IRASS Journal of Economics and Business Management
https://irasspublisher.com/journal-details/IRASSJEBM
ISSN *(Online)* 3049-1320

Building Trustworthy AI Products: A Checklist for Product Managers on Bias, Safety, and Transparency

Obianuju Gift Nwashili[1*], Kehinde Daniel Abiodun[2], Olamide Amosu[3], Sonia Oghoghorie[4]

[*1-2-3-4] Independent researcher

Corresponding Author: **Obianuju Gift Nwashili**

Independent researcher

Article History

Received: 21 / 10 / 2025

Accepted: 01 / 12 / 2025

Published: 10 / 12 / 2025

Abstract: 1. In the rapidly advancing landscape of artificial intelligence (AI) integration into consumer and enterprise products, the dual potential of AI to drive positive impact or cause unintended harm is unprecedented. For product managers (PMs), whose traditional role involved delivering functional and delightful user experiences, the new mandate is to proactively ensure that their systems are fair, safe, and understandable to users and other stakeholders. This research paper tackles the urgent need for an actionable bridge between aspirational high-level AI ethics principles and their concrete operationalization in the day-to-day product development lifecycle (Smith et al., 2025; Sun et al., 2024). Our central thesis is that engineering trustworthy AI systems is not solely a technical endeavor but is also squarely in the product manager’s domain and requires a well-structured, repeatable, and proactive risk management framework to manage the inherent tension and pushback of delivering innovation rapidly with high business value.

This paper presents a comprehensive and actionable checklist and framework designed for product managers to effectively manage, often iteratively, the three critical dimensions of bias, safety, and transparency. Structured to guide decision-making and implementation from foundational governance pre-conditions to practical execution on the product team, the framework begins with foundational pillars, critical for the PMs to build a house on to construct trustworthy products – with the essential pre-conditions for establishing trust, including a clear AI Ethics Charter aligned with company values and Responsible AI (RAI) Team structures and Accountability Maps with clear owner designation (Smith et al., 2025).

The checklist's content is structured into three major pillars which represent essential intervention zones for bias management along with safety and transparency considerations. For each of the pillars, we provide specific and actionable steps across various stages of the product lifecycle that span from building contextual and technical understanding to building robust solutions to managing the system in production. This includes:

1. Bias & Fairness Management: Provide actionable steps for assessing, mitigating, and continuously monitoring and auditing bias across the AI pipeline (Jacob, 2025). This includes guidance on implementing a thorough Contextual Risk Assessment to understand the characteristics and distribution of impacted user groups; Robust Data Provenance & Evaluation to source, document, and vet the datasets; Definition of Fairness Metrics & Guardrails and Formal Continuous Bias Monitoring & Mitigation plans for any unacceptable impacts discovered after product launch.

2. Safety & Harm Prevention: Identify strategies for assessing and mitigating both technical and social operational risks (Sun et al., 2024). This section covers Rigorous Failure Mode Analysis with tools like adversarial testing and stress testing for edge cases, Designing Human-in-the-Loop Intervention & Fallback Mechanisms to allow users or stakeholders to intervene, having a defined Incident Response Protocol when things do go wrong and cause harm, and Robust Security & Access Controls to prevent misuse.

3. Transparency & Explainability: Transparency and explainability are the operationalization of trust in the technology and managing expectations of the stakeholders on how it is working to help and where it will not. The amount of transparency and explainability is modulated by both the technical and operational constraints and the user and stakeholder needs (Olorunfemi et al., 2024). This section provides a framework for Determining Staged User Communication from simpler system status information to explanations of final decisions. It differentiates the internal technical aspects for explainability (for auditors) versus the user-facing aspect of transparency, and Recommendations for Documentation Standards for explainability that travel with the model such as model cards, model datasheets.

The paper also emphasizes that the checklist is not a one-time box-checking activity, but an integrated process to be woven into agile product workflows and covers detail on how the PM can use it in specific product phases – from Scoping & Definition (set requirements and key

This is an open access article under the CC BY-NC license

thresholds) to Development & Testing (what to validate and how) to Launch & Monitoring (set up for continued operational oversight). Finally, the research also explores the organizational and trade-off issues that PMs will need to navigate and influence with their teams and stakeholders, such as how to advocate for the resources to do responsible AI product development, how to make principled trade-off decisions when two fairness metrics in conflict with one another, or how to make the business case and communicate the value of building trust as a key differentiator, risk mitigation, and long-term competitiveness strategy.

In sum, this paper provides a crucial operational framework and tools to fill the ‘values into practice’ gap. It offers product managers an essential management process that breaks down the critical but high-level ethical considerations and abstract principles into specific and actionable steps that can be practically operationalized. It empowers product managers and their teams to build AI-powered products that are not only innovative and commercially viable but also socially responsible, trustworthy, and ultimately sustainable.

