ a collection of indices. Contributions are numbered sequentially; when split into multiple perspectives, letters are added (e.g., 144A, 144B). Brackets indicate sets of contributions, and "..." denotes additional elements not shown. Coding and analysis were performed using Atlas.ti software (Friese 2012). Coding and analysis were performed using Atlas.ti software (Friese 2012).

**Abductive Analysis** This study employs an abductive analysis, which combines elements of deductive and inductive reasoning (Thompson 2022). Abduction, rooted in pragmatist philosophy (Peirce et al. 1985), enables researchers to iteratively explore how empirical findings interact with theoretical understanding (Atkinson, Coffey, and Delamont 2003; Timmermans and Tavory 2012). Rather than testing predefined hypotheses (deduction) or developing theory solely from data (induction), abductive reasoning fosters a dynamic interplay between empirical insights and conceptual frameworks (Jo Reichertz 2013; Thompson 2022).

This approach is well suited to this study, as both the cultivation of virtues in AI practitioners and the use of Socratic Dialogue for virtue development are underexplored areas. The analysis examined how the Socratic Dialogue fosters virtues essential for ethical AI, drawing on the virtue-based framework of Hagendorff (2022), the interpretation of the Socratic Dialogue by Boers (2022), and the operationalization of practical wisdom by Rego et al. (2025). The coding and thematic analysis was grounded in the dialogue itself, with thematic development based on participant perspectives rather than predefined theoretical categories. The theoretical perspectives served as an additional lens to verify and enrich the interpretation of what was happening in practice.

# 5 The Socratic Dialogue: Method and Case Analysis

This section presents an integrated account of the Socratic Dialogue method and its application to a real-world case involving AI systems for fraud and terrorism detection. It examines how the dialogue cultivates virtues while providing a detailed account of the method in practice. The process followed five key phases, with some aligning to the dimensions of practical wisdom: inquiring, judging, and acting, introduced earlier. These dimensions frame the analysis of how participants engaged with the moral complexities of relying on opaque, automated systems in high-stakes decisions about crime prevention and customer treatment. Conducted according to Kessels' method, the dialogue focused on a concrete, morally charged experience, integrating emotional, cognitive, and normative reflection (Boers 2022, p. 40). An overview of key insights derived from the dialogue can be found in Table 1. The following account describes each phase and illustrates its application in the case study.

## 5.1 Phase 1. Formulating the Question of Conscience

The dialogue began with the formulation of a central question of conscience: a moral question that cannot be answered simply by appealing to facts or legislation. Such questions express normative uncertainty and invite reflection on what ought to be done, rather than merely what can be done. In this case, the question was: "***To what extent do we want a machine to track down crime?***". The deliberate use of *want* instead of *can* avoided a purely legal or technical framing and encouraged participants to explore underlying values and societal goals. Similarly, the phrase *to what extent* invited them to identify a pivotal point of moral acceptability. By searching for these boundaries, the question stimulates exploring the factors and considerations that shaped participants' judgments about the ethical use of automation in high-stakes contexts like fraud and terrorism detection. The facilitator played an important role in guiding the participants to identify and formulate moral tensions.

## 5.2 Phase 2a. The Experience: *Inquiring*

After formulating the central question of conscience, participants grounded the moral exploration in a lived experience. This prevents the dialogue from drifting into abstract speculation, which often obscures what participants, including oneself, actually think. By working from a concrete, situated event, all participants share the same frame of reference, making the inquiry more tangible.

Each participant was invited to share a personal experience that embodied the moral tension of the central question. Such experiences must involve normative uncertainty: situations in which one was genuinely unsure what to do or what to think. They cannot be hypothetical; instead, they should present the real-world ambiguity in which ethical judgments are made. In line with (Boers 2022, p. 41), the experience is not merely illustrative but pivotal, offering a gateway into the deeper ethical tensions raised by the question. In the

