BRIGHT AND DARK IMAGINING: HOW CREATORS NAVIGATE MORAL CONSEQUENCES OF DEVELOPING IDEAS FOR ARTIFICIAL INTELLIGENCE

LYDIA PAINE HAGTVEDT
Akamai Technologies
145 Broadway
Cambridge, MA, 02142
Tel: (857) 227-4027
Email: lydiahagtvedt@gmail.com

SARAH HARVEY
UCL School of Management
University College London
One Canada Square, London E14 5AA, United Kingdom
Tel: +44 (0)20 3108 6000
Email: sarah.r.harvey@ucl.ac.uk

OZUMCAN DEMIR-CALISKAN
Imperial College Business School
Imperial College London
Exhibition Road, London SW7 2AZ, United Kingdom
Tel: +44 (0)20 7589 5111
Email: odemirca@ic.ac.uk

HENRIK HAGTVEDT
Carroll School of Management
Boston College
Fulton Hall, 140 Commonwealth Avenue
Chestnut Hill, MA, 02467
Tel: (617) 552-4034
Email: hagtvedt@bc.edu

ACKNOWLEDGEMENTS

We would like to thank our associate editor, Elizabeth (Bess) D. Rouse, as well as the three anonymous reviewers, who provided thorough, insightful, and constructive feedback. We also thank the first author's dissertation committee, Spencer Harrison, Mike Pratt, and Teresa Amabile, for their wisdom, encouragement, and feedback on earlier versions of this paper. In addition, we thank many current and former members of the Management and Organization Department at Boston College, who also encouraged the first author with their thoughtful questions and support from the beginning of this project. Finally, we would like to thank the AI creators who shared their diverse experiences and perspectives with us.

Abstract

Despite an emerging stream of work on negative behaviors associated with engaging in creativity, research on the consequences of creativity has largely focused on unleashing the proximal success of new ideas. Both approaches overlook the downstream potential for creative ideas to directly cause harm. Through an inductive, qualitative study of individuals creating artificial intelligence (AI) technologies, this study shifts the conversation to how workers navigate potential distal moral consequences of ideas while engaging in creative work. Our study unveils that surprises during creative work catalyze a process of imagining future consequences of ideas, which shapes the way creators engage with moral issues and approach idea development. A key insight of our study is that imagining unfolds in two ways: *Bright imagining* is associated with disconnecting moral issues from idea development, so that creators develop ideas in the relative absence of constraints and moral issues are addressed through systematized safeguards. *Dark imagining* is associated with integrating moral issues into idea development, transforming morally-motivated constraints into creative forces with potential to shape the nature of ideas themselves. Our study recasts interacting with moral consequences intertwined with creative ideas as itself a creative, constructive process.

> The past 15 years (of AI development) have been remarkable… You put that alongside big developments in areas such as quantum computing, and the possibilities start to get intriguing and frightening at the same time. …The breadth and variety of the ways that AI is impacting our lives, I think is really fascinating. Everybody has so much exposure to it now every day and we don't even think about it. So that's really cool… [But] like anything and everything, there's always going to be bad actors who can get hold of things and use them for purposes other than their original intention. (Ellen, artificial intelligence [AI] creator)

Creative ideas—ideas that are novel and useful (Amabile, 1983; Woodman, Sawyer, & Griffin, 1993)—can change the world, for better or for worse. As the quotation above from AI creator Ellen illustrates, creative ideas are often anticipated to have "intriguing" and "fascinating" effects with the potential to generate long-term value far into the future, providing a starting point for scientific and social progress (Amabile, 1988; Cronin & Loewenstein, 2018; Simonton, 1984). Yet, Ellen's comment reveals that creative ideas also have the potential for negative—even "frightening"—downstream consequences for users and broader society. The development of complex technologies that will be integrated into many people's lives, such as nuclear technology, gene editing, nanotechnology (Hecht, 2010; Jasper, 2010), and more recently artificial intelligence (Fleming, 2019), illustrate the great potential for new ideas to cause harm.

To date, organizational creativity research has emphasized the good that may come from new ideas, assuming that the consequences of creativity are inherently positive (Amabile & Pratt, 2016; Anderson, Potočnik, & Zhou, 2014; Runco, 2010; Yong, Mannucci, & Lander, 2020). When creative ideas are successfully implemented, they are expected to produce economic or social value (Hua, Harvey, & Rietzschel, 2022). In the short term, that value is captured by the usefulness criterion for assessing new ideas (Amabile, 1983; George, 2007). Usefulness is the extent to which an idea solves a problem in a feasible or practical way, typically assessed once an idea or product has been generated (Harvey & Berry, 2023). Positive consequences are assumed to unfold over time for more distal stakeholders as new ideas generate value, shift

paradigms, or spur innovation (Amabile, 1988; Cronin & Loewenstein, 2018). Beyond the direct value of ideas themselves, creative behavior can also lead to indirect positive outcomes in the short or longer term, such as reducing a psychological burden (Goncalo, Vincent, & Krause, 2015) or enhancing well-being (Acar, Tadik, Myers, van der Sman, & Uysal, 2021). From the positive perspective on creativity, a core challenge for creators and decision-makers involved in developing ideas is to accurately forecast the value of ideas to capture their benefits (Berg, 2016; Berg, 2019; Harvey & Mueller, 2021; Mueller, Melwani, Loewenstein, & Deal, 2018).

More recently, research has begun to explore the negative consequences of creativity. Direct and proximal negative consequences arise from poor implementation of ideas, such as monetary or reputational costs (Mueller et al., 2018; Venkatamarani, Derfler-Rozin, Liu, & Mao, 2023; Harrison & Nurohamed, 2023) and the opportunity cost of overlooking a useful idea due to its novelty (Mueller, Melwani, & Goncalo, 2012; Mueller et al., 2018; Berg, 2016). Research on the "dark side" of creativity has further revealed indirect negative outcomes that arise from creators' subsequent behaviors, rather than the direct effects of ideas. That work shows that the flexibility and disinhibition associated with a creative disposition or identity can spill over into undesirable and potentially unethical behavior on unrelated tasks or in relationships outside the creative process (Khessina, Goncalo, & Krause, 2018; Kim, Vincent, & Goncalo, 2013; Harrison & Wagner, 2016; Mainemelis & Sakellariou, 2022). Research on the negative consequences of creativity presents creators and decision-makers with the challenge of making relatively low-risk choices about which ideas and actions to pursue.

What has not yet been explored in either stream of work is the potential for ideas *themselves* to produce negative distal consequences, particularly when those consequences involve others and so raise the question of morality (Jones, 1991). Because such consequences

unfold over time, only after the idea's successful implementation, they may impact stakeholders far beyond the creative process. The question of how creators navigate possible negative downstream consequences of ideas inverts the approach taken in prior research, in that it involves assessing potential distal harms rather than positive effects, and those harms emerge directly from ideas rather than indirectly from engaging in creative behavior. Our focus on ideas' potential for harmful distal consequences is distinct from prior research in two theoretically important ways. First, because many of those who bear an idea's consequences are not involved in its development and so cannot protect themselves as the idea develops, pursuing that idea raises a moral issue. Moral issues arise when one's freely performed actions or decisions (e.g., those of a creator developing an idea) can feasibly lead to consequences for others (Jones, 1991), which we refer to as moral consequences. Second, negative distal consequences of creative ideas are uncertain (George, 2007; Huang & Pearce, 2015); they may occur because users and other stakeholders apply ideas in ways that are unexpected, and they may happen to some, but not others. In other words, creators face a high degree of uncertainty about the moral consequences of their ideas, making it more difficult to judge what is morally right (Sonenshein, 2007). Judging moral issues requires imagining ideas in a more distant future and envisioning how people may interact with them in more complex ways—akin to a generative process (Keem, Shalley, Kim, & Jeong, 2018; Whyte, Comi, & Mosca, 2022), rather than a more straightforward assessment of usefulness, risk, or negative indirect effects. Engaging with negative distal consequences may therefore be deeply entwined with the process of developing ideas.

The present research explores whether, when, and how creators navigate moral issues arising from their ideas during the course of creative work in a context where creative advances are happening rapidly and the potential for widespread harm is heightened: artificial intelligence

(AI). Understanding AI development has been identified as a frontier of creativity research (Amabile, 2020; Berg, Duguid, Goncalo, Harrison & Miron-Spektor, 2023). AI development requires substantial creativity in the form of new ideas and technologies; AI creators leverage machine learning to create algorithms that learn from data to solve challenging problems, from detecting hate speech to classifying facial expressions to predicting natural disasters. New AI tools have the potential to help solve society's most difficult problems, including disease (Hsieh et al., 2019), poverty (Bennington-Castro, 2017), and global warming (Cho, 2018). However, the same AI tools pose major threats to the public, such as enabling cyberattacks (Maxwell, 2020), fueling biased decision-making (Obermeyer, Powers, Vogeli, & Mullainathan, 2019), and displacing human workers (Huang & Rust, 2018; Jia, Luo, Fang, & Liao, 2024; Semuels, 2020). Tellingly, "Godfather of AI" Geoffrey Hinton, who won the 2018 Turing Award for work spurring the current AI boom, recently left his job at Google to speak out about the risks of AI (Ricker, 2023).

Through an inductive study of 69 professionals engaged in creative AI work, we find that creators navigate potential distal moral consequences of their ideas through *bright* and *dark imagining.* We conceptualize imagining as a sensemaking process in which creators project the future of their ideas and their distal consequences. Bright and dark imagining sustain and support different ways of *engaging with moral issues in idea development*, shaping both how creators think about moral issues and how they engage in the process of developing ideas. Bright and dark imagining thus go right to the heart of *how* creators create and so have the potential to impact creators' emerging ideas. A key insight from our data is that these processes are catalyzed by surprises that arise during the creative process itself, so that how people engage in creative work influences the way that they address moral issues during idea development. Our study thus

reveals that creators often engage deeply with potential distal outcomes of their ideas, including negative consequences, and shows how that unfolds as a constructive, creative process (Sonenshein, 2007) rather than an assessment or forecast. Our study further lays a foundation for understanding how and when creators engage with moral issues that arise from creative work, moving beyond explanations for moral behavior that rely on the personal characteristics of creators (Kapoor & Kaufman, 2022). Finally, our study unpacks processes through which an understanding of the consequences of ideas can be captured in constraints that spur the creative process (Acar, Tarakci, & van Knippenberg, 2019). In addition to those contributions, our study informs practice about the development of novel technologies like AI and sheds light on the processes that may give rise to product constraints and regulation.

NAVIGATING MORAL CONSEQUENCES OF IDEAS DURING CREATIVE WORK

Research on the receiving side of creativity has typically framed the consequences of creative ideas as being captured by assessments of an idea's novelty or usefulness. Creators and decision-makers involved in developing ideas assess how other decision-makers, gatekeepers, or direct consumers of their creative outputs will respond to ideas (Zhou, Wang, Song, & Wu, 2017) in light of performance goals and specific task criteria, and adjust ideas to achieve better proximal outcomes (Amabile & Pratt, 2016; Lonergan, Scott, & Mumford, 2004; Perry-Smith & Mannucci, 2017). Proximal outcomes have also been conceptualized as being contained in product or output constraints that "define the end result of the creative processes" (Acar et al., 2019: 99; Rosso, 2014), such as product or design specifications. Research suggests that novelty and usefulness assessments and constraints may also shape how ideas or products develop through a reciprocal process of adjusting ideas in response to constraints (Moreau & Dahl, 2005; Harrison & Rouse, 2014; Goldenberg, Mazursky, & Solomon, 1999; Amabile & Pratt, 2016).

In assessing and deciding whether to pursue ideas based on possible consequences, creators and especially decision-makers emphasize relatively short-term, proximal outcomes rather than more distal ones (Berg, 2019; Mueller et al., 2018). Creators and decision-makers often under-value novelty (Mueller et al., 2012), which is inherently uncertain and so entails a longer-term, more distant focus (Mueller, Wakslak, & Krishnan, 2014), while preferring ideas with short-term, tangible benefits for the greatest number of people (Blair & Mumford, 2007). The risks of ideas have similarly been construed primarily as short term and associated with implementation uncertainties, including financial risks (Haselhuhn, Wong, & Ormiston, 2022) and the reputational risk of an individual whose idea may fail (Lu, Bartol, Venkataramani, Zheng, & Liu, 2019; Venkatamarani et al., 2023).

Research that views consequences of ideas in terms of assessment and choice leaves two gaps in understanding how creators engage with the potential consequences of their ideas. First, whereas research has shown that creators usually consider the relatively proximal outcomes of their ideas while ignoring the distant future, it has said relatively little about those situations in which creators *do* consider distal consequences of their ideas during creative work. Second, although the literature has developed a deep consensus to assess potential outcomes of creativity in terms of novelty and usefulness (Harvey & Berry, 2023), it has not explored when or how other dimensions for assessing novel ideas are considered during that process, including other dimensions on which creative ideas and other outputs may be constrained. It therefore says little about how creators may engage with the potential moral consequences of their ideas, which are likely to unfold over the longer term, as an idea is implemented, used, and adopted by an increasingly wide set of stakeholders. In other words, it cannot explain whether or when creators will choose to constrain ideas based on possible moral consequences.

Projecting Distal Consequences of Ideas as a Constructive, Creative Process

Some research shows that creators may engage in relatively longer-term projections of the consequences of their ideas. For example, creators engage in creative forecasting to predict the consequences that will follow new ideas within particular settings (Berg, 2016; Byrne, Shipman, & Mumford, 2010; Kettner, Guilford, & Christensen, 1959; Wilson, Guilford, Christensen, & Lewis, 1954), such as the moderately distal outcomes of audience responses (Berg, 2016) or economic value (Fuchs, Sting, Schlikel, & Alexy, 2019). When considering distal consequences, people tend to move away from a cycle of assessing ideas and towards projecting them into the future. Evaluative criteria and how creators may constrain ideas with respect to them are not fixed; creators can shift their assessments of outcomes to a more distant future (Harvey & Berry, 2023). For instance, Kaplan and Orliskowski (2013) found that breakdowns in collective understanding of the value of an organizational strategy could prompt a search for new understandings of future possibilities. Similarly, Harvey and Mueller (2021) showed that healthcare policy groups who disrupted their assumptions about criteria for evaluating novel ideas tended to establish a new understanding of the long-term value of those ideas. From this perspective, considering distal consequences of ideas is itself a creative process that entails shifting perspectives about how to assess future potentials. This notion is supported by the finding that creators, who are more likely than managers or decision-makers to engage in divergent, generative processes, also tend to be more accurate at predicting audience responses to new ideas (despite also being subject to errors in forecasting; Berg, 2016) and that more extensive forecasting during idea evaluation and implementation planning boosts creativity (Byrne et al., 2010). Similarly, engaging with the potential outcomes of ideas implied by product constraints can stimulate creativity (Acar et al., 2019; Rosso, 2014).

These insights suggest reorienting our understanding of how creators assess the potential consequences of ideas, recognizing that it is itself a creative, constructive process. Ideas have a virtually infinite number of possible futures (Thompson, 2018; Thompson & Byrne, 2022), and gathering information about novel ideas in the absence of implementing them cannot reveal which one will succeed (Harvey & Mueller, 2021). Therefore, the future of ideas cannot simply be planned (Weick, 1995; Wenzel, Krämer, Koch, & Reckwitz, 2020). Instead, creators have to project ideas into the future to imagine potential consequences (Whyte et al., 2022; Thompson & Byrne, 2022). Moreover, the way that people project the future shapes their current understanding and action (Kaplan & Orlikowski, 2013; Vaara, Sonenshein, & Boje, 2016). By extension, creators may generate alternative ways of assessing the consequences of ideas, which may shape their process for developing and constraining those ideas.

In sum, whereas research on proximal outcomes finds that people often do not choose novel ideas, it hints that when creators do consider more distal consequences, it may trigger a creative process of elaborating and expanding creators' understanding of the possible consequences of ideas. Yet, it is not known when and why creators recognize the distal downstream consequences of ideas or how that understanding affects their creative process.

Engaging with Moral Consequences of Ideas as a Constructive, Creative Process

Emerging research suggests that creators focus on novelty and usefulness to the exclusion of moral issues, revealing a "dark side" in which creators may ignore, dismiss, or disengage from moral issues (Bandura, 2002; Cropley, 2010; Lubart, 2010; Richards, 2010). Noting that creativity and unethical behavior share the common roots of thinking flexibly and challenging social norms (Kim et al., 2013; Keem et al., 2018; Kapoor & Kaufmann, 2022), and that creative individuals may feel that they are entitled to more than they deserve (Vincent & Kouchaki,

2016), that research suggests that creators may push the boundaries of expected conduct both in terms of their ideas and their ethical decisions (Khessina et al., 2018; Vincent & Goncalo, 2014). Retrospective accounts of the psychological dynamics of engaging with potential harm in new domains support this possibility (e.g., Bethe, 1968). For example, post-hoc reports of the scientists who developed nuclear technology indicate that they may have proceeded despite major risks due to an intentional blindness to negative consequences (Hecht, 2010), a "fascination or unquestioning belief in what they [were] doing" (Cropley, 2010: 360), or simply "complacency or hubris" (Jasper, 2010: 111). Similarly, Thomas Edison actively pursued electric lighting despite serious concerns about its feasibility and a host of safety issues (Hargadon & Douglas, 2001). In parallel, research in this vein suggests that considering moral issues may distract creators from their core work, reducing creativity (Kundro, 2023).

Research on moral and ethical issues in creativity has prioritized characteristics of individual creators, including having a creative disposition (Bierly, Kolodinsky, & Charette, 2009; Keem et al., 2018) or a creative role identity (Vincent and Kouchaki, 2016; Kim et al., 2013). That work has examined choices that are unrelated to creative ideas and that have clear moral consequences. For instance, Vincent and Kouchaki (2016) examined whether people with a creative identity would act in dishonest ways, such as lying and stealing, and Keem and colleagues (2018) asked respondents whether they engaged in behaviors such as taking advantage of company resources. Their findings echo the accounts of scientists described above, implying that creators act in unethical ways despite obvious moral consequences.

Focusing on the effects of creators' characteristics on unrelated ethical decisions overlooks the possibility that perceiving, understanding, and making judgments about the moral consequences of developing ideas is itself a creative, constructed process. Moral issues may be

"problems in need of creative solutions" (Runco, 1993), and considering moral issues during creative work may promote creativity (Kundro, 2023). Moral issues are ambiguous: Effects on different stakeholders are unclear and may conflict, and decision-makers need to weigh potential benefits and harms; the correct choice is uncertain (Reinecke & Ansari, 2015). In such situations marked by equivocality and uncertainty, people engage in a sensemaking process to construct an understanding of moral consequences that may unfold and make judgments about how to deal with them (Sonenshein, 2007). Contrary to the darker view of creators' ethicality, creators may actually have an advantage in that process; divergent thinking underlies creators' ability to imagine the potential consequences of ideas, consider a wide group of stakeholders, and integrate paradoxical moral issues (Bierly et al., 2009; Keem et al., 2018; Whitaker & Godwin, 2013). Supporting that view, Mumford and colleagues (2010) found that creative thinking skills were positively related to ethical decision making in doctoral students, and Keem and colleagues (2018) found that people with a creative disposition and strong moral identity engaged in more generative processes around moral issues and correspondingly behaved more ethically. Underlining the importance of creativity to making judgments about morality, philosophers and psychologists have used the term "moral imagination" to describe a process in which decision-makers project potential outcomes for different stakeholders (Johnson, 1985; Keem et al., 2018).

Research thus hints that considering the moral issues raised by ideas may be a creative process. What remains unclear is how projecting the moral consequences of ideas intertwines with the processes of developing and constraining ideas.

METHODS

To explore when and how creators navigate moral issues arising from the potential negative consequences of their developing ideas, we adopted an inductive, qualitative approach

well suited for understanding the lived experiences of informants and for theorizing processes (Gioia, Corley, & Hamilton, 2013; Langley, 1999), particularly when relatively little is known about the phenomenon under study (Creswell, 1998; Edmondson & McManus, 2007; Locke, 2001).

Research Context and Exploratory Work

The context for our study is artificial intelligence (AI) development—the development of tools and technologies that simulate autonomous intelligent processes based on computer science, machine learning, and deep learning. Developing AI is a creative process (Amabile & Pratt, 2016; Amabile, 2020) that entails generating, elaborating, and producing new ideas for technologies and selecting some to move forward, prototyping or implementing them. If successful, that process results in creative technologies fueled by algorithms. For instance, participants in our study developed algorithms to power tools such as sensors in a healthcare robot for the elderly, online shopping recommendations, a GPS tracker that collected data from and controlled aspects of long-distance transport trucks (e.g., fuel level, engine performance), an app to streamline physician-patient interactions, and a banking fraud detection system. At the beginning of the study, we were broadly interested in exploring those creative processes.

The first author undertook preparatory work to gain familiarity with the context. She attended a three-day AI conference as a non-participant observer, where she learned from presentations, engaged with attendees, and took notes on her experiences and interactions. That helped us to develop an understanding of the nature of AI and the dynamics of the field (Feldman, Bell, & Berger, 2004). In parallel, she conducted exploratory interviews with three AI creators and five executives from two AI organizations. These interviews were intentionally

broad in scope, covering informants' creative and problem-solving processes, interactions with other AI creators, and views about the future of AI.

During this preparatory work, moral issues related to the unknown potential outcomes of ideas emerged as highly salient in AI development. For example, Alice, an executive at a large AI organization, called out the challenge of dealing with ethical issues: "The challenges we need to overcome, it's the ethical part. I think within maybe the next five years, I am sure we'll run into questions like, well, the technology is there, so we *could*, but *should* we?" Lance, an AI startup executive, pointed to the ambiguity of the future: "It is sort of like driving home at night. Like you can only see 30 feet in front of you where your headlights are… Like you're driving along and at any point in time, we can only see 2-ish years in the future." The insight that moral issues loom large in AI development led us to narrow our research to exploring how individuals navigate moral issues during the creative process.

Our exploratory work further pointed to the lack of field or organization-level guidance on how to navigate moral issues. Data for the study were collected in 2018-2019, when AI was an emerging field and ethical standards and norms had yet to develop, enhancing the value of the setting for developing theory about how creators navigate moral issues. Indeed, AI development has been marked by a proliferation of debates regarding ethical dilemmas (Bossman, 2016; Liebrenz, Schleifer, Buadze, Bhugra, & Smith, 2023; Neri, Coppola, Miele, Bibbolino, & Grassi, 2020; Stahl, 2021), including what constitutes ethical AI (Munn, 2023; Rainie, Anderson, & Vogels, 2021), the role that ethical oversight should play (Hecht et al., 2018; Katyal, Liepold, & Iyengar, 2020; Vincent, 2019; Wearn, Freeman, & Jacoby, 2019), and how best to understand and control AI (Arrieta et al., 2020; Clarke, 2019; Lebovitz, Lifshitz-Assaf, & Levina, 2020).

The exploratory work refined our data collection approach by focusing us on AI creators who were actively involved in the process of creating new AI tools, such as software engineers, machine learning engineers, and research scientists, to whom we refer as AI creators. Further, we sampled our informants across different domains and organizations. While some big companies (e.g., Microsoft) have begun to work on creating ethical guidelines for developing AI, none of the organizations in our sample had established practices for addressing consequences of AI development at the time of our study, forcing AI creators to navigate moral issues on their own.

Data Collection

Sampling. We used two strategies to recruit participants. First, we invited AI organizations and authors of recent AI papers to participate in a research study. Second, at the end of each interview resulting from that strategy, we asked informants to connect us with colleagues who may be willing to be interviewed (e.g., Fayard, Stigliani, & Bechky, 2017; Petriglieri, Ashford, & Wrzesniewski, 2019). This approach was particularly helpful because, with growing public concerns regarding the dangers of AI, AI creators tend to be wary of outsiders who might bring negative attention to their work. Being referred by someone in the field helped us build rapport and made informants more open. To mitigate the potential bias of relying on informants' professional connections, we asked informants to connect us with those who had different perspectives and were working in different domains.

Our final sample consists of 69 interviews (64 AI creators and 5 AI executives) and includes researchers, scientists, engineers, and professors who were using machine learning techniques to develop AI in a range of industries. Appendix A provides descriptions of the informants. Our sample includes AI creators from both academia and industry.

Semi-structured interviews. The interviews were conducted by the first author and lasted approximately one hour. Interviews were recorded with informants' consent and transcribed verbatim. Consistent with many qualitative studies, the interview protocol evolved as data collection and analysis progressed to capture emerging concepts (Spradley, 1979). For example, initial interviews indicated that moral issues were associated with prospective thinking. This realization led us to revise our interview questions to capture past experiences that shaped informants' perspectives and project decisions that were associated with their thoughts about the future. The final interview protocol is presented in Appendix B.

Data Analysis

We used a bricolaged analytic strategy (Pratt, Sonenshein, & Feldman, 2022) that combined elements of a grounded approach (Strauss & Corbin, 1998) with elements of process theorizing (Langley, 1999). Below, we describe the analysis as a series of phases that moved from initial grounded analysis to building a data structure to process and comparative analysis of cases. As is typical in qualitative research, work was not as linear as described within each phase, and boundaries between phases were blurry (Charmaz, 2014).

Phase 0[1]: Analysis began during data collection, following a grounded approach of constant comparison (Locke, 2001; Strauss & Corbin, 1998) to capture emerging themes and iterate them with research literature. The first author read and reflected on transcripts in batches and created contact summary sheets that captured informant details and emerging themes from interviews (Miles & Huberman, 1994). The first author also drafted memos about themes and emerging insights (Creswell, 1998; Strauss & Corbin, 1998) and a data table with quotations and

[1] We refer to this phase as Phase 0 as it is the process that led to the development of the specific research question addressed in the present manuscript.

notes from each interview, which allowed her to identify patterns across informants and categorize concepts in multiple ways (Wolcott, 1994).

A key insight from the initial analysis was that AI creators had different ways of engaging with potential benefits and harms of their ideas and that they engaged differently with their work in response. The first author validated this insight through member checks during the latter batches of interviews (Creswell & Miller, 2000; Lincoln & Guba, 1985).

Phase 1: This insight formed a foundation that sensitized the next phase of analysis. Consistent with a grounded approach, we revisited the literature with this insight in mind to refine the research question (Gioia et al., 2013). In particular, we engaged with literature on ethics and morality in creativity (e.g., Khessina et al., 2018) and forecasting consequences of creative ideas (e.g., Zhou, Wang, Bavato, Tasselli, & Wu, 2019). That led us to shift the research question towards how creators deal with the moral issues associated with potential harms that may come with creative ideas.

To address that question, we followed the Gioia method of building a data structure (Gioia et al., 2013) by systematically searching the interview transcripts for instances where creators experienced moral issues during their creative process. Two authors, who were fresh to the context and data, coded transcripts separately. We began with open coding, drawing on informants' language to capture how AI creators thought about the future of AI, their ideas, and their creative work. For instance, we captured terms like "the next Terminator" or "a companion," and that creators felt "terrified" or "excited." The two authors coded interviews in batches, meeting regularly to discuss emerging themes. During these discussions, the other two authors acted as sounding boards, providing context-specific information and asking provocative questions that helped reveal what was interesting in the data.

Once the open coding revealed no new codes, we moved on to axial coding, where we looked for similarities and differences between first-order codes to develop more abstract themes (Gioia et al., 2013). For example, it became clear that AI creators developed positive and negative visions of the future of AI, and that emotional responses were tied to those accounts. We therefore introduced second-order themes of "constructing positive or negative images of the future" to capture them. We then further iterated the data structure with relevant literature to develop aggregate dimensions. For instance, we observed that constructing positive or negative images and distinct ways of making sense of ideas were components of "bright and dark imagining." We also observed ways that creators described engaging in idea development (e.g., Harrison, Rouse, Fisher, & Amabile, 2022) associated with the way they imagined the future.

Phase 2: We then used tables as "analytical devices" (Cloutier & Ravasi, 2021) to compare themes from Phase 1 across informants. A key insight was that, while many creators tended to engage in one type of imagining more than another, many also engaged in both. That insight prompted a further round of coding to uncover an explanation for the variation we observed, which revealed that surprises encountered during creative work catalyze forms of imagining. Our search for systematic differences in surprises resulted in two categories: "experiencing surprises during deep technical work" and "experiencing surprises when broadening the context." We found that many creators had a combination of those experiences and exhibited a mix of bright and dark imagining, but that one or the other tended to dominate their experience and lead to a relative focus on bright or dark imagining. We iterated our data structure by adding this category. The resulting structure is illustrated in Figure 1.

Insert Figure 1 about here

In the next stage, we aimed to "understand the patterns" to develop a process theory (Langley, 1999) by relating the constructs that emerged in the data structure to one another (Gioia et al., 2013). To aid that process, we used visual artifacts (Langley & Ravasi, 2019), such as hand drawn figures and models generated using software tools, that mapped the relationship between second-order themes. Creating and collaborating through these artifacts helped us better understand how different codes related to one another. For example, we initially coded "internalizing" and "systematizing safeguards" as separate parts of the process. Visualizing their relationship made us realize that they both constituted ways of "engaging with moral issues in idea development" (see Figure 1). To further theorize the relationship between constructs, we also constructed and visualized project summaries where it was possible to extract complete project processes from the data. The summaries illustrated the relationships between key constructs as creators engaged with specific creative ideas. We applied our data structure to those cases to relate experiencing surprises, forms of imagining the consequences of creative work, and ways of engaging with moral issues and idea development to test and confirm the way that AI creators' technologies and projects developed.

Working with analytical artifacts in a systematic way also enabled us to explore alternative explanations for our observations, such as how the characteristics of AI creators shaped the way they engaged with moral issues. Appendix A details those characteristics and Appendix C summaries their relationship to our main findings. We are cautious about drawing conclusions from these characteristics because in the field of AI, the lines between academia and industry are often blurred, with many companies housing fundamental research and many academics applying their work in for-profit companies. However, the data suggest that AI creators may develop more complex ways of navigating moral issues over time, with a higher

percentage of creators with over 10 years of experience engaging in both approaches than those with less experience. The differences in approach based on informants' jobs and industry also suggest that different catalyzing surprises may be associated with different forms of work.

FINDINGS

> I'm very optimistic about this job thing (job displacement) that people talk about. Because I actually have a vision in the future that people don't have to work, because the wealth will be grown by automation and all the advanced technologies. So, not having a job is not a bad thing. (Jenny)

> It's possible that [my AI] could be used to generate spam or to generate definitely, clearly-not-okay honeypot bots. Like, "Let's get one million people to talk to it. Let's make sure that people who are online are getting radicalized in horrible ways." Like, "If you Tweet about a break-up, I want to make sure that I quickly send you these five awful videos about how it's all the woman's fault." …It is often more evil than we intended it to be. (Joy)

AI creators in our study painted vivid and detailed images of the future. Informants were united in their motivation to contribute to scientific progress and improve people's everyday lives through their work, but they also acknowledged the potential for AI to harm large numbers of people, and they experienced uncertainty and equivocality about how those distal outcomes may unfold. Adam, a research scientist working on transportation and navigation algorithms, described how he grappled with the moral issues raised by his ideas: "It was unclear the valence of its impact. So maybe I make algorithms like 10% better, but in one scenario, some government then uses that to do something terrible, then will that really help the world?" Adam added that these "prosocial…concerns were really gnawing at [him,]" illustrating the struggle expressed by AI developers in our sample.

We found that within this context, creators engaged with the potential moral consequences of their work through a dynamic practice of *imagining*, exemplified by the quotations above, in which they projected the distal consequences of ideas. Imagining took place

as creators made sense of *surprises* encountered through their creative work, and it shaped the way that creators *engaged with moral issues* and in *idea development* itself. Specifically, our analysis revealed that surprises, imagining, engaging with moral issues, and engaging in idea development could unfold in two ways (illustrated in Figure 2). We characterize the first process as *disconnecting moral issues from idea development through bright imagining*, which entailed a magical belief in the good that would come from AI (Jenny). This enabled creators to separate moral issues from creative work, developing ideas in an unconstrained manner while *systematizing safeguards* to apply across projects. We characterize the second process as *integrating moral issues with idea development through dark imagining*, through which creators came to see their ideas as entailing potential downstream harms (Joy) but also as being controllable. That encouraged creators to internalize responsibility for moral issues while embedding constraints directly into their ideas. That transformed moral issues into creative challenges for developing AI technologies themselves. We elaborate these in detail below and provide additional examples in Table 1.

Insert Figure 2 and Table 1 about here

Figure 2 shows that creators tended to be oriented towards either bright or dark imagining, but they could engage in both and sometimes switched between them, depending on the nature of surprises they encountered in their creative work. As we discuss in detail below, bright imagining was catalyzed by surprises during routine creative work that triggered creators' curiosity and motivation. Correspondingly, bright imagining developed through creative work for most creators in our sample. In contrast, dark imagining was catalyzed by surprises that occurred when creators broadened out to the user context or a wider social interaction. These

surprises were less frequent but salient. If such surprises were sufficiently salient or if a creator experienced several over time, their dark imagining could eclipse bright imagining altogether. A small percentage of creators in our sample engaged primarily in dark imagining (about 30%, versus about 50% in bright imagining), and several informants did not recount any surprises that prompted dark imagining, whereas almost everyone displayed some bright imagining. We elaborate these relationships below.

Disconnecting Moral Issues from Idea Development Through Bright Imagining

AI creators in our study often expressed wonder and awe about AI when experiencing surprises during *deep technical work.* To make sense of those surprises, creators engaged in *bright imagining*, portraying the future as filled with promise. We further found that bright imagining motivated creators to pursue novelty and develop ideas without constraint, while viewing potential harms as distant. They separated moral issues from that work while systematizing safeguards across projects. In this way, bright imagining *disconnected moral issues from idea development*. We provide extended examples of this approach in Table 2.

Insert Table 2 about here

Experiencing surprises during deep technical work. As creators developed their ideas into AI technologies, they often discovered that their AIs produced unexpected results. This elicited surprise and wonder, accompanied by a sense that they did not fully understand the technology. For example, AI developer Eric described writing a program for computers to play chess together as part of a competition and witnessing surprising outputs:

> I write it, so basically, I know everything it can do, but sometimes it surprised me, like, oh, what? Why did it make that move? I didn't think that it could make that move, and it did! And that kind of, "Oh wow," I mean something that's outside of my expectations, and it can happen with this simple algorithm.

Another informant, Barry, described how his drug binding neural network experiments produced outputs he did not anticipate:

> It was supposed to predict if the drug was binding or not binding, and what you saw was that some groups of drugs had probability of 80 for binding, some had a probability of 60, some had a probability of 20. And it was super cool, because it was like, holy crap, this isn't just telling me whether or not it binds, yes or no, it's actually telling me what class the thing binds to just with this probability. It's, like, grouping them into these bands, and I didn't ask it to do that, it just happened kind of naturally.

Barry was animated as he elaborated, "That was a really exciting experience because…I didn't think it was gonna work in the first place, and not only does it work, but it actually gives you this…clustering that you didn't even ask for, and it was free." Barry's comment illustrates how the routine, everyday work of running the algorithm generated novelty above what he expected the network to be able to do. Informants described many instances when, after spending hours and hours developing their idea, they discovered that it did something unexpected. AI creator Jim noted that "something like that regularly happens," and Adam emphasized that surprises often entailed "day-to-day…kind of mundane things." Another informant, Gary, claimed that he was "positively surprised almost every time I look at the explanations" from a model. As technical surprises arose often through delving deeper into the details of ideas, a bright image of the future built up and strengthened through routine creative work.

Bright imagining. Encountering such surprises fueled an interpretive process through which AI creators came to understand the nature of their ideas and their potential consequences. Since creators were deeply entrenched in their work when technical surprises arose, those surprises gave creators a sense of amazement and a belief that emerging ideas were a kind of magical "black box" that could not be fully understood, building awe and wonder for their ideas and emphasizing positive distal consequences. Richard, a Chief Technology Officer and engineer

at an image classification and natural language processing startup, explained how experiencing surprising outputs produced this interpretation: “The relationship between what you put in and what you get out is so complicated that you don’t really have a rigorous understanding of what’s going on. …(It) starts to feel like magic.” Similarly, Jacob, who was developing AI for design systems, explained, “It surprises you in ways you don’t expect. …It’s the kind of thing where the solution that it gets is so unlike yours that it just, it’s mind-blowing to me.” These quotations illustrate that when creators observed surprises during deep technical work, they came to see the technology as not fully understood or understandable, yet filled with promise.

This perspective highlighted the potential benefits that could come from creators’ ideas. Creators portrayed a future where the “magic” of AI would change the world for the better, while bracketing its potential harms. For example, Thomas, who worked on optimizing evolutionary algorithms, generated a number of ideas for how AI would benefit people:

> Assuming we can handle the policy issues, [human-level AI] would be an unprecedented level of productivity. Human-level AI means that everybody could have an individualized assistant capable of handling the massive amount of information. …Does someone want to schedule an appointment with me? It’ll be our AIs arranging a time according to my schedule preferences and maybe even the subtle irritation in my voice in setting up the meeting. AIs to handle where my money gets invested. AIs to watch what I’m coding and make suggested changes about structure… Customized medical care based on my specific expressed genes. AI teachers that adapt to each student to teach them at their own pace. Those are just a few ways I think AI will be expressed.

At the same time, Thomas minimized “policy issues”—jobs replaced, money invested in unethical stocks, or diagnoses delivered in an uncaring way. Another AI creator, Jack, similarly prioritized a positive vision of the future while bracketing potential harms:

> People maybe don’t retrain and lose their jobs, and gosh, that’s just tragic for them. But sort of in the long term…things improve. …We’ve automated all this stuff, and people just don’t need to do that crap anymore. So I think AI is maybe going to be one of the most powerful examples of that. It’s going to be different than the internet or the car. This is going to get to the point where it can do 90% of what an average person can do, and that’s going to free up a lot of hopefully productivity for people.

Bright imagining thus captured the positive distal consequences of ideas that were generated through a search for understanding what the technology could do.

Disconnecting moral issues from idea development. Bright imagining formed a backdrop that sustained AI creators in compartmentalizing moral issues and disconnecting them from idea development. That disconnection shaped both the way that they worked on ideas and how they intended to manage the moral consequences that could arise from those ideas.

Engaging with moral issues by systematizing safeguards. AI creators recognized and expressed interest in the implications of their ideas; they did not entirely dismiss negative distal consequences. Instead, because bright imagining made benefits of AI salient while bracketing potential harms, creators distanced those harms into a far-off future, away from ideas themselves. Matthew describes pushing the "wide impact" of AI on human workers into the future:

> I don't feel that guilty that much right now. …At this point in the evolution of AI, the displacement is small, and it's going to stay small for a long time. …It'll be a tiny percentage every year forever. It will have a wide impact. It's that old thing about boiling the frog–the frog doesn't notice that he's in boiling water because it's only raising a tenth of a degree every minute. So it isn't gonna be a sharp spike in displacement, it's just gonna happen slowly over time.

John similarly suggested that AI would not "suddenly" take over the world, but that instead there would be a "steady, steady slope upwards" that would unfold over time. Thomas, too, noted that an "apocalyptic" future for AI was "really far away" because of AI's complexity and black-box nature. Viewing AI as broadly "good" also allowed creators to distance negative consequences from ideas themselves and associate them with users of those ideas. Nick compared AI to other technologies: "Bad people can do bad things with everything. If you invent a 3D printer, then people can 3D print a gun. Every time you come up with something new, people with bad

intentions can use it in a bad way." Attributing risks to people rather than the technology itself allowed creators to further distance the potential harms of AI.

In turn, creators who engaged in bright imagining allocated moral responsibility—the responsibility to address moral issues—to policy makers, ethicists, or organizational decision makers, and they *systematized safeguards* by applying general rules to guide AI behavior across their projects. Systematizing safeguards enabled creators to respond to potential consequences without having to constrain their idea development process. One approach was to call for industry-level regulation and guidelines. Data scientist Sam suggested that a question such as who to blame if an autonomous car crashes is a "fundamental question that needs to be answered—not by AI scientists, it needs to be answered by society." Creators hoped that others would eventually regulate these matters.

Since little regulation yet existed, creators also developed personal norms for their creative work. Those guidelines were external to the core idea itself and tended to apply at the beginning or end of projects. For instance, creators determined what to work on by applying self-generated guidelines to avoid certain projects or domains. Jeremy, who was working on self-driving cars, explained that he did not want to make "AI related to military stuff," because, "I want to be part of this AI generation, but I also want to do it properly." Another informant, Zack, expressed the same sentiment towards "advertising companies, which I personally don't think is a really great thing. …Because if you optimize this, you're not going to make people more free. You're going to manipulate them." Systematically safeguarding the fields in which their ideas would be applied helped creators to develop the ideas they did work on more freely. Another safeguard was to apply a "button" to "switch off" a technology that brought about adverse effects. Colton explained, "At (a) high level, can we think of, you know, like, a big red button

that we can push when things go wrong?" Many informants applied this safeguard to their own work and called for it to be systematized as a principle through regulation. Eve, who worked on recommendation systems, argued that regulations could help them develop technologies more freely: "I don't think we should just stop advancing the technology just because people can potentially use it for bad purposes. I think we should be more on the law side of how we regulate people using it." Creators thus viewed safeguards as external to idea development.

Developing ideas without constraints. Creators' positive views of the future oriented them towards pursuing novelty and unconstrained idea development, aided by systemized safeguards that were distanced from any specific creative project. Bright imagining engendered strong positive affect as creators expressed excitement about the promise of AI and the motivation of scientific discovery. Ryan described "rare moments where I had really, really good ideas" as "epiphanies" that gave him an emotional boost and motivated him to continue his scientific work: "When I realized it made sense, then the emotional effect is huge, huge. It was an amazing feeling, very high." Richard described the potential of the "young and fast-paced" "machine learning space" as "quite honestly mind-blowing." He went on to express "awe and surprise" as well as a sense of "accomplishment" about developing new technologies. Bright imagining thus oriented creators towards novelty and engaged them deeply and emotionally in discovery, deriving from their experiences but also sustaining their bright visions of the future.

The pursuit of novelty motivated creators to develop their ideas unfettered, foregrounding the novelty of ideas in the abstract over practical concerns. John described working on abstract problems unconstrained by practical realities, rather than engaging with real use cases: "Computer scientists frequently abstract things down into weird, alternate universes of strange mathematics and hyperplanes, and you know, we often think about unreal spaces and solving

problems in those unreal spaces." Bob, who was developing AI for use in medical systems, demonstrates how he zeroed in on technical challenges without concern for distal harms:

> For researchers like me, we do not care too much about what the future might bring. Instead, we are looking into the high-impact or the burning questions we are having today and trying to answer it. Where it will lead us, we actually do not care too much. …Personally, I do not worry too much.

Creators thus compartmentalized creative work and allocated moral issues to others. Jack, who was working on protein-docking models, explained how he saw moral issues as separate from his work: "It's more in the back of my mind. If I was letting it influence things, I might not be working in this field." Similarly, Nick, who was developing AI to automate organizational processes, noted, "Discussions about AI and ethics, and what's fair and what's not, and where we should set a limit and shouldn't…that's really not my field." And, when asked whether he saw possible dangerous applications as connected to his work, Nathan, a professor of computer science, thought, "It's actually not a technology or scientific question. …I try to leave it to the actual experts." These informants moved ideas forward by shielding them from moral issues.

Integrating Moral Issues with Idea Development Through Dark Imagining

In addition to frequent surprises during deep technical work, many AI creators also *experienced surprises when broadening the context* in which they were developing or thinking through their ideas. Creators made sense of the latter type through *dark imagining*—painting a picture of a future filled with harms that could occur due to AI. When creators engaged in dark imagining, they tended to internalize moral issues as their personal responsibility and to embed constraints into their developing ideas. This *integrated moral issues* into idea development. Please see Table 2 for extended examples.

Experiencing surprises when broadening the context. AI creators in our study encountered surprises when they broadened their view of their AI tools, whether through

working with potential users in a particular domain to design and test the idea, context-driven testing, or socialization beyond their immediate technical colleagues. Those experiences occurred less frequently than technical surprises, arising in later stages of a project when testing an idea or through encounters with non-technical colleagues, but they were salient to creators.

Broader contextual surprises revealed the human impact of AI, opening creators up to a different future than they had previously considered. These surprises thus shifted the way creators envisioned the future of AI. For instance, Josh, who was creating navigation and control algorithms for a healthcare robot, recalled that during his first user testing session, he saw the "fear" in his users' "eyes," and that changed his priority from "accuracy and performance to being user friendly." He went on to comment that "having fully self-controlled AI" would be a "nightmare." As another example, Christopher, a machine learning intern at an image classification and natural language processing startup, described becoming concerned with "the perils of deploying" his image classification AI when he realized through testing, "We weren't getting the most accurate counts for pedestrians and bikes." Observing his AI fail to identify someone who "isn't part of the majority class" on which the model had been trained, such as "a kid on scooter, or someone in a wheelchair," gave him empathy for those people. In addition to attuning to potential human impacts during idea development, creators also encountered surprises when they broadened their interactions to a wider community, including people from other disciplines, such as philosophy, sociology, and public policy. Creators also described these as less frequent but salient experiences that influenced their thinking about the future. For example, developer Alan explained that meeting a Marxist sociologist and discussing with him how AI tools could affect the kind of work that people do "shaped my thinking…about the

impact of research." These interactions with a broader community thus also opened creators to the diverse possibilities for their ideas.

Surprises in the broader context moved creators towards dark imagining, either as a "gradual awakening" (Adam) that occurred as surprising experiences accumulated, or as a "pivotal moment" (Ellen). Adam recalled several experiences that opened his eyes to potential distal consequences over time: a policy fellowship that revealed a lack of AI expertise in government; stepping back to design the incentives for his AI systems; and reading the book "Cat's Cradle," with its "image of the celebrated scientist who is just not a very solid human being who ended up precipitating the end of the world." And, illustrating a single turning point, professor Ellen described "one of the most remarkable experiences" in her career taking place during a discussion of ethics in a zoology professor's cognitive science doctoral course, where her thinking shifted away from "human-level intelligence, wahoo…that would be so cool!" towards "look[ing] past the shininess" of AI and "into what the implications actually were." Such surprises could thus change how creators thought about the future consequences of ideas.

Dark imagining. Surprises that arose when AI creators broadened out to a wider context introduced new aspects of ideas for creators to explore that they had not previously considered. To make sense of those surprises, they tried to unpack even the darkest black boxes and thereby demystify the technology, developing a view of their ideas as explainable and controllable. For example, Harry worked on language translation AI, aiming to create AI capable of sophisticated communication. He described, "Before I started working on [my machine translation AI], it seemed kind of magical…that this is even possible… There's this…illusion of the machine learning." However, when he became aware of "medium-term problems," like the proliferation of "fake stories" using a competitor's generative model, it made the issue concrete and

potentially imminent, since it could occur even now. He began digging into the functionality of his algorithms and working to improve them by applying them to different problems. This revealed the nature of his AI as controllable machine rather than magic: "After I…worked on them and improved them and actually used them to do these things and understood how that machine did that, it looked more like…a machine. …It was more like a relative system." Steven similarly described a need to reverse engineer AI to provide transparency and ultimately control:

> We need to…make sure we understand what AI is, then be transparent with regards to the algorithmic processes of AI, so we can actually reverse-engineer if necessary, right? Otherwise, if [it's] sitting in a black box that we don't understand, then it's not necessarily a good idea, in the long run. …We need transparency, as much as possible. Then, we would be able to situate AI where it belongs, or where we want it to belong, by controlling the procedures.

Demystifying the technology showed creators that the consequences of their ideas did not arise through "magic" and could instead be predicted and controlled.

Making sense of broader contextual surprises by building an understanding of how their ideas worked in the real world provided creators with information to develop negative images of the future. Those images entailed how AI could cause harm to individuals as well as society at large, making potential harms concrete and immediate. For instance, Ellen, who was developing AI for robotics systems, envisioned the future with "great concern": "When you no longer have to send soldiers into harm's way to wage warfare, when you can just send your robots…or just use your robotic drones to make the air strikes…that makes it that much easier to do completely horrible things to each other." Dark imagining was complex and nuanced; rather than "prophesizing doom" (John), extreme narratives were typically dismissed or discounted by AI creators. Instead, dark imagining elaborated project-specific harms that could emerge from the downstream effects of ideas. Jordan's thought experiment about how his game content generator could be "malicious" extrapolated how negative effects could unfold:

> Imitation is a concern of mine because it is not really creative. But, of course, it's also malicious. Because it's really not giving credit to whatever is out there. So, those are things that I care about and those are things that I should care about if I want to move forward, for example, with things such as game design which uses existing designs. Do we show something like, you know, someone else's design? And if we do, is it okay with them?

As another example, Marie noted that if her credit card fraud algorithm was optimized to capture every fraud instance, it would also misclassify some non-fraudulent cases, which "could be really serious." She went on to describe "lots of negatives" from how her work could be applied:

> Applied technology also means that we have less and less privacy. For example, we put the video [technology] everywhere, so people just expect it all the time. And also all that we do from the internet, all the information are (sic) recorded. So we can say that all of our daily life…will be collected and recorded. This is something that I don't like a lot.

Hence, dark imagining followed from a deep understanding of the technology and generated predictions for AI that could include both general and project-specific distal harms.

Integrating moral issues with idea development. By spotlighting the potential harms of ideas, dark imagining supported *internalizing* responsibility for moral issues and *embedding* constraints directly into specific ideas with customized solutions intended to limit the harm of their AI tools. Creators therefore integrated moral issues into creative work.