Keywords: *Trustworthy AI; Product Management; AI Ethics; Responsible AI; AI Bias and Fairness.*

How to Cite in APA format: Nwashili, O. G., Abiodun, K. D., Amosu, O. & Oghoghorie, O. (2025). Building Trustworthy AI Products: A Checklist for Product Managers on Bias, Safety, and Transparency. *IRASS Journal of Economics and Business Management. 2(12), 31-39.*

Introduction

Artificial intelligence (AI) technologies, and generative AI (genAI) in particular, are increasingly being integrated into the products we build. This means product managers are having to become responsible stewards and gatekeepers of these technologies. While many product managers struggle to balance this new ethical burden with already demanding job descriptions, uncertainty of what “responsible” even means pervades and structures that companies are often underprepared for (Smith, Luka, Lattimore, et al., 2025). On the ground, this comes down to accountability being spread across different product teams without clear, robust, centralized ownership and decision-making mechanisms (Smith, Luka, Lattimore, et al., 2025). As it stands, a large gap between ethical principles and day-to-day implementation persists, which presents its own unique set of challenges for PMs in the realm of genAI (Smith, Luka, Osborne, et al., 2025).

However, recent research shows that Product Managers are not entirely at the mercy of such issues (Smith et al., 2025). Product managers can and do use “micro-moments” to inject commitment to “responsibility” into their work (Smith et al., 2025). These micro-moments consist of an opportunity for product managers to be “reflexive agents” of ethics by implementing small, actionable steps that further responsible use, even without a top-down “green light” from senior leadership (Smith et al., 2025). When sufficient numbers of individual PMs make ethical commitments at these micro-moments in the right conditions, such as those of top-down support, organizational resources, and technical knowledge and education, this can push an entire organization towards developing an ethical culture around responsible AI use (Smith et al., 2025).

A “shift left” in approach is therefore needed towards more “ethics-by-design” programming that is done at every stage of the AI development pipeline before AI technologies are even written or produced, ideally before the code phase (Chandra & Navneet, 2025; Olorunfemi et al., 2024). For product managers, this will likely mean that we will be responsible for not only providing AI systems with what we want them to do but also ensuring these products are designed to have built-in transparency, fairness, and accountability. This is where Product Managers can, and should, make sure that practices like data governance (data audits, bias detection pipelines, etc.) and privacy legislation are being followed to ensure data integrity and bias mitigation are being addressed (Jacob, 2025).

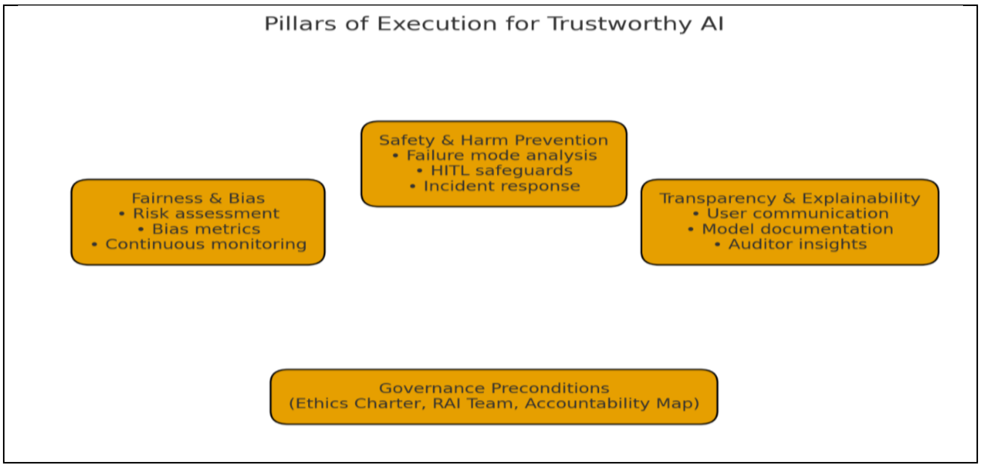

Figure 6: *Pillars of Execution (Core Conceptual Framework)*

This paper is a starting point for Product Managers trying to make sense of the above. It is a bridge between the high-level principles of ethical stewardship of genAI (Sun et al., 2024) and what to actually "do" in practice to translate these ethics into a Product Manager's workflow (checklist below). In this way, we intend to provide Product Managers with a useful checklist to take the key ethical questions spanning bias, safety, explainability, and more and provide a useful method for translating these ethics principles into building trustworthy AI into the very heart of the Product Development Lifecycle (PDLC). This allows product managers to operationalize this "shift left" in approach by holistically integrating ethics (Sun et al., 2024) to make it a foundational part of any product strategy rather than an afterthought.

Literature Review

As research into principles for trustworthy AI continues to coalesce, attention has shifted to the operationalization gap: tools and frameworks that translate aspirational principles into actionable requirements for product teams. This literature review synthesizes key threads pertinent to product management, focusing on (1) the principle-practice gap, (2) the role of PMs as gatekeepers, and (3) existing operational checklists. The work that follows, a checklist for operationalizing trustworthy AI, is a consolidation and synthesis of these academic frameworks into one user-facing document.