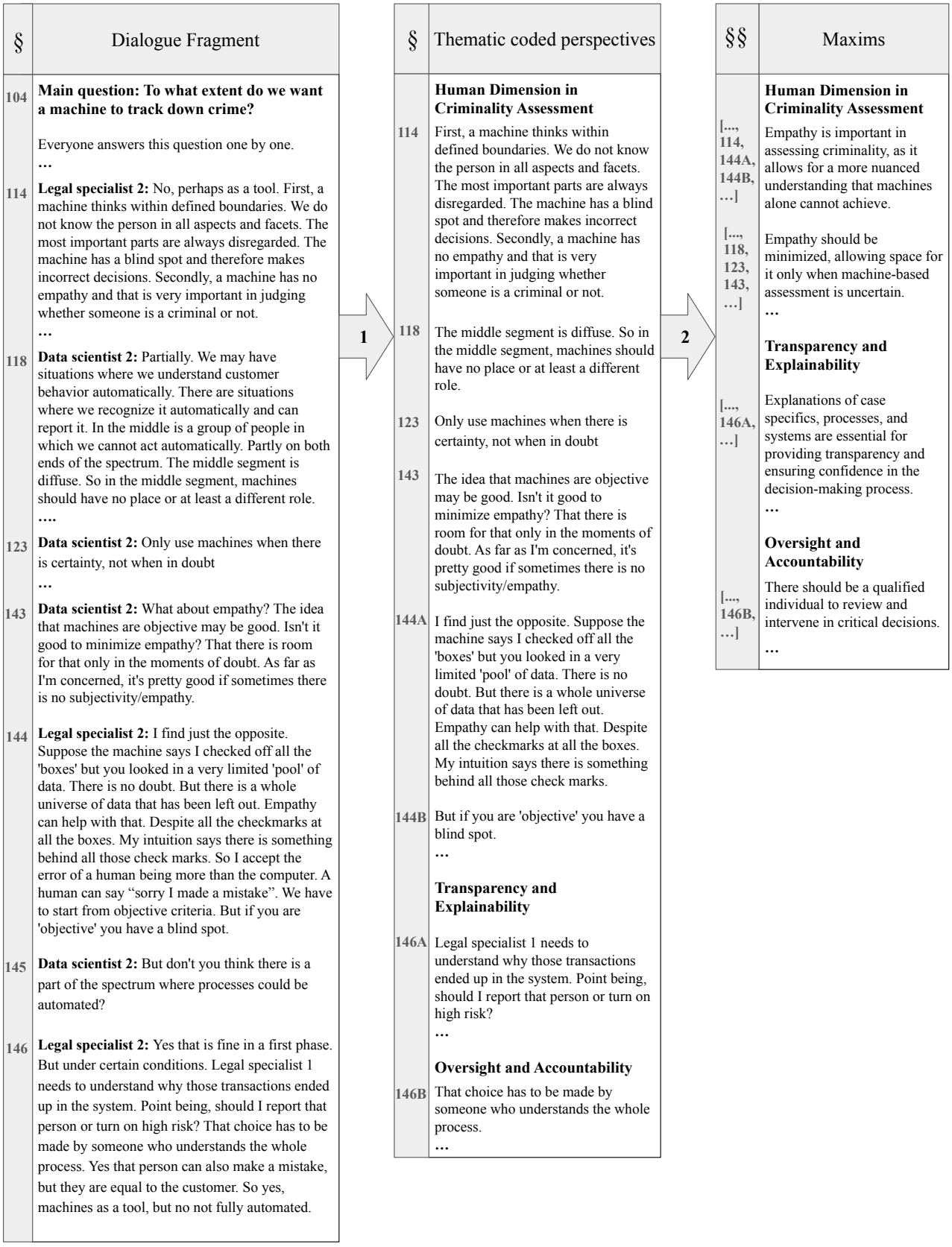


Figure 1: Analytical Process from Dialogue to Maxims (1) Thematic Coding, (2) Essence Extraction.

context of practical wisdom, this phase corresponds to *inquiring*, where one's moral perception is trained by paying attention to the specific, often ambiguous, details of lived experience.

After all experiences were shared, the group voted on which one to explore further. The chosen experience served as a lens through which to investigate the central question. The facilitator ensured that proposed experiences were suitable, capturing the essence of the moral dilemma while remaining specific enough to anchor the discussion.

In this dialogue, the selected experience was contributed by a legal specialist who previously worked in the general banking department. While the full story included many details, its essence was as follows:

*The participant had received a call from a customer whose bank account had been frozen due to transactions flagged under anti-money laundering and anti-terrorism regulations. The customer had an Arabic surname and had made donations to an Islamic foundation. According to the automated system and accompanying standard procedures, the employee was required to enter specific codes and immediately transfer the call to the fraud and anti-terrorism department, informing the customer that the freeze was terrorism-related. Even after consulting a manager, the directive remained unchanged. Although the customer's explanation seemed plausible and non-threatening, and the employee felt a strong discomfort about the course of action, they were unable to intervene in the process. The system did not allow space for discretion or explanation. Only later did the employee learn that the customer had done nothing wrong.*

### 5.3 Phase 2b. Reconstructing the Experience: The Hot Spot

During the reconstruction phase, participants questioned the experience giver in detail to ensure everyone understood exactly what had happened. The questioning began with factual details (What did the woman on the phone say? What was her reaction? What was the procedure?) and gradually moved toward motives and emotions (What doubts arose? What did the participant feel when transferring the call to the terrorism department?). Through this process, the group identified the *hot spot*: the emotionally and morally charged moment when ethical tension became most acute (Boers 2022, p. 174). This is where perception, emotion, and action intersect, offering a focal point for reflection.

