Responsible AI Principles in Product Management Practices

MASTER THESIS

to obtain the Erasmus Mundus Joint Master Degree in Digital Communication Leadership (DCLead)

of

Faculty of Social Sciences

Paris Lodron University of Salzburg, Austria

Technical Faculty of IT and Design

Aalborg University in Copenhagen, Denmark

Submitted by

Mudassir Akber Khan

s1093172

mak.ce14@gmail.com

Primary Supervisor: Anders Hansen Henten

Secondary Supervisor: Sergio Sparviero

Department of Communication Studies

Salzburg, 31st July 2024

Table of Content

1. Executive Summary

In the wake of rapid advancements and investments in AI technologies, integrating AI into software products has become widespread. Product Manager, also called "Mini-CEO", plays a crucial role in driving this innovation. However, as ethical considerations gain prominence, it is imperative for Product Managers to incorporate responsible AI principles into their product management practices for "responsible innovation". Ethical issues of AI are well-documented, but there is a notable scarcity of research on integrating responsible AI into product management. This gap leaves Product Managers ill-equipped to ensure ethical and sustainable AI development, risking unintended consequences for users and society.

This study aims to fill this gap by answering two research questions: 1. How do organizations currently integrate responsible AI into their product management? 2. What are the challenges of integrating responsible AI into product management? For this purpose, a semi-systematic review of the literature is conducted. A conceptual framework is also developed, incorporating the Responsible Innovation and ISPMA SPM frameworks. For data collection, 20 semi-structured interviews were conducted with AI product managers, consultants, and researchers in responsible AI space. A thematic analysis of interviews was conducted to gain insights from the experiences and perspectives of the interviewees.

Research findings indicate that understanding market needs for ethical values, incorporating responsible AI into practices such as roadmapping and product lifecycle management, and continuous performance and risk assessments are crucial for ethical AI products. Effective stakeholder coordination and comprehensive documentation to meet ethical standards are also essential for ensuring AI products align with regulatory requirements and user expectations. The study also identified several challenges in responsible AI integration, such as increased overheads, unrealistic customer expectations, complexity, skill shortages, and negligence in the early stages. Strategies to address these include automated risk assessments, effective customer communication, thinking beyond regulations, forming dedicated RAI teams, adopting holistic approaches, and engaging in external collaboration. This research provides insights for AI product managers to integrate responsible AI and for researchers and policymakers to develop informed regulations and policies for safer AI products.

Keywords: Responsible AI, Responsible Innovation, Ethical Issues of AI, Product Management, AI Product Management, AI Products

2. Introduction

2.1. Research Problem

Artificial intelligence has been around for decades, accompanied by a long-standing array of ethical issues. Recent advancements in generative AI have introduced unprecedented ethical challenges. Despite ethical challenges, AI products are increasingly emerging across various industries, catering to both consumer and business markets. Investors are injecting substantial funds into AI startups, and major corporations are committing billions to embed AI into their existing and new product lines. In response, governments are working on regulations, while international organizations are developing various frameworks to address these concerns.

In this evolving landscape, the role of an AI product manager is crucial in driving AI innovation and addressing ethical concerns. Product managers, also called "Mini-CEO", are tasked with guiding products from discovery to launch, encompassing planning, development, and deployment. They serve as an interface between the top management, development teams, and the customers playing a pivotal role in the success or failure of a software product.

Considering the significant influence product managers have on the trajectory of products, it is essential to examine how they integrate responsible AI principles into their practices and the challenges they encounter. However, this topic has not been adequately explored in existing literature. This gap highlights the need for an empirical study to provide valuable insights into navigating the ethical challenges of AI products. These insights will also assist researchers and policymakers by highlighting the perspectives of AI product managers, enabling more informed and effective regulation and policy development. Ultimately, this exploration will contribute to the creation of safer AI products that benefit society as a whole.

2.2. Research Objectives

This research study has the following objectives:

1. Examine the current practices of organizations in integrating responsible AI into product management.

2. Identify the challenges organizations face in integrating responsible AI into product management and the strategy to deal with those challenges.

2.3. Research Questions

This study aims to answer the following research questions:

1. *How do organizations currently integrate responsible AI into their product management?*
2. *What are the challenges of integrating responsible AI into product management?*

2.4. Structure of Thesis Report

Chapter 1 provides an executive summary of the research. Chapter 2 introduces the research problem, research objectives, research questions, and the structure of the report.

Chapter 3 explores the relevant literature covering artificial intelligence (AI), ethical issues of AI, responsible AI (RAI), its challenges, and product management (PM).

Chapter 4 discusses the conceptual framework, based on the Responsible Innