AI and Ethics (2025) 5:607–616
https://doi.org/10.1007/s43681-023-00381-7

ORIGINAL RESEARCH

Engaging engineering teams through moral imagination: a bottom-up approach for responsible innovation and ethical culture change in technology companies

Benjamin Lange[1] · Geoff Keeling[1] · Amanda McCroskery[1] · Ben Zevenbergen[1] · Sandra Blascovich[1] · Kyle Pedersen[1] · Alison Lentz[1] · Blaise Agüera y Arcas[1]

Received: 17 October 2023 / Accepted: 10 November 2023 / Published online: 19 December 2023
© The Author(s) 2023

Abstract
We propose a 'Moral Imagination' methodology to facilitate a culture of responsible innovation for engineering and product teams in technology companies. Our approach has been operationalized over the past two years at Google, where we have conducted over 60 workshops with teams from across the organization. We argue that our approach is a crucial complement to existing formal and informal initiatives for fostering a culture of ethical awareness, deliberation, and decision-making in technology design such as company principles, ethics and privacy review procedures, and compliance controls. We characterize some distinctive benefits of our methodology for the technology sector in particular.

Keywords Ethical culture · Responsible innovation · AI ethics · Ethics in technology · Moral imagination · Culture change management · Ethical awareness · Ethical deliberation · Ethical-decision-making · Bottom-up methodology · Practitioner perspective

1 Introduction

The norms and values of technology teams shape which technologies are produced and how.[1] But these norms and values are rarely made explicit and subjected to critical appraisal, leading to limited ethical reflection and potentially reinforcing biases in technological development. In response, there are growing calls to change the culture that shapes the production of technologies. These include calls for greater governance such as government regulation and industry self-regulation by internal ethics review committees[2] and governing principles.[3] Also included are calls for engineer education in computer science curricula and industry training modules,[4] alongside technical best practice

[1] Winner [56], Ihde [22], Feenberg [8], see also Weinstein, Reich and Sahami [51].

✉ Benjamin Lange
benjamin.lange@lmu.de

Amanda McCroskery
amccroskery@google.com

Ben Zevenbergen
benzevenbergen@google.com

[1] Google Research, Mountain View, USA

[2] Jackman and Kanerva [23], Blackman (2022: 151–57), Shneiderman (2021: 34), Prunkl et al. [34].

[3] Whittlestone et al. [54].

[4] Grosz et al. [18], Fiesler, Garrett, and Beard [9], Garret, Beard, and Fiesler [9, 14].

developments such as anticipatory or value sensitive design,[5] technical approaches to value alignment[6] and algorithmic auditing.[7]

These approaches, while critical for promoting socially and ethically responsible technological development, have under-addressed a central cultural dynamic of technology production, namely, the group forums where critical decisions about product and research direction are made day-to-day. In the technology industry in general, these decisions are often negotiated within engineering, product, and research teams, where largely autonomous, entrepreneurially driven groups decide which problems are addressed through technology and how, before elevating recommendations to managers and executives. To be sure, the autonomy afforded to engineering teams is a key aspect of how technology companies drive innovation. For example, in a recent study on the determinants of innovation in a Swedish software company, Andersén and Ljungkvist [2] found that '[a]lthough managers play a key role in top-down oriented innovation processes [2], innovation is often achieved by smaller groups at the operational level.' Indeed, this observation corroborates an earlier finding by Peerasit Patanakul, Jiyao Chen, and Gary S. Lynn [31] that '[r]elative to other team structures, autonomous teams are more effective in addressing projects [26, 27] with high technology novelty or radical innovation.'[8]

In this paper, we describe a "Moral Imagination" methodology to drive a culture of responsible innovation, ethical awareness, deliberation, decision-making, and commitment in technology organizations.[9] This approach aims to prompt a *role obligation shift* among teams that makes the consideration of the moral implications of their work an inherent part of their self-conception and day-to-day decision-making. As practitioners, we have developed and tested this capability building approach over the past two years at Google through over 60 workshops involving a range of research and product teams [36].

Our primary aim is to make the conceptual case for our approach, and not to present an empirical study on the efficacy of the proposed methodology, which we are pursuing in other work. Neither do we intend to suggest that our approach is superior compared to other initiatives such as traditional ethics and compliance controls, review boards, or ethics committees. Rather, we see Moral Imagination as an important complementary effort that can serve as one part of a portfolio approach to responsible innovation at technology companies [36, 38].

Our discussion makes three contributions to existing scholarship on responsible innovation and technology ethics. First, it highlights a neglected gap in the current arsenal of instruments to manage responsible innovation at technology companies—the shaping of norms within engineering culture. Second, it spells out a concrete method for filling this gap, in a way that builds on existing responsible innovation frameworks, while at the same time tailoring the methodology to the distinctive cultural and organizational features of technology companies. Third, it aims to detail concrete real-world cases of application by illustrating the operationalization of this approach based on our practitioner experience.

The paper is structured as follows. Section 2 details the specific challenge that we are concerned with: equipping teams with the skills to responsibly navigate the increasing social and ethical requirements in developing their technology and products. We argue that a comprehensive culture of responsible innovation at tech companies requires interventions that work to adjust tech team norms. Against this backdrop, Sect. 3 then develops our Moral Imagination approach. We suggest that there are three key capabilities that should be fulfilled to enable robust ethical culture change among teams within technology companies: Ethical Awareness, Ethical Deliberation and Decision-making, and Ethical Commitment. We then show how our framework enhances these capabilities. Section 4 concludes.