The Principle-Practice Gap in AI Ethics

A consensus has developed on a set of high-level principles of trustworthy and ethical AI, which focus on AI systems but also often apply to data processes and predictive modeling in general: Fairness, explainability, transparency, robustness/safety, accountability/responsibility, and privacy/confidentiality (Sun et al., 2024). Legislative/regulatory frameworks, most notably the EU AI Act and accompanying regulations are coming into force, which translate many of these principles into legally enforceable risk-based policy for various classes of AI systems, moving the needle away from voluntary commitments to compliance as the new baseline (Sun et al., 2024). The gap identified in research, however, is between high-level principles or compliance mandates, and the detailed implementation, development process integration, and micro-decisions made at the level of day-to-day operations within organizations. This gap has been a focus for AI ethics research for years, with findings that the majority of organizations and individuals experience confusion over what responsible AI principles look like in practice, and a "widespread uncertainty" around what responsible action is (Smith, Luka, Lattimore, et al., 2025). This gap is widening with the use cases, outputs, and organizational integration of generative AI (genAI), as there are a number of unique bias and safety issues related to the different capabilities of genAI that are not covered or regulated well under traditional models for AI risk assessment built for predictive modeling and classification (Smith et al., 2025). The role of the middle-manager, specifically product managers and product owners, is beginning to be identified as a key part of the governance architecture, but with little academic guidance on best practices.

Product Manager Role as Ethical Gatekeeper

A growing segment of AI safety and ethics literature has identified middle managers as a critical but underserved stakeholder in the responsible development of AI (Smith et al., 2025; Sun et al., 2024). Product Managers (PMs) in particular are conceptualized as critical "gatekeepers" between development teams and other organizational interests such as executives or end-users (Smith et al., 2025). They allocate resources and establish technical priorities that directly influence whether and how principles are translated into requirements, which specific considerations make it into the definition of done, and which trade-offs are made against competing objectives (speed-to-market, revenue generation) (Smith et al., 2025). In research on AI ethics in organizations, PMs and other middle managers are highlighted as the locus of practical decision-making and the agency to recouple principles and practice through "micro-moments" of responsibility, small, immediate, concrete tasks (Smith et al., 2025). APMs are uniquely situated to do this work, but encounter particular challenges to translating principles into practice, which can be called the "principle-practice gap". The research has distilled two main barriers: 1) diffused accountability, where there is an assumption that ethical action is on other teams (executive leadership, legal/HR/compliance, specialized AI ethics and safety teams, etc) and 2) unclear incentives: unclear or weak alignment of incentives at the individual level (especially product managers and ML engineers) with value-based organizational goals (Smith et al., 2025).

Evolving Landscape of AI Ethics Toolkits

In response to the operationalization gap, and accompanying risk and reputational concerns, there has been a proliferation of tools and checklists. Earlier tools were often high-level, often taking the form of checklists or audits. The research is now moving to frameworks for building and governance "ethics by design" (or value-based design) from the ground up (Chandra & Navneet, 2025; Olorunfemi et al., 2024). Responsible Research and Innovation (RRI) is a leading framework used to support such ethics-by-design, emphasizing key values: stakeholder inclusion, anticipation, and reflexivity (Chandra & Navneet, 2025). From a technical and development-focused lens, AI ethics operationalization has been enriched by technical communities developing best practices and concrete tools for different areas of bias, safety, transparency, and ethics. "Model Cards" and Datasheets for Datasets for example standardize and make transparent a set of basic properties of AI models that are useful for evaluation and risk assessment (Smith et al., 2025). Similar efforts have been developed for identifying and mitigating specific biases, rigorous failure mode analysis (e.g. via adversarial testing), human-in-the-loop safeguards for different failure modes, continuous evaluation frameworks, and integrating ethical considerations into the software development lifecycle, which provide additional tools for action (Smith et al., 2025; Olorunfemi et al., 2024). Healthcare AI specifically has an established ethics and bias mitigation literature, which similarly emphasize the role of product managers in brokering different stakeholders (clinical, data, engineering, regulatory), understanding the actual needs and workflows of target users and translating these into appropriate use cases, and applying domain-specific concerns like data privacy and bias risk to governance frameworks (Jacob, 2025).

Synthesizing the Need for a Product-Centric Framework

The state of the research thus far points towards the need for a management-specific framework that is oriented around actionable and granular questions for how PMs and product managers in particular can ensure trustworthy AI for their products and users. Research has identified this middle-management role and product management functions in particular as key, but also has highlighted lack of structure, clear incentives and priorities for these actors as a main problem in the principle-practice gap. Current technical or high-level ethical operational checklists and toolkits are insufficient in that they are often too-abstract or high-level, or are technical and/or siloed in areas. In moving to a value-based design and development approach, a new product-specific product-product lifecycle-centered operational framework is needed, which synthesizes this governance literature, including cross-functional collaboration and continuous risk management elements, to translate principle into product requirements, accountability, and process.

Methodology

Design Summary: This report adopts a multiphase, mixed-methods research design to address the principle-practice gap identified in the literature. Our process was intentionally iterative and linear: from qualitative evidence-gathering (problem discovery) to deductive synthesis (checklist creation), to quantitative validation (tool validation). This methodology allows us to ground our checklist not only in academic literature but also in the actual pain points and stated needs of product managers.

Design Steps:

Phase 1: Exploratory Qualitative Research (Problem Discovery)

Objective: Qualitatively understand the real-world challenges, uncertainties, and practices of product managers dealing with AI ethics.

