

Product Manager Practices for Delegating Work to Generative AI: “Accountability must not be delegated to non-human actors”

Mara Ulloa
Computer Science
Northwestern University
Evanston, Illinois, USA
Microsoft Research
Redmond, Washington
USA
mara.ulloa@u.northwestern.edu

Jenna L. Butler
Microsoft Research
Redmond, Washington
USA
jennbu@microsoft.com

Sankeerti Haniyur
Microsoft Research
Redmond, Washington
USA
sahaniyur@microsoft.com

Courtney Miller
Computer Science
Carnegie Mellon University
Pittsburgh, Pennsylvania
USA
Microsoft Research
Redmond, Washington
USA
cemiller@andrew.cmu.edu

Barrett Amos
Microsoft Research
Redmond, Washington
USA
barrett.amos@microsoft.com

Advait Sarkar
Microsoft Research
Cambridge, Cambridge
United Kingdom
advait@microsoft.com

Margaret-Anne Storey
Computer Science
University of Victoria
Victoria, British Columbia
Canada
Microsoft Research
Redmond, Washington
USA
mstorey@uvic.ca

Abstract

Generative AI (GenAI) is changing the nature of knowledge work, particularly for Product Managers (PMs) in software development teams. While much software engineering research has focused on developers’ interactions with GenAI, there is less understanding of how the work of PMs is evolving due to GenAI. To address this gap, we conducted a mixed-methods study at Microsoft, a large, multinational software company: surveying 885 PMs, analyzing telemetry data for a subset of PMs (N=731), and interviewing a subset of 15 PMs. We contribute: (1) PMs’ current GenAI adoption rates, uses cases, and perceived benefits and barriers; (2) a value-based Selective Delegation Framework capturing how PMs assess which tasks to delegate to GenAI; (3) PMs adaptation practices for integrating GenAI into their roles and perceptions of how their role is evolving. We end by discussing implications on the broader GenAI workflow adoption process and software development roles.

ACM Reference Format:

Mara Ulloa, Jenna L. Butler, Sankeerti Haniyur, Courtney Miller, Barrett Amos, Advait Sarkar, and Margaret-Anne Storey. 2026. Product Manager Practices for Delegating Work to Generative AI: “Accountability must not be delegated to non-human actors”. In *2026 IEEE/ACM 48th International Conference on Software Engineering (ICSE-SEIP ’26)*, April 12–18, 2026, Rio de Janeiro, Brazil. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3786583.3786849>

1 Introduction

Generative AI (GenAI) tools are rapidly reshaping software engineering (SE) workflows. Code generation systems such as GitHub Copilot and Gemini’s Code Assist now enable developers to produce code faster than ever before [3, 28, 30, 31, 51, 70, 76]. Yet, focusing solely on the developer’s tasks risks overlooking the larger transformation underway on the end-to-end process of software creation which involves not just software engineers (SWEs), but also Product Managers (PMs). PMs are a type of knowledge worker who connect business goals, user needs, and technical execution in software development [19, 24, 25]. Despite their pivotal role in shaping and ensuring successful software outcomes, PMs have not been studied much in the SE research literature [32], and scholars have called for more research and frameworks to capture the nature of PM work [52]. While there is a fast-growing body of work examining how SWEs use GenAI [1, 2, 17, 49, 50], little is known about how GenAI is transforming the PM role. Moreover, this is an especially critical time for studying PMs: early evidence suggests that GenAI is reshaping roles and task distribution, with SWEs’ work shifting [17, 49] and knowledge workers more broadly experiencing changes in both the perception and execution of work [36, 39, 41, 53].

Studying how PMs delegate tasks to GenAI offers valuable insight into how delegation in general arises, since the PM role positions them as delegators by design [25, 66]. Their role naturally requires balancing strategic oversight with effective distribution of responsibilities [23], making them a compelling group to survey and interview in building an understanding of delegation values. Understanding how tasks delegation occurs within PMs workflows can inform delegation practices in broader GenAI knowledge work.



This work is licensed under a Creative Commons Attribution 4.0 International License. *ICSE-SEIP ’26, Rio de Janeiro, Brazil*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2426-8/26/04
<https://doi.org/10.1145/3786583.3786849>

Recognizing this research gap in this moment of rapid technological change, we conducted a large-scale mixed-methods study with Microsoft PMs. As one of the world leaders in GenAI development, Microsoft operates at the frontier of adoption, with leadership actively encouraging use and employees gaining early access to tools, like GitHub Copilot and Microsoft Copilot. This distinct context provides a window into how GenAI may reshape the PM role more broadly as these tools and tools like them diffuse across the industry. Through our study, we surveyed 885 PMs and interviewed a subset of 15, exploring their usage patterns, levels of expertise, beliefs, barriers, and lessons learned. Through our research, we aim to answer the following questions:

- **RQ1:** What are GenAI adoption rates, use cases, and perceived benefits and barriers for GenAI use among PMs?
- **RQ2:** What values shape which tasks PMs delegate to GenAI?
- **RQ3:** How are PM delegation practices and the PM role itself evolving as GenAI becomes integrated into their workflows?

We offer the following contributions: (1) the current state of PM GenAI use (RQ1, 4.1) (2) a value-based Selective Delegation Framework describing PMs task delegation to GenAI (RQ2, 4.2), (3) PM workflow changes, such as viewing GenAI as a necessary skill and their role evolving (RQ3, 4.3). We situate our study within broader genAI-for-work trends, advancing understanding of AI workflow adoption shaped by sociotechnical dynamics. As PMs are taking on more developer-like tasks, we urge PM organizations and the SE research community to reimagine the software development lifecycle (SDLC). Thus, we provide foundational insights into the relationship between GenAI and PMs.

2 Background and Related Work

2.1 The Software Product Management Role

PMs are key internal stakeholders within software development teams. PMs generally follow either an individual contributor (IC) or managerial track (supervising others). ICs focus on specific product tasks, while managers oversee team performance and broader goals.

Across both IC and manager tracks, responsibilities and expectations of PMs vary, often reflecting their diverse backgrounds and the needs of their organizations [35, 55]. Unlike their peers, such as SWEs who often have formal training in computer science (CS), PMs often come to their roles from a wider range of disciplines, e.g., business administration, psychology, and user experience [24, 35].

PMs fill cross-functional gaps in SE teams, owning the product lifecycle from ideation through delivery and iteration [19, 25]. Blending technical and non-technical expertise, PMs act as liaisons who align products with user needs and market demands. To support their work, PMs use planning and evaluation tools such as roadmaps and business cases [24]. Prior work has also examined how PMs balance ethical considerations with business goals and user satisfaction [40]. Estes et al. highlight PMs' essential competencies and their value within development teams [25].

While the software product management research area is relatively young [32], some research describes: the role of PM in software startups [55]; challenges PMs encounter [63]; the specific activities and roles PMs fill [44, 45]; the PM role in different environments [62]; and the impact PMs can have on SE overall [23].

Some studies have examined PMs' use of GenAI within broader knowledge worker research. One study included three PMs alongside consultants, journalists, and graduate students, highlighting challenges in synthesizing unstructured data and calling for AI that supports collaborative transparency and reduces over-reliance [75]. Other work explored how industry professionals, including PMs, engage in prompt engineering during GenAI prototyping [67].

Although recent work has explored how PMs use GenAI (e.g., gatekeeping responsible AI use or manage AI-first products to drive key features [54, 61]), little is known about how PMs use AI more broadly or how they use it in AI-forward companies. Our study provides novel empirical insight into PMs' engagement with GenAI and potential role evolution as these tools see wider adoption.

2.2 GenAI Adoption and Delegation Practices

2.2.1 GenAI Adoption in knowledge work. Knowledge workers represent a nebulous (and disputed) professional designation that includes workers in medicine, research, and engineering (e.g., SWEs and PMs) [20, 43]. The accelerated adoption of GenAI, exceeding that of both the personal computer and the internet, underscores its potential to fundamentally reshape and adapt knowledge work [9].

SWEs are integrating GenAI tools such as GitHub Copilot, Gemini, Goose, and GitLab Duo, into existing workflows to support various aspects of their coding process [3, 27–30]. Moreover, SWEs are adopting GenAI to author code, understand large codebases, and conduct code testing and review contributions [3, 6, 51, 70, 76].

While a growing body of literature highlights the potential benefits of adopting GenAI for knowledge work [5, 42, 59, 60, 69, 73], many scholars caution against adoption without careful consideration. They highlight GenAI's limitations: the potential for labor displacement and the devaluation of human expertise [7, 13, 58]. In SE, concerns have emerged around overstating GenAI's programming capabilities and the risk of lower-quality code, including the introduction of security vulnerabilities [34]. In response to these concerns, researchers have proposed frameworks for evaluating both the benefits and risks of GenAI integration [65, 69] and have argued that GenAI assessments should be grounded in real-world usage environments. Building on these insights, we examine PMs' perceptions, adoption, barriers, use cases, workflow changes, and values guiding GenAI delegation.