2 The need for norm shift

2.1 Overview

Our argumentative strategy in this section is to show that—given a plausible assumption about the responsibility of technology companies in general—a crucial lever for enabling responsible innovation is currently not fully realized, and to then argue in the subsequent section that our approach

[5] Umbrello [47] and Umbrello et al. [48] as well as Friedman [11] and Friedman and Hendry [12].

[6] Amodei et al. [1], Kenton et al. (2021), Gabriel [13].

[7] Brown, Davidovic, and Hasan [5] and Hasan et al. [20].

[8] For further empirical treatments of the role of teams and interpersonal dynamics in software engineering see Scott and Einstein [40], Caldwell and O'Reilly [6], Karn [26], Glynn et al. [16], Somech and Drach-Zahavy [43], Robbins and O'Gorman [36], Gerrard and Lockett [15], and Hoffman et al. [21].

[9] Some terminological clarifications. First, we use the term "engineer" to cover a wide spectrum of roles involved in tech development, including but not limited to software engineers, data scientists, research scientists, product managers, engineering leads, UX researchers, UX designers, or managers. We use the term 'engineering teams' inclusively. Second, throughout the discussion, we refer to the concept of "ethical culture" understood, roughly, as "the shared values, beliefs, norms, policies, procedures, systems, and artifacts that shape the behaviors of members of an organization and support ethical conduct." We consider ethical culture as the encompassing construct within which initiatives such as responsible innovation or value-sensitive design that specifically aim at embedding the values or interests of broader stakeholders in technology research and development can fall.

can fill this gap. More specifically, our argument consists of the following four key claims:

1. **Tech's Practical Responsibility Requirement:** in light of the policy vacuum in which technology is developed, technology companies have a practical responsibility to consider how to produce technologies that are sensitive to the ethical and sociotechnical contexts in which those technologies will be deployed**.**
2. **Importance of Shaping Norms:** shaping engineers' team culture and prevalent norms is a key element in responding to *Tech's Practical Responsibility Requirement.*
3. **Gap in Current Measures:** typically employed hard and soft controls are necessary but insufficient to fully shape predominant team norms in the design and development stages.
4. **New Opportunity for Intervention:** therefore, there is an opportunity to devise new interventions that can meet the specific requirements distinctive engineering team culture poses and complement existing measures.

2.2 Elaboration of argument

1. **Tech's Practical Responsibility Requirement:** in light of the policy vacuum in which technology is developed, technology companies have a practical responsibility to consider how to produce technologies that are sensitive to the ethical and sociotechnical contexts in which those technologies will be deployed.

We take this assumption as given. Our claim here draws on what has been called the "pacing problem" for technology.

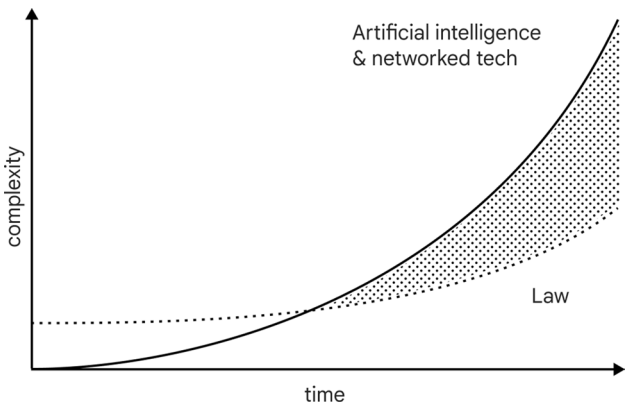


Illustration 1. Pacing problem in technology

The pace of technological innovation exceeds that of regulation.[10] As a result, existing laws and policies do not necessarily provide guidance to technical companies about how to align their work with the needs and interests of society—what James Moor [29] has also referred to as a "policy vacuum"[11].

The pacing problem presents two *ethical gray areas*. That is, ambiguities about how to proceed ethically when developing technology given that rules, laws, and policies do not straightforwardly provide guidance. The first gray area concerns the absence of appropriate regulation for new technology, in the sense that legislation is not in place to set relevant boundaries for technological systems. The second gray area concerns the *interpretation* of existing laws. This means that when regulation does exist, it has likely been designed to be appropriately broad to address foreseeable new technology. By necessity, this level of generality requires that engineers need to be able to interpret it meaningfully for the relevant context of their own work or that its interpretation may be ambiguous in the context of novel and technological developments.

Effectively navigating these ethical gray areas requires that tech companies understand and address the sociotechnical context in which their products will be deployed, in particular, pre-empting various risks of harm that the technology might pose to affected stakeholders (customers, local communities, society at large, among others) [45, 46].

'Responsible Research and Innovation' as a concept, literature, and set of processes can be seen as a response to this practical reality. Methods such as anticipatory governance,[12] technology assessment,[13] upstream stakeholder engagement,[14] and value sensitive design,[15] have arisen with the goal of embedding stakeholder values into the design of technology accordingly.

(2) **Importance of shaping norms:** shaping predominant engineers' team culture norms is a key element in responding to *Tech's Practical Responsibility Requirement*.