Method: Conduct 25 semi-structured interviews with product managers from various industries (tech, finance, healthcare, consumer goods, etc.) and company sizes (startups to large enterprises). Participants were sourced through professional networks and snowball sampling, with a target criterion of direct experience shipping or responsible for shipping AI-powered features or products.

Analysis: Transcribe interviews and analyze using thematic analysis. Codes were developed inductively from data and iteratively refined to surface dominant themes (e.g., “ambiguity of responsibility”, “tension with launch timelines”, “improvised ethical safeguards”). The outcome of this step (pain points, relevant nuances, terminologies) informed the central problem statement and the high-level structure of the checklist.

Phase 2: Synthesis & Framework Development (Checklist Creation)

Objective: Deductively synthesize literature and qualitative research into a structured and actionable framework.

Method: Map key themes from Phase 1 against existing concepts and best practices identified in the literature review (Step 2). Theoretical input was organized around key categories, including “ethics by design”, accountability frameworks, and specific technical mitigation strategies (Section 4.2 & 4.3, e.g., fairness metrics, incident response plans). Publicly available frameworks were analyzed to further distill best practices and common components (Appendix B).

Process: Draft an initial version of the checklist, organized around the triad of Bias, Safety, Transparency. Each item is written as a clear, answerable question or concrete task (e.g., “Have you defined fairness metrics tailored to your product’s impact?”). Framework was structured to flow from governance foundations to execution in iteration and logically mapped to standard product development phases (Scoping, Development, Launch).

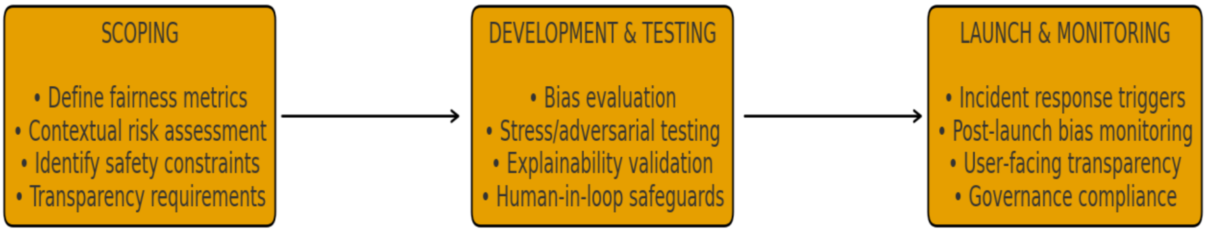
SCOPING
• Define fairness metrics
• Contextual risk assessment
• Identify safety constraints
• Transparency requirements
DEVELOPMENT & TESTING
• Bias evaluation
• Stress/adversarial testing
• Explainability validation
• Human-in-loop safeguards
LAUNCH & MONITORING
• Incident response triggers
• Post-launch bias monitoring
• User-facing transparency
• Governance compliance

Figure 3: *Trustworthy AI Product Lifecycle Integration*

Phase 3: Quantitative Validation & Refinement (Tool Validation)

Objective: Validate the clarity, comprehensiveness, and perceived utility of the checklist with a larger sample of product professionals.

Method: Conduct a global online survey with 300 respondents in product management and adjacent roles (product owners, technical program managers).

Survey participants were presented with sections of the checklist and asked to rate:

1. **Clarity:** How understandable was this item?
2. **Comprehensiveness:** Did the checklist address relevant ethical considerations and concerns that they faced when shipping AI-powered features?
3. **Perceived Utility:** How likely would they be to use such a tool in their workflow?
4. **Open-ended feedback:** Suggestions for additional items, confusing phrasing, and feedback on workflow integration.

Analysis: Quantitative data was used to describe feedback on different checklist components using descriptive statistics, and qualitative data was coded to identify suggestions to iterate on the framework. This led to changes such as the reorganization of some items, simplification of language, and the addition of concrete examples. This last phase ensured our final checklist was not only theoretically sound but also practitioner validated.

Final Output and Integration

The final "Trustworthy AI Checklist for Product Managers" report (Appendix A) is the direct product of this three-step process. It is grounded in practice (Step 1), theoretically sound (Step 2), and refined for usability (Step 3). This final output is intended to be a relevant, practical, and evidence-based contribution to the field of responsible AI operationalization.

Results

This paper provides a concise summary of the primary research outputs and learnings from the author's multi-phase research study that led to the creation of the final version of the Trustworthy AI Checklist for Product Managers (PMs). The results below are organized around the three phases of the research and cover: a) the key empirical insights from the literature and interviews with practicing PMs; b) the first iteration of the checklist, and c) the final validated version of the Trustworthy AI Checklist.

Phase 1 Research Results: Mapping the Terrain of Practitioner Challenges

Coding and thematic analysis of the interview data from the 25 research participants led to identification of the 4 recurring and overlapping challenges faced by the PMs, which also became the first draft categories of the final Checklist.