2.2.2 GenAI Delegation vs Usage. Our study specifically aims to understand *what values influence which tasks PMs delegate to GenAI*. This understanding in turn can be used to inform organizational practices, and potentially the design of GenAI tools for PMs.

It is important to distinguish between *delegation* and *usage*. **Delegation** occurs when a worker decides “to transfer rights and responsibilities for execution of a task or decision-making and the associated outcomes” [66] to an AI system—prior to use. In contrast, **usage** is the act of applying GenAI to a task after such a decision has been made. A recent systematic review highlights that human-agent delegation research remains in its early stages [66].

Research has started mapping how workers use or envision GenAI, including creating, searching, manipulating information, seeking advice, or integrating AI with minimal context, e.g., in SE, LLMs can infer if code review comments match pull request changes. [11, 12, 31, 70, 76]. Beyond capturing GenAI usage patterns,

it is also key to understand how knowledge workers reason about *which tasks* to delegate, something we aim to address.

The decision to delegate a task to GenAI—which precedes actual usage—involves a complex judgment process. Decision-making typically requires methodically narrowing options, weighing benefits and drawbacks, and arriving at an intentional choice [16, 33]. Deciding whether to delegate or use GenAI is a nuanced form of problem-solving: workers must navigate the “*jagged technological frontier*”, where GenAI performs some tasks easily while struggling with others of comparable complexity [21].

Prior work described how knowledge workers use GenAI [11, 12], and other scholars have highlighted that usage also involves emotional, metacognitive, and social dynamics [56, 68]. These dynamics often exert their influence during the deliberation and delegation phase that precedes actual usage. However, we lack an empirical basis for reasoning about how these dynamics influence delegation to GenAI for the PM role. We address this gap with our study, deriving a framework for selective delegation (Sec. 4.2).

3 Methods

We used a concurrent mixed methods approach. Across Microsoft, we distributed a survey and in parallel conducted one-on-one semi-structured interviews. The survey allowed us to understand GenAI adoption patterns at scale, while the interviews allowed us to probe deeper into PMs’ GenAI usage, motivations, concerns, and workflows. We collected telemetry data from consenting participants to confirm self-reported GenAI usage and to supplement demographic information. Our study follows best practices for mixed-methods empirical SE research [64] and was approved by Microsoft’s ethics board. The study was conducted from April to June 2025.

3.1 GenAI-Early Adopter PM Environment

We conducted this study at Microsoft, a large, multi-national software company with over 100,000 employees. Microsoft is an “AI forward” company, an early adopter of and advocate for GenAI tools. Microsoft also designs GenAI tools, provides licenses for its employees and usage is encouraged company-wide. This licensed access had been in place for over three years during our study. Our population included both Individual Contributors (ICs), who handle day-to-day PM work, and People Managers (hence referred to as “Managers”), who combine PM duties with managerial tasks (planning, managing others) and thus do tasks like performance reviews.

3.2 Instrument Design and Data Collection

3.2.1 Survey Design, Recruitment, and Data Collection. To design our survey, we first conducted informal interviews with ten PMs, recruited through convenience sampling. These interviews helped us understand cultural and contextual factors, e.g., common jargon, challenges, and varying levels of GenAI use—informing the key topics, concerns, and values mirrored in our survey questions.

Since managers and ICs have different perspectives, we branched the survey based on role [71].¹ **IC questions** focused on daily usage, beliefs and questions on four categories: Current GenAI Usage (5 questions), Attitudes & Perceptions (5 questions), Barriers & Opportunities (7 questions), Future Outlook & Feature Requests (6

questions). **Manager questions** focused on observed team dynamics and organizational strategies, specifically Team GenAI Adoption & Barriers (5 questions), Managerial Responsibilities & Decision-Making (5 questions), Future Outlook for Team (3 questions), Leadership & Culture (6 questions).

Additionally, both groups were asked about challenges: e.g., ICs were asked “*Is there a technical or workflow limitation that makes using GenAI difficult for you*” while managers were asked “*What are the biggest challenges you and your direct reports face when trying to use GenAI for your team?*”

We piloted the survey with the PM team in our company division (44 ICs and managers). Feedback included moving demographics to the end to emphasize core questions, grouping Likert items into blocks, and making minor wording edits. The pilot data was not included in the final dataset. We shared the survey with four Microsoft SE researchers (none authors in this study); one suggested adding two Work Trend Index questions [47]—on Copilot as a thought partner and as a tool—both were added to the survey. We emailed the survey to 5,111 PMs in one Microsoft division. Over 21 days, 885 PMs (17%) took an average of 13 minutes to complete it. While all 885 PMs consented to and completed the survey, only 731 consented for us to examine their telemetry data, thus, in the supplemental material table shows demographics for these 731 PMs [71].

3.2.2 Interview Design, Recruitment, and Data Collection. In parallel to deploying the survey, we conducted 15 semi-structured interviews with early consenting survey respondents. Because 621/885 survey participants expressed interest in being interviewed, we applied purposive sampling to ensure diversity across geography, GenAI usage levels, job level, education, and gender (see the table in supplemental material[71]). Interviews lasted on average 30 minutes, guided by a semi-structured script [71] covering role, workflow, GenAI usage, barriers, and future needs. We conducted the interviews and generated automatic transcriptions using Microsoft Teams.

3.2.3 Telemetry Data Collection. We collected telemetry data for survey respondents who consented to share it. This allowed us to keep the survey demographic section brief, as telemetry captures: the (1) individual’s title, and (2) country the individual worked in. This data is routinely collected by the company, but individual consent is needed for researcher access. For those 731 PMs who consented, we linked their survey responses with existing demographic data to identify sample characteristics such as location. The telemetry data also included (3) frequency of interaction with GenAI tools. In the survey we asked participants to self-report, “*How often do you use Generative AI tools for your PM tasks?*”, allowing us to compare self-reported usage to actual usage.

3.3 Data Analysis

3.3.1 Qualitative Data: We analyzed the interviews before fully analyzing survey responses. To analyze the qualitative data from 15 semi-structured interviews and open-ended survey responses, we used reflexive thematic analysis [14]. We opted for this interpretive approach as our goal was to generate rich, nuanced insights rather than quantify responses. In reflexive thematic analysis, researcher subjectivity and reflexivity are considered integral to theme development, making inter-rater reliability metrics incompatible with

¹<https://doi.org/10.5281/zenodo.18157873>

the method [14, 15]. Instead, we ensured rigor through consensus validation: codes were collaboratively reviewed and refined during bi-weekly team meetings until verbal agreement was reached. This iterative process prioritized depth of understanding and reflexivity over statistical measures of agreement.

For interview analysis, one author, who also conducted the interviews and is trained in qualitative analysis, began by randomly selecting five transcripts to ensure variation and performed open coding while reviewing transcripts for accuracy and summarizing key points. This inductive approach allowed salient participant perspectives to emerge without imposing a priori categories. The same author then used affinity diagramming to cluster related observations and document emerging patterns and analytical insights [8, 48], producing a preliminary codebook of high-level codes. All authors reviewed and refined this codebook through bi-weekly discussions, reflecting on positionality, assumptions, and biases, and aligning themes with study objectives (e.g., usage patterns, delegation behaviors, role expectations). Codes offering novel insights relative to prior literature were prioritized. Using the evolving codebook as a guide, the lead researcher coded the remaining interviews, adding new codes as needed. We iteratively refined the codebook until no new codes emerged, indicating saturation.

For survey responses, three authors each selected one open-ended question and independently performed open coding for all responses. The team then met to review initial codes, resolve discrepancies, and consolidate overlapping codes. Three questions were double-coded by two authors; the remaining questions were coded by one researcher. All codes were subsequently reviewed in team meetings, where disagreements were discussed until consensus was reached. This process produced a shared codebook that integrated insights from both interviews and surveys.

3.3.2 Quantitative Data: Statistical Analysis of Survey and Telemetry Data. The survey included both quantitative and qualitative items. For quantitative responses (e.g., Likert-scale, multiple-choice), we computed descriptive statistics (mean, standard deviation) and applied inferential tests (ANOVA, Mann–Whitney U) to examine relationships such as self-reported versus actual GenAI usage as indicated from the telemetry data. Analyses were conducted using Visual Studio, Jupyter Notebook, and Python scripts.

3.3.3 Integrating Survey and Interview Data: Exploratory Sequential Analysis. Although interviews and surveys were conducted in parallel, we analyzed interviews first to help us interpret survey findings, revealing the context behind participants' responses. Interview insights informed our coding of open-ended survey questions and guided interpretation of quantitative patterns. For example, survey data indicated that many participants first engaged with GenAI through self-exploration, which aligned with interview themes emphasizing personal experimentation as a source of value. Conversely, divergence between data sources helped contextualize and deepen our understanding of participants' perspectives. For example, the survey showed some managers were not concerned about job displacement, but our interviews revealed this was due to impending retirement. While not concerned for themselves, they expressed concern for their teams and emphasized preparing them with skills to avoid risks from GenAI use. Finally, the Selective

Delegation Framework emerged as an insight from the study and subsequent analysis, rather than being a predetermined objective.