Engaging with moral issues by internalizing. Dark imagining moved the potential harms of AI closer, making them concrete and immediate. Ray, who was creating AI to map biological systems for healthcare, noted an acute need to prevent the concrete harms he envisioned if AI learned to simulate human emotion, commenting that AI was "not this mysterious science fiction creation," and its "consequences…(were) more concrete" and "real." Similarly, Jordan, in developing a game design AI, emphasized that he "should care about" the possible harms before moving forward with a particular design.

Dark imagining thus gave rise to a sense of urgency and personal responsibility. When engaging in dark imagining, creators equipped themselves with the knowledge to address the moral issues pertaining to their work. They internalized agency over the future of AI, expressing a sense of duty to prevent or minimize the likelihood of the harms they envisioned. Christian described his responsibility for the navigational AI he was developing and its outcomes: "It wasn't a sensation of guilt. It was more a feeling of responsibility…for what I was doing. When you develop a system…you carry at least part of the responsibility of what this technology is going to become." Rather than compartmentalizing and assigning responsibility to others, these informants placed responsibility squarely within the AI community, including researchers:

> Some gang or maybe bad guy or bad girl can own some kind of technology, right? …And they can make some kind of dangerous AI. …Definitely we need to think about that. And not only the research community but the AI community also have (sic) a lot of debate about that, the risks of AI, and how can we deal with that? (Eric)

Harry indicated his trust that AI developers would take on the challenge of limiting potential harms and would "get better and better at this." Adam similarly expressed his emerging sense of responsibility: "It's very hard to predict the downstream effects. But the more I think about it, the more I think there is some responsibility to make sure that you've at least thought a little bit about that. …Otherwise, where does the moral responsibility lie?"

To further internalize moral issues, creators engaging in dark imagining explored moral theories, sometimes in great detail. Christian described his thought experiments about how to navigate moral issues with the self-driving car technology he worked on:

> You can say, okay, let me adopt a utilitarian approach, which means for example that I will always try (to) maximize the number of people saved. Or I can…decide, for example, that I will put more priority on, say, young children…or on elderly people…or black women, whatever. I'm not saying it's good or bad of course here, I'm saying it's a different moral theory. You can say also that I'm not going into this utilitarian approach at all and I'm not going to decide to kill one person to save five. …Let the vehicle, if it's going forward, for example, and one person is going to be killed, well, too bad…I'm not

> going to change the course of action. And then the question comes to us, to the designers, to the programmers: What would be an acceptable decision?

This process sometimes led creators to question their work and its goals. For instance, comparing the treatment of people to the treatment of AI led Ellen to question the morality of her work:

> As a researcher, I deeply question whether that (human-level AI) should be our goal at all at this point. And that's purely from my own ethical standpoint. The way I think about it is, we can't treat other human beings particularly well and equitably yet across the board as a species, so how would we treat an intelligence that was completely different?

Adam, who was initially "driven" by "intellectual passion and just trying to understand," re-oriented his work to tackle moral issues: "These days I've switched my focus of research to a subfield called AI safety. …I'm also getting more into the ethics of AI and making sure that the AI behaves ethically." Creators engaged in dark imagining thought explicitly and directly about the possible moral consequences of their work and internalized responsibility for them.

Developing ideas by embedding constraints. The examples further illustrate how dark imagining oriented creators towards the usefulness of ideas, sustained by negative affect associated with pessimistic visions of the future and a motivation to avoid those outcomes. Harry noted that the "end goal" of his work was "to be useful to people." Ray elaborated:

> There's a motivation that you want to build something that's useful, because you know that it can be used. It may not be you who actually takes it out there, but somebody can immediately. …And then there's a motivation because you have actually built systems, real systems in the real world. There's an entire world of motivations that come in and problems that come in that otherwise would not.

Underlying this motivation was the fear and anxiety that dark imagining elicited from creators. Just as Ray tempered his motivation to build "systems in the real world" with "thinking of the consequences," Samantha, who was building a speech interface for an "intelligent nutritionist" tool, said, "It's scary to think what happens if all this data got accessed," and Joy described feeling "uncomfortable" about how her AI could be used to generate "awful" spam and worrying

that people were unprepared "to fight against what's coming." Those negative emotions were intertwined with a desire to develop useful ideas for end users and society more broadly.

Correspondingly, dark imagining fundamentally reshaped how creators engaged in their work: Creators generated constraints designed to limit the potential harms of a specific AI while developing their ideas, and rather than applying tool-agnostic rules, they developed solutions that were customized for each AI. An embedding approach began with considering concrete users and sometimes involving them directly during the design process. For example, Brian, who was developing a virtual assistant, was concerned about job displacement due to AI in "competition against a human" or "replac[ing] the human." He explained:

> I always try to kind of work on paths that I thought started with the human. What do they want to do? What would the best experience possible be for them, so that they're in control, they're empowered to think… You know, to augment their capability? …I think it's more of this conscious choice inspired by [my colleague] to…start at the human.

Similarly, as he built his game design AI, Jordan worked with "a human designer" end user to understand what it was like to collaborate with AI to create games, focusing on how to make the actions of AI more explainable to end users. In another case, Josh, who was developing care tools for the elderly, held design sessions to understand how elderly people wanted robots to look and act, as well as to make users more comfortable with the technology.

That process gave creators detailed information about how users would interact with their ideas, which they could use to develop ways to constrain their ideas. One way they aimed to embed constraints was by devising techniques to eliminate or reduce bias in their AIs—a creative challenge itself. Ray reflected on the dangers of creating biased AIs: "These AI systems can very easily become tools for enforcing and promoting stereotypes and biases. Because [biases] exist in data, then they become engrained in the AI." In his work creating AI in healthcare, Ray argued that there was no longer "an excuse that, 'Hey, that's in the data, so what can we do?' No, we

can do things about it. …And it's a challenge." He explained, "We have to build technology that prevents that from happening." Applying strict criteria to data that went into the model forced his team to think more creatively about what data to use and how to debias the model, such as by identifying more fine-grained relationships between variables used to make predictions or by developing "stepping-stone" techniques "that explicitly explore and promote diversity." Creators thus incorporated developing constraints for a particular idea into their creative process.

Creators also transformed constraints into creative challenges to limit the functionality of their AIs, such as by restricting their capabilities, autonomy, or efficiency. Constraining functionality thereby reshaped the way ideas developed. For example, Alan described how he prevented his game design AI from replacing human workers, saying that he was "actively working against the ability that it could ever replace humans. So now I consider part of the goals of my research are to actually make my work resistant to that function." He described working toward this goal by requiring his game design AI to "maintain a dependence on human beings" and thereby restricting the speed at which it produced output: "It's not just that it's dependent on humans; it has to wait for humans. So it can't, like, make 300 games in a day, because it has to wait, like, a week for the artist to email it back. So I'm kind of engineering slowness into the system." This approach went well beyond simply providing a way for humans to override the system, forcing the technology to interact with humans throughout its processes. By considering what they did not want their tools to accomplish, creators taking this approach devised specific techniques to reduce their tools' abilities.

DISCUSSION

Creativity is a source of organizational, scientific, and societal progress. Yet, the same ideas that are responsible for that progress may have unintended distal consequences that extend

harm to a wide group of stakeholders over time (Khessina et al., 2018). To the extent that prior literature has studied such negative consequences, it has done so retrospectively and has tended to treat creators as focused on immediate outcomes rather than ideas' potential distal harms. By exploring how creators deal with potential downstream consequences while developing ideas, our study finds quite a different view: Creators engaged deeply with the possible distal outcomes of ideas. How they projected that future, through bright and dark imagining, prompted both different ways of dealing with moral issues and different ways of engaging in idea development.

An Emergent Model of Navigating Moral Issues in Creative Work

Our model, summarized in Figure 2, shows how creators navigate the potential distal consequences of their ideas. We conceptualize the model as an imaginative, prospective sensemaking process: A search for understanding is initiated by encountering unexpected and ambiguous effects of developing ideas. As creators rework their understanding of the distal consequences of their ideas, this, in turn, shapes their creative activity (Sonenshein, 2007; Sandberg & Tsoukas, 2014; Thompson & Byrne, 2022; Wenzel et al., 2020). Bright and dark imagining are thus projections for the future through which creators make sense of their ongoing creative work (Kaplan & Orlikowski, 2013). Because these imaginings emerged in response to surprises during creative work, we theorize that they are in a continual state of construction and evolution, updating as new experiences and interactions emerge (Thompson, 2018).

How surprises shape imagining. We observed that most creators had a dominant orientation (Harrison, Askin, & Hagtvedt, 2023) towards either bright or dark imagining, though few engaged exclusively in one or the other. Deep technical surprises were frequent events; we theorize that they were likely to begin early in one's career and occur consistently over time, forming an important part of creators' motivation towards and curiosity about new ideas. Indeed,

only one informant did not express any degree of bright imagining (see Appendix D). Broader contextual surprises were less common and predictable because they depended on whether creators engaged with users in design, tested their ideas in context, or interacted about their ideas with a broader community. Correspondingly, more of our informants (nine) showed no evidence of dark imagining. As surprises from the broader context were highly salient, they shifted creators towards dark imagining and shaping the human consequences of their ideas.

A broader-context surprise could be a turning point where creators moved from bright to dark imagining, but routine technical surprises could continue to fuel bright imagining, complicating the process of making sense of the future. Most creators in our study exhibited some degree of both forms of imagining. Exploring those cases in our data leads us to conceptualize a process in which broader experiences could accumulate and shift creators from bright to dark imagining gradually, with most creators existing somewhere on a spectrum that blended the two. Broader surprises may color a creator's view without tipping them entirely into dark imagining, because routine technical surprises can sustain magical thinking about ideas. Technical surprises may also draw creators more deeply into the details of their work through unconstrained idea generation, leading to more routine surprises from deep technical work. Alternatively, creators may shift from bright to dark imagining, such as Adam, who claimed "moral responsibility" and changed the nature of his work after several contextual surprises: "Over time that sort of (intellectual) passion sort of died down, and I had an existential crisis… I didn't really feel like what I was doing was useful to anyone." The flexibility to engage in both bright and dark imagining and to shift over time may have reflected and supported the moral pluralism (e.g., Graham et al., 2013; Hecht, 2010) that is likely to exist around highly novel ideas.

The importance of surprising experiences in shaping sensemaking is further underscored by exploring negative cases in our data. Creators encountering apocalyptic predictions in the broader community tended to dismiss them as unrealistic, so these predictions did not prompt sensemaking. And, if an AI failed to perform a simple task, creators responded by thinking that the idea was fallible and therefore harmless, thereby overlooking those incidents as well. We suggest that sensemaking results when surprises challenge a creator's understanding of ideas.

How imagining shapes idea development and engagement with moral issues. The model also shows that the processes involved in imagining the consequences of ideas fundamentally shaped the way creators engaged in creative work. The way that people make sense of ambiguous situations can give rise to practices that help enact the future (Thompson & Byrne, 2022; Wenzel et al., 2020; Kaplan & Orlikowski, 2013; Bartel & Garud, 2009). We observed that bright and dark imagining led to different ways of navigating moral issues and engaging in idea development. The way that creators made sense of ideas shaped how they engaged with moral issues by affecting the distance at which they perceived potential harms; viewing ideas as magical and unexplainable made them more abstract and pushed potential harms into a more distal future (Trope & Lieberman, 2010), while calling for constraints through generalized regulation (Acar et al., 2019). Viewing ideas as controllable and unpacking how they worked made harms more concrete and immediate (Trope & Lieberman, 2010), transforming potential moral consequences into product constraints that stimulated new creative solutions (e.g., Rosso, 2014). How they made sense of ideas further shaped the way creators engaged in the process of developing ideas by influencing the content of their priorities, such that bright imagining facilitated the pursuit of novelty through unconstrained idea generation (Acar & van den Ende, 2016), and dark imagining supported a process of developing useful solutions based

on human needs more similar to an iterative design-led approach (Ravasi & Stigliani, 2012). We suggest that creators' orientations towards bright and dark imagining supported and sustained their approaches to moral issues and idea development by helping them make sense of their creative work in light of imagined distal consequences (Bartel & Garud, 2009).

Boundaries of the Model

In focusing on how creators navigate moral issues in the creative process, our study does not reveal whether one approach is more effective than another in reducing harms. Further, our work focuses on how creators wrestle with moral issues and allows the collective dynamics that may facilitate that process to fade into the background. Studying the emerging industry of AI provided an extreme example that enabled this focus (Bamberger & Pratt, 2010); standards had not yet emerged to guide AI creators' choices, and the cutting-edge and cross-domain nature of participants' work meant that peers provided limited guidance. Our qualitative data also do not provide definitive insight into the personalities or underlying characteristics of creators, so we are not able to say whether particular characteristics led a creator to pursue pure scientific discovery work versus applied work that enabled them to broaden beyond the immediate creative task. Nevertheless, because we were able to identify surprises as catalysts for bright versus dark imagining, our study reveals the crucial role of experiences during the creative process in shaping how creators engage with potential moral consequences of their ideas.

AI development was an excellent setting for studying our research question, because it entails a creative process where uncertainty about moral consequences is high due to the emergent state of the field and the salience of potential harms. It therefore surfaced many instances of our phenomenon of interest, providing a good setting for developing theory (Glaser & Straus, 1967). The logic of analytic generalizability (Yin, 2009) suggests that our core insights will translate to other domains of creativity, because both scholars and anecdotal evidence

suggest that almost any novel idea has the potential to create some harm. For instance, a clever marketing campaign may encourage people to act in ways that are unhealthy, such as eating too much sugar or overspending (McLaren, 1993), and automation may alter someone's job, impairing their performance or displacing them altogether (Beane, 2019). However, boundary conditions of our theorizing may alter the precise nature of the processes we observed. One boundary is the degree of harm that an idea may cause. Producing a piece of art that one person finds offensive, for instance, may cause less of an internal struggle for a creator than making technological advances that result in a weapon that could kill thousands. Another boundary is the industry's maturity. Although novel ideas push boundaries even in mature industries, some domains may have a strong consensus around values that enables creators to overlook potential harms. Those boundaries may alter facets of the processes we observed.

Theoretical Contributions of the Model

Our study presents a grounded model of how creators engage with the potential moral consequences of their ideas. Because creative contexts are characterized by equivocality and uncertainty (George, 2007; Mueller et al., 2012; Sonenshein, 2014), the outcomes of novel ideas are unknown and their consequences cannot simply be assessed through a decision-making process. Instead, they must be projected through a constructive, imaginative process (Berg, 2016; Sonenshein, 2007). Whereas past work has examined the effectiveness of shorter-term projections (e.g., Berg, 2016; 2019), our study unpacks how more distal projections unfold. Such projections are triggered by surprising experiences that prompt creators to make sense of the consequences of their ideas by constructing a vision of how an idea will move into the future. Underlying that process is a view of ideas as unpredictable black boxes (see also Lebovitz et al., 2022) or as understandable and controllable. That parallels alternative approaches for engaging

in creative work: The disconnecting approach emphasizes pure scientific discovery and idea generation, whereas the integrating approach follows a design logic in which creators take inspiration from enacting ideas and observing users (Harvey & Berry, 2023).

Our model provides a framework for how creators integrate potential consequences of ideas, beyond their short-term novelty and usefulness, into their work. Our study adds moral consequences to the set of criteria that creators consider when they project their ideas into the future, showing how, for example, moral issues may prompt deciding whether to work on a project or making a feature safer for users. Although novelty and usefulness have been conceptualized and measured in a variety of ways (Harvey & Berry, 2023), their corollary of potential harms has rarely been explicitly highlighted. Our research therefore provides a starting point for expanding the very definition of creativity. Future research may consider how creators develop ideas that are novel and useful, *and* that have minimal potential for harm. Although we focused on moral issues, we observed that these were linked with the usefulness and novelty of ideas in creators' projections. Further research may consider how creators incorporate other criteria into those projections, such as sustainability (Harrison & Nurohamed, 2023), aesthetic dimensions (Stigliani & Ravasi, 2018), or long-term value (Harvey & Mueller, 2021), and what catalyzes those considerations. Our findings hint that engaging explicitly and directly with moral issues focused creators on usefulness by shifting them from an individualistic motivation towards a more collectivistic one, which may also underlie other criteria for creative ideas.