We think that a key element in ensuring that engineers are adequately sensitive to the ethical and sociotechnical contexts of their deployed technology is changing the established informal rules and beliefs that govern the behavior of engineering teams. There are several reasons for this, which all concern the structural and organizational features of engineering teams, the deeply pervasive norms of engineering culture, and the agile nature of technology development.

[10] Garrett, Beard, and Fiesler 9.

[11] Moor [29].

[12] Guston [19].

[13] Schot and Rip [39].

[14] Wilsdon and Willis [55].

[15] Friedman [11] and Friedman and Hendry [12].

Firstly, large technology companies like Google are, from an organizational standpoint, best understood as networks of autonomous cross-functional teams, rather than as single, aggregate entities that act in accordance with unified sets of goals.[16] These technology companies typically have 'distributed' as opposed to 'centralized' organizational structures that are organized around key research, products, and services. This means that teams typically possess a great deal of autonomy, and they often define, frame, and develop technology solutions before elevating recommendations to executives or senior managers. This mode of operation applies even in moments where technology companies undertake a concerted push toward a general goal [50, 51].

Secondly, there are deeply ingrained norms that govern engineering culture itself. Values that relate to technical systems such as simplicity, efficiency, scalability, and elegance tend to be the focus, without explicit reference and acknowledgement of a wider set of ethical values that are also embedded in the work. Of course, there are good reasons for the cultural centrality of these values, including the ability to innovate rapidly and at scale. Engineers tend to be familiar with productive critique, optimization, and trade-offs for these types of values; however, the same know-how for negotiating these value tensions does not apply straightforwardly to social and ethical values. In practice, when individuals begin to discuss a wider set of ethically significant considerations, we find that teams are less familiar with ways to facilitate discussion about them in a productive, meaningful, and actionable way. This is also not to deny that individual engineers care a great deal about the social impact of their work; in fact, we experience that engineers often consider their work as part of a larger "life project of creating good in the world."[17] However, existing norms are often not conducive to enabling the required awareness, information-seeking, deliberation, and follow through on complex ethical issues.

Thirdly, in the early stages of technology development, there are often multiple paths that teams can pursue. Plans are actively in development, and concepts for technology research and products change rapidly, moving from a state of ambiguity into concrete formation over a period of weeks or months. In this state, the implicit beliefs, attitudes and social norms of teams exert preeminent influence over what and how technologies are built. Here social norms can determine how deliberative conversations proceed and engineers' perceptions of which activities are necessary to shape a development process.

For these reasons, the ethical gray areas underwrite the need for norm change in responding to tech's practical responsibility requirement. Since engineers inevitably develop technologies in domains where the application of existing policies and laws is ambiguous or not applicable, and because they possess a great deal of autonomy in designing technology, they must–*as teams* be able to rely on well-developed norms of recognizing ethical issues, identifying when more information is needed, being able to reason through the different moral considerations at stake in a situation, and then practically acting on these in a way that translates them into robust commitments.

(3) **Gap in current measures:** currently available hard and soft controls are necessary but insufficient to fully shape predominant team norms in the design and development stages.

Technology companies typically have a large arsenal of mechanisms at its disposal to support an ethical culture and drive responsible innovation within the unique operating context of the technology sector.

Traditionally the ethical culture of companies can be distinguished along two dimensions: formal and informal elements.[18] The first, so-called "hard controls", refer to the concrete and explicit plans, policies, and procedures within an organization [24]. Of these formal systems, many attempt to influence company culture through intervention at an individual level. Ethics training programs are a good example of this, including the code of conduct, whistle-blowing and speak-up training, privacy and data security training, diversity, equity, and inclusion training, as well as training for responsible corporate citizenship among others.[19] In addition to these measures, at Google there exist multiple review boards that operate at the project level, assessing research and products against Googles' AI principles, security, and privacy standards, for example.[20] The second element of ethical culture is informal. These so-called "soft controls" include the implicit, intangible elements, such as the values, expectations, beliefs, myths and assumptions that prevail in the organization that are not explicitly formalized through policies and processes. These informal elements also greatly matter in shaping ethical culture since the implicit norms and beliefs are key drivers of ethical conduct.

[16] See, for example, Birhane et al. [3], Andersén and Ljungkvist [2], and Patanakul et al. [31].

[17] Smith [42].

[18] See Kaptein [25].

[19] See Treviño and Weaver (2003).

[20] Google [17].

These ethics and compliance mechanisms, education and training programmes, and review boards are central to fostering a culture of responsible innovation within technology companies and our aim is not to criticize them or question their importance. However, we think that such mechanisms leave a gap with regards to the sustained promotion of a strong ethical culture in which technologies are consistently created with appropriate ethical foresight. While others have highlighted some[21] of the reasons for this apparent "principles to practices gap,"[22] in our experience, a main reason for this gap is that existing hard controls do not sufficiently influence the organization at a technology team level. Engineering team norms and culture can vary somewhat widely even within a company, and these have substantial and direct impact over ethically significant technological design decisions.