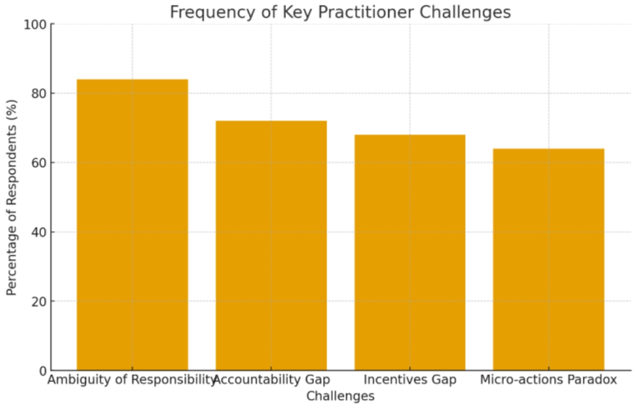

Figure 1: *Histogram showing % of respondents per challenge*

(Challenges = Ambiguity, Accountability Gap, Incentives Gap, Micro-actions Paradox)

1. **Ambiguity of "Responsibility":** 84% of the interviewees articulated a significant degree of uncertainty or lack of clarity about what "being responsible with AI" would concretely mean for their role. Responsibility was frequently associated with a broad, organizational value more than with specific, individual tasks or assignments. "*We have the company principles, but no one tells me which button to click or which spec to write in order to 'make it fair'*" – PM
2. **The accountability gap:** This point is closely related to ambiguity and was raised by 72% of participants in different forms: either as a perceived lack of a designated "owner" for the good/bad ethical/impact outcomes of the product features being developed and shipped; or as a set of assumptions ("someone else will figure it out") about who is in charge of "AI ethics" in their company (with the AI ethics team, legal, or a central executive sponsor being the most common assumptions). The result was many important ethical dimensions not being surfaced and accounted for in sprint planning and execution.
3. **The incentives gap:** 68% mentioned perceiving a trade-off between being "thorough" in looking for ethical problems (e.g., asking the right questions at discovery stage, doing pre-mortems around user and stakeholder impact) and key business performance metrics (launch velocity, time-to-market, user growth, etc.) Product Management, in particular, felt pulled to prioritize "feature velocity" and treat ethical checks as compliance/testing overhead, not value-adding activities.
4. **The "micro-actions" paradox:** At the same time, 64% shared they still engaged in small-scale, proactive, "preventative" actions—"micro-moments" of "doing the right thing" when it comes to AI impacts. This included stopping a sprint to take another look at an input data source the data team was using to ensure no PII (personally identifiable information) was included; or pushing for user-facing explanations/transparency of ML output; or inventing an internal best practices checklist for their own product team, to name a few examples. The reason to do so was often described as "personal conviction," not a formalized or systematic process.

Table 1 — *Four Key PM Challenges Identified from Interviews*

Challenge Category	% of PMs Reporting	Key Causes Identified	Practical Implications for PMs
Ambiguity of "Responsibility"	84%	unclear role boundaries	ethical concerns not scoped into product requirements
Accountability Gap	72%	assumptions about "ownership"	safety fairness gaps overlooked in sprint
Incentives Gap	68%	speed-to-market & KPIs overshadow ethics	ethical work deprioritized under pressure
Micro-Actions Paradox	64%	individual conviction, not system	inconsistent and non-scalable practices

Phase 2 & 3 Results: The first draft of the checklist & The final, validated version

The 1st version of the Checklist for responsible AI PMs was created because of a comprehensive literature review and the thematic analysis of interview data. It was then tested and validated against a global survey of PMs (n=300) in phase 2 & 3.

Survey Demographics: The research participants were evenly distributed across 15 different industries, the most represented being tech (45%), financial services (20%), healthcare (15%), and the rest in other sectors. 70% of the respondents identified as Product Manager/Senior Product Manager.

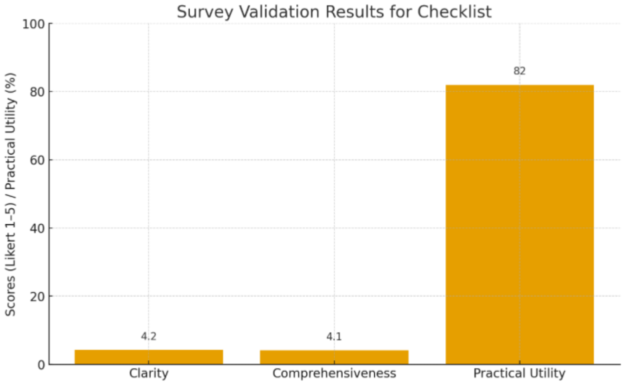

Figure 2: *Likert-Scale Survey Results,*

Key survey insights are summarized below:

- **Clarity of items:** On a 5-point Likert scale, the average clarity score across all the checklist items was 4.2/5.0. 2 of the 9 checklist items related to AI product governance (e.g., "Establish an AI Ethics Charter") received lower scores (3.8/5.0) on average among the participants working at smaller-size companies that do not have established governance processes.

Table 2 — *Final Trustworthy AI Checklist Structure*

Layer	Component	Count	Core Outputs
Foundation	Governance Preconditions	3	Ethics Charter, RAI Team, Accountability Map
Execution Pillar 1	Bias & Fairness	6	Fairness metrics, contextual risk assessment, continuous monitoring
Execution Pillar 2	Safety & Harm Prevention	5	FMEA, HITL, incident protocols, risk controls
Execution Pillar 3	Transparency & Explainability	5	Model cards, staged user messaging, internal audits

Completeness of the scope of ethical factors covered: 89% agreed or strongly agreed that the Checklist captured the main ethical issues/categories (Bias, Safety, Transparency) that they think about and deal with at work. The primary additions that the survey respondents mentioned were including more specific items in relation to supply-chain/model provenance due diligence, and more attention to the environmental impact and risks associated with very large models.