4 Findings

	Identity	Accountability	Personal & Social Dynamics
	<i>Alignment between task and PM's sense of role</i>	<i>Perceived responsibility/risk for task outcome</i>	<i>Personal, social, or cultural influences on delegation</i>
Individual	Task reflects professional expertise	PM feels personally responsible for task quality or consequences	Self-perception: cognitive responses to delegating tasks
Team	Existing team norms shape which tasks PMs delegate	Peer accountability expectations affect delegation	Team openness, stigma, and transparency shape delegation
Organization	Org identity with GenAI shapes delegation	Alignment with org values (e.g., privacy & compliance)	Stance on experimentation and tool adoption

Table 1: PMs identified values at the individual, team, and organizational levels influencing task delegation to GenAI.

We present findings across our three research questions. **RQ1:** What are GenAI adoption rates, use cases, and perceived benefits and barriers for GenAI use among PMs? Sec. 4.1 sets the stage by describing how PMs currently engage with GenAI and what they believe about its use. Building on this foundation, through **RQ2:** What values shape which tasks PMs delegate to GenAI? In Sec. 4.2, we identify the value shaping delegation decisions. Finally, for **RQ3:** How are PM delegation practices and the PM role itself evolving as GenAI becomes integrated into their workflows? In Sec. 4.3, we show how PMs develop the values depicted in the framework.²

4.1 PM GenAI adoption rates, use cases, and perceived benefits and challenges (RQ1)

As little is known about how PMs use GenAI for their work, we started by trying to understand the current usage of GenAI by PMs including self-reported usage amounts, what tasks they used it for, any benefits they find from using it, what challenges they experience in using it, and their overall beliefs about GenAI. Here, please refer to Figure 1: the survey includes 735 Individual Contributors (ICs) and 150 Managers, all of whom are Product Managers (PMs).

4.1.1 Adoption and use cases. A majority of ICs reported using GenAI daily or almost daily (75%, 554/735), with 19% (142/735) indicating usage 1–3 times per week, 4% (28/735) reporting usage 1–3 times per month, and 1% (11/735) reporting either never using it or using it less than once per month. To assess the accuracy of

²Survey quotes are labeled PX, and interview quotes PXi; X = participant number.

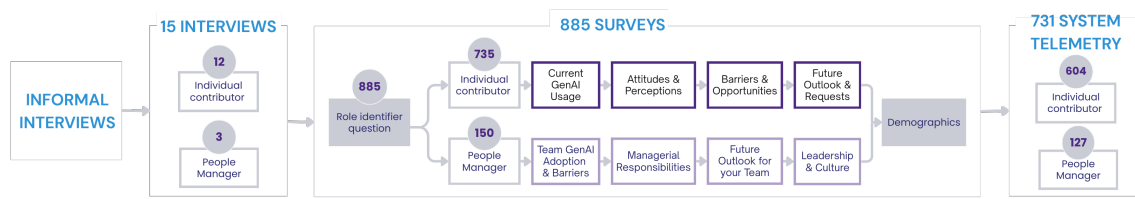


Figure 1: We conducted informal interviews, 15 interviews, and gathered 885 survey responses. 731/885 survey participants consented to telemetry data analysis.

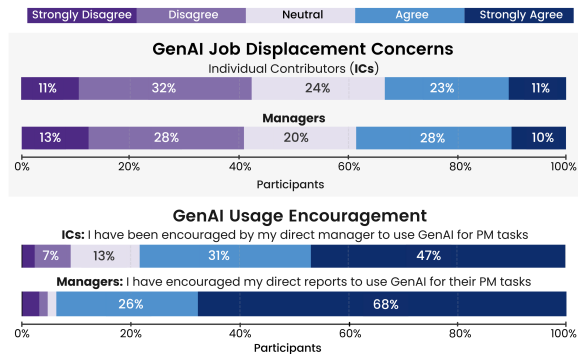


Figure 2: ICs and Managers reported GenAI job displacement concerns and high encouragement to use GenAI.

self-reported usage, we accessed participants' M365 Copilot usage logs. We compared self-reported usage frequency with actual query counts using Spearman rank correlation, which yielded a positive correlation coefficient of 0.37. This suggests that higher self-reported usage is moderately associated with higher actual usage. Additionally, we conducted a one-way ANOVA to test whether mean query counts differed significantly across self-reported usage groups. The analysis produced an F-value of 6.58 with a p-value < 0.00001 , indicating statistically significant differences in actual usage between the groups. We also examined whether participants were encouraged to use GenAI tools in their work. Seventy-eight (573/735) percent of ICs reported being encouraged by their direct manager to use GenAI, while 94% (141/150) of Managers reported encouraging their direct reports (i.e., ICs) to do so (Figure 2).

ICs also reported the **specific PM tasks** they use GenAI for with most use cases involving documents and writing. ICs could pick three options from a list, so the total will not always add up to 100%. 74% (547/735) of respondents reported using it for drafting or refining documents, 53% (391/735) use it for summarizing documents and 32% (236/735) use it for writing emails or external communications. Others used it to brainstorm ideas (35%, 254/735), conduct user research or data analysis (22%, 159/735), prototype ideas and write code (12%, 91/735), and learn a new skill (9%, 65/735).

For Managers, we focused less on individual productivity and usage and instead asked for open text examples of how they have seen their team use it, as well as how they use it specifically for managerial tasks. We asked Managers how they have used GenAI for **managerial tasks**; most reported using it to support performance evaluation of their employees in some way. Some reported that it helped them summarize peer feedback to share with their direct report ("I used it to organize [feedback] my directs received to identify trends in strengths and improvements." P76); others shared that they

use it as a way to potentially mitigate any bias they might have ("For [performance reviews], I use it to confirm delivery against core priorities to mitigate my own bias" P63). Some Managers reported using it to help them ensure accuracy in their writing, such as P210 who reports, "Since both tone and wording significantly influence how feedback is perceived, I leverage LLMs to rephrase my feedback (direct or supportive), aligning it with my intended tone and goals."

In addition to performance evaluation and communication, Managers reported using it for team planning ("understanding and organizing team planning, drafting OKRs, allocating resources" P66), meeting management ("Organizing trainings sessions, including agenda and presentation. Organizing brainstorming and team building activities. [...] Taking meeting notes." P852), and presentations and writing ("Rewriting summary mails/status updates and even generating a draft from scratch" P457).

4.1.2 Perceived benefits of GenAI. On a 5-point Likert scale, 81% of ICs selected 'Strongly Agree' or 'Agree' in response to the statement *Using GenAI often saves me time* (Figure 4). Next, we had an optional open text question, *In your own words, can you share a brief example of when GenAI saved you time or provided a clear benefit in your PM work?* We coded these open text questions; the most common response involved saving time in writing documents or presentations. One respondent said, "GenAI has helped me [turn] brainstorming notes into product spec requirements. It has been able to draft problem statements and potential solutions based on my inputs." (P84). The next most reported area of benefit was in summarizing and recapping meetings, as illustrated by P21, "Meeting recaps have been incredibly useful, ability to draft comms or updates as well." and P88, "GenAI recaps of Teams meetings and automated curation of action items is the best benefit." Some reported that prototyping was the biggest benefit, drastically reducing the time from idea to product, with P110 reporting, "Prototyping ideas is a lot quicker; I prototyped ideas on Failure Bucketing following a research paper. From paper to working prototype, it took me less than two hours. This would have taken 1-2 days of engineers' or my time to read and build a working prototype" Seven participants explicitly wrote that they could not provide an example of GenAI saving them time, such as P328 who reported, "It has not. Every time I've used GenAI it feels like I've asked an incompetent assistant to do my job badly."

Again, with Managers, we focused less on individual benefits and instead asked for examples of **how GenAI influenced a team deliverable**. Some reported a decrease in human effort for important tasks ("Time to closure and approvals has significantly decreased, and we have improved the auditability of our team's work without a lot of extra human effort." P13), while others reported ways they had empowered their team to be less dependent on themselves, as

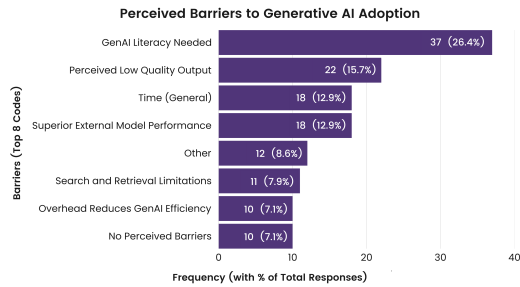


Figure 3: ICs’ coded responses to the optional, open-text question #17, *Is there anything else you’d like to tell us about barriers to using GenAI more effectively in your role?*

illustrated by P66 *“I built an agent [...] focused on exec communication and attention, brevity, clarity – I save a lot of time because I ask folks to run content through this tool before I review.”* Some benefits were similar to those of ICs, such as improving writing and saving time with written documents: *“By using GenAI, we were able to generate a very well written first draft of a [product release] blog post [...]. This is a key use case - AI is allowing us to go from 0-80% on documents with just a few ideas and some links [...]. It prevents typical writers block situations [...] This allows us to advance ideas so much faster, universally accelerating our product thinking.”* P142.