The facilitator's role was crucial. Participants were often eager to offer judgments, advice, or arguments in favor of automation, sometimes embedding these in leading questions ("Don't you think that…?"). Such contributions were cut short or were asked to be reformulated, as this phase was devoted solely to understanding the particulars.

Some found it challenging to postpone judgment, yet this delay was essential: rushing to evaluate can obstruct genuine understanding. Learning to withhold initial judgment served as an exercise in developing moral perception, a key aspect of practical wisdom.

In this case, the hot spot was the moment when the employee, despite feeling that the customer posed no threat, had to follow a rigid system protocol that labeled the situation terrorism-related. The discomfort of the participant revealed a moral friction: the desire to respond empathetically versus the mandate to comply with opaque and procedural automation. This moment of tension offered a shared reflection on moral complexity, especially in situations where institutional logic and personal moral insight diverge.

### 5.4 Phase 3. Imagining Oneself in the Hot Spot: *Inquiring*

A core phase in the Socratic Dialogue method developed by Kessels is role-taking. In this step, participants imaginatively place themselves in the shoes of the person who shared the experience, reliving the *hot spot* moment as if it were their own (Boers 2022, p. 174). This practice shifts the focus from detached analysis to embodied moral perception, helping participants engage with the emotional and normative weight of the situation rather than discussing it from a distance. Each of the participants tries to answer the questions:

- What would I feel?
- What would I think?
- What would I do?

These reflections create space for moral dissensus (Boers 2022, p. 45): differences in response that reveal the diverse intuitions and value priorities among participants. Such divergence is not a problem to be resolved but a resource for deepening ethical reflection. It challenges participants to make their implicit assumptions explicit and to consider how others might perceive the same situation differently. The role of dissensus in shaping moral insight is explored further in the next phase.

In the dialogue, this exercise revealed contrasting moral responses. One participant, a data scientist, reflected:

*"I would experience stress over the message to be given. I would be understanding toward the lady. I would also understand the notification that something should be investigated. I would be supportive of that communication and so, unlike [the original storyteller], would not feel so bad about it."*

Another participant, also a data scientist, offered a different reaction:

*"I would lose confidence in the system and find it harder to go along with it. I would want to protect that woman from the system. I would want to know how the system works. How could I change it? How often is something really wrong in these cases? I would want to communicate it as nicely as possible to that lady."*

These reflections surfaced differing stances on moral responsibility, system trust and empathetic engagement. While one participant leaned toward accepting the system's procedural authority, the other expressed a protective

instinct and a desire for transparency and reform. Both responses underscored how lived ethical engagement depends not only on values but on how those values are prioritized and enacted under pressure.

In this phase, *inquiring* takes on a richer meaning. It is not simply about identifying moral tensions, but about embodying them, reflecting on emotional, cognitive, and action-oriented responses, and cultivating the self-awareness needed to navigate them. This aligns with Aristotle's notion of phronesis: ethical insight grounded in experience rather than abstract rules (Crisp 2014). Through role-taking, participants also gain exposure to alternative responses to the same situation, enriching their moral understanding.

### 5.5 Phase 4. Formulating Rules: *Judging*

In the final stage of the Socratic Dialogue, participants move from *inquiring* to *judgment*: from reflecting on the moral aspects of the situation to deliberating on how to respond. While participants did not explicitly frame their reflections in terms of named virtues, their reasoning touched on themes that were later interpreted through the lenses of virtues essential for ethical AI:

- **Responsibility**: When does delegating decisions to machines diminish human accountability? What is the ethical cost of removing the possibility for apology, ownership, and empathy?
- **Justice**: Does automation enhance fairness through consistency, or does it risk reinforcing systemic bias under the guise of neutrality?
- **Care**: What is lost and what is gained when empathetic human understanding is sidelined by procedural automation? How can systems balance the dignity of the individual with the collective imperative to prevent harm through fraud and terrorism detection?
- **Honesty**: How transparent must systems be to earn trust? What obligations do institutions have to ensure that those affected understand how and why decisions are made?

Judgments that emerge from this process did not function as universal moral laws, but rather as maxims: situated judgments that reflect the moral significance of the dialogue (Boers 2022, p. 105).

For example, when participants deliberated on the role of empathy in criminality assessment, divergent but meaningful perspectives emerged:

– *"Empathy is important in assessing criminality, as it allows for a more nuanced understanding that machines alone cannot achieve."*

– *"Empathy should be minimized, allowing space for it only when machine-based assessment is uncertain."*

Other maxims reflected concerns about fairness and accountability:

– *"A human mistake is more acceptable than a mistake made by a machine."*

– *"Machines can assess more consistently and objectively than humans."*

A full overview of the maxims, organized by theme, is provided in Table 1.