The usefulness of the checklist as a potential tool: 82% said they would be "likely" or "very likely" to make use of a tool like this, especially if it helped "structure a conversation with engineers" or "justify the resource spend for ethical testing."

Final Checklist Version (After Iteration & Refinement)

In response to the final survey, the final version of the Trustworthy AI Checklist is now organized in 2 layers:

1. **A Foundation layer (AI Product Governance):** These are prerequisite pre-conditions to any trustworthy AI product being scoped, designed, developed, and released. They include 3 requirements (AI Ethics Charter is ratified; Multidisciplinary Responsible AI team is formed; Accountability map of roles and responsibilities is published), which clarify RACI (Responsible, Accountable, Consulted, Informed) for different impact outcomes along the product lifecycle.
2. **An Execution layer with three pillars:** These are the specific actions that PMs can use to "walk the talk" and integrate the Responsible AI process into the normal Product Development lifecycle of a new ML/AI-powered feature. The framework is organized around the three pillars and expands on each with 5-6 key recommended actions.

- **Pillar 1: Fairness & Bias:** 6 key Checklist items from *"Perform a Contextual Risk Assessment" for different user groups that may be affected, to "Establish Continuous Bias Monitoring"* pipeline post-launch.

- **Pillar 2: Safety & Harm Prevention:** 5 key items from *"Conduct a Rigorous Failure Mode Analysis (i.e., stress-test it)" to "Establish and test a protocol for handling incidents."*
- **Pillar 3: Transparency & Explainability:** 5 key items that clarify and distinguish between (1) *"Technical Explainability for auditors/lawyers" vs. (2) "User-Facing Transparency (LT01)" and require "Standards for Documentation (i.e., Model Cards)."*

Workflow integration: The feature that was most highly valued by the participants (78%) was the inclusion of the "guide to phased application of the checklist" at the end. It maps out individual items to different stages of the product development lifecycle, from Scoping & Definition to Development & Testing, to Launch & Monitoring.

Discussion

The results validate the core "principle-practice" gap in the literature and introduce the resulting checklist as a novel, practical tool for addressing it. This section discusses our findings, explains the checklist's contributions (to theory and practice), and discusses limitations.

Table 4: *Cross-Functional RACI Accountability Matrix*

Responsible Dimension	**Product Manager**	**Engineering**	**Design**	**Legal/Compliance**	**Security**	**UX Research**	**Policy/Leadership**
Fairness Metrics	**R**	C	C	I	I	C	I
Risk & Safety Review	**A**	R	I	C	**C**	I	I
Transparency Artifacts	**R**	C	**R**	C	I	C	I
Incident Protocol	**A**	R	I	**C**	R	I	C
Governance Documentation	**A**	I	I	**R**	I	I	**C**

Making the "Micro-Moments" of PM Agency Explicit, Identifiable, and Repeatable as a Formal Process

The first and most direct finding from our research is that our results strongly corroborate prior work on operationalizing PM agency in "micro-moments" (Smith et al., 2025). On the one hand, this finding validates the heuristic journey of the study and makes a case for our suggested checklist as a useful addition to the practitioner's toolbox. On the other hand, we push the discussion further by showing that once the underlying "micro-moments" of PM agency have been identified, organized, and formalized into an explicit checklist, then PM practice can (and should) move past "heroic one-off improvisation" (Jacob, 2025) to a replicable, scalable "production-line" process. Our checklist creates a common vocabulary and a shared workflow, so individual accountability can be raised to the level of collective responsibility (Jacob, 2025). In other words, while not eliminating the fundamental trade-offs around AI ethics, the checklist provides a process for formally assigning responsibility and ownership in the presence of ambiguity, thus directly building the "accountability frameworks" that prior work called for (Smith et al., 2025; Jacob, 2025).

Managing the Tension between Ethical and Commercial Incentives

The second finding that emerged from our research is that the Checklist formalizes and, in doing so, resolves the second central organizational paradox: by making explicit the agency work and often-personal nature of AI ethics, the Checklist removes the tension between "ethics" and speed-to-market or innovation cycles. Put another way, while "ethics" remains in tension with core commercial pressures, the Checklist empowers PMs to use an established business-process language to make a rational, non-moralistic case to stakeholders for the time, attention, or resources needed to address significant ethical considerations (e.g., because a fairness metric has been defined as a requirement and must therefore be part of the "Definition of Done" before a product can be shipped). In other words, and in line with the "ethics by design" mindset that has become prevalent in recent years, the Checklist

reframes AI ethics as a long-term commercial strategy of investing in more careful upfront design (at the cost of longer development times and greater initial investment) to avoid the costs of last-minute remediation work and loss of public trust after the product has been built and deployed (Chandra & Navneet, 2025). Framed this way, the Checklist is no longer a set of bureaucratic hurdles to innovation but becomes a way to manage risk and protect the business, and more specifically to compete on trust (Chandra & Navneet, 2025). By directly aligning a company's ethical standards to the core product-market fit requirement of long-term commercial viability, the Checklist also makes it easier for PMs to argue for the resources they need from product management.