4.1.3 Perceived barriers of using GenAI. We asked ICs what limits their GenAI use (question #15 in supplemental material). From a list of options they could pick up to three items, so the total will not always equal 100%. One quarter (185/735) selected “none,” indicating satisfaction with current usage. However, 47% (345/735) of ICs cited concerns about output quality or inaccuracies, and 35% (259/735) reported insufficient time to experiment. Other barriers included lack of relevant use cases (19%, 142/735), unclear data privacy/security guidelines (16%, 119/735), lack of knowledge or training (14%, 101/735), and unsatisfactory initial experiences leading them to abandon the tool (14%, 100/735). We followed up with an optional, open-ended survey question (#17 in supplemental material): *“Is there anything else you’d like to tell us about barriers to using GenAI more effectively in your role?”* and coded 165 responses (Figure 3 shows the top 8 codes). The most common theme was the lack of GenAI literacy, as P869 explained needing learning resources: *“I need clear examples of what I can use GenAI for and how. Not to the PM role in general but to the implementation I have of the PM role.”* Additionally, 22 participants said quality isn’t good enough: *“Its accuracy (or lack thereof) is the single biggest barrier.”* (P145). Other issues included lack of time to learn, perception that external models (unusable for work) are superior, and high overhead (e.g., validating output) making use impractical.

We also asked Managers about issues their teams faced using a multi-select question. The most reported was *difficulty integrating GenAI tools into existing workflows*, cited by 77/150 Managers (51%). Next was *not enough time to experiment / adopt new processes* (47%, 70/150), followed by *concern about accuracy or bias* (46%, 69/150). Less common were lack of training (24%, 36/150), unclear company policy (24%, 36/150), and not knowing where to start (23%, 35/150). As participants could select multiple options, totals exceed 100%.

4.1.4 Beliefs and Feelings about GenAI. Finally, we analyzed PMs perceptions and sentiments about this new way of working. We

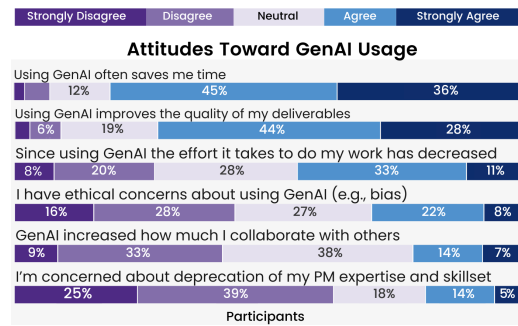


Figure 4: A selection of responses to Likert questions posed to 735 ICs about their thoughts and beliefs on GenAI.

asked ICs a series of Likert questions to learn more about their views of GenAI, such as whether they trust it, if they think it improves work quality, and if they have ethical concerns (Figure 4).

In addition to believing that GenAI saves them time, the majority of ICs (72%, 529/735) believe that GenAI improves their work’s quality. However, the effort required to do their work hasn’t decreased for most ICs (56%, 412/735). When asked if they have ethical concerns over the usage of GenAI, almost a third of ICs (30%, 221/735) responded that they do. When asked to elaborate on these concerns, people report being worried about not knowing how to handle bias (*“I know there are some biases in GenAI but I wouldn’t know what to do to course correct of fix it.”* P26), worries about the environment, and concerns about deskilling, replacement or lack of growth opportunities (*“I worry about becoming a data entry monkey with limited ability to grow. This tool that could easily remove the need for the middle structure of our company where people learn and grow in preparation for advancement all while it locks in top management and data entry/entry level personnel in their existing roles.”* P355). Lastly, we asked how they perceive GenAI as a whole. Eighty-four percent (617/735) said they view Copilot as a tool that can help them and augment their work, while 64% (473/735) view it has a thought partner or teammate (some chose both conceptualizations).

4.2 Selective Delegation Framework: Deciding to delegate PM tasks to GenAI (RQ2)

Participants’ willingness to delegate tasks to GenAI was shaped by assessment of various values. Before using GenAI, PMs’ decisions were guided by: (1) alignment with their **identity** (who they are as a person), (2) a sense of **accountability** (what they or others expect of them), and (3) **personal & social dynamics** (how they or others perceive them). Moreover, PMs described values influencing GenAI task delegation as operating at the individual, team, and organizational levels. We present a Selective Delegation Framework (Table 1) that illustrates these relationships.

4.2.1 Individual task delegation to GenAI. Tasks that participants did not strongly **identify** with, such as formatting or summarizing meeting notes, were easily delegated to GenAI. However, when a task was central to their professional identity (e.g., writing for those with a literature background), participants were more reluctant to delegate to GenAI. For example, P815 shared, *“I’m a journalist before this. I value my own voice and am not going to dilute it with AI, but*

also means it is not faster or more effective for me.” Similarly, P40i described written communication as core to the PM role, saying, “being a good communicator is important for everybody, but it’s very important for PM. I take great pride on my written communication ability as a craft. If I let AI write something for me then they’re no longer my words. It’s not my craft anymore.”

This reluctance extended beyond writing to other tasks tied to PM identity. As P440 put it, “I’d like to see a shift in focusing on AI assistants that streamline menial tasks (task completion tracking, reminders, status updates), rather than our current focus on using GenAI to replace PM thought work (using GenAI for specs, strategies).” Participants felt GenAI should support, not replace, the cognitive and expressive aspects of their work. P365 emphasized this distinction, saying, “We need to create robots that do mundane tasks, not create systems that make us complacent and use less of our cognitive thinking and vocabulary.” Delegating high-identity tasks, for such participants, felt like losing a part of themselves.

Participants more readily delegated low-**accountability** tasks to GenAI, but hesitated with high-accountability work such as writing specifications or journey maps. As P661 said, “Accountability is one of the key aspects in a PM role. GenAI can [never] be used as a silver bullet because of this.” Even when using GenAI, PMs still felt accountable for outcomes. P547 shared, “As a PM, I used to write a newsletter myself. [However,] I have delegated this task to GenAI and it’s now responsible but I am still accountable [...] it’s posted under my name, any mistakes made by GenAI reflect poorly upon me.” Similarly, P805 noted, “Due to the nature of my job and the consequences & risks of distributing incorrect information, I can’t trust the outputs unless I have a way to quickly validate that the work done by AI is correct.” Participants emphasized the consequences tied to human accountability vs AI: “If a human REALLY messes something up... you go to jail. But what do you do when an AI really messes something up? It doesn’t care.” (P40). Another participant, P640, explained why accountability must remain human-held “GenAI can augment the decision-making and... offer output to help decisions [but] **accountability must not be delegated to non-human actors.**” In essence, participants calibrated their use of GenAI to maintain a strong sense of accountability, even when delegating tasks.

Finally, PMs described a **personal dynamic** grounded in self-perception, reflecting on their cognitive responses when delegating tasks to GenAI. These reflections included emotions such as dissatisfaction, unease, increased productivity, or enhanced capability. For example, the thought of using AI alone elicited feelings of dissatisfaction: “When people suggest that I use AI to help me with my job, even for just drafting a document, I don’t feel as satisfied as if I had done the thinking/writing/problem-solving on my own.” In contrast, some participants described GenAI as supportive, producing positive feelings, reducing overwhelm and increasing productivity: “As someone with a neurodivergent brain, GenAI has been beneficial in managing cognitive load and reducing overwhelm. [...] This support has been invaluable in maintaining clarity and focus, allowing me to leverage my strengths and contribute more effectively to my team.” (P368). Consistent with this perspective, (P202) reported feeling more productive and found GenAI to be beneficial: “I feel more productive, almost like a fellow team mate, a peer programmer is sitting next to me who is always there to work together.” Overall, PMs experienced a range of emotional reactions when using GenAI.

In sum, at the individual level, participants described a non-linear, three-step approach to GenAI. They considered their professional **identity** (Does using GenAI for this task encroach on my personal or professional identity?). Next, they calibrated **accountability** (Is the GenAI accountable, or am I?). Finally, they reflected on the emotional impact of the decision (How does this impact my self-perception).

4.2.2 Team level task delegation to GenAI. In addition to individual values, PMs described their team as strongly influencing decisions to delegate tasks to GenAI. Values included existing team norms and identity, the need to maintain accountability, and the extent to which teammates supported GenAI usage.