In identifying the nature of creative work as a source for stimulating the way that creators think about the consequences of ideas, our work also contributes to our understanding of the morality of creativity and an emerging interest in its "dark side" (Cropley, 2010; Khessina et al., 2018)—issues that are becoming increasingly urgent for researchers to understand with recent

technological developments such as AI (Amabile, 2020). Our study goes beyond prior work that has tended to attribute moral behavior to the personal characteristics of creators (Kapoor & Kaufman, 2022; Keem et al., 2018; Vincent & Kouchaki, 2016), giving an impression that those creators may disengage from moral issues altogether. We highlight the inherent ambiguity of those moral consequences, which require creators to construct an understanding of moral issues in order to respond to them. By conceptualizing creators as engaging with moral issues in alternative ways, we capture a nuanced understanding that allows for flexibility and learning over time. Additionally, while much research has emphasized the importance of positive affect for idea generation (Amabile & Pratt, 2016), our study shows how dark imagining can arise during the creative process and drive creators to tackle the additional creative challenge of embedding customized constraints into their ideas. This enables them to move forward with their ideas despite having experienced negative emotions about their possible future. We further show that even disconnecting from moral issues does not mean that creators altogether eschew such concerns; indeed, creators in our study welcomed the development of regulations that would free their generative capacity, trusting that others would develop knowledge to ensure safety. This finding builds on that of Goncalo and colleagues (2015), who found that constraining creators to act in a politically correct way could free them to engage fully in creative idea generation. Ironically, the deep desire to pursue scientific curiosity in a domain that is potentially harmful or stigmatized (e.g., Piazza & Perretti, 2015) may actually make creators more open to recognizing and embracing a broader set of limitations within the domain.

Finally, our study unpacks how constraints can become a stimulus spurring greater creativity, showing how criteria for evaluating the potential moral consequences of new ideas also act as constraints that guide their development (see also Acar et al., 2019). Whereas prior

research has recognized that constraints may fuel creativity (Goldenberg et al., 1999), our study shows how that occurs in morally-charged domains. Our work also shifts where constraints and assessments occur in the creative process; rather than an iterative process of generating ideas then assessing and refining constraints (Amabile & Pratt, 2016), we observed that considering consequences and constraining ideas could unfold continually in the background of creative work. In particular, creators can use their projections of ideas' moral consequences to constrain certain aspects of ideas as a creative stimulus, transforming the problem of potential harms into a challenge to be solved within the idea itself. Counterintuitively, this suggests that engaging with the practical issues faced by users of ideas can stimulate greater scientific curiosity. At the same time, we found that some specific insights may also become generalized across ideas within a field. Further research is needed to explore how these alternative forms of constraint work together to prevent the negative distal consequences that may result from ideas within a domain.

Practical Implications

The potential harms of AI are at the forefront of both practitioner and scholarly debate (Amabile, 2020; Berg et al., 2023; Ricker, 2023). Our study informs these debates by shedding light on the processes that may give rise to customized product constraints versus regulation. A substantial majority of the AI creators we studied were in favor of limiting AI's potential harm, but they diverged in their preferred approaches. On one hand, calls for regulation came strongly from those who emphasized the benefits AI could bring, because such regulation would allow them to generate new tools and technologies relatively uninterrupted. But from a societal perspective, it may be counterproductive to isolate the development of regulations from the core work of developing AI; whence will ideas for effective ways of controlling and constraining AI come, if not from those involved in its creation? On the other hand, creators who emphasized the

harms of AI were more active in integrating limitations directly into their technologies. But from a practical perspective, customized solutions for each AI product are likely not feasible or desirable, in that they rely on the agency and creativity of individual scientists and engineers. Further research is needed to unpack the relative effectiveness of alternative means of constraining AI. We propose, however, that a mix of both approaches is likely to be needed: product constraints integrated into specific technologies, with core insights from those constraints captured through broader guidelines and safeguards.

More broadly, it may behoove managers in various industries to be aware of the two forms of imagining illustrated here, the catalyzing surprises that encourage them, and the possibility that they may form self-reinforcing pathways during creative work. For instance, discoveries of technical surprises spark enthusiasm, signal opportunities on the horizon, and encourage disconnection from moral issues. This sequence of events may benefit the development of new and possibly pathbreaking products, but managers could still maintain procedures to remind creators of potentially negative distal consequences. In doing so, these managers may not only help to head off disasters, but they may also spur additional creativity along the way. At the same time, overemphasizing possible negative consequences may hinder the development of products and procedures with the potential to greatly benefit humankind that exist only in the minds of visionaries. For example, fears associated with inoculation procedures could have prevented the development of lifesaving vaccines. Consider the countless unpredicted breakthroughs of which nobody has yet conceived. Clearly, the distal implications of creativity are profound and far-reaching, and managers can do well to consider the roles of both dark and bright imagining.

REFERENCES

Acar, O. A., & van den Ende, J. 2016. Knowledge distance, cognitive-search processes, and creativity: The making of winning solutions in science contests. ***Psychological Science*****,** 27(5): 692-699.

Acar, O. A., Tarakci, M., & van Knippenberg, D. 2019. Creativity and innovation under constraints: A cross-disciplinary integrative review. ***Journal of Management***, 45(1): 96-121.

Acar, S., Tadik, H., Myers, D., van der Sman, C. & Uysal, R. 2021. Creativity and well-being: A meta-analysis. ***The Journal of Creative Behavior***, 55(3): 738-751.

Amabile, T. M. 1983. The social psychology of creativity: A componential conceptualization. ***Journal of Personality and Social Psychology***, 45: 357-376.

Amabile, T. M. 1988. A model of creativity and innovation in organizations. In B. M. Staw & L. L. Cummings (Eds.), ***Research in organizational behavior*** (vol. 10, pp. 123-167). Greenwich, CT: JAI.

Amabile, T. M. 2020. Creativity, artificial intelligence, and a world of surprises: Guidepost Letter for Academy of Management Discoveries. ***Academy of Management Discoveries***, 6(3): 351-354.

Amabile, T. M., & Pratt, M. G. 2016. The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. ***Research in Organizational Behavior***, 36: 157-183.

Anderson, N., Potočnik, K., & Zhou, J. 2014. Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. ***Journal of Management***, 40(5): 1297-1333.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. ***Information Fusion***, 58: 82-115.

Bamberger, P. A., & Pratt, M. G. 2010. Moving forward by looking back: Reclaiming unconventional research contexts and samples in organizational scholarship. ***Academy of Management Journal****, 53*(4), 665–671.

Bandura, A. 2002. Selective moral disengagement in the exercise of moral agency. ***Journal of Moral Education***, 31: 101-119.

Bartel, C. A., & Garud, R. 2009. The role of narratives in sustaining organizational innovation. ***Organization Science***, 20(1): 107-117.

Beane, M. 2019. Shadow learning: Building robotic surgical skill when approved means fail. ***Administrative Science Quarterly***, 64(1): 87-123.

Bennington-Castro, J. 2017. AI is a game-changer in the fight against hunger and poverty. Here's why. ***NBC News***, August 16.

Berg, J. M. 2016. Balancing on the creative highwire: Forecasting the success of novel ideas in organizations. ***Administrative Science Quarterly***, 61(3): 433-468.

Berg, J. M. 2019. When silver is gold: Forecasting the potential creativity of initial ideas. ***Organizational Behavior and Human Decision Processes***, 154: 96-117.

Berg, J. M., Duguid, M. M., Goncalo, J. A., Harrison, S. H., & Miron-Spektor, E. 2023. Escaping irony: Making research on creativity in organizations more creative. ***Organizational Behavior and Human Decision Processes***, 175: 104-235.

Bethe, H. 1968. J. Robert Oppenheimer, 1904-1967. ***Biographical Memoirs of Fellows of the Royal Society***, 14: 391-416.

Bierly, P. E., Kolodinsky, R. W., & Charette, B. J. 2009. Understanding the complex relationship between creativity and ethical ideologies. ***Journal of Business Ethics***, 86(1): 101-112.

Blair, C. S., & Mumford, M. D. 2007. Errors in idea evaluation: Preference for the unoriginal? ***The Journal of Creative Behavior***, 41(3): 197-222.

Bossman, J. 2016. Top 9 ethical issues in artificial intelligence. ***World Economic Forum***, October 21.

Byrne, C. L., Shipman, A. S., & Mumford, M. D. 2010. The effects of forecasting on creative problem-solving: An experimental study. ***Creativity Research Journal***, 22: 119-138.

Charmaz, K. 2014. ***Constructing grounded theory*** (2nd ed.). Thousand Oaks, CA: Sage.

Cho, R. 2018. Artificial intelligence—A game changer for climate change and the environment. ***Columbia Climate School***, June 5.

Clarke, R. 2019. Principles and business processes for responsible AI. ***Computer Law & Security Review***, 35(4): 410-422.

Cloutier, C., & Ravasi, D. 2021. Using tables to enhance trustworthiness in qualitative research. ***Strategic Organization***, 19(1): 113-133.

Creswell, J. W. 1998. ***Qualitative inquiry and research design: Choosing among 5 traditions***. Thousand Oaks, CA: Sage.

Creswell, J. W., & Miller, D. L. 2000. Determining validity in qualitative inquiry. ***Theory into Practice***, 39(3): 124-130.

Cronin, M. A., & Loewenstein, J. 2018. ***The craft of creativity***. Stanford, CA: Stanford University Press.

Cropley, D. H. 2010. The dark side of creativity: A differentiated model. In D. H. Cropley, A. J. Cropley, J. C. Kaufman, & M. A. Runco (Eds.), ***The dark side of creativity*** (pp. 360-374). Cambridge, UK: Cambridge University Press.

Edmondson, A. C., & McManus, S. E. 2007. Methodological fit in management field research. ***Academy of Management Review***, 32(4): 1246-1264.

Fayard, A. L., Stigliani, I., & Bechky, B. A. 2017. How nascent occupations construct a mandate: The case of service designers' ethos. ***Administrative Science Quarterly***, 62(2): 270-303.

Feldman, M., Bell, J., & Berger, M. 2003. ***Gaining access***: ***A practical and theoretical guide for qualitative researchers.*** Walnut Creek, CA: Rowman Altamira.

Fleming, P. 2019. Robots and organization studies: Why robots might not want to steal your job. ***Organization Studies***, 40(1): 23-38.

Fuchs, C., Sting, F. J., Schlickel, M., & Alexy, O. 2019. The ideator's bias: How identity-induced self-efficacy drives overestimation in employee-driven process innovation. ***Academy of Management Journal***, 62(5): 1498-1522.

George, J. M. 2007. Creativity in organizations. ***Academy of Management Annals***, 1: 439-477.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. 2013. Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. ***Organizational Research Methods***, 16(1): 15-31.

Glaser, B., & Strauss, A. 1967. ***The discovery of grounded theory: Strategies for qualitative research***. Chicago, IL: Aldine.

Goldenberg, J., Mazursky, D., & Solomon, S. 1999. The fundamental templates of quality ads. ***Marketing Science***, 18(3): 333-351.

Goncalo, J. A., Chatman, J. A., Duguid, M. M., & Kennedy, J. A. 2015. Creativity from constraint? How the political correctness norm influences creativity in mixed-sex work groups." ***Administrative Science Quarterly***, 60(1): 1-30.

Goncalo, J. A., Vincent, L. C., & Krause, V. 2015. The liberating consequences of creative work: How a creative outlet lifts the physical burden of secrecy. ***Journal of Experimental Social Psychology***, 59: 32-39.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. 2013. Moral foundations theory: The pragmatic validity of moral pluralism. ***Advances in Experimental Social Psychology***, 47: 55-130. New York: Academic Press.

Hargadon, A. B., & Douglas, Y. 2001. When innovations meet institutions: Edison and the design of the electric light. ***Administrative Science Quarterly***, 46: 476-501.

Harrison, S. H., Askin, N., & Hagtvedt, L. P. 2023. Recognition killed the radio star? Recognition orientations and sustained creativity after the Best New Artist Grammy nomination. ***Administrative Science Quarterly***, 68(1): 97-145.

Harrison, S. H., & Nurohamed, S. 2023. Dirty creativity: An inductive study of how creative workers champion new designs that are stigmatized. ***Organizational Behavior and Human Decision Processes***, 175: 104-224.

Harrison, S. H., & Rouse, E. D. 2014. Let's dance! Elastic coordination in creative group work: A qualitative study of modern dancers. ***Academy of Management Journal***, 57(5): 1256-1283.

Harrison, S. H., Rouse, E. D., Fisher, C. M., & Amabile, T. M. 2022. The turn toward creative work. ***Academy of Management Collections***, 1(1): 1-15.

Harrison, S. H., & Wagner, D. T. 2016. Spilling outside the box: The effects of individuals' creative behaviors at work on time spent with their spouses at home. ***Academy of Management Journal***, 59(3): 841-859.

Harvey, S., & Berry, J. 2023. Toward a meta-theory of creativity forms: How novelty and usefulness shape creativity. ***Academy of Management Review***, 48(3): 504-529.

Harvey, S., & Mueller, J. S. 2021. Staying alive: Toward a diverging consensus model of overcoming a bias against novelty in groups. ***Organization Science***, 32(2): 293-314.

Haselhuhn, M. P., Wong, E. M., & Ormiston, M. E. 2022. Investors respond negatively to executives' discussion of creativity. ***Organizational Behavior and Human Decision Processes***, 171: 104-155.

Hecht, B., Wilcox, L., Bigham, J. P., Schöninb, J., Hoque, E., Ernst, J., Bisk, Y., De Russis, L., Yarosh, L., Anjum, B., Contractor, D., & Wu, C. 2018. It's time to do something: Mitigating the negative impacts of computing through a change to the peer review process. ***Future of Computing Academy***.

Hecht, D. K. 2010. Imagining the bomb: Robert Oppenheimer, Nuclear Weapons, and the Assimilation of Technological Innovation. In D. H. Cropley, A. J. Cropley, J. C. Kaufman, & M. A. Runco (Eds.), ***The dark side of creativity***: 72-89. Cambridge, UK: Cambridge University Press.

Hsieh, T. C., Mensah, M. A., Pantel, J. T., Aguilar, D., Bar, O., Bavat, A., Becerra-Solano, L., Bentzen, H. B., Biskup, S., Borisov, O., & Braaten, O. 2019. PEDIA: Prioritization of exome data by image analysis. ***Genetics in Medicine***, 21: 2807-2814.

Hua, M., Harvey, S., & Rietzschel, E. F. 2022. Unpacking "ideas" in creative work: A multidisciplinary review. ***Academy of Management Annals***, 16(2): 621-656.

Huang, L., & Pearce, J. L. 2015. Managing the unknowable: The effectiveness of early-stage investor gut feel in entrepreneurial investment decisions. ***Administrative Science Quarterly***, 60(4): 634-670.

Huang, M-H., & Rust, R. T. 2018. Artificial intelligence in service. ***Journal of Service Research***, 21(2): 155-172.

Jasper, J. M. 2010. The innovation dilemma. In D. H. Cropley, A. J. Cropley, J. C. Kaufman, & M. A. Runco (Eds.), ***The dark side of creativity*** (pp. 91-113). Cambridge, UK: Cambridge University Press.

Jia, N., Luo, X., Fang, Z., & Liao, C. 2024. When and how artificial intelligence augments employee creativity. ***Academy of Management Journal***. https://doi.org/10.5465/amj.2022.0426.

Johnson, M. 1985. Imagination in moral judgment. ***Philosophy and Phenomenological Research***, 46(2): 265-280.

Jones, T. M. 1991. Ethical decision making by individuals in organizations: An issue-contingent model. ***Academy of Management Review***, 16: 366-395.

Kaplan, S., & Orlikowski, W. J. 2013. Temporal work in strategy making. ***Organization Science***, 24(4): 965-995.

Kapoor, H., & Kaufman, J. C. 2022. The evil within: The AMORAL model of dark creativity. ***Theory & Psychology***, 32(3): 467-490.

Katyal, V., Liepold, C., & Iyengar, S. 2020. Artificial intelligence and ethics: An emerging area of board oversight responsibility. ***Harvard Law School Forum on Corporate Governance***, June 25.

Keem, S., Shalley, C. E., Kim, E., & Jeong, I. 2018. Are creative individuals bad apples? A dual pathway model of unethical behavior. ***Journal of Applied Psychology***, 103(4): 416.

Kettner, N. W., Guilford, J. P., & Christensen, P. R. 1959. A factor-analytic study across the domains of reasoning, creativity, and evaluation. ***Psychological Monographs: General and Applied***, 73(9): 1-31.

Khessina, O. M., Goncalo, J. A., & Krause, V. 2018. It's time to sober up: The direct costs, side effects and long-term consequences of creativity and innovation. ***Research in Organizational Behavior***, 38: 107-135.

Kim, S. H., Vincent, L. C., & Goncalo, J. A. 2013. Outside advantage: Can social rejection fuel creative thought? ***Journal of Experimental Psychology: General***, 142(3): 605.

Kundro, T. G. 2023. The benefits and burdens of work moralization on creativity. ***Academy of Management Journal***, 66(4): 1183-1208.