For example, while review boards, committees, and ethical commitments provide vital guardrails and checks on a product or research against critical ethical risks of harm, they mostly come into effect at later stages of development, and thus fail to encourage explicit reflection on the ethical costs and benefits of different design strategies at earlier stages of a product development, including product ideation. Often, these stages require foundational help in developing a mindset and vocabulary for ethical analysis that enables teams to become aware of the concrete moral implications of their work, how design choices relate to trade-offs between important values and can have ethical implications for key stakeholders, how the team can then deliberate through these in an ethically sound manner, and take concrete action for the further development of their product. To do so would require alignment with a collaborative and bottom-up approach where teams build crucial ethical capabilities that are tailored to their specific issues and requirements at a fine-grained level.

Ethics trainings that are offered at an individual level encounter a different set of limitations for influencing ethically relevant design decisions among technology teams. While individual education can influence individuals' beliefs, this doesn't guarantee that the individual can successfully convince others of similar beliefs or to influence technology direction by a team. These require team solutioning and commitment. Individuals whose moral intuition has directed them to broach the topic of the ethical dimensions of their work, are confronted with fears of analysis paralysis, lack of shared understanding of vocabulary and concepts, lack of confidence in moving through difficult conversations productively, and lack of understanding about how to integrate the moral dimensions into concrete technical design decisions. Indeed, often entrenched norms "keep opinions and behaviors in place even if individuals no longer privately support them, a phenomenon known as pluralistic ignorance."[23]

(4) **New Opportunity for Intervention:** therefore, there is an opportunity to devise new interventions that complement existing measures and can meet the specific requirements that distinctive engineering team culture poses.

What is consequently lacking is to influence which and how technologies are built, and to complement existing initiatives, are measures that directly address the culture and norms about how technology teams produce their work in the context of their work. Formats must be flexible and able to adapt to the nature of a team's work, the various stages of their projects, and the idiosyncrasies of particular team cultures given embedded personalities and existing power dynamics. Addressing team norms directly in discussion about their work yields the opportunity to weaken existing norms and replace them with a new social contract that explicitly incorporates follow through on team responsibilities in light of agreed-upon ethical commitments.

3 Moral Imagination

In this section we propose a 'Moral Imagination' methodology that aims to promote a role-obligation shift among engineers by influencing the norms of behavior, rules, best practices, and beliefs of engineering culture at a team level. The methodology builds upon the Moral Imagination literature in business ethics, alongside ideas from the philosophy of technology and the responsible innovation literature.[24] We first articulate what a Moral Imagination approach amounts to and its function in the context of technology companies (3.1). We then outline a framework that specifies three key ethical capabilities around which our approach is structured (3.2). Last, we propose a method for strengthening those capabilities based on our practitioner-experience of conducting more than 60 workshops with teams at Google (3.3).

[21] Schiff et al. [37].

[22] Mittelstadt [28].

[23] Prentice and Paluck [33].

[24] For some key frameworks in the responsible innovation literature, see Owen et al. [30], Stilgoe et al. [44], and Van Oudheusden [49]. See Werhane (1999; 2008 on moral imagination. Fisher, E. et al. (2006) "Midstream Modulation" approach is similar in spirit to our developed method here. See also Umbrello's [47] imaginative value sensitive design method.

3.1 Moral imagination for engineering teams

We define Moral Imagination as,

> **Moral Imagination:** the ability to *i)* register that one's perspective on a decision-making situation, including the available options and the normative factors relevant to adjudicating those options is limited; and to *ii)* creatively imagine alternative perspectives that reveal new approaches to that situation or new considerations that bear on the competing approaches.

Crucial to this is "becoming aware of one's context, understanding the conceptual scheme or "script" dominating that context, and envisioning possible moral conflicts or dilemmas that might arise in that context or as outcomes of the dominating scheme."[25] What developing Moral Imagination allows engineers to do is recognize the limitations of their pre-theoretic mental models about how their technology impacts the world, what the costs and benefits of that technology are, and what *their* role is in ensuring responsible technological development.

Thus a central aim of our approach is to facilitate a *role obligation shift* among engineers. It aims to shift teams' self-conception away from a mindset where ethical considerations are removed from perceived responsibilities – something that "falls outside of the job description" – toward a mindset where the consideration of the moral implications is an inherent part of the research and development process. It aims to prompt teams to realize what they do not yet understand about how their technologies impact users, and more broadly the sociotechnical dynamics and value tensions of their technologies as well as to empower teams to create a map for information gathering about the issues and topics the team did not consider before.[26]

In terms of concrete applications, we have worked with teams to make explicit and negotiate the actual values that guide their work, help them better deliberate through relevant trade-offs inherent in their design process, including responsibility objectives in Objectives and Key responsibilities (OKRs) as well as learn to use their ethical commitments and values to develop frameworks for principled ethical-decision making.[27]

3.2 Three key ethical capabilities for moral imagination

Our approach focuses on three ethical capabilities that we consider central to realizing Moral Imagination and to enhance these capabilities among teams to foster meaningful and productive norm change.

What undergirds our focus on these capabilities is a conception of teams as moral group agents who have the ability to reach informed moral judgements through awareness and reasoning, act with intent, and to be held accountable for their own actions.[28] These focal points also relate to our prior discussion insofar as the reality of autonomous technology teams operating in ethical gray areas requires an enablement approach that builds teams' ability to navigate complex ethical challenges and translate this into concrete actions and change along their product or research lifecycle.