Participants described how their **team’s established identity**, including norms that promote experimentation or maintain traditional non-GenAI practices, shaped their GenAI delegation decisions. For example, one participant stated “We are a data science team, so we’ve explored a lot of projects with GenAI, but I haven’t seen it change how we work at a team culture level. Innovation and experimentation was already the norm.” (P475) Another participant offered a similar perspective, noting that despite widespread use of meeting summarization and organizational encouragement, their team still expected manual note-taking: “Leadership that has been around a long time like things a certain way. Rather than just using Copilot for meeting notes they still expect notes to be captured and sent out the old fashioned way.” (P553) Participants emphasized that team norms consistently influenced their delegation behavior.

Participants identified accountability as a core PM skill requiring continuous effort. When delegating to GenAI, they considered its impact on **team accountability**. One participant suspected a teammate used GenAI without factoring in accountability: “I worry it’ll make people more complacent. I have worked with colleagues that have very clearly used GenAI without revising, and there is a lot of labor (on my part) involved in improving the quality of their contributions.” (P128) This increased their own workload and lowered team morale. PMs emphasized that ineffective delegation—without plans to preserve PM accountability—risked undermining customer-driven, data-informed product goals. Others noted that automation can erode essential skills: “I worry that we’ll get to the point of AI talking to each other... and PMs no longer building the product judgment and core skills.” (P329) Overall, PMs stressed delegating only when clear team strategies uphold accountability, ensuring GenAI does not degrade judgment or shared responsibility.

Participants described how **teammate dynamics** shaped their GenAI delegation decisions. Negative peer attitudes discouraged use, while support from peers and managers reinforced it. For example, strong anti-AI views from teammates made some hesitant: “There are ‘anti-AI’ folks within my team, which makes me feel discouraged to use GenAI.” (P613) In contrast, recognition for AI use generated positive energy and encouragement: “The team, partnering with [other] teams, proposed a project during the last [day of learning], and won first place. This has created a positive energy among the team, and they to learn more about [GenAI].” (P204) Participants also described team strategies to fine-tune AI delegation and improve workflows: “My team has shared commitments in how we will a) advance our AI knowledge and industry awareness b) adopt AI tooling to improve our workflows / increase productivity.” (P825). Managers highlighted the combination of strategy and encouragement as a method for helping direct reports leverage GenAI:

“Continue to encourage my team to learn and experiment, then share with the broader org what they are doing and how it’s helping.” (P29) Overall, participants emphasized that social dynamics, whether judgment or support, significantly influenced delegation decisions.

4.2.3 Organizational level task delegation to GenAI. At the organizational level, PMs highlighted the company’s identity as a GenAI leader, the importance of maintaining established guiding principles, and a culture that strongly encourages GenAI use.

Participants often spoke proudly of their **organization’s identity** as “customer zero”, developing and actively using GenAI in-house. Regular engagement with GenAI was seen as essential for building cutting-edge tools for customers. For example, one participant encouraged: “Embrace being customer zero. Use the GenAI tools as frequently as possible. We are able to be the best technical trusted advisor when we can make the solutions relatable to a customer’s mission, use cases, and pain points.” (P451) Others highlighted the many GenAI initiatives underway: “there are already sooooo many big initiatives here across the company.” (P28) A PM pursuing a Master’s degree noted that while GenAI use was discouraged at university, it was strongly encouraged at their company: “For school, they don’t want you to use it. But at the same time, everyone in industry is using it.” P04i. Overall, participants depicted their organization’s evolving identity as a key motivator for expanding GenAI adoption.

Organizational identity strongly shapes **social dynamics**. Employees are encouraged to use GenAI daily. ICs reported generally feeling supported, while 1 in 4 managers cited leadership pressure to drive adoption. One manager noted the future value justifies early use: “If AI isn’t able to produce positive results, I encourage my team to try again in a few weeks. AI experiences and quality are increasing rapidly.” (P247) P871 observed that encouragement sometimes overlooked pitfalls: “We are not just encouraged but we have no choice but to use GenAI at work without discussing ethical concerns or possible implications to how it could affect our creativity and PM strengths.”. Another participant remarked that the strong push might promote performative over utility: “There’s too much of a visibility game going on where PMs are flooding channels with the slightest possible AI experience to appear relevant. The current environment makes folks afraid for their jobs, so we are getting more performative posturing than actual innovation.” (P59) However, some PMs felt that the AI push needed a more deliberate approach and should align with concrete use cases: “The use of AI should be handled as an Organizational Change Management initiative [...] While the technology is great, if an org or a person doesn’t use it, it has no value to them.” (P299) Participants were generally receptive to organizational encouragement of GenAI but raised concerns about skill atrophy and stressed thoughtful integration into workflows.

As GenAI adoption grows, participants reflected on core organizational priorities. Many emphasized the need to uphold **long-standing values** such as security, privacy, and customer trust. One participant noted: “More focus and willingness to slow down in the name of privacy and security, even if that means we lose initial market leading status. Improve the models so the work actually gives us what we need and not wastes time.” (P345) A few participants expressed trust in the organization to safeguard data and interactions with proprietary GenAI tools: “I’m not super worried about AI risks such as privacy or content moderation issues. I feel Microsoft has solid privacy and security controls around our AI tools and potential content

issues just don’t bother me at this point.” (P836) P828 also highlighted that the organization was integrating GenAI responsibly and hoped it would continue to mitigate risks: “We’re doing a great job having a Responsible AI focus. I have done lots of work in this space, and trust and see we are driven by a great mission as a company, approaching this area with great care. At the same time, the risks in this space are unpredictable. This is the biggest change in tech in a couple of generations, maybe longer!”

4.3 How PM delegation practices and role is evolving as GenAI is adopted (RQ3)

In Sec. 4.2 we present the Selective Delegation Framework describing how PMs decide to delegate tasks to GenAI. But how did PMs develop this set of values? What processes might be causing not just individual delegation decisions in terms of their **identity** and **accountability**, but the **social dynamics** aspects of the delegation framework to evolve?

To explore PM processes for the *development of the skill of GenAI delegation*, we asked, *If you could share one piece of advice with fellow PMs about using GenAI, what would it be?* Participants described two ways they evolve delegation: **direct experimentation to evaluate accountability**, and, for Managers, **leading by example to navigate social dynamics**. They also saw a need to continuously adapt, viewing GenAI use as a **broadening of their PM identity** (e.g., engaging in technical work and blurring PM–SWE roles).

4.3.1 Direct experimentation. The most common advice from Managers and ICs was to “**just start trying**.” Participants encouraged PMs to experiment without hesitation and to persist even when initial attempts fail. P275 emphasized “Use it to make you better / it’s amazing” and P427 added, “Try it, try it again, and keep trying until you break it.” Beyond starting, PMs highlighted the value of returning to the tool and maintaining a growth mindset—focusing on learning rather than perfection: “Don’t worry about not getting the prompt right the first time, just keep trying.” (P37) and “Jump in and try - don’t let perfection be the enemy of good.” (P657). Finally, participants noted GenAI’s rapidly evolving nature, encouraging PMs to revisit it over time: “Keep trying it, it’s always evolving so don’t assume your experience today is the definitive way it will work a week from now. [...] whenever you’re working on some task, think, is there some way I could leverage AI to make this better?” (P839).

In addition to starting with GenAI, PMs advised experimenting to understand its capabilities and discover what works best. P128i suggested, “Explore with it, play with it, and then celebrate the learnings.” Another participant, P10, added, “Use multiple GenAIs every day. Prompt every day. Evaluate what you liked/didn’t like about the output. Compare them.” Some also recommended trying different tools and using experiments to guide future usage: “Continue using the tool to see how inputs produce certain outputs, learn how to dictate what you are looking for it to return to you.” (P91). Experimentation is key to refining GenAI delegation practices and understanding its evolving capabilities, which is essential for exercising judgement about **accountability** at individual, team, and organizational levels.

4.3.2 Managers can serve as models of GenAI adoption for their team. For Managers, a key piece of advice was to **lead by example**. Some stated it plainly: “Be bold and lead with example.” (P55). Others offered more concrete guidance: “Lead by example, use it in your

workflows and share the pros/cons with the team/learnings, etc." (P76) Managers believe their own usage and enthusiasm will rub off, as P536 mentioned *"Start with yourself. If the manager is using and talking about how helpful the tools are the teams will naturally try to help."* Finally, Managers suggested sharing personal experiences can normalize experimentation and reinforce the importance of just starting: *"Share what you've tried with your team. I use Copilot and other products at least once an hour. I share the successful, silly, confusing, etc results with my team to spark thinking for their own use cases and reinforce how easy it is to just jump in and try."* (P393). Just as direct experimentation allows PMs to develop and exercise judgments about accountability, Managers attempting to model GenAI use allows individuals and teams to collectively develop their sense of judgments of **social dynamics**.