Langley, A. 1999. Strategies for theorizing from process data. ***Academy of Management Review***, 24(4): 691-710.

Langley, A., & Ravasi, D. 2019. Visual artifacts as tools for analysis and theorizing. In T. B. Zilber, J. A. Amis, & J. Mair (Eds.), ***The production of managerial knowledge and organizational theory: New approaches to writing, producing and consuming theory*** (pp. 173-200). Bingley, UK: Emerald Publishing Limited.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. 2022. To engage or not to engage AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. ***Organization Science***, 33(1): 126-148.

Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D., & Smith, A. 2023. Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. ***The Lancet Digital Health***, *5*(3), e105-e106.

Lincoln, Y. S., & Guba, E. G. 1985. ***Naturalistic inquiry***. Thousand Oaks, CA: Sage.

Locke, K. 2001. ***Grounded theory in management research***. Thousand Oaks, CA: Sage.

Lonergan, D. C., Scott, G. M., & Mumford, M. D. 2004. Evaluative aspects of creative thought: Effects of appraisal and revision standards. ***Creativity Research Journal***, *16*(2-3): 231-246.

Lu, S., Bartol, K. M., Venkataramani, V., Zheng, X., & Liu, X. 2019. Pitching novel ideas to the boss: The interactive effects of employees' idea enactment and influence tactics on creativity assessment and implementation. ***Academy of Management Journal*****,** 62(2): 579-606.

Lubart, T. 2010. Cross-cultural perspectives on creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), ***The Cambridge handbook of creativity*** (pp. 265-278). Cambridge, UK: Cambridge University Press.

Mainemelis, C., & Sakellariou, E. 2022. Creativity and the arts of disguise: Switching between formal and informal channels in the evolution of creative projects. ***Organization Science***, 34(1): 380-403.

Maxwell, P. 2020. Artificial intelligence is the future of warfare (Just not in the way you think). ***Modern War Institute at West Point***, April 20.

McLaren, R. B. 1993. The dark side of creativity. ***Creativity Research Journal***, 6(1-2): 137-144.

Miles, M., & Huberman, A. M. 1994. ***Qualitative data analysis*** (2nd ed.). Thousand Oaks, CA: Sage.

Moreau, C. P., & Dahl, D. W. 2005. Designing the solution: The impact of constraints on consumers' creativity. ***Journal of Consumer Research***, 32(1): 13-22.

Mueller, J., Melwani, S., Loewenstein, J., & Deal, J. J. 2018. Reframing the decision-makers' dilemma: Towards a social context model of creative idea recognition. ***Academy of Management Journal***, 61(1): 94-110.

Mueller, J. S., Melwani, S., & Goncalo, J. A. 2012. The bias against creativity: Why people desire but reject creative ideas. ***Psychological Science***, 23(1): 13-17.

Mueller, J. S., Wakslak, C. J., & Krishnan, V. 2014. Construing creativity: The how and why of recognizing creative ideas. ***Journal of Experimental Social Psychology***, 51: 81-87.

Mumford, M. D., Waples, E. P., Antes, A. L., Brown, R. P., Connelly, S., Murphy, S. T., & Devenport, L. D. 2010. Creativity and ethics: The relationship of creative and ethical problem-solving. ***Creativity Research Journal***, 22(1): 74-89.

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

Neri, E., Coppola, F., Miele, V., Bibbolino, C., & Grassi, R. 2020. Artificial intelligence: Who is responsible for the diagnosis? ***La Radiologia Medica***, 125: 517-521.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. ***Science***, 366(6464): 447-453.

Perry-Smith, J. E., & Mannucci, P. V. 2017. From creativity to innovation: The social network drivers of the four phases of the idea journey. ***Academy of Management Review***, 42(1): 53-79.

Petriglieri, G., Ashford, S. J., & Wrzesniewski, A. 2019. Agony and ecstasy in the gig economy: Cultivating holding environments for precarious and personalized work identities. ***Administrative Science Quarterly***, 64(1): 124-170.

Piazza, A., & Perretti, F. 2015. Categorical stigma and firm disengagement: Nuclear power generation in the United States, 1970-2000. ***Organization Science***, 26(3), 724-742.

Pratt, M. G., Sonenshein, S., & Feldman, M. S. 2022. Moving beyond templates: A bricolage approach to conducting trustworthy qualitative research. ***Organizational Research Methods***, 25(2): 211-238.

Rainie, L., Anderson, J., & Vogels, E. A. 2021. Experts doubt ethical AI design will be broadly adopted as the norm within the next decade: 1. Worries about developments in AI. ***Pew Research Center***, June 16.

Ravasi, D., & Stigliani, I. 2012. Product design: A review and research agenda for management studies. ***International Journal of Management Reviews***, 14(4): 464-488.

Reinecke, J., & Ansari, S. 2015. What is a "fair" price? Ethics as sensemaking. ***Organization Science***, 26(3): 867-888.

Richards, R. 2010. Everyday creativity: Process and way of life—four key issues. In J. C. Kaufman & R. J. Sternberg (Eds.), ***The Cambridge handbook of creativity*** (pp. 189-215). Cambridge, UK: Cambridge University Press.

Ricker, T. 2023. 'Godfather of AI' quits Google with regrets and fears about his life's work. ***The Verge***, May 1.

Rosso, B. D. 2014. Creativity and constraints: Exploring the role of constraints in the creative processes of research and development teams. ***Organization Studies***, 35(4): 551-585.

Runco, M. A. 1993. Creative morality: Intentional and unconventional. ***Creativity Research Journal***, 6(1-2): 17-28.

Runco, M. A. 2010. Creativity has no dark side. In D. H. Cropley, A. J. Cropley, J. C. Kaufman, & M. A. Runco (Eds.), ***The dark side of creativity*** (pp. 15-32). New York, NY: Cambridge University Press.

Sandberg, J., & Tsoukas, H. 2015. Making sense of the sensemaking perspective: Its constituents, limitations, and opportunities for further development. ***Journal of Organizational Behavior***, 36(S1): S6-S32.

Semuels, A. 2020. Millions of Americans have lost jobs in the pandemic—and robots and AI are replacing them faster than ever. ***Time Magazine***, August 6.

Simonton, D. K. 1984. Artistic creativity and interpersonal relationships across and within generations. ***Journal of Personality and Social Psychology***, 46(6): 1273-1286.

Sonenshein, S. 2007. The role of construction, intuition, and justification in responding to ethical issues at work: The sensemaking-intuition model. ***Academy of Management Review***, 32(4): 1022-1040.

Spradley, J. 1979. ***The ethnographic interview***. New York: Holt, Rinehart & Winston.

Stahl, B. C. 2021. Ethical issues of AI. In ***SpringerBriefs in research and innovation governance: Artificial intelligence for a better future***: 19-33. New York: Springer.

Stigliani, I., & Ravasi, D. 2018. The shaping of form: Exploring designers' use of aesthetic knowledge. ***Organization Studies***, 39(5-6): 747-784.

Strauss, A., & Corbin, J. 1998. ***Basics of qualitative research: Techniques and procedures for developing grounded theory*** (2nd ed.). Thousand Oaks, CA: Sage.

Thompson, N. A. 2018. Imagination and creativity in organizations. ***Organization Studies***, 39(2-3): 229-250.

Thompson, N. A., & Byrne, O. 2022. Imagining futures: Theorizing the practical knowledge of future-making. ***Organization Studies***, 43(2): 247-268.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. ***Psychological Review***, 117(2): 440-463.

Vaara, E., Sonenshein, S., & Boje, D. 2016. Narratives as sources of stability and change in organizations: Approaches and directions for future research. ***Academy of Management Annals***, 10(1): 495-560.

Venkataramani, V., Derfler-Rozin, R., Liu, X., & Mao, J. Y. 2023. Keep off my turf! Low-status managers' territoriality as a response to employees' novel ideas. ***Organization Science***. https://doi.org/10.1287/orsc.2021.15132.

Vincent, J. 2019. The problem with AI ethics: Is Big Tech's embrace of AI ethics boards actually helping anyone? ***The Verge***, April 3.

Vincent, L. C., & Goncalo, J. A. 2014. License to steal: How the creative identity entitles dishonesty. In S. Moran, D. Cropley, & J. C. Kaufman (Eds.), ***The ethics of creativity*** (pp. 137-151). Basingstoke, United Kingdom: Palgrave Macmillan.

Vincent, L. C., & Kouchaki, M. 2016. Creative, rare, entitled, and dishonest: How commonality of creativity in one's group decreases an individual's entitlement and dishonesty. ***Academy of Management Journal***, 59(4): 1451-1473.

Wearn, O. R., Freeman, R., & Jacoby, D. M. P. 2019. Responsible AI for conservation. ***Nature Machine Intelligence***, 1: 72-73.

Weick, K. E. 1995. ***Sensemaking in Organizations***. Sage: Thousand Oaks, CA.

Wenzel, M., Krämer, H., Koch, J., & Reckwitz, A. 2020. Future and organization studies: On the rediscovery of a problematic temporal category in organizations. ***Organization Studies,*** 41(10): 1441-1455.

Whitaker, B. G., & Godwin, L. N. 2013. The antecedents of moral imagination in the workplace: A social cognitive theory perspective. ***Journal of Business Ethics***, 114: 61-73.

Whyte, J., Comi, A., & Mosca, L. 2022. Making futures that matter: Future making, online working and organizing remotely. ***Organization Theory***, 3: 1-20.

Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. 1954. A factor-analytic study of creative-thinking abilities. ***Psychometrika***, 19(4): 297-311.

Wolcott, H. F. 1994. ***Transforming qualitative data: Description, analysis, and interpretation***. Thousand Oaks, CA: Sage.

Woodman, R. W., Sawyer, J. E., & Griffin, R. W. 1993. Toward a theory of organizational creativity. ***Academy of Management Review***, 18(2): 293-321.

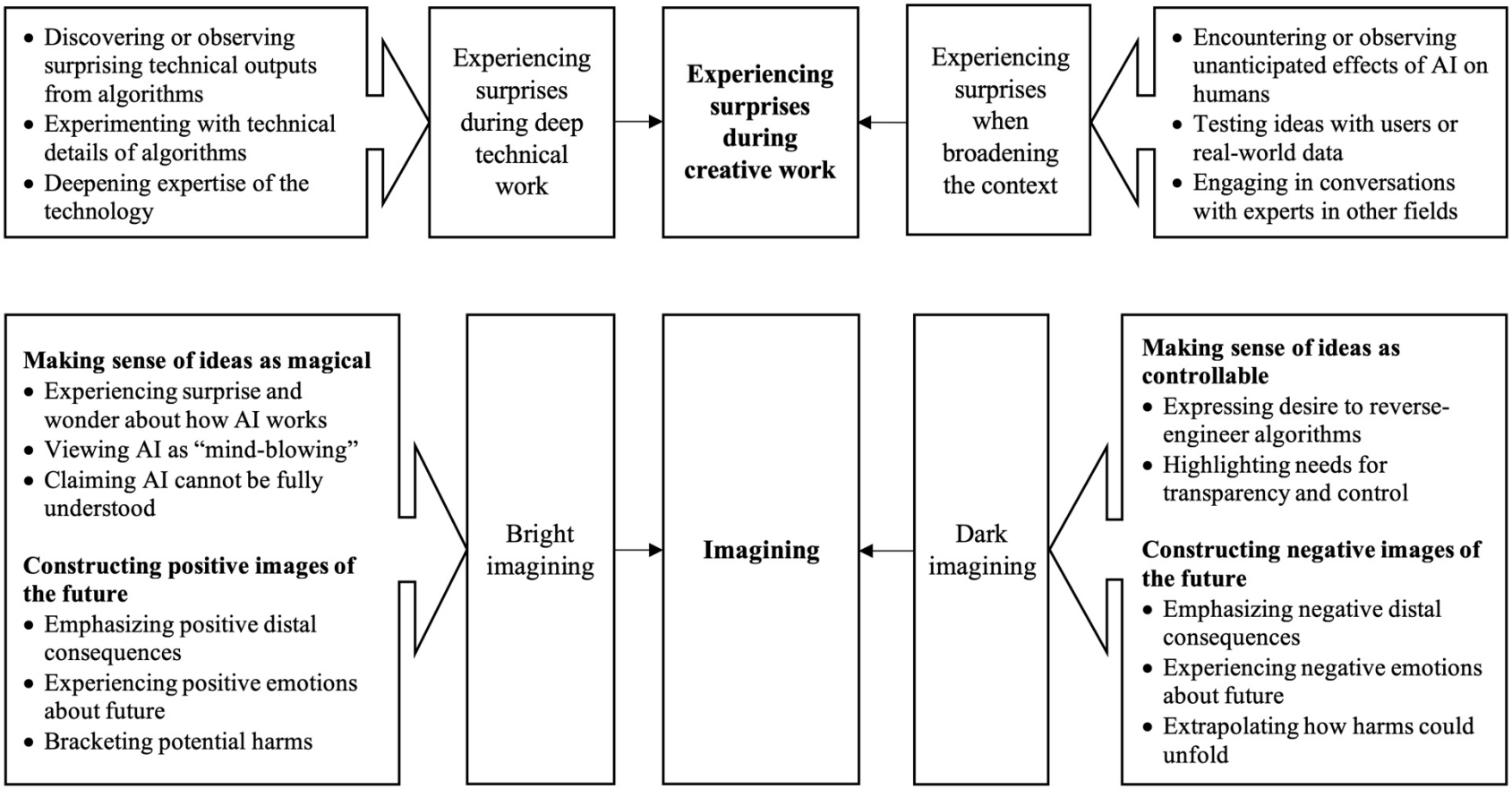
Yin, R. K. 2009. ***Case study research: Design and methods*** (4th ed.). Thousand Oaks, CA: Sage.

Yong, K., Mannucci, P.V., & Lander, M. W. 2020. Fostering creativity across countries: The moderating effect of cultural bundles on creativity. ***Organizational Behavior and Human Decision Processes***, 157: 1-45.

Zhou, J., Wang, X. M., Song, L. J., & Wu, J. 2017. Is it new? Personal and contextual influences on perceptions of novelty and creativity. ***Journal of Applied Psychology***, 102(2): 180-202.

Zhou, J., Wang, X. M., Bavato, D., Tasselli, S., & Wu, J. 2019. Understanding the receiving side of creativity: A multidisciplinary review and implications for management research. ***Journal of Management***, 45(6): 2570-2595.