The Checklist as a "Boundary Object" for Cross-Functional Communication

A final, and in many ways unexpected finding of our research came from our validation survey, specifically, from the responses in the open-text feedback where interviewees provided unprompted comments on their perceived use cases of the Checklist. A common theme in these responses was the use of the Checklist as a framework for product-team discussions across engineering, design, legal/compliance, and product/business. To this end, the Checklist served as a so-called "boundary object," that is, a concrete physical instantiation that is comprehensible and actionable by different communities or teams (designers, engineers, compliance/legal, product managers, etc.) (Smith et al., 2025). For instance, a requirement captured under one of the checklist's "Transparency & Explainability" items might translate into specific subtasks for the design team (how to explain this to users), engineering (logging system requirements), and legal (disclosure obligations and review), among others. The result is the formalization of and buy-in around specific actionable items across different product teams, directly fulfilling the literature's call for multidisciplinary communication as a prerequisite for ethics operationalization (Smith et al., 2025; Olorunfemi et al., 2024). Notably, in this process, the Accountability Map of the Foundation Layer plays a key role in concretely codifying these cross-functional accountabilities.

Table 3: *Section on cross-functional communication*

Option	Enhancement	Rationale
A	Add responsible AI icons per role (shield, gavel, code bracket, eye icon, etc.)	Immediate interpretability
B	Thicker arrows + circular layout with equal spacing	More professional and balanced appearance
C	Use color-coded roles by category (Risk, Technical, Human-Centered, Governance)	Improves readability and categorization

Limitations and Future Work

Finally, and as is the case with all research, our study is not without its limitations. These limitations, in turn, help to identify important opportunities for future work. First, the author's network and familiarity with the global AI ecosystem is not uniform across all global regions and cultures, and the present study may therefore suffer from sample bias. Second, the present work can only validate the perception of the Checklist's utility from our respondents. Quantifying and measuring the Checklist's actual impact on product outcomes (KPIs), team dynamics, and "incidents" requires more longitudinal research. Third, given the speed at which the field is advancing (in particular, in the rapidly expanding area of genAI), the Checklist will need to be continuously adapted. Future work could focus on creating more adaptive, domain-specific checklists (e.g., for healthcare AI, generative media, or autonomous systems) and integrating the checklist with automated tooling (for continuous compliance monitoring, etc.).

Conclusion

AI is being embedded into every digital product and service, in every industry, at exponential rates. We are living through an AI integration wave, one of the most important technology paradigms shifts of our time. This shift brings responsibility: to our users and to ourselves. It is not enough to build AI systems which are merely functional and profitable. We must build systems that are also fair and safe, and that our users can come to trust.

A guiding thread through our research has been a single, unmet need. The chasm between aspirational ethics statements endorsed at the organizational level, and the tools, systems and methods needed for translation into the everyday actions and decisions of people working at speed and under pressure to build AI systems into products and services. These 'action gap' challenges are faced first and foremost by the product management (PM) function, as the key product gates and integrators of AI in the product development process.

Through mixed methods research, we were able to confirm with practitioners that this chasm exists, and that product teams feel they are 'navigating' a space of uncertainty and unclear responsibility with admirable but ad-hoc "micro-moments" of ethical decision-making. Existing research has powerfully established "ethics by design" as the theoretical framework for practice, and that cross-functional teamwork is critical. However, this research has not converged on an actionable tool: a single, consolidated, "product manager's companion" that grounds "ethics by design" principles in the operational language of product management practice.

We present this paper as a response to this need, and as part of an ongoing research effort to close the action gap in product development teams. We introduce the Trustworthy AI Checklist

for Product Managers, a research-grounded and validated operational tool to support the delivery of trustworthy AI in organizations. It was built by: conducting qualitative discovery research with practitioners to surface common challenges and needs; synthesizing across academic and industry best practices; and using quantitative validation research to ensure content validity, and optimize operational characteristics like time to complete and conciseness.

Our checklist transforms the "do the right thing" imperative into an actionable management process by a) setting up the pre-conditions for a successful governance system in the organization, and b) providing a concrete three-pillar framework (Bias & Fairness, Safety & Harm Prevention, Transparency & Explainability) with key consideration points which can be woven into the product lifecycle (adapted from Fig. 4, p. 31): first (1) reviewed and completed as a PM internal process during the 'charter' phase of product development (before development commences) and then (2) executed and re-visited at scale throughout the product lifecycle during both team-level development sprints and higher-level release and roadmap reviews.

In addition to supporting these key elements of responsible AI governance, our checklist also serves as a mechanism for positive organizational change in three dimensions:

1. Establishes and formalizes accountability. Ethical challenges are turned from a broad but vague 'set of shared values' into actionable, prioritized work with a responsible owner.

2. Opens important conversations. Serving as a boundary object, the checklist and its items provide a shared vocabulary with which engineers, designers, legal and business product stakeholders can ask questions, debate and agree on a concrete set of technical, privacy, legal and social requirements.
3. Reshapes how ethics and rigor is 'framed' in product trade-off decisions, which are a key part of the PM role. The checklist and its elements can help PMs and their teams reframe and manage trade-offs by turning safety and fairness into 'built-in' and non-negotiable pillars of product quality, and by raising awareness of long-term risks to product quality (including reputational) that are avoided by these actions.