4.3.3 PMs perceive the need for continuous personal improvement. Many PMs acknowledged widespread concerns—among knowledge workers and in the media—about GenAI replacing jobs, but chose to focus on its potential benefits. One participant advised, *"Don't be scared – it's an exciting time for technology – embrace AI, it's going to speed up a bunch of mundane tasks and give you the intelligence and a soundboard to discuss and sharpen your ideas."* (P651)

While replacement fears were less pronounced, participants emphasized the risk of falling behind. As one put it, *"AI won't replace humans, but humans with AI will replace humans without AI."* (P376) A manager framed this similarly: *"I firmly believe my staff won't be surpassed by AI, but we will be surpassed by someone who uses AI effectively."* (P879) Another participant echoed, *"Someone not using AI can't compete with someone using AI, as someone with a 1990 flip phone can't compete with someone with an iPhone."* (P444)

Several PMs described adapting to this shift as both natural to the PM role and essential for innovation. P833 reflected *"It's a necessary area of upskilling to keep up with where the industry is going – the PM job isn't going away, but it's changing because of AI tools that will augment how we approach traditional tasks."* Similarly, P279 framed this evolution as core to PM identity: *"A great PM should have the needed competencies to evolve in their role to what the upcoming roles that drive the creation of new technologies & use of GenAI are."*

4.3.4 Perceived blurring of PM and SWE roles. A primary manifestation of this **shift in identity** is the blurring of PM and SWE roles. With GenAI, PMs felt capable of undertaking tasks traditionally performed by SWEs, such as writing scripts, developing prototypes, and creating agents, e.g., P187 said GenAI *"helped me write SQL queries for data analysis that I wouldn't be able to do myself"*. In the survey, 91 (12%) ICs reported using GenAI for prototyping and coding. One Manager used GenAI to prototype, bridge expertise, and communicate more effectively with developers: *"I've been using AI for UI prototypes, see the concepts we want to get in front of users, improve language I use with devs."* (P13i) An IC described feeling empowered as *"PMs can take more of a front-line role working with data and building prototypes, building agentic experiences."* (P66)

Several PMs envisioned further blurring of traditional PM and SWE boundaries. P629 suggested, *"it would be incredibly valuable to have sessions that help PMs get more familiar with dev-focused AI tools as well—especially since the division of responsibilities between PM and Engineering may evolve in this new AI-driven landscape."* While this requires stepping beyond the PM comfort zone, it offers an opportunity to develop new skills: *"Get out of your comfort zone.*

Stop using AI to draft specs or to create messages. Try AI for areas that your peers work in (coding, designing, user research) and start learning how you can be great in some of those areas too. We all need to become more of a Swiss army knife." (P236) Yet, rather than necessarily seeing this as broadening their official responsibility scope, some PMs viewed this horizontal expertise as strengthening their core PM role: *"I used GenAI during the planning process; I've been vibing coding to help me be stronger as a PM."* (P624)

5 Discussion

GenAI integration can reshape work by enhancing productivity, evolving roles, and enabling new human–computer interactions. Realizing this potential, however, requires a human-centered understanding of workers' values and delegation practices. To examine PMs' perspectives, we conducted a mixed-methods study with 885 survey responses, 15 interviews, and machine telemetry (GenAI usage and PM demographics) from 731 PMs. We analyzed (1) current GenAI usage and beliefs, (2) values shaping delegation, and (3) perceived adaptations to workflows and role evolution. Below we discuss a broader view of AI workflow adoption and propose rethinking of current software development roles.

5.1 Broadening our understanding of AI workflow adoption

We proposed a value-based Selective Delegation Framework (4.2) describing PMs' GenAI delegation behavior. PMs consider identity, accountability, and personal and social dynamics when deciding which tasks to delegate or retain. The framework builds on prior research on AI workflow adoption across knowledge work [22, 56, 57, 68, 72], showing that integration challenges often arise from social dynamics rather than technical gaps, emphasizing the need for sociotechnical approaches over purely algorithmic optimization.

Previous work in AI workflow adoption has focused on the types of tasks individuals use AI for: creation, information seeking, or advice support [11, 12]. The Selective Delegation Framework complements this work by furthering our understanding of how individuals first decide to delegate tasks to GenAI. *Delegation* can be understood as a process of deciding to entrust a task to a technological system prior to usage. In contrast, *usage* is the action taken after a decision to delegate has been reached.

Similarly, the Selective Delegation Framework extends prior work on decisions knowledge workers make about AI adoption. For instance, work on GenAI's metacognitive demands (monitoring and controlling one's own thought processes) describes the challenge of knowing whether, when, and how to incorporate GenAI into workflows as a worker's *'automation strategy'* [56, 68]. These prior studies observe knowledge work changing from content production to *'critical integration'* of AI output into ones workflow. This integration requires the application of an essential human skill, critical evaluation. Rather than blanket adoption, PMs in this study performed *'critical integration'*, assessing GenAI's usefulness for tasks, weighing its value and implications at multiple levels.

Furthermore, previous work has speculatively connected Argyris' *'double loop model of how organizations learn from errors'* to models of GenAI workflow adoption [4, 56, 57]. In the outer loop, workers go beyond simply modifying GenAI output to reflecting and making modifications to their workflows based on experience working with GenAI. However, such outer loop shifts have not been

empirically documented in prior work. Our study demonstrates that, at least for the PM profession, GenAI adoption can result in broader adaptation of workflows. More plainly, PMs described both (1) workflow adaptations (delegation practices); and (2) an evolution towards learning skills atypical to their PM role (Sec. 4.3).

Other work describes an *'iterative goal satisfaction'* GenAI framework, where the user attempts to satisfy a series of goals in stages with GenAI assistance. While Drosos et al. [22] outlined the *'iterative goal satisfaction'* framework, they primarily focused on illustrating challenges at each stage via a short-term laboratory study. In contrast, we offer empirical evidence for the framework, as PMs demonstrated and discussed going through each stage, thereby extending their work and contributing concrete cases grounded in real-world PM practice. While previous studies allude to a broader set of decisions around AI workflow adoption than those identified thus far, we have studied these decisions with a novel empirical basis, founded in the experiences of PMs using GenAI, enhancing our understanding of the integration of GenAI into workflows.

Our findings suggest that plans to adapt workflows in response to GenAI ought to be staged, prioritizing a deep understanding of worker values and GenAI's limitations, so workers can iteratively build trust in delegating tasks and assess its long-term impact on work practices. In industry, purchasing GenAI licenses without evaluating workers' situated practices reflects a short-sighted approach; treating AI as a "magic bullet" without a clear integration strategy poses a major challenge for executives [74].

Thus, while research notes a GenAI-driven shift from content creation to *critical integration* and an inherently human ability to adapt (workflows), workers like PMs, intellectually creative by nature, are justifiably concerned about its impact on their roles and expertise identity. This makes it clear that GenAI is no panacea, and calls for further research aimed at anchoring GenAI to PM values and its limitations for PM work specifically, as we've seen with SWEs [49, 73], extending the groundwork laid by our study.

5.2 Rethinking Software Development Roles

Our findings reinforce a critical point: understanding the evolving PM role is essential for the SE community. Traditionally, PMs have been peripheral teammates in the SDLC, e.g., translators between business needs and technical execution [24]. With GenAI, this view may no longer strongly hold, as PMs are taking on technical tasks previously reserved for SWEs. PMs in our study reported experimenting with prototyping, querying, and completing tasks that used to be reserved for SWEs.

Although some PMs we surveyed engaged in SWE-like tasks, the PM role is highly variable, leading to diverse GenAI adoption challenges. While much research and training target code generation, PMs reported a major barrier: limited GenAI literacy tailored to their responsibilities (Figure 3). They viewed GenAI literacy as essential for unlocking unrealized value from GenAI in their workflows. This gap between potential and realized value led to PMs hesitating and demonstrating only strategic adoption, as seen in the **Selective Delegation Framework**. Ironically, GenAI adoption, meant to empower developers, can create tension by encroaching on tasks they find meaningful. PMs note a similar impact of GenAI on their work, i.e., in our study GenAI was perceived as a threat to their **identity**. This growing tension between empowerment and

the erosion of professional **identity** highlights the need for more intentional development, not just of GenAI tools, but also of roles, by facilitating clearer GenAI task **delegation** and thoughtful **workflow adaptations**. Recognizing early implications of role shifting and GenAI delegation behaviors, we call on both the PM's organization and the SE research community to critically reimagine the SDLC, not by simply reshuffling tasks, but by rethinking traditional assumptions about *how*, and by *whom*, software is engineered.

Specifically, we invite other researchers to investigate the impact of the blurred role demands we saw and what kinds of organizational support is needed. As one participant put it: *"We need to invest more resources & time into defining guardrails for each PM role & task so creativity can flourish. Like a trapeze artist, we need the confidence that there's a safety net underneath to encourage exploration of what's possible."* (P679). Organizational support for sharing PM-focused GenAI use cases enables others to realize its value, while sustained use fosters understanding and self-efficacy [37].