FIGURE 1: Data Structure



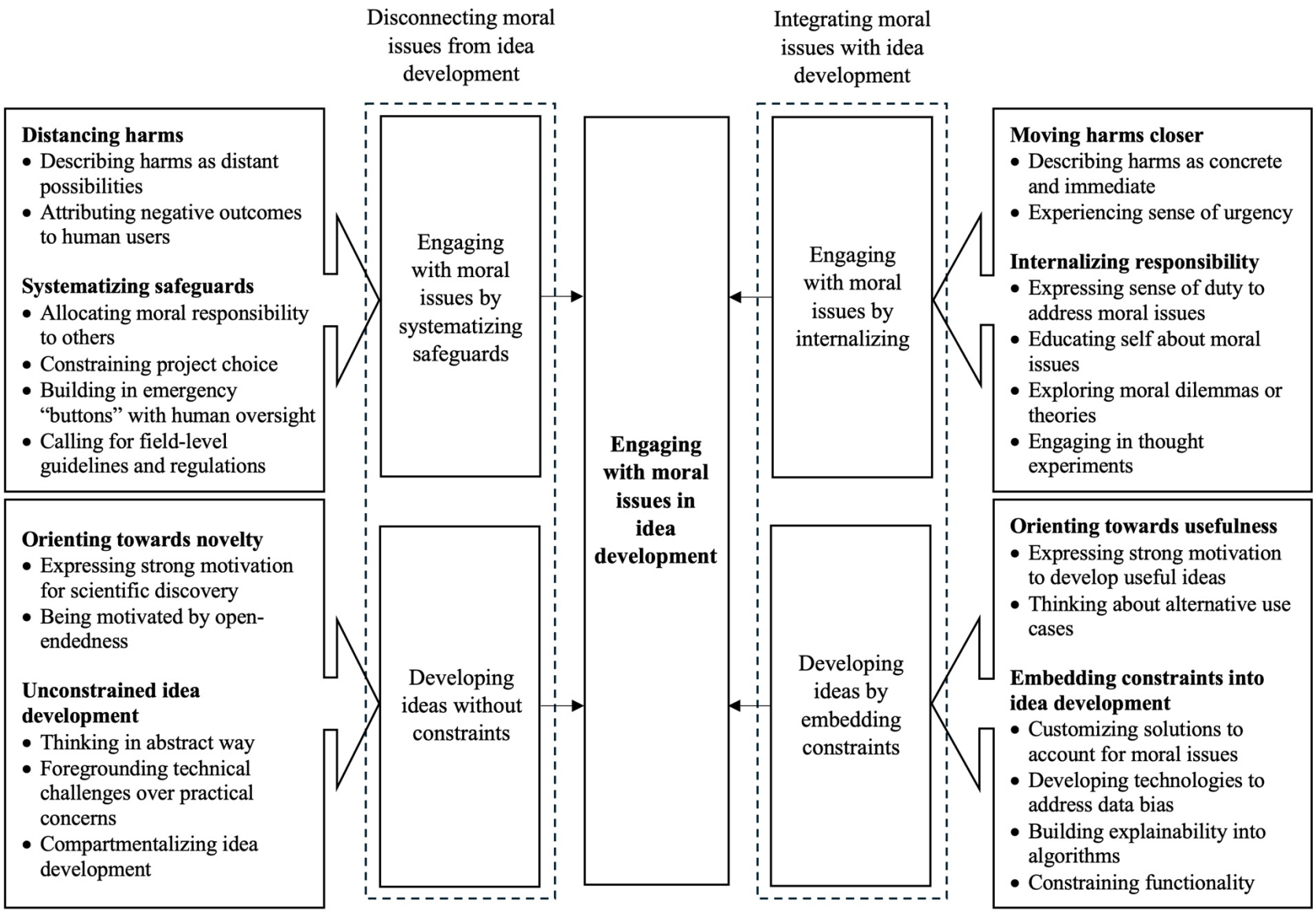

Disconnecting moral issues from idea development
Integrating moral issues with idea development
Distancing harms
• Describing harms as distant possibilities
• Attributing negative outcomes to human users
Systematizing safeguards
• Allocating moral responsibility to others
• Constraining project choice
• Building in emergency "buttons" with human oversight
• Calling for field-level guidelines and regulations
Engaging with moral issues by systematizing safeguards
Moving harms closer
• Describing harms as concrete and immediate
• Experiencing sense of urgency
Internalizing responsibility
• Expressing sense of duty to address moral issues
• Educating self about moral issues
• Exploring moral dilemmas or theories
• Engaging in thought experiments
Engaging with moral issues by internalizing
Engaging with moral issues in idea development
Orienting towards novelty
• Expressing strong motivation for scientific discovery
• Being motivated by open-endedness
Unconstrained idea development
• Thinking in abstract way
• Foregrounding technical challenges over practical concerns
• Compartmentalizing idea development
Developing ideas without constraints
Orienting towards usefulness
• Expressing strong motivation to develop useful ideas
• Thinking about alternative use cases
Embedding constraints into idea development
• Customizing solutions to account for moral issues
• Developing technologies to address data bias
• Building explainability into algorithms
• Constraining functionality
Developing ideas by embedding constraints

FIGURE 2: Emergent Model of Navigating Moral Issues in Idea Development

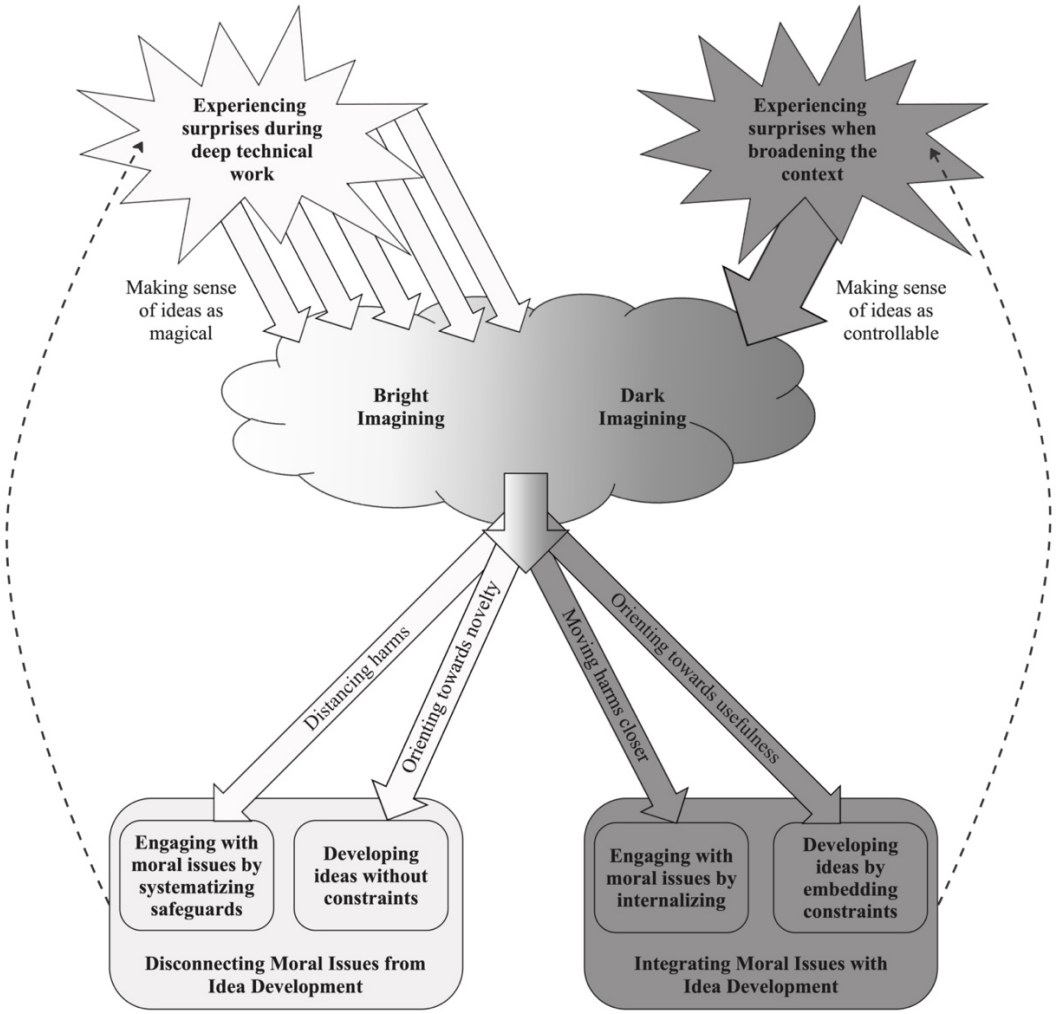


TABLE 1: Illustrative Comments from Informants

EXPERIENCING SURPRISES DURING CREATIVE WORK	
Experiencing surprises during deep technical work	***Experiencing surprises when broadening the context***
You can spend a lot of time trying to come up with specific building blocks that should be important for the solution. But evolution on some problems can find better, more robust solutions than you thought about. This amazed me many times. Specific situations? …I mean many, many times. (Nelson)	I actually had a number of totally independent, really super interesting conversations with Uber drivers at these conferences going from one place to another about their anxieties of self-driving cars and stuff. …What struck me is just the genuine sort of sense of anxiety that these people are feeling. Like whether or not their fears will actually come to pass is not perhaps as important as the fact that it really is stressing a lot of people out. And in that sense, I felt like one of the other ways—or the broader scale way I can perhaps try and make this work better is by advocating where I can for sensible rules on how automation can be used to avoid exacerbating inequality. (Gabriel)
You will get surprises. I mean that's what AI scientists live for. I can tell you that's part of the joy of working in AI, is when your system actually produces something that you did not expect, something smart and interesting. (Sam)	So, in my case, for instance, I was doing sound recognition on robots, and I eventually had pushed all the software and code online, as an open-source library, and then I got people from everywhere in the world to use it, with other robots and with different hardware. And then it would come back with feedback and say, "Oh, you know, like, I'm facing this issue, here." And then, "Okay. Interesting. Send me the recording. I'm gonna look at this." (Randy)
We want systems that do more for us than what we put in. And it happens all the time. In just about every paper, there is something like that. It actually has become more or less a standard. If you write a paper about a system, if it does what you set out to do, it's not really interesting. The interesting piece is, what else does it do that you did not actually put in? And every good paper has a surprise result like that. I can talk about that for hours. (Ray)	When I was working with [Game Design AI], which was the main focus of my PhD thesis, I had an early prototype which just let the designer create a game level, in painstaking detail, which took about 30 minutes. And then the AI would go in and, you know, change everything. And you could not really undo it. You could just select among options that it gave you. And users felt completely—well, two things. Felt like they did not have control—so, that's important. But also, they were so invested in what they did that they were very prone on basically fussing about even the smallest detail. And the AI, to be fair, it wasn't particularly smart. So, it obviously made mistakes. (Jordan)

IMAGINING			
Bright imagining		***Dark imagining***	
Making sense of ideas as magical	There are a lot of these settings that going into training a neural network, and one of them is this step size or learning rate. And it's sort of a black box. No one knows what a good learning rate is for a given neural network and a given data set, and you train them together. Some people have some guesses, but it's sort of black magic, and there's no equation that says, this is the best learning rate to use. (Jack)	*Making sense of ideas as controllable*	Some of the concerns with AI safety are very practical, about making sure that algorithms do what we expect them to do. So like kind of actually curtailing some of the surprise. …I think a good scientist will deliberately try to peel away at least some things to validate that they actually understand what's happening. …So it's like coming to a deeper understanding about the system that you're designing, which can be related to that black box nature, but it's more often related to how the black box is interacting with the greater apparatus that you do actually have control over—the incentive, the particular knobs you're twisting, how the experiment is set up, the experimental design that you're testing it within. …You understand all those things and how they relate to setting the weights of the black box or whatever. I guess that's how I think about it. (Adam)
	Sometimes, when you get something unexpected out of your experiment, it's a bit magical when something you did not expect [occurs]. You run your code and you see a pattern occurring. And this is quite—you say, "Okay, I did not think about it. Amazing!" (Dylan)		If you get into the black box, I think, somebody has developed it, right? There is not a black box, actually. So I think, as a technical person, I have been working on AI things, and I try to understand the technical details inside that, not considering that as a black box. (George)
	"There is some math in it, so you have to scale the HPs (hyperparameters) according to whatever you do. But I think in large part it's mostly like black magic, random tries—you start with an interval and then try to move up and down around that interval, and whatever works. I think this is quite representative of how the deep learning community works in terms of HP tuning at this point." (Andrea)		The first time I got an algorithm working that really did a great job, I was like, "Wow, this is incredible." But at the same time, it evoked a little skepticism and suspicion—I'm like trying to think, how is this actually working? What does it really know? So I'd go and try to devise tricky tests for it to see where it would do well and where it would fail. So it was sort of a skeptical curiosity. Once I got past that "Holy shit, this is awesome, this works" moment, it's like, "Well, does it really?" That was probably the strongest reaction that it evoked in me, is really, "What's going on here?" (Justin)

Constructing positive images of the future	AI will be our companion, our digital sibling. Living on the other side of the digital looking glass, it will make the digital universe accessible to us. By doing so, it will help make both our physical and digital lives easier and more manageable. (John)	*Constructing negative images of the future*	These language models are gonna pick up on a lot of bias that we write, now, as humans. Not we, like you or me, but just anything on the Internet, if your entire existence and all you were and everything you ever knew was only the things written on the Internet, maybe if you think about it like that, maybe these models are gonna be amplifying a lot of the things that we already don't like about certain things we do. And that's a real problem. (Harry)
	I'm very excited about all this collective decision making, collective decision aiding, so helping people to fairly divide resources, to fairly exchange goods. …So from issues like also democracy, how to get people involved in the democratic process, new ways of voting, maybe different ways to maybe re-fund our societies, public and city engagement. So all of that is very exciting to me. I see an amazing potential. …It could be used in very nice ways. (Dylan)		What's going to happen is that whoever owns a supercomputer, and whoever has a team of 20 ML experts, can now produce fakes of the most famous celebrities or presidents doing whatever they want, and it's impossible to prove that that's not an original. So, what does this do? This destabilizes news, politics right? Well, some people are going to say, "Well, I won't trust anything I see on the news at all, because…" It just breaks down all kinds of things in society. (Jack)
ENGAGING WITH MORAL ISSUES IN IDEA DEVELOPMENT			
Disconnecting moral issues from idea development		**Integrating moral issues with idea development**	
Engaging with moral issues by systematizing safeguards		***Engaging with moral issues by internalizing***	
Distancing harms	I feel it's not going to be this sudden, suddenly AI takes off and takes over the world. I think there's a fairly gradual process with a lot of I guess intermediate episodes where we can get some additional experience with what AGI might look like and get more informed of how we need to deal with it. (James)	*Moving harms closer*	A different project we worked on which I think has more immediate potential for causing harm is—so some of this earlier work was focused on fooling image classifiers… But another question might be, how do you attack an online system? Say somebody has a web service where they'll classify images for you; is it possible to attack such a system? …These kinds of image classification services are being used on web forums to prevent certain types of, say, bad images from being posted. …Well, now you can get a picture of drugs and just manipulate it in a certain way so it looks like a puppy; and oh, the web service needs to post a picture of a puppy. (William)

	Who knows what could happen a long time from now? But the quote I like is, "It's like worrying about overpopulation on Mars." It's such a far-off problem that sure, it could be an issue one day, but there's so many things in between that it's not on the forefront of what I'm worried about. (Doug)		There's that kind of long-term risk side of the story. But then there is I think a much more serious and much more concrete issue of bias in machine learning systems which we definitely need to overcome. (Sean)
Systematizing Safeguards	There are a whole lot of other people who are worrying about the regulations, the security, about fairness, but they are in different and separate research areas. I do agree that these topics should be taken care of, but it's probably by someone else. (Bob)	*Internalizing responsibility*	You know, I've spent a lot of time trying to work on making these things better—you know, we have to think about the flipside of what are those impacts and second-order effects. So, safety is very concerned with the algorithmic question of like, "Well, we have these tools that we're trying to build to do specific things, but how do we actually make sure that they're doing what we want?" (Andrew)
	The way I typically choose the applied field is looking for something that I think is a positive contribution to society. …Certainly the choices to pick applied projects that are inherently and undoubtedly good for society, I hope that I would have done already, but certainly having a constant reminder of the pros-cons list I think definitely helps nudge things in that direction. (Doug)		We are moving on the right direction by increasing the awareness that the more AI is capable of doing things, the more we need to be cautious about controlling what it does. …So, what I'm trying to say is that we need to be a bit more careful, in general, about these things in our studies. And, of course, we need to be affected, more than we do. We should be doing a better job with data biases, including myself. (Steven)

Developing ideas without constraints		***Developing ideas by embedding constraints***	
Orienting towards novelty	The research I'm interested in is much more open-ended and much more exploratory, and less defined by absolute perfection or optimization. That is not what drives me. Open-endedness or an algorithm that can produce unexpected results is actually the thing that I'm interested in. (Thomas)	*Orienting towards usefulness*	A lot of my work goes into the support work surrounding the creation of machine learning models, so things like, how do you host these machine learning models? How do you make available to end users? How do end users interact with machine learning models? What interfaces make sense, like how do we make it intuitive to an end user, how to train and iterate on a machine learning model, without having a lot of machine learning expertise? It's that sort of task that takes up a lot of my time. (Richard)

	Just doing something that you have never done before. So after a while, you need to be creative, because you have exhausted all the simple things to do. You start doing complicated things, and then there is nothing new with these complicated things, so you start doing even more complicated things. So that's one way of trying to search for new things without having to think about the goal. …Which might be good, but in the name of practical applications, it's actually very hard to use this for practical applications. (Daniel)		I try to work on stuff that's relevant, where I can kind of see that it's important that people understand the answer to this question. Or I see that this is an important problem that needs to be solved and the solution might actually be used by people in the real world, maybe not immediately but maybe in the near future. …We're still in the pretty early stages of figuring out what this machine learning stuff is useful for. …I try to make things that are useful to people. (William)
Unconstrained idea development	You just abstractly think about, well if I had a data set that looked like this, like maybe a sequence of numbers over time, and then I wanted to compute this sort of thing, like find an anomaly in that sequence, what would I do? What would be the best way—what would be an efficient algorithm for doing that? And that's kind of its own style of work because you never necessarily touch any real data. It's just kind of thinking abstractly about what the problems are and what's the most efficient way or best way to solve them according to some metric you've defined. (Colton)	*Embedding*	We can, say, collect all the data from YouTube that were recorded by a lot of people with different phones and different environment, with a lot of different noise, and we sort of get a representative sampling of a wide range of test cases. And then, once we have that, we can train a model that is much more representative. If you have recordings for millions of people, with a different accent and different environments, and then you retrain your model and try to get something more and more accurate, then, eventually, you'll get something that covers most cases. (Randy)
	Right now I think AI is only progressing because of researchers, so if they start to feel threatened about these kind of scenarios (about who would be blamed if an AI went wrong), there will be no progress. (Andrea)		One good way to do so which I was trying to emphasize is by making the robot unusable in a bad way in any sense. You make people more comfortable to be around it. And this is one way to counter the critics and commands about, "Okay, this robot is just pushing objects, but you put a gun in his hand, and he would be a killing robot." (Kyle)