Rather than seeking consensus, the Socratic Dialogue values dissensus as a source of moral learning. Participants tested their intuitions against those of others, sharpening their reasoning through careful listening and questioning. Divergent views became openings for deeper reflection. The maxims served both as records of shared moral inquiry and as provocations for continued ethical deliberation. In this way, the dialogue helped cultivate practical wisdom not only in individuals but also within professional communities committed to responsible AI.

This phase also helped to cultivate fortitude, as the dialogue created a space where questioning others' ideas, values, and beliefs was not only permitted but expected. Participants were encouraged to state their judgments clearly while remaining open to other perspectives. Fortitude was expressed through both self-expression and attentiveness to others, making it possible to disagree without dismissal and to hold convictions without closing inquiry. Such sustained, open engagement is central to moral development and essential for addressing the complex value tensions in AI ethics.

### 5.6 Phase 5. Finding the Essence: From Experience to Principle

The final step of the Socratic Dialogue is to distill the moral essence of the issue as the underlying value tension or shared moral insight that gives the experience its significance. This essence is not a universal truth imposed from outside, but a situated understanding that emerges from the dialogue itself. It is shaped by the group's concrete deliberation and reflects the deeper meaning that participants collectively discover through the exchange of perspectives.

What ultimately emerged as the essence was to moral obligation to protect human dignity. Before the dialogue, this may have seemed self-evident, something everyone could agree on in the abstract. But through collective engagement, grounded in a concrete and morally complex experience, participants developed a richer and more nuanced understanding of what human dignity entails in this context. They came to see how dignity is implicated at multiple levels of the problem, see Table 1.

Importantly, participants did not just affirm the importance of human dignity. They explored it, questioned it, and generated insight around it. They learned how others in their professional community interpret this concept, where their views align or diverge, and what trade-offs it demands in practice. This shared inquiry transformed human dignity from a vague ethical ideal into a lived, context-sensitive principle; something that could actively guide reflection and action in future decisions.

## 6 Reflection on Outcomes and Virtues

The Socratic Dialogue fosters the cultivation of moral virtues by enabling professionals to engage deeply with ethically charged experiences. In the case examined, participants reflected on the use of AI systems in fraud and terrorism detection. While virtues such as justice, care, honesty,

| Theme | Maxims |
|---|---|
| **Human Dimension in Criminality Assessment** | • Empathy is important in assessing criminality, as it allows for a more nuanced understanding that machines alone cannot achieve.<br>• Empathy should be minimized, allowing space for it only when machine-based assessment is uncertain.<br>• It's important to empathize with those affected, while also supporting necessary investigations.<br>• The individual undergoing criminal assessment must have the opportunity to be heard.<br>• A human mistake is more acceptable than a mistake made by a machine.<br>• Being accused carries a significant stigma.<br>• Ensuring human dignity is an unconditional requirement. |
| **Machine Dimension in Criminality Assessment** | • Automation is essential for handling large volumes.<br>• Machines can assess more consistently and objectively than humans.<br>• Effectiveness does not necessarily equate to fairness.<br>• In assessments, assumptions based on questionable factors should be handled cautiously, especially in sensitive contexts to avoid discrimination.<br>• Technology's binary nature can be insufficient for complex, nuanced situations. |
| **Oversight and Accountability** | • Crime detection affects individuals, who should have the right to human judgment.<br>• There should always be an option for qualified human review in critical decisions.<br>• Humans must control the machine, with explainability as a key component.<br>• Humans must remain accountable for machine decisions.<br>• Justice is not deterministic therefore we need to leave control in the hand of man. |
| **Authority and Acceptance** | • In sensitive matters, orders from humans are more accepted than those from machines.<br>• Voicing an opinion and feeling heard makes a significant difference in accepting orders on sensitive matters.<br>• It is important to be heard when expressing concerns about the system performance.<br>• Feeling powerless happens when your input is ignored from the start. |
| **Transparency and Explainability** | • Explaining systems and processes supports justice and fairness.<br>• Explanations on case specifics are essential for providing transparency and ensuring confidence in the decision-making process.<br>• In cases of perceived injustice, understanding the system becomes important. |

Table 1: Themes and Maxims resulting from the analysis of observations of the Socratic Dialogue on Criminality Assessment.

and responsibility were not explicitly named during the dialogue, they clearly emerged through the participants' moral reasoning and emotional engagement.

Cultivation, in this context, means moving beyond merely knowing ethical principles to developing the perceptual and emotional capacities to recognize and respond to value tensions in practice.

*Justice* emerged as a central concern, linking fairness with impartiality. Participants valued AI's potential to reduce human bias and ensure consistent decisions, but also stressed the right to be heard and treated with dignity, seeing fairness as both equal treatment and human recognition. While automation was seen as essential for efficiency at scale, concerns about systemic bias in data and algorithms revealed a core tension: justice demands both impartial systems and attention to individual circumstances.