> **Ethical awareness**: Ability to recognize normatively significant factors and implications (*e.g.,* moral values, ethical risks of harms, constraints and rights violations) in situations, decisions, and other relevant choice scenarios.

A precondition for robust ethical deliberation, decision-making, and commitment is to expand the team's perceptual paradigm beyond that of established engineering norms, while also sensitizing the team to moral discourse.[29] Developing an understanding of moral values, their normative force, action-guidingness, appropriate definitions of ethical terms for work-contexts, and how these relate to the technology and products that a team is developing are all crucial elements of this capability. In addition to shaping participants' understanding of moral values, ethical awareness also pertains to risks of harm to various stakeholder groups, especially in a sociotechnical and not just technical context.

> **Ethical deliberation and decision-making**: Ability to engage in reasoning and deliberation in relevant choice scenarios, including tensions between value and other moral commitments, conflicts, moral dilemmas, and trade-offs.

Once the team has a better understanding of the ethical dimensions of their work, alongside a grasp of key ethical vocabulary, teams can be introduced to conceptual tools which allow them to understand and negotiate situations in which the competing normative considerations come into conflict. This may encompass covering conceptual distinctions concerning 'pro-tanto' and 'all-things-considered', the gradeability of normative concepts and values, including different degrees to which conflicts can occur, the notion of

[25] Werhane [52], p. 3.

[26] For a detailed exploration and description of the various elements and content of the Moral Imagination workshop see [redacted for peer review].

[27] See [redacted for peer review] for a detailed case study.

[28] See Rest's [35] ethical-decision-making model.

[29] Clarkeburn [7].

weighing different moral factors that may relate to a choice situation for a team in a way that is ethically rigorous and robust, and the idea that which moral factors are apparent or significant may vary based on perspective. This point is relevant because the status quo of engineering culture is often primarily consequentialist and can accordingly be broadened by being introduced to different moral considerations besides outcome-oriented utility calculations.

> **Ethical commitment**: Ability to derive and set concrete plans to guide further product development and/or research.

Increased ethical awareness and decision-making capacities enables teams to navigate complex ethical challenges as part of their work. But building these capabilities will miss their mark if there is no commitment and accountability with respect to translating these insights into practical change. To that end, teams need to address what it means to act ethically and with integrity in their product or research context. This may mean deviating from widely accepted norms about the content, sequence, and pace of design, development, and release activities. What it means to operationalize ethical commitment varies depending on organizational structures. At Google, we co-develop with teams a set of actionable responsibility objectives that can inform Product Requirement Documents (PRD) and individual or team Objectives and Key Results (OKRs).

3.3 Methodology

In the previous subsection we discussed the key ethical capabilities that the Moral Imagination approach intends to influence to facilitate an ethical role obligation shift among engineering teams. In this subsection, we detail a practical four-step workshop method to strengthen these capabilities based on our experience of conducting Moral Imagination workshops with teams at Google.

Our method expands upon existing responsible innovation frameworks such as those developed by Owen et al. [30] and Stilgoe et al. (2012), as well as Fisher et al. [10], by providing a tailored methodology for facilitating responsible innovation for product and research teams that engage in software development for at-scale technologies including artificial intelligence.[30] For instance, while these approaches provide frameworks for facilitating responsible innovation in broad strokes, our approach is specifically tailored to the day-to-day realities of engineering teams in technology companies on the ground. Similarly, we adapt [52, 53] notion of moral imagination within this overall approach in a way that is tailored to the specific needs of engineering teams.

Our approach is operationalized through a series of workshops that are facilitated by a multidisciplinary team with academic backgrounds in ethics and/or practical experience with ethics in the technology industry. The workshop provides a structured engagement forum to assist teams typically at early stages of their work, for example, during the ideation, experimentation, prototyping, piloting, or re-imagining phases.[31] Workshops are designed to draw attention toward the salient dimensions of engineers' work, and model and support how they can work through them together while building a shared capacity for ethical awareness, productive debate, solution finding, and planning. The workshops are specifically adapted and tailored to a team's progress and work: they are modular and involve content that is customized for relevance to the dilemmas teams face—though it is always centrally focused on the key ethical capabilities of Awareness, Deliberation and Decision-Making, and Commitment through Moral Imagination.

The workshops employ a non-didactic approach to ethics, in the sense that the aim is not to lecture participants about key moral principles and considerations. Nor is it to impose a particular ethical framework. Rather, our approach is to construct exercises that enable engineers to reframe their work through an ethical lens, and then re-envision their work and its corresponding responsible development process. In doing so, our goal is to align with the technology industry's culture of autonomy and entrepreneurship, while building momentum from many engineers' expressed desire to drive their innovations toward socially beneficial ends.

At a high level, the Moral Imagination workshops involve the following four-step structure.

1. **Reflection** Externalization of a team's current moral intuitions, beliefs, and convictions about their work.

   Norms and beliefs about a team's work have to be made explicit to be challenged and altered. This first step therefore aims to surface and understand the particular ethical paradigm with which a team is operating in their day-to-day work by enabling teams to reflect on, articulate and clarify the values they feel are currently motivating or inherent in their work. Semi-structured discussions are used to surface the values that the team brings to bear in their work, including personal motivations, beliefs about the technological benefits, and envisioned characteristics of a world where the technology has been successfully deployed and is ubiquitous, aiming to formulate a positive vision for their technology. Building upon these discussions, facilitators introduce the concept

[30] Here our aim is to sketch conceptually how the workshop format fosters the relevant capabilities. We elaborate on the practical details of the workshop in a separate practice-based piece of work.