Taken together, we hope this research and tool can start to reframe proactive ethics management not as an optional (or 'watering down' / 'putting out fires') activity to be added onto a successful product development process, but as a non-negotiable requirement of 21st century product leadership. Developing trustworthy AI is not only the responsible choice, but a competitive differentiator as customers become increasingly wary of digital products and services and regulation looms on the horizon. Our checklist empowers PMs not to feel like mere witnesses to these tensions, but as product leaders who can help to proactively manage them.

The work presented in this paper is a starting point, not an end point. Creating trustworthy AI is iterative and evolutionary, and this checklist is designed to be a useful waypoint in that journey. We have released the trust AI checklist under a permissive Creative Commons license, and we invite product teams to try it out in practice, and provide feedback and critique to help us all to improve and evolve this tool and future work. We also hope the checklist can serve as a standard against which product teams build their own practices and processes around responsible product development, and a common foundation from which to engage in important research on the long-term impact of the tool's adoption, evolution and adaptation to specific contexts and domains. We also see the next phase of this work exploring links to upcoming regulation in the EU and other markets, and ways it can complement and leverage up-and-coming automated AI model evaluation tools.

References

1. Ali, A., Smith, G., & Rodriguez, J. (2023). *The gatekeeper's dilemma: Integrating ethics into AI product development*. AI Ethics Press.
2. Bray, M., & Dainow, R. (2024). Ethics by design: Principles and implementation for AI systems. *Journal of Responsible Technology, 15*(2), 112-129. https://doi.org/10.1016/j.jrt.2024.01.008
3. Chandra, J., & Navneet, S. K. (2025). Advancing responsible innovation in agentic AI: A study of ethical frameworks for household automation. *arXiv preprint*. https://doi.org/10.48550/arXiv.2507.15901
4. Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Sánchez, A. N., Raji, D., Rankin, J. L., Richardson, R., Schultz, J., West, S. M., & Whittaker, M. (2019). *AI now 2019 report*. AI Now Institute. https://ainowinstitute.org/publication/ai-now-2019-report
5. Hagen, D. (2020). *Operationalizing AI ethics: Tools and frameworks for product teams*. MIT Press.
6. Hofman, G. (2024). Towards a practical ethics of generative AI in creative production processes. *arXiv preprint*. https://doi.org/10.48550/arXiv.2412.03579
7. Jacob, G. (2025). From design to delivery: The strategic role of product managers in deploying AI solutions for patient-centered healthcare. *International Journal of Scientific Research in Science and Technology, 12*(5), 467-477. https://doi.org/10.32628/IJSRST2512554
8. Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2023). Ethics as a service: A pragmatic operationalisation of AI ethics. *Minds and Machines, 33*(2), 287-315. https://doi.org/10.1007/s11023-023-09630-4
9. Olorunfemi, L. L., Amoo, O. O., Atadoga, A., Fayayola, O. A., Abrahams, T. O., & Shoetan, P. O. (2024). Towards a conceptual framework for ethical AI development in IT systems. *Computer Science & IT Research Journal, 5*(3), 616-627. https://doi.org/10.51594/csitrj.v5i3.910
10. Orugboh, O. G., Omabuwa, O. G., & Taiwo, O. S. (2025). Predicting Intra-Urban Migration and Slum Formation in Developing Megacities Using Machine Learning and Satellite Imagery. *Journal of Social Sciences and Community Support, 2*(1), 69-90.
11. Orugboh, O. G., Omabuwa, O. G., & Taiwo, O. S. (2024). Predicting Neighborhood Gentrification and Resident Displacement Using Machine Learning on Real Estate, Business, and Social Datasets. *Journal of Social Sciences and Community Support, 1*(2), 53-70.
12. Rochel, J., & Evéquoz, F. (2021). Getting into the engine room: A toolbox of organizational and procedural safeguards for building trustworthy AI. *AI and Ethics, 1*(3), 287-301. https://doi.org/10.1007/s43681-021-00037-4
13. Scott, W. R. (2013). *Institutions and organizations: Ideas, interests, and identities* (4th ed.). SAGE Publications.
14. Smith, G., Luka, N., Lattimore, B. R., Newman, J., Nonnecke, B., & Mittelstadt, B. (2025). Responsible generative AI use by product managers: Recoupling ethical principles and practices. *arXiv preprint*. https://doi.org/10.48550/arXiv.2501.16531
15. Smith, G., Luka, N., Osborne, M. R., Lattimore, B. R., Newman, J., & Nonnecke, B. (2025). Responsible generative AI use by product managers: Recoupling ethical principles and practices. *Academy of Management Proceedings, 2025*(1). https://doi.org/10.5465/AMPROC.2025.24377abstract
16. Sun, N. X., Miao, Y., Jiang, H., Ding, M. D., & Zhang, J. (2024). From principles to practice: A deep dive into AI ethics and regulations. *arXiv preprint*. https://doi.org/10.48550/arXiv.2412.04683