TABLE 2: Extended Examples of Each Approach

Disconnecting through Bright Imagining	Integrating through Dark Imagining
Ryan, founding member of a navigational AI lab, exemplified the disconnecting approach. After discovering the surprise that AI models "could do something crazy" in "edge cases" that humans would handle well (e.g., getting out of the way if a horse flew out of the back of a car in front of you), he concluded that AI was too complicated for anyone to fully understand. Despite being concerned about the ways AI could go wrong, he imagined a future of "artificial systems proliferating with cool, interesting stuff forever." He disconnected moral issues from his idea development suggesting that others should specialize in developing safeguards for ethical issues, claiming, "It is just not my job. My job is to do basic research."	Samantha, who was building a speech interface for an "intelligent nutritionist" tool to be used in a medical study, illustrates the integrating approach to moral issues. Early in her career, she observed the effects on people experiencing food issues of receiving highly critical reviews of their food diaries. This prompted Samantha to think expansively about the potential negative consequences of her ideas. She expressed concern for both the risks of AI providing incorrect nutritional advice and more distal risks regarding the privacy of patients' data. Samantha took on these moral issues directly, educating herself about healthcare and good nutritional practice and working with experts to embed that knowledge directly into her technology. She described this as having "an expert somewhere in the pipeline," rather than the human oversight approach involving an off-switch in the event of a negative outcome.
James, researcher working on face recognition systems, illustrates the disconnecting approach. He described "one of the most fun things about doing research in AI" as "the black box nature—that sometimes you get solutions that you didn't expect." Surprises he encountered during deep technical work prompted him to see AI solutions as not fully understandable. James distanced any potential harms AI could have by describing them as "sci-fi" and being "very far away." Instead, he imagined a brighter future where conversational AI and 3D printed robots assist humans. He disconnected moral issues from his idea development by keeping moral issues at "the back of [his] mind" yet not allowing them to "affect [his] day-to-day work." He argued that moral issues such as job replacement are "political questions" suggesting others should come up with "rules and regulations for those who may be affected by these advancements."	Joe, who was building a digital tracker for trucks, exemplified the integrating approach to moral issues. As he broadened his focus from technical issues to the context, he started thinking about how users of this system, including the truck drivers, who generate the data yet have little control over its use, engage with his digital tracker. He realized that "the truck drivers would be happy if they did not have this system which helps their boss to know what they are doing." This prompted him to imagine an even darker future for them where "trucks will become autonomous." which he described as "the worst news for them." He emphasized the urgency of the issue by saying "it's not the future, it's today, where truck drivers are more controlled." Joe accepted the responsibility he had as an AI creator and educated himself about the link between data analytics and society by attending local meetups. During idea development, Joe used real-life databases and worked closely with customers, constantly "adapt[ing] to the changing requirements."

APPENDIX A: Overview of Informants

	Code Name	*Organization*	*AI Application Domain*	*Role/Title*	*Years in AI*
			Executive Interviews		
1	Alice	IT & Consulting Company	IT/Work processes	Senior VP of Future of Work	N/A
2	Elizabeth	Marketing AI Startup	Marketing & e-commerce	Director of Marketing	N/A
3	Lance	Marketing AI Startup	Marketing & e-commerce	Chief Revenue Officer	N/A
4	Alex	Marketing AI Startup	Marketing & e-commerce	CEO	N/A
5	Brandon	IoT Startup	Robotics	Co-founder	N/A
			Exploratory AI Creator Interviews		
6	Eric	Marketing AI Startup	Marketing & e-commerce	Data scientists	8
7	Evan	Marketing AI Startup	Marketing & e-commerce	Software Engineer	13
8	Charles	Enterprise Process Automation Startup	IT/Work processes	Software Engineer	5
9	Richard	Enterprise Process Automation Startup	IT/Work processes	Software Engineer	5
10	Christopher	Enterprise Process Automation Startup	Transportation & navigation	Software Engineer	4
			AI Creator Interviews		
11	Brad	IT & Consulting Company	IT/Work processes	Head of R&D	7
12	Marie	IT & Consulting Company	IT/Work processes	Researcher	9
13	Joe	IT & Consulting Company	Transportation & navigation	Data Scientist	18
14	Bob	University	Healthcare	Professor	6
15	John	University	IT/Work processes	Professor	27
16	Nelson	University	Healthcare	Professor	27
17	Jack	Supercomputer Manufacturer	Digital content generation	Software Engineer	8
18	Barry	Supercomputer Manufacturer	Healthcare	Software Engineer	3
19	Kevin	Supercomputer Manufacturer	IT/Work processes	Software Engineer	2
20	Jim	Supercomputer Manufacturer	IT/Work processes	Software Engineer	4
21	Andrea	Supercomputer Manufacturer	IT/Work processes	Software Engineer	4
22	Nick	Supercomputer Manufacturer	IT/Work processes	Research Engineer	14
23	Ryan	Transportation Network Company	Transportation & navigation	Research Scientist	21
24	Adam	Transportation Network Company	Transportation & navigation	Research Scientist	10
25	James	Transportation Network Company	Transportation & navigation	Research Scientist	13
26	Jenny	Transportation Network Company	Transportation & navigation	Research Scientist	10
27	Thomas	Transportation Network Company	Transportation & navigation	Research Scientist	8
28	Daniel	Research Institute	Robotics	Research Scientist	14
29	Matthew	University	Healthcare	Software Engineer and Adjunct Professor	3
30	Mark	University	Game content generation	Professor	17
31	Andrew	Research Institute	Image classification	Research Scientist	5

32	Kenneth	University	Robotics	Postdoctoral Researcher	10
33	Josh	University	Healthcare	Postdoctoral Researcher	10
34	George	University	Healthcare	Postdoctoral Researcher	11
35	Brian	Virtual Assistant Startup	Virtual assistants	Software Engineer	32
36	Jacob	Web Design Startup	Web design	Software Engineer	1
37	Gary	University	Digital content classification	PhD Candidate	6
38	Justin	AI Smart Home Startup	Smart-home consumer product	Data Scientist	6
39	Frank	Marketing & E-commerce Startup	Marketing & e-commerce	Software Engineer	10
40	Colton	Healthcare Startup	Healthcare	Software Engineer	4
41	Greg	AI Website optimization Startup	Web design	Software Engineer	5
42	Sam	IT Services & Consulting Company	Marketing & e-commerce	Data Scientist	30
43	Ray	University	Healthcare	Professor	32
44	Dennis	University	Robotics	PhD Candidate	5
45	Ellen	University	Robotics	Professor	18
46	Henry	University	Optical systems	Professor	20
47	Doug	University	Agriculture	Professor	6
48	Nathan	University	Healthcare	Professor	10
49	Zack	University	Robotics	Professor	18
50	Steven	University	Game content generation	Professor	13
51	Kyle	University	Robotics	PhD Candidate	5
52	Jeremy	Global Automative Supplier	Transportation & navigation	Software Engineer	9
53	Sean	Social Media & Networking Company	Digital content classification	Research Scientist	11
54	Christian	University	Transportation & navigation	Professor	40
55	Noah	University	Healthcare	Professor	6
56	Jordan	University	Game content generation	Professor	14
57	Dylan	University	Resource allocation	Professor	20
58	Alan	University	Game content generation	Postdoctoral Researcher	8
59	Amanda	University	Healthcare	PhD Candidate	6
60	Joy	University	Game content generation	PhD Candidate	7
61	Samantha	University	Healthcare	PhD Candidate	7
62	Bruce	University	Robotics	Professor	15
63	Randy	University	Auditory processes	Postdoctoral Researcher	9
64	Eve	Enterprise Software Company	Marketing & e-commerce	Research Engineer	6
65	Caleb	Enterprise Software Company	Virtual assistants	Research Scientist	6
66	Gabriel	Enterprise Software Company	Virtual assistants	Research Scientist	2
67	Harry	Enterprise Software Company	Digital translation	Research Scientist	7
68	William	University	IT/work processes	PhD Candidate	3
69	Logan	University	Digital translation	Postdoctoral Researcher	5

APPENDIX B: Semi-Structured Interview Protocol

A. Background/Basic Information

1. What is your educational background?
2. What is your professional background?
3. How long have you been at [organization]?
4. What is your role at [organization]?

B. Daily Work Experiences, Motivations, and Problem Solving

1. What initially led you to pursue work in machine learning and artificial intelligence?
2. What is (are) your current focal project(s)?
3. How do you decide which projects to pursue?
4. What motivates you or drives you in your work?
5. What is your typical process for developing a new machine learning model?
6. Think of a recent time when you had to work through a challenge. How did you work through it?
7. What role does the AI community play in these processes, if any?

C. Potential Outcomes

1. As you're doing your work, what sorts of future applications or ultimate uses do you have in mind? What do you see your work building toward?
2. Do you ever think of more downstream outcomes that might unfold? If so, what types of outcomes do you think about? If not, why not?
    a. [Prompt if needed] For instance, imagine that your creation is very successful and widely adopted. What effects do you think that will have?
    b. [If negative outcomes discussed] What can be done about these issues, if anything?
        i. Does your work play a role in this process? If so, how? If not, why not?
    c. Have you always held these views? If so, why? If not, what made you aware of these issues?
3. How do these considerations affect you and your work, if at all? If they don't, why not?
    a. [Prompt if needed] For instance, are you changing aspects of your work based on these factors? If so, in what ways? If not, why not?
        i. Did any particular experiences lead you to approach your work in this way? If so, please describe them.
    b. [If constraints described] How did you think about your work when you weren't implementing these sorts of boundaries?
    c. [If constraints described] How do you think about your work as a result of placing these sorts of boundaries on what you're doing?

D. Broader Future of AI

1. What do you see as the future of AI in general?
    a. [Prompt if needed] How do you think society will be affected, if at all?
2. Do you think that we should have any particular concerns about how AI might develop? If so, please describe them. If not, why not?
    a. [If yes] What do you think should be done about these issues, if anything?
        i. How do you see this relating to your own work, if at all?
3. Do your views on the future of AI affect you or your work in any way? If so, how? If not, why not?
    a. [Prompt if needed] Have you changed anything about your work based on these views? If so, how? If not, why not?
4. Have you always thought this way about the future of AI? If so, why? If not, what changed?

E. Closing

Is there anything else related to these topics that you'd like to mention?

APPENDIX C: Percentage of Informants by Organizational Setting, Experience, and Job Role

		Disconnecting through Bright Imagining	*Integrating through Dark Imagining*	*Mix of Orientations*
Years of Experience in AI	Less than 10	66%	60%	33%
	10 or more	34%	40%	67%
Organizational Setting	Academic	48%	70%	27%
	Industry	52%	30%	73%
Job Role	Research	66%	75%	60%
	Engineer	34%	25%	40%

Note. Each pair of rows within each column represents the total group of creators classified as either disconnecting, integrating, or a mix of both.

APPENDIX D: Coding of Data Structure by Informant

#	Code Name	Experiencing surprises during deep technical work	Bright Imagining	Disconnecting moral issues from idea development	Experiencing surprises when broadening the context	Dark imagining	Integrating moral issues with idea development
1-5	Alice, Elizabeth, Lance, Alex, and Brandon: Executive Interviews						
6	Nick	✓	✓	✓	•	•	•
7	Gary	✓	✓	✓	•	•	•
8	Colton	✓	✓	✓	•	✓	•
9	James	✓	✓	✓	✓	✓	•
10	Jim	✓	✓	✓	✓	✓	•
11	Kenneth	✓	✓	✓	✓	✓	•
12	Doug	✓	✓	✓	✓	✓	•
13	Andrea	✓	✓	✓	•	✓	•
14	Jacob	✓	✓	✓	•	•	•
15	John	✓	✓	✓	✓	✓	•
16	Marie	✓	✓	✓	•	✓	•
17	Bob	✓	✓	✓	✓	✓	•
18	Jack	✓	✓	✓	•	✓	✓
19	Daniel	✓	✓	✓	•	✓	•
20	Thomas	✓	✓	✓	✓	✓	•
21	Matthew	✓	✓	✓	✓	✓	•
22	Dennis	✓	✓	✓	✓	✓	•
23	Barry	✓	✓	✓	✓	✓	•
24	Bruce	✓	✓	✓	•	✓	•
25	Zack	✓	✓	✓	•	✓	✓
26	Evan	•	✓	✓	•	•	•
27	Nathan	✓	✓	✓	•	✓	•
28	Noah	•	✓	✓	•	•	•
29	Eve	•	✓	✓	✓	•	•
30	Caleb	•	✓	✓	•	•	✓
31	Charles	✓	✓	✓	✓	•	✓
32	Nelson	✓	✓	✓	•	✓	✓
33	Amanda	✓	✓	✓	•	✓	✓
34	Brad	✓	✓	•	✓	•	•
35	Adam	✓	✓	✓	✓	✓	✓
36	Ellen	✓	✓	✓	✓	✓	✓
37	Richard	✓	✓	✓	✓	✓	✓
38	Dylan	✓	✓	✓	✓	✓	✓
39	Christopher	✓	✓	✓	✓	✓	✓
40	Sam	✓	✓	✓	✓	✓	✓
41	Mark	✓	✓	✓	✓	✓	✓
42	Harry	•	✓	✓	•	✓	✓

43	Jenny	✓	✓	•	✓	✓	•
44	Eric	✓	✓	•	•	✓	✓
45	Sean	•	✓	✓	•	✓	✓
46	Ray	✓	✓	✓	✓	✓	✓
47	Jeremy	•	✓	✓	✓	✓	•
48	Joe	✓	✓	✓	✓	✓	✓
49	Ryan	✓	✓	✓	✓	✓	✓
50	Henry	•	✓	✓	✓	✓	✓
51	William	•	•	•	✓	✓	✓
52	Andrew	✓	•	•	✓	✓	✓
53	Alan	•	✓	•	✓	✓	✓
54	Brian	•	✓	•	✓	✓	✓
55	Jordan	•	✓	•	✓	✓	✓
56	Christian	•	✓	✓	✓	✓	✓
57	Joy	✓	✓	•	✓	✓	✓
58	Josh	✓	✓	•	✓	✓	✓
59	Randy	•	✓	✓	✓	✓	✓
60	George	✓	✓	•	✓	✓	✓
61	Kyle	•	✓	•	✓	✓	✓
62	Samantha	•	✓	•	✓	✓	✓
63	Kevin	•	•	✓	✓	✓	✓
64	Steven	✓	✓	•	✓	✓	✓
65	Gabriel	•	✓	✓	✓	✓	✓
66	Logan	•	✓	✓	✓	✓	✓
67	Frank	✓	✓	•	✓	✓	✓
68	Justin	✓	✓	•	✓	✓	✓
69	Greg	•	✓	✓	✓	✓	✓

Notes. The table shows which categories from the data structure were coded from each informant interview. Informant names in light grey indicate the participant displayed a dominant orientation towards disconnecting through bright imagining; dark grey indicates a dominant orientation towards integrating through dark imagining; medium grey indicates no dominant orientation. Since many informants exhibited a mix that involved categories in their non-dominant orientation, we performed the same assessment by looking at the overall sentiment from each interview. The results were highly similar and overlapping.

Author Biographical Sketches

Lydia Paine Hagtvedt (lydiahagtvedt@gmail.com) is an independent scholar and a senior user experience researcher at Akamai Technologies. She earned her PhD and MS from the Boston College Carroll School of Management. Her research investigates creativity, curiosity, and the development, application, and user experience of digital technologies.

Sarah Harvey (sarah.r.harvey@ucl.ac.uk) is a professor in the Organisations & Innovation group at the UCL School of Management. She received her PhD in Organisational Behaviour from the London Business School. Her research focuses on creativity and the dynamic processes through which ideas develop as individuals and groups engage in creative work, particularly the way that people work together to synthesize knowledge, identify creative ideas, and pursue some ideas over others.

Ozumcan Demir-Caliskan (odemirca@ic.ac.uk) is an assistant professor of creativity and innovation at Imperial College London. She received her PhD from University College London. Leveraging her background in industrial design, she explores the intersection of creative processes, emerging technologies, and new work practices.

Henrik Hagtvedt (hagtvedt@bc.edu) is an associate professor of marketing at the Carroll School of Management at Boston College. He received his PhD from the University of Georgia. His primary research focus is aesthetics and visual marketing, including topics such as digital displays, visual art, product and promotional design, and luxury branding.