[31] The benefits of engagement with innovation teams at an early to mid stage of their work has been outlined in Fisher E et al. (2006).

of values in ethics, groups negotiate the most important values, and work to clarify and interpret them in the particular context of their technology. Teams also reflect on whether and in what respects current plans instantiate or fail to instantiate the stated values, and surface tensions between values that require tradeoffs.

2 **Expansion** Challenging a team's perspective for the purpose of reflecting on their moral intuitions. Envisioning possibilities for the acquisition of ethically relevant information to inform a team's approach, and for the work itself.

Once teams' moral intuitions and beliefs have been made more explicit among the group, the next step is to *challenge* those intuitions and facilitate the internalization of ethical considerations beyond those that were initially surfaced. As part of building ethical awareness, the focus at this stage is to challenge the teams' paradigm from an ethical point of view to help the group consider the key ethical implications that their work contains and also surface relevant knowledge gaps.

The centerpiece of this section is a bespoke technomoral scenario that extends the underlying logic of each team's technology 5 or 10 years into the future. The scenario complicates the interplay of technology and society, ends on a cliffhanger, and emphasizes the importance of gaining different points of view as a means to anticipate ethical considerations. Participants role-play in small groups and argue the case against each other, putting to practice their ability to interpret values, argue for or against them in technological design, and build comfort with critical evaluation of their work. Throughout this section, further value tensions are solicited, documented, and described. An inclusion-focused exercise then aids participants in understanding the needs and interests of multiple stakeholder groups and how to include their voices to improve decision making. Participants are invited to a 'veil of ignorance' scenario where they are encouraged to envision, articulate and elaborate on the issues that might arise for the stakeholder ecosystem.[32] This exercise alerts participants to the possibility their team's perspective is limited, enumerates an initial set of perspectives from which the work would benefit, and emphasizes diverse perspectives collected equitably must be a high priority. Other exercises include anticipation of sociotechnical harm, in which teams are exposed to a taxonomy of harm and brainstorm a number of concrete adverse impacts their work could potentially have, alongside alternative paths for the work in light of those possible impacts.

3. **Evaluation** Reasoning through a number of ethical perspectives about the team's work.

Reflection and *Expansion* aim to build the ethical awareness of teams, specifically with an eye towards enabling a better grasp of ethically relevant factors including risks of harm. Once these have been surfaced and internalized in the context of the teams' own technologies, the next step focuses on helping the team learn to deliberate and reason through concrete ethical choice scenarios that are relevant to them. So, after a team's ethical paradigm has been made explicit and challenged by the team itself, ethical reasoning tools are successively introduced to enable the team to learn to reason through trade-offs and choice scenarios that they have identified as arising in the context of their work.

During Moral Imagination workshops at Google, moral theories are introduced schematically, and presented as a set of reasoning tools that enable participants to approach a problem from different angles. Key notions such as the "weighing" of competing moral values, trade-offs, and gradability of moral commitments are introduced to teams. Exercises then involve participants responding to arguments, formulating responses from multiple perspectives, and discussing among each other to reach a consensus on how best to resolve a particular value tension. The elements described aim to provide teams with a shared foundation to have ethical conversations in a pluralistic manner that goes beyond entirely deontological or consequentialist paradigms.

4. **Action** Translation of insights and learnings into concrete team practices.

This last step focuses on supporting teams in taking actions based on their learnings. Participants reflect on prior discussions and articulate ethical focus areas that can inform the technology concept and design in future work. Moderators work with participants during and after the workshop to shape these focus areas into responsibility objectives, which serve as actionable statements that can shape OKRs or be included in PRDs. The workshop's focus on discussion, clarification, and negotiation ensures that the responsibility objectives are broadly supported and considered as legitimate North Stars for the team.

Moral Imagination workshops enable participants to challenge beliefs and begin to reshape the norms that guide decision-making and planning in their team. Importantly, the norm change at issue here is shared and co-constructed. Furthermore, in our experience, workshops render participants more aware of the value of seeking accurate information about how their work will function as a sociotechnical artifact, and also empower participants to interpret this information in the context of research and development processes. To that end,

[32] See Weidinger, L., McKee, K. R., Everett R., et al. (2023) for a recent discussion of using Rawls' veil of ignorance to align AI systems with principles of justice.

participants are able to proactively identify and mitigate ethical risks, which supports them in making better use of other ethics controls such as review boards, and in particular empowers teams with a degree of moral autonomy when engaging with these other ethical controls. This holistic approach, on which the Moral Imagination methodology complements more traditional hard controls, ultimately enables engineering teams to develop technologies in a way that is morally informed and which better meets the challenges for technology companies that we articulated in Sect. 2.