Finally, organizations mandating GenAI use [46] can ease the burden on individuals by clarifying for PMs (1) **how their shifting role will be evaluated**, aligning professional growth with company goals, and (2) **how to mentor direct reports and early-career PMs**, strengthening leadership skills noted as a need in our study. Proactive efforts also help organizations anticipate shifts in SE **team social dynamics** and evolving notions of **accountability**, as P821 noted *"The SWE role will move 'up the stack' to be a bit more high level. The PM job will move 'down the stack' and a bit closer to code. We will meet in the middle and swap a lot of accountabilities"*.

6 Limitations

Like all scientific studies, our work has some limitations. First, it was conducted within a single company, which may limit generalizability. However, single case studies can yield important insights transferable to other contexts [26], and the company's global, multilingual teams provide greater diversity than typical single-site research. Also, the company culture strongly encourages GenAI use; a study where this is not the case would offer a valuable complementary perspective. Our study also shares limitations common to qualitative research, including potential issues in question design [10], overlooked insights due to researcher inexperience [38], and self-reporting bias in interviews [18]. To mitigate these risks, interviews and primary qualitative analyses were conducted by researchers trained in qualitative methods, and we avoided mandatory open-ended survey questions and leading interview prompts.

7 Conclusion

GenAI is reshaping knowledge work, particularly for PMs—key contributors in SE teams and early adopters of these tools. Given their pivotal role, their experiences offer valuable guidance for designing GenAI workplace tools. Here, we presented PMs' current GenAI usage and beliefs; the Sec. 4.2 framework, outlining how PMs navigate identity, accountability, and personal and social dynamics when delegating tasks to GenAI across individual, team, and organizational levels; and their workflow adaptations and role shift perceptions. We discussed how our findings expand understanding of AI workflow adoption, the impact on software development, and recommendations for organizations and SE research. We aim to foster inclusive dialogue that empowers PMs to shape, rather than simply adapt to, GenAI workflows.

References

- [1] Silvia Mara Abrahao Gonzales, John Grundy, Mauro Pezzè, Margaret-Anne Storey, and Damian Andrew Tamburri. 2025. Software Engineering by and for Humans in an AI Era. (2025).
- [2] Iftekhhar Ahmed, Aldeida Aleti, Haipeng Cai, Alexander Chatzigeorgiou, Pinjia He, Xing Hu, Mauro Pezzè, Denys Poshyvanyk, and Xin Xia. 2025. Artificial Intelligence for Software Engineering: The Journey so far and the Road ahead. *ACM Transactions on Software Engineering and Methodology* 34, 5 (2025), 1–27.
- [3] Zhamri Che Ani, Zauridah Abdul Hamid, and Nur Nazifa Zhamri. 2023. The recent trends of research on GitHub copilot: A systematic review. In *International Conference on Computing and Informatics*. Springer, 355–366.
- [4] Chris Argyris. 1977. Double loop learning in organizations. *Harvard business review* 55, 5 (1977), 115–125.
- [5] Jinheon Baek, Sujay Kumar Jauhar, Silviu Cucerzan, and Sung Ju Hwang. 2024. Researchagent: Iterative research idea generation over scientific literature with large language models. *arXiv preprint arXiv:2404.07738* (2024).
- [6] Joel Becker, Nate Rush, Elizabeth Barnes, and David Rein. 2025. Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity. *arXiv preprint arXiv:2507.09089* (2025).
- [7] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.
- [8] Hugh Beyer and Karen Holtzblatt. 1999. Contextual design. *interactions* 6, 1 (1999), 32–42.
- [9] Alexander Bick, Adam Blandin, and David J Deming. 2024. *The rapid adoption of generative AI*. Technical Report. National Bureau of Economic Research.
- [10] Peter Birmingham and David Wilkinson. 2003. *Using research instruments: A guide for researchers*. Routledge.
- [11] Michelle Brachman, Amina El-Ashry, Casey Dugan, and Werner Geyer. 2024. How Knowledge Workers Use and Want to Use LLMs in an Enterprise Context. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '24). Association for Computing Machinery, New York, NY, USA, Article 189, 8 pages. doi:10.1145/3613905.3650841
- [12] Michelle Brachman, Amina El-Ashry, Casey Dugan, and Werner Geyer. 2025. Current and Future Use of Large Language Models for Knowledge Work. *arXiv preprint arXiv:2503.16774* (2025).
- [13] Virginia Braun. 2025. Dr Braun(love) - or: how I learned to stop worrying and love GenAI (yeah nah). <https://www.youtube.com/watch?v=3zFQC3c-eQk>. Accessed: 2025-08-01.
- [14] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [15] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.
- [16] Elizabeth Bruch and Fred Feinberg. 2017. Decision-making processes in social contexts. *Annual review of sociology* 43, 1 (2017), 207–227.
- [17] Jenna Butler, Jina Suh, Sankeerti Haniyur, and Constance Hadley. 2025. Dear Diary: A randomized controlled trial of Generative AI coding tools in the workplace. In *2025 IEEE/ACM 47th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 319–329.
- [18] George Chalhouh and Ivan Flechais. 2022. Data protection at a discount: Investigating the ux of data protection from user, designer, and business leader perspectives. *Proceedings of the ACM on Human-computer Interaction* 6, CSCW2 (2022), 1–36.
- [19] Ellen Chisa. 2014. Evolution of the product manager. *Commun. ACM* 57, 11 (2014), 48–52.
- [20] James Cortada. 2009. *Rise of the knowledge worker*. Routledge.
- [21] Fabrizio Dell'Acqua, Edward McFowland III, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayner, François Candelon, and Karim R Lakhani. 2023. Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* 24-013 (2023).
- [22] Ian Drosos, Advait Sarkar, Xiaotong Xu, Carina Negreanu, Sean Rintel, and Lev Tankelevitch. 2024. "It's like a rubber duck that talks back": Understanding Generative AI-Assisted Data Analysis Workflows through a Participatory Prompting Study. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–21.
- [23] Christof Ebert. 2007. The impacts of software product management. *Journal of Systems and Software* 80, 6 (2007), 850–861. doi:10.1016/j.jss.2006.09.017
- [24] Christof Ebert. 2014. Software product management. *IEEE Software* 31, 3 (2014), 21–24.
- [25] Hannah Estes, Katie Hollowell, John Cheek, Shiv Patel, Hanna Reese, and Kathryn T. Stolee. 2025. Advancing Software Product Management Education: Insights from an Industry Survey. In *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering* (Clarion Hotel Trondheim, Trondheim, Norway) (FSE Companion '25). Association for Computing Machinery, New York, NY, USA, 999–1009. doi:10.1145/3696630.3727256
- [26] Bent Flyvbjerg. 2006. Five misunderstandings about case-study research. *Qualitative inquiry* 12, 2 (2006), 219–245.
- [27] GitHub Block. 2024. Goose: Context-aware AI assistant. GitHub Pages. <https://block.github.io/goose/>. Accessed: 08-Sept-2025.
- [28] GitHub Copilot. 2025. *GitHub Copilot*. <https://copilot.github.com/>. Accessed: 2025-09-08.
- [29] GitLab. 2024. GitLab Duo: Your AI companion for the entire DevSecOps lifecycle. <https://about.gitlab.com/gitlab-duo/>. Accessed: 8 September 2025.
- [30] Google. 2025. Gemini Code Assist. <https://cloud.google.com/gemini/docs/codeassist>. Accessed: 2025-09-08.
- [31] Vincent J Hellendoorn, Jason Tsay, Manisha Mukherjee, and Martin Hirzel. 2021. Towards automating code review at scale. In *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 1479–1482.
- [32] Sami Hyrynsalmi, Arho Suominen, and Marko Seppanen. 2021. A Bibliographical Study of Software Product Management Research. In *2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. 1–8.
- [33] David H Jonassen. 2012. Designing for decision making. *Educational technology research and development* 60, 2 (2012), 341–359.
- [34] Mansi Khemka and Brian Houck. 2024. Toward Effective AI Support for Developers: A survey of desires and concerns. *Commun. ACM* 67, 11 (2024), 42–49.
- [35] Hans-Bernd Kittlaus and Samuel A Fricker. 2017. Software product management. Berlin: SpringerVerlag GmbH Germany 298 (2017).
- [36] Charlotte Kobiella, Yarhy Said Flores López, Franz Waltenberger, Fiona Draxler, and Albrecht Schmidt. 2024. "If the Machine Is As Good As Me, Then What Use Am I?" – How the Use of ChatGPT Changes Young Professionals' Perception of Productivity and Accomplishment. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 1018, 16 pages. doi:10.1145/3613904.3641964
- [37] Charlotte Kobiella, Teodora Mitrevska, Albrecht Schmidt, and Fiona Draxler. 2025. When Efficiency Meets Fulfillment: Understanding Long-Term LLM Integration in Knowledge Work. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work*. 1–15.
- [38] Benjamin Koskei and Catherine Simiyu. 2015. Role of interviews, observation, pitfalls and ethical issues in qualitative research methods. *Journal of Educational Policy and Entrepreneurial Research* 2, 3 (2015), 108–117.
- [39] Matthew Law and Rama Adithya Varanasi. 2025. *Generative AI and Changing Work: Systematic Review of Practitioner-Led Work Transformations Through the Lens of Job Crafting*. Springer Nature Switzerland, 131–152. doi:10.1007/978-3-031-92823-9_10
- [40] Eilat Lev Ari, Maayan Roichman, and Eran Toch. 2024. Strategies of Product Managers: Negotiating Social Values in Digital Product Design. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 227, 17 pages. doi:10.1145/3613904.3642409
- [41] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 381, 18 pages. doi:10.1145/3613904.3642114
- [42] Weixin Liang, Yuhui Zhang, Hancheng Cao, Binglu Wang, Daisy Yi Ding, Xinyu Yang, Kailas Vodrahalli, Siyu He, Daniel Scott Smith, Yian Yin, et al. 2024. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI* 1, 8 (2024), A10a2400196.
- [43] Fritz Machlup. 1962. *The production and distribution of knowledge in the United States*. Vol. 278. Princeton university press.
- [44] Andrey Maglyas, Uolevi Nikula, and Kari Smolander. 2013. What are the roles of software product managers? An empirical investigation. *Journal of Systems and Software* 86, 12 (2013), 3071–3090. doi:10.1016/j.jss.2013.07.045
- [45] Andrey Maglyas, Uolevi Nikula, Kari Smolander, and Samuel A. Fricker. 2017. Core software product management activities. *Journal of Advances in Management Research* 14, 1 (02 2017), 23–45.
- [46] Bernard Marr. 2025. Microsoft Makes AI Mandatory For Employees: What It Means For Your Career. <https://www.forbes.com/sites/bernardmarr/2025/07/08/microsoft-makes-ai-mandatory-for-employees-what-it-means-for-your-career/>. Forbes, Jul 08.
- [47] Microsoft. 2025. The year the Frontier Firm is born. <https://www.microsoft.com/en-us/worklab/work-trend-index/2025-the-year-the-frontier-firm-is-born>. , Jul 08.
- [48] Matthew B Miles and A Michael Huberman. 1994. *Qualitative data analysis: An expanded sourcebook*. sage.
- [49] Courtney Miller, Rudrajit Choudhuri, Mara Ulloa, Sankeerti Haniyur, Robert DeLine, Margaret-Anne Storey, Emerson Murphy-Hill, Christian Bird, and Jenna L Butler. 2025. "Maybe We Need Some More Examples:" Individual and Team Drivers of Developer GenAI Tool Use. *arXiv preprint arXiv:2507.21280* (2025).