*Care* was reflected in valuing empathy and human judgment for understanding the emotional and situational context of decisions, especially in high-stakes settings. Participants saw empathy as important for ensuring human dignity, yet also linked care to preventing societal harms such as fraud and terrorism. These aims can conflict, as empathy for individuals may need to be balanced against protecting the broader public, revealing the complex demands of care in socio-technical systems.

*Honesty* was reflected in calls for transparency and ex-

plainability. Participants stressed that understanding how systems function and why certain decisions are made is essential for maintaining fairness and trust. Honesty also extended beyond technical clarity to include a moral commitment to open dialogue, both with those affected by the system and with colleagues who raise concerns internally.

*Responsibility* was expressed in the insistence on human accountability for automated decisions and the need for meaningful oversight. While automation is vital for handling large data volumes, participants warned against ethical detachment. Some preferred human error, which allows for ownership, apology, and empathy. Others, especially data scientists, felt responsible for system outcomes despite being far from final decisions. These views highlight the need to navigate the boundary between human and machine responsibility with carefully.

Across Justice, Honesty, Care, and Responsibility, participants faced tensions that revealed virtues as dynamic guides rather than fixed rules, requiring continual interpretation and negotiation in context.

Beyond these virtues, the dialogue fostered the Aristotelian virtue of practical wisdom (phronesis). This involves perceiving what is morally significant, making sound judgments, and preparing to act accordingly. Each phase of the dialogue aligned with these aspects. Rather than relying on abstract theories, participants reasoned through concrete dilemmas drawn from lived experience, symbolizing Aristotle's claim that ethical understanding arises from experience instead of abstract rules (Crisp 2014).

Practical wisdom was cultivated as a collective effort. It was sharpened through dissensus, as participants had to articulate and explain their views under scrutiny. It was enriched through diverse perspectives, which challenged assumptions and exposed blind spots. This interplay of perspectives helped transform general ideals, like human dignity, into context-sensitive principles that could guide future action. In this way, the dialogue not only supported individual moral insight but also strengthened the group's shared capacity for ethical reasoning.

Fortitude, a companion virtue to phronesis, was equally essential. In the dialogue, fortitude emerged not as a rigid defense of one's views but as the courage to engage in open moral inquiry. It requires strength to express one's ethical position, humility to subject it to critique, and resilience to remain in dialogue even under disagreement. Participants practiced this by revisiting assumptions, asking clarifying questions, and exploring others' value priorities. In doing so, they exercised the perseverance needed to sustain moral integrity in professional environments where reflection is often sidelined. This form of fortitude supports personal development and the conditions for collective ethical reasoning in AI teams. It helps build a professional culture in which dissent is not punished, but cultivated as part of responsible AI practice. This emphasis on cultivating dissent resonates with arguments in the literature highlighting the importance of dissent mechanisms for navigating trade-offs and normative uncertainty in AI development. Such mechanisms can support more deliberative decision-making and strengthen public accountability (Dobbe, Gilbert, and Mintz 2021).

# 7 Relevance of Results in AI Ethics

As we pointed out in the Introduction, the literature on virtue ethics in AI development offers mostly conceptual accounts and no empirical or practical accounts on how virtues can be understood, embodied, built and developed in practice. This gap extends to relevant adjacent fields studying the normative dimensions of AI system design, including FAccT, Human-Computer Interaction, Design for Values, Value Sensitive Design and Participatory Design. This gap is alarming given the many ways in which undesirable and harmful outcomes from AI systems may be prevented or addressed in AI system development. On the other hand, the gap can be partly understood by the known limitations occurring in organizations. As Green (2021) shows, tech ethics more broadly is often "vague and toothless, has a myopic focus on individual engineers and technology design, and is subsumed into corporate logics and incentives.". As such, as Wagner (2018) argues, "it is important to have common criteria based on which the quality of ethical and human rights commitments made can be evaluated."

While we subscribe to the importance of distinguishing ethical reflection from the need to regulate and uphold fundamental rights, we pose that empirical research and practice with Socratic Dialogue approaches as explored here, are necessary to understand what criteria may effectively foster, guide and guarantee tangible and effective virtue development in AI system design. In turn, virtues and practical wisdom, are known to be crucial for effective ethical decision-making and "necessary in the proper interpretation and implementation of human rights" (Sison 2018).

Dialogue alone cannot undo structural barriers that inhibit ethical action such as power dynamics in the workplace, misaligned incentives, or resource constraints that may impede conditions for ethical reflection (Manders-Huits 2011; Lauer 2021). However, it can help practitioners name and navigate these tensions, laying a foundation for more systemic change. To have impact, dialogue must be embedded within organizational cultures that value ethical inquiry. This includes leadership that stimulates moral courage, governance structures that reward ethical action and operational norms that protect space for dissent, establishing a *just culture* (Dobbe 2022).