4 Conclusion

In this paper, we introduced the Moral Imagination approach as a method for driving ethical culture change within technology companies. The approach is a "soft control" method that emphasizes externalization and multiperspectival evaluation of the norms and values that precipitate innovation within teams through semi-structured deliberation and negotiation, alongside co-development of action-oriented ethical commitments (for example, through OKRs and PRDs). The Moral Imagination approach has been executed over 60 times at Google, and is positioned alongside "hard controls" such as ethics and privacy reviews that together make up Google's portfolio approach to fostering a culture of responsible innovation in line with Google's AI Principles [32]. We have argued that Moral Imagination is uniquely well-positioned to complement and address the limitations of more traditional "hard controls" in the context of technology companies, where team norms and values exhibit substantial influence over research and product decisions given the bottom-up and highly autonomous engineering culture that drives innovation within these companies. Our hope is that the Moral Imagination approach can serve as a template to foster a culture of responsible innovation across the industry.

While we are encouraged by the early results of the Moral Imagination approach, we continue to refine the approach and develop new tools and resources to scale the program within Google. This includes a dedicated empirical research track focused on measuring the efficacy of the approach as a method for ethical culture change, alongside the development of new workshop modules that aim to further upskill teams on topics such as ethical reasoning, critical reflection on metrics, and many other topics. Furthermore, we aim to contribute to and enrich the social conversation around ethical culture change within technology companies by publishing case studies alongside the findings of our empirical research, and also by externalizing the methodology to solicit participation and critical input from a broad range of stakeholders.

Author contributions BL and GK are joint first authors. AM and BZ are creators of the program and joint second authors. SB and KP are contributors to the program and contributing authors. AL and BAA are executive supporters and contributing authors.

Declarations

Competing interests The author(s) are current or former employees of Google LLC and own stock as part of the standard compensation package.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Amodei, D., Chris O., Jacob S., Paul C., John S., Dan M.: "Concrete problems in AI safety." (2016). arXiv preprint arXiv:1606.06565.
2. Andersén, J., Ljungkvist, T.: Resource orchestration for team-based innovation: a case study of the interplay between teams, customers, and top management. R&D Manag. **51**(1), 147–160 (2021)
3. Birhane, Abeba, et al.: "The values encoded in machine learning research." 2022 ACM Conference on Fairness, Accountability, and Transparency (2022)
4. Blackman, R.: Ethical Machines. Harvard Business Review Press, Boston (2022)
5. Brown, S., Davidovic, J., Hasan, A.: The algorithm audit: scoring the algorithms that score us. Big Data Soc. **8**(1), 2053951720983865 (2021)
6. Caldwell, D. F., O'Reilly III, C. A.: "The determinants of team-based innovation in organizations: The role of social influence." Small Group Res **34.4**, 497–517 (2003)
7. Clarkeburn, H.: A test for ethical sensitivity in science. J Moral Educat **31**(4), 439–453 (2002)
8. Feenberg, A.: The critical theory of technology. Capital. Nat. Social. **1**(5), 17–45 (1990)
9. Fiesler, C., Garrett, N., & Beard, N.: What do we teach when we teach tech ethics? A syllabi analysis. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education, pp. 289–295 (2020)
10. Fisher, E., Mahajan, R.L., Mitcham, C.: Midstream modulation of technology: governance from within. Bull. Sci. Technol. Soc. **26**(6), 485–496 (2006)
11. Friedman, B.: Value-sensitive design. Interactions **3**(6), 16–23 (1996)
12. Friedman, B., Hendry, D. G.: Value sensitive design: Shaping technology with moral imagination. Mit Press (2019)
13. Gabriel, I.: Artificial intelligence, values, and alignment. Minds Mach. **30**(3), 411–437 (2020)