- [50] Ike Obi, Jenna Butler, Sankeerti Haniyur, Brian Hassan, Margaret-Anne Storey, and Brendan Murphy. 2025. Identifying factors contributing to “bad days” for software developers: A mixed-methods study. In *2025 IEEE/ACM 47th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 1–11.
- [51] David OBrien, Sumon Biswas, Sayem Mohammad Imtiaz, Rabe Abdalkareem, Emad Shihab, and Hriday Rajan. 2024. Are prompt engineering and todo comments friends or foes? an evaluation on github copilot. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 1–13.
- [52] Chalani Oruthotaarachchi and Janaka Wijayanayake. 2025. Aligning Software Product Management with Software Engineering Concepts: A Systematic Literature Review. *Journal of Information Systems Engineering and Business Intelligence* 11 (Jul. 2025), 143–159. doi:10.20473/jisebi.11.2.143-159
- [53] Srishti Palani and Gonzalo Ramos. 2024. Evolving Roles and Workflows of Creative Practitioners in the Age of Generative AI. In *Proceedings of the 16th Conference on Creativity & Cognition (Chicago, IL, USA) (C & C '24)*. Association for Computing Machinery, New York, NY, USA, 170–184. doi:10.1145/3635636.3656190
- [54] Nishant Ashokkumar Parikh. 2025. Managing AI-First Products: Roles, Skills, Challenges, and Strategies of AI Product Managers. *IEEE Engineering Management Review* (2025), 1–11. doi:10.1109/EMR.2025.3530942
- [55] Frederic Pattyn. 2024. The Critical Role of Product Managers and Their Responsibilities in Software Startups: A systematic Literature Review. *American Journal of Engineering and Technology Management* (2024).
- [56] Advait Sarkar. 2023. Exploring perspectives on the impact of Artificial Intelligence on the Creativity of Knowledge Work: Beyond Mechanised Plagiarism and Stochastic parrots. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*, 1–17.
- [57] Advait Sarkar. 2024. AI Should Challenge, Not Obey. *Commun. ACM* 67, 10 (2024), 18–21.
- [58] Hope Schroeder, Marianne Aubin Le Quéré, Casey Randazzo, David Mimno, and Sarita Schoenebeck. 2025. Large Language Models in Qualitative Research: Uses, Tensions, and Intentions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 481, 17 pages. doi:10.1145/3706598.3713120
- [59] Hamid Shamszare and Avishek Choudhury. 2023. Clinicians’ perceptions of artificial intelligence: focus on workload, risk, trust, clinical decision making, and clinical integration. In *Healthcare*, Vol. 11. MDPI, 2308.
- [60] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2023. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–57.
- [61] Genevieve Smith, Natalia Luka, Merrick Osborne, Brian Lattimore, Jessica Newman, Brandie Nonnecke, and Brent Mittelstadt. 2025. Responsible Generative AI Use by Product Managers: Recoupling Ethical Principals and Practices. (2025). On Arxiv.
- [62] Olga Springer and Jakub Miler. 2018. The Role of a Software Product Manager in Various Business Environments. In *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. 985–994.
- [63] Olga Springer and Jakub Miler. 2022. A comprehensive overview of software product management challenges. *Empirical Software Engineering* 27 (2022).
- [64] Margaret-Anne Storey, Rashina Hoda, Alessandra Maciel Paz Milani, and Maria Teresa Baldassarre. 2025. Guiding principles for mixed methods research in software engineering. *Empirical Software Engineering* 30, 5 (2025), 138.
- [65] Margaret-Anne Storey, Daniel Russo, Nicole Novielli, Takashi Kobayashi, and Dong Wang. 2024. A Disruptive Research Playbook for Studying Disruptive Innovations. *ACM Trans. Softw. Eng. Methodol.* 33, 8, Article 195 (Nov. 2024), 29 pages. doi:10.1145/3678172
- [66] Jobin Alexander Strunk, Leonardo Banh, Anika Nissen, Gero Strobel, and Stefan Smolnik. 2024. To delegate or not to delegate? Factors influencing human-agent IS interaction. *UNK* (2024).
- [67] Hari Subramonyam, Divy Thakkar, Andrew Ku, Juergen Dieber, and Anoop K Sinha. 2025. Prototyping with prompts: Emerging approaches and challenges in generative ai design for collaborative software teams. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.
- [68] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The metacognitive demands and opportunities of generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–24.
- [69] Bianca Trinkenreich, Fabio Calefato, Geir Hanssen, Kelly Blincoe, Marcos Kalinowski, Mauro Pezzè, Paolo Tell, and Margaret-Anne Storey. 2025. Get on the Train or be Left on the Station: Using LLMs for Software Engineering Research. In *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering*. 1503–1507.
- [70] Rosalia Tufano, Simone Masiero, Antonio Mastropaolo, Luca Pascarella, Denys Poshyvanyk, and Gabriele Bavota. 2022. Using pre-trained models to boost code review automation. In *Proceedings of the 44th international conference on software engineering*. 2291–2302.
- [71] Mara Ulloa, Jenna Butler, Sankeerti Haniyur, Courtney Miller, Barrett Amos, Sarkar Advait, and Margaret. Storey. 2026. Supplementary Material for “Product Manager Practices for Delegating Work to Generative AI: “Accountability Must Not Be Delegated to Non-human Actors”. Zenodo. doi:10.5281/zenodo.17822096
- [72] Mara Ulloa, Blaine Rothrock, Faraz S Ahmad, and Maia Jacobs. 2022. Invisible clinical labor driving the successful integration of AI in healthcare. *Frontiers in Computer Science* 4 (2022), 1045704.
- [73] Justin D Weisz, Shraddha Vijay Kumar, Michael Muller, Karen-Ellen Browne, Arielle Goldberg, Katrin Ellice Heintze, and Shagun Bajpai. 2025. Examining the use and impact of an ai code assistant on developer productivity and experience in the enterprise. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–13.
- [74] Shixian Xie, John Zimmerman, and Motahhare Eslami. 2025. Exploring What People Need to Know to be AI Literate: Tailoring for a Diversity of AI Roles and Responsibilities. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [75] Bhada Yun, Dana Feng, Ace S Chen, Afshin Nikzad, and Niloufar Salehi. 2025. Generative ai in knowledge work: Design implications for data navigation and decision-making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [76] Jiyang Zhang, Sheena Panthaplackel, Pengyu Nie, Junyi Jessy Li, and Milos Gligoric. 2022. Codit5: Pretraining for source code and natural language editing. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*. 1–12.