As such, the Socratic Dialogue may complement existing responsible AI practices, such as the determination of appropriate fairness criteria by surfacing the moral reasoning that these metrics often presuppose (Mitchell et al. 2021). While fairness metrics describe what can be measured, dialogue can help clarify what should matter morally and why. It encourages participants to articulate, contest, and refine their value commitments and potentially gain deeper insight into the inevitable trade-offs these involve (Kleinberg, Mullainathan, and Raghavan 2016). In this way, dialogue could play a valuable role in socio-technical AI governance, helping stakeholders to navigate the ethical complexity that metrics alone cannot capture. As regulatory measures evolves away from tech-centric standards, to centering development practices and organizational mechanisms, a more robust body of studies in the Socratic Dialogue may well inform associated criteria and standards.

More broadly, the virtue-oriented approach of Socratic Dialogue may help to address some of the limitations of participatory and value-sensitive design practices, which have been criticised for lacking normative grounding or reproducing extractive dynamics between those who make design decisions and those who merely 'participate' in the design process (Manders-Huits 2011; Bødker and Kyng 2018; Sloane et al. 2022). By cultivating a lasting capacity for ethical deliberation within a team or organisation, Socratic Dialogue may stimulate a culture in which such reflective and inclusive practices are supported in ways that meaningfully empower marginal communities and other actors whose voices are otherwise undervalued in system design practices.

## 8 Limitations

While the Socratic Dialogue has potential for cultivating virtues, four limitations should be acknowledged. Firstly, one of the most immediate concerns is the time investment required. A well-conducted dialogue typically takes two to three hours, which may seem excessive in fast-paced environments where quick decision-making is prioritized. However, this time commitment should be weighed against potential long-term benefits. By fostering structured ethical reflection, the Socratic Dialogue can lead to more streamlined decision-making within organizations. Rather than addressing ethical concerns in a fragmented or reactive manner, organizations that integrate the Socratic Dialogue into their processes may ultimately save time by avoiding costly system redesigns (Bevilacqua et al. 2023) or societal costs of failures caused by poor system considerations. Seen in this light, the time required for deep deliberation is not merely an operational cost but an investment in the robustness and foresight of AI decision-making.

Secondly, another challenge lies in the need for skilled facilitation. The effectiveness of a Socratic Dialogue depends on the facilitator's ability to guide discussion, encourage open inquiry, and create a psychologically safe space for participants (van Baarle and Verweij 2008). Without proper facilitation, discussions risk becoming superficial or unstructured, limiting their value. Yet the need for expertise extends to other methods, such as Value Sensitive Design, algorithmic auditing, or fairness impact assessments. Ethical deliberation does not emerge spontaneously but is cultivated through structured methodologies that demand training and skill. The need for expertise in Socratic Dialogue, therefore, is not a unique limitation but a common requirement for any serious attempt to embed ethical considerations into AI development.

Thirdly, resistance to ethical questioning can also be a challenge. Socratic Dialogue asks people to reflect critically on their own views, which can feel uncomfortable, especially when it challenges personal beliefs or company routines. However, this discomfort can lead to valuable reflection and learning. If participation is forced, some may disengage, turning the dialogue into a box-ticking exercise. Organizations need to sustain a culture where ethical reflection is seen as a normal and important part of working with AI, not as something imposed from the outside.

Fourthly, power differences in hierarchical settings can make open discussion harder and have been shown to plague most AI ethics endeavors (Birhane et al. 2022). People in lower positions might not feel comfortable speaking up or challenging those in authority. This can lead to safe, surface-level conversations that avoid deeper issues. One way to reduce this is to be intentional about who gets invited. For example, leaving out high-level managers in early sessions, or making sure the group is balanced, can help others speak more freely. Good facilitation also plays an important role in making sure everyone is heard and the conversation doesn't just follow the usual power lines. While you can't remove all power dynamics, dialogue can still create a space where different perspectives are taken seriously.

Taken together, these limitations highlight the challenges of integrating Socratic Dialogue into AI ethics discussions. However, rather than dismissing these challenges as insurmountable, they should be viewed as areas that require thoughtful implementation. The effectiveness of any ethical deliberation method depends on how well it is embedded into an organization's practices, and Socratic Dialogue is no exception. When implemented with care, it has the potential to cultivate deeper ethical reasoning and lead to more responsible AI systems.

## 9 Conclusion

This paper explored how participation in the Socratic Dialogue supports the cultivation of key moral virtues among AI system developers. Grounded in a case study from the financial sector, the dialogue revealed that structured ethical inquiry offers more than reflective insight; it fosters the development of practical wisdom and fortitude, virtues essential for navigating the moral complexity of AI development. Through sustained engagement with conflicting values, participants practiced ethical discernment, learned to navigate ambiguity, and developed the courage to uphold their convictions even under institutional or peer pressure.