14. Garrett, N., Beard, N., Fiesler, C.: More Than" If Time Allows" The Role of Ethics in AI Education. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp. 272–278 (202)
15. Gerrard, B., Lockett, A.: Team-specific human capital and performance. Br. J. Manag. **29**(1), 10–25 (2018)
16. Glynn, M.A., Kazanjian, R., Drazin, R.: Fostering innovation in complex product development settings: the role of team member identity and interteam interdependence. J Product Innovat. Manag. **27**(7), 1082–1095 (2010)
17. Google (2022) 2022 AI Principles Progress Update. https://ai.google/static/documents/ai-principles-2022-progress-update.pdf
18. Grosz, B.J., Grant, D.G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A., Waldo, J.: Embedded ethics: integrating ethics across CS education. Commun. ACM. ACM **62**(8), 54–61 (2019)
19. Guston, D.H.: Understanding 'anticipatory governance.' Soc. Stud. Sci. **44**(2), 218–242 (2014)
20. Hasan, A., Brown, S., Davidovic, J., Lange, B., Regan, M.: Algorithmic bias and risk assessments: lessons from practice. DISO **1**, 14 (2022)
21. Hoffmann, M., et al.: The human side of Software Engineering Teams: an investigation of contemporary challenges. IEEE Transact. Software Eng. **49.1**, 211–225 (2022)
22. Ihde, D.: Technology and the lifeworld: from garden to Earth (Indiana Series In The Philosophy Of Technology). Indiana University Press, Bloomington (1990)
23. Jackman, M., Kanerva, L.: Evolving the IRB: building robust review for industry research. Wash. Lee L. Rev. Online **72**, 442 (2015)
24. Kaptein, M.: Developing and testing a measure for the ethical culture of organisations: the corporate ethical virtues model. J. Organisat. Behav. **29**, 923–947 (2008)
25. Kaptein, M.: Understanding unethical behaviour by unraveling ethical culture. Hum. Relat.Relat. **64**, 843–869 (2011)
26. Karn, J.: "An ethnographic study of conflict in software engineering teams." J. Inform. Inform. Technol. Organiz. (2008). https://doi.org/10.28945/133
27. Kenton, Z., Everitt, T., Weidinger, L., Gabriel, I., Mikulik,V., Irving, G.: "Alignment of language agents." arXiv preprint arXiv:2103.14659 (2021)
28. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intellig. **1**(11), 501–507 (2019)
29. Moor, J.H.: What is computer ethics? Metaphilosophy **16**(4), 266–272 (1985)
30. Owen, R., Macnaghten, P., Stilgoe, J.: Responsible research and innovation: from science in society to science for society, with society. Sci. Public Policy **39**, 751–760 (2012)
31. Patanakul, P., Chen, J., Lynn, G.S.: Autonomous teams and new product development. J. Prod. Innov. Manag.Innov. Manag. **29**(5), 734–750 (2012)
32. Pichai, S.: AI at Google: our principles. The Keyword **7**(2018), 1–3 (2018)
33. Prentice, D., Paluck, E.L.: Engineering social change using social norms: lessons from the study of collective action. Curr. Opin. Psychol. Opin Psychol. **35**, 138–142 (2020)
34. Prunkl, C.E.A., Ashurst, C., Anderljung, M., Webb, H., Leike, J., Dafoe, A.: Institutionalizing ethics in AI through broader impact requirements. Nat. Mach. Intellig. **3**(2), 104–110 (2021)
35. Rest, J. R.: "moral development: advances in research and theory." (1986)
36. Robbins, P., O'Gorman, C.: Innovating the innovation process: an organisational experiment in global pharma pursuing radical innovation. R&D Manag. **45**(1), 76–93 (2015)
37. Schiff, D., Rakova, B., Ayesh, A., Fanti, A., Lennon, M.: Principles to practices for responsible AI: Closing the gap. (2020). arXiv preprint arXiv:2006.04707
38. Schiff, D., Rakova, B., Ayesh, A., Fanti, A., Lennon, M.: Explaining the principles to practices gap in AI. IEEE Technol. Soc. Mag. **40**(2), 81–94 (2021)
39. Schot, J., Rip, A.: The past and future of constructive technology assessment. Technol. Forecast. Soc. Chang. **54**(2–3), 251–268 (1997)
40. Scott, S.G., Einstein, W.O.: Strategic performance appraisal in team-based organizations: one size does not fit all. Acad. Manag. Perspect.Perspect. **15**(2), 107–116 (2001)
41. Shneiderman, B.: Responsible AI: bridging from ethics to practice. Commun. ACM. ACM **64**(8), 32–35 (2021)
42. Smith, J. M.: Extracting accountability: Engineers and corporate social responsibility. Mit Press **9** (2021)
43. Somech, A., Drach-Zahavy, A.: Translating team creativity to innovation implementation: the role of team composition and climate for innovation. J. Manag. **39**(3), 684–708 (2013)
44. Stilgoe, J., Owen, R., Macnaghten, P.: Developing a framework for responsible innovation. Res. Policy **42**(9), 1568–1580 (2013). https://doi.org/10.1016/j.respol.2013.05.008
45. Treviño, L.K., Butterfield, K.D., McCabe, D.L.: The ethical context in organizations: Influences on employee attitudes and behaviors. Bus. Ethics Q. **8**(3), 447–476 (1998)
46. Trevino, L.K., Weaver, G.R.: Managing Ethics in Business Organizations: Social Scientific Perspectives. Stanford University Press, Stanford (2003)
47. Umbrello, S.: Imaginative value sensitive design: using moral imagination theory to inform responsible technology design. Sci. Eng. Ethics **26**(2), 575–595 (2020)
48. Umbrello, S., Bernstein, M.J., Vermaas, P.E., Resseguier, A., Gonzalez, G., Porcari, A., Adomaitis, L.: From speculation to reality: enhancing anticipatory ethics for emerging technologies (ATE) in practice. Technol. Soc. **74**, 102325 (2023)
49. Van Oudheusden, M.: Where are the politics in responsible innovation? European governance, technology assessments, and beyond. J. Respons. Innov. **1**(1), 67–86 (2014)
50. Weidinger, L., McKee, K.R., Everett, R., Gabriel, I.: Using the Veil of Ignorance to align AI systems with principles of justice. The Proceedings of the National Academy of Sciences (PNAS) **120**(18), e2213709120 (2023)
51. Weinstein, J., Reich, R., Sahami M.: System error: Where big tech went wrong and how we can reboot. Hachette UK (2021)
52. Werhane, P.: A note on moral imagination. SSRN Electron. J. Electron. J. (2008). https://doi.org/10.2139/ssrn.908386
53. Werhane, P.H.: Moral imagination and management decision-making. Oxford University Press, New York (1999)
54. Whittlestone, J., Nyrup, R., Alexandrova, A., Cave, S.: The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. 195–200. (2019) https://doi.org/10.1145/3306618.3314289
55. Wilsdon, J., Willis, R.:See-through science: Why public engagement needs to move upstream. Demos (2004)
56. Langdon. W.: "Do artifacts have politics?." Daedalus 121–136 (1980)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.