By foregrounding the lived experiences of practitioners, the dialogue illuminated how core virtues justice, care, honesty, and responsibility are not static principles but dynamic and sometimes internally contested commitments. Rather than resolving these tensions through rigid rules, participants learned to inhabit them thoughtfully, cultivating the resilience needed to make balanced judgments. Importantly, this growth was not limited to individual reflection; it was made possible through the collective nature of the dialogue, which encouraged empathy, mutual accountability, and the articulation of shared norms.

The findings suggest that Socratic Dialogue can function as a developmental space where AI practitioners are not merely trained in compliance but actively formed in their moral capacities. When supported by skilled facilitation and embedded in a culture that values inquiry, this method can strengthen practitioners' ability to resist conformity, question directives, and advocate for ethically sound system design. In doing so, it offers a valuable contribution to the cultivation of virtue within the socio-technical ecosystems that shape AI today.

## Positionality Statement

This statement reflects the perspective of the first author. Having grown up with a parent who is a mediator and narrative therapist, I developed an appreciation for exploring diverse perspectives and uncovering underlying assumptions in conflict. My later involvement with a Socratic Dialogue hobbyist group strengthened this orientation and motivated my interest in researching its application to ethical AI practice. These experiences inevitably shape how I evaluate the method's potential.

## Acknowledgments

This research was conducted as part of the Gravitation research program Hybrid Intelligence, funded by the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) under file number 024.004.022.

## References

Atkinson, P.; Coffey, A.; and Delamont, S. 2003. *Key themes in qualitative research: continuities and changes*. Walnut Creek: Altamira Press. ISBN 978-0-7591-0126-5 978-0-7591-0127-2.

Bankins, S.; Ocampo, A. C.; Marrone, M.; Restubog, S. L. D.; and Woo, S. E. 2024. A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*, 45(2): 159–182.

Bevilacqua, M.; Berente, N.; Domin, H.; Goehring, B.; and Rossi, F. 2023. The Return on Investment in AI Ethics: A Holistic Framework. arXiv:2309.13057.

Birhane, A.; Ruane, E.; Laurent, T.; S. Brown, M.; Flowers, J.; Ventresque, A.; and L. Dancy, C. 2022. The Forgotten Margins of AI Ethics. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 948–958. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9352-2.

Bødker, S.; and Kyng, M. 2018. Participatory design that matters—Facing the big issues. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25(1): 1–31.

Boers, E. 2022. *From Science to Conscience. The Socratic Dialogue Reconsidered*. Ph.D. thesis, Radboud University, Nijmegen.

Constantinescu, M.; Voinea, C.; Uszkai, R.; and Vică, C. 2021. Understanding responsibility in Responsible AI. Dianoetic virtues and the hard problem of context. *Ethics and Information Technology*, 23(4): 803–814.

Crisp, R. 2014. *Aristotle: nicomachean ethics*. Cambridge University Press.

Dobbe, R.; Gilbert, T. K.; and Mintz, Y. 2021. Hard Choices in Artificial Intelligence. arXiv:2106.11022.

Dobbe, R. I. J. 2022. System Safety and Artificial Intelligence. Version Number: 1.

Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; and Srikumar, M. 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*.

Friese, S. 2012. *Qualitative Data Analysis with ATLAS.ti*. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd. ISBN 978-0-85702-131-1 978-1-5297-9959-0.

Green, B. 2021. The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice. *Journal of Social Computing*, 2(3): 209–225.

Hagendorff, T. 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1): 99–120.

Hagendorff, T. 2022. A Virtue-Based Framework to Support Putting AI Ethics into Practice. *Philosophy & Technology*, 35(3): 55.

Hayes, P.; Fitzpatrick, N.; and Ferrández, J. M. 2024. From applied ethics and ethical principles to virtue and narrative in AI practices. *AI and Ethics*.

Heckmann, G. 1981. *Das sokratische Gespräch: Erfahrungen in philosophischen Hochschulseminaren*. Hannover: Schroedel. ISBN 978-3-507-39014-0.

Jo Reichertz. 2013. Induction, Deduction, Abduction. In *The SAGE Handbook of Qualitative Data Analysis*, 123–136.

Jobin, A.; Ienca, M.; and Vayena, E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9): 389–399.

Johnny Saldaña. 2015. *The Coding Manual for Qualitative Researchers*. SAGE. ISBN 978-1-4739-4359-9.

Kazim, E.; and Koshiyama, A. S. 2021. A high-level overview of AI ethics. *Patterns*, 2(9): 100314.

Kessels, J. 1997. *Socrates op de markt: filosofie in bedrijf*. Amsterdam: Boom. ISBN 978-90-5352-350-6.

Kessels, J.; Boers, E.; and Mostert, P. 2009. *Free space: field guide to conversations*. Boom. ISBN 90-8506-834-7.

Kleinberg, J.; Mullainathan, S.; and Raghavan, M. 2016. Inherent Trade-Offs in the Fair Determination of Risk Scores. arXiv:1609.05807.

Knezic, D.; Wubbels, T.; Elbers, E.; and Hajer, M. 2010. The Socratic Dialogue and teacher education. *Teaching and Teacher Education*, 26(4): 1104–1111.

Kraut, R. 2022. Aristotle's Ethics. In Zalta, E. N.; and Nodelman, U., eds., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2022 edition.

Lauer, D. 2021. You cannot have AI ethics without ethics. *AI and Ethics*, 1(1): 21–25.

Maaike Harbers; Komala Mazerant; Jan Ewout Ruiter; and Rudolf Kampers. 2019. Socratic Dialogue as a Method for Moral Inquiry in HCI. Glasgow Scotland UK: ACM. ISBN 978-1-4503-5971-9.

Manders-Huits, N. 2011. What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics*, 17(2): 271–287.

McNamara, A.; Smith, J.; and Murphy-Hill, E. 2018. Does ACM's code of ethics change ethical decision making in software development? In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software*

*Engineering*, 729–733. Lake Buena Vista FL USA: ACM. ISBN 978-1-4503-5573-5.

Mitchell, S.; Potash, E.; Barocas, S.; D'Amour, A.; and Lum, K. 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application*, 8(Volume 8, 2021): 141–163. Publisher: Annual Reviews.

Mittelstadt, B. 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11): 501–507.

Munn, L. 2023. The uselessness of AI ethics. *AI and Ethics*, 3(3): 869–877.

Naughton, M. 2017. *Practical Wisdom as the Sine Qua Non Virtue for the Business Leader*, 189–197. Dordrecht: Springer Netherlands. ISBN 978-94-007-6510-8.

Nelson, L. 1929. Die sokratische Methode. *Abhandlungen der Fries'schen Schule: Neue Folge*, 1: 21–78.

Peirce, C. S.; Hartshorne, C.; Weiss, P.; and Peirce, C. S. 1985. *Principles of philosophy: two volumes in one*. Number Vol. 1/2 in Collected papers of Charles Sanders Peirce / ed. by Charles Hartshorne. Cambridge, Mass.: Belknap Press of Harvard Univ. Press, 5. [printing] edition. ISBN 978-0-674-13800-1.

Qian, Y.; Siau, K. L.; and Nah, F. F. 2024. Societal impacts of artificial intelligence: Ethical, legal, and governance issues. *Societal Impacts*, 3: 100040.

Raquib, A.; Channa, B.; Zubair, T.; and Qadir, J. 2022. Islamic virtue-based ethics for artificial intelligence. *Discover Artificial Intelligence*, 2(1): 11.

Regan, R. J. 2005. *The Cardinal Virtues: Prudence, Justice, Fortitude, and Temperance*. Hackett Publishing Company.

Rego, A.; Meyer, M.; Júnior, D. R.; and Cunha, M. P. E. 2025. Wise leaders fostering employees' speaking up behaviors: developing and validating a measure of leader-expressed practical wisdom. *Review of Managerial Science*, 19(1): 157–195.

Saran, R.; and Neisser, B., eds. 2004. *Enquiring Minds: Socratic Dialogue in Education*. Stylus Publishing, LLC.

Schwitzgebel, E. 2009. Do ethicists steal more books? *Philosophical Psychology*, 22(6): 711–725.

Schwitzgebel, E.; and Rust, J. 2014. The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philosophical Psychology*, 27(3): 293–327.

Sison, A. J. G. 2018. Virtue Ethics and Natural Law Responses to Human Rights Quandaries in Business. *Business and Human Rights Journal*, 3(2): 211–232.

Sloane, M.; Moss, E.; Awomolo, O.; and Forlano, L. 2022. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–6.

Stahl, B. C.; Antoniou, J.; Ryan, M.; Macnish, K.; and Jiya, T. 2022. Organisational responses to the ethical issues of artificial intelligence. *AI & SOCIETY*, 37(1): 23–37.

Thompson, J. 2022. A Guide to Abductive Thematic Analysis. *The Qualitative Report*.

Timmermans, S.; and Tavory, I. 2012. Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis. *Sociological Theory*, 30(3): 167–186.

Vallor, S. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press. ISBN 978-0-19-049851-1.

van Baarle, E.; and Verweij, D. 2008. Inquiry, Criticism and Reasonableness: Socratic Dialogue as a Research Method? *Practical Philosophy*.

Wagner, B. 2018. Ethics as an escape from regulation. From "ethics-washing" to ethics-shopping? Publisher: Amsterdam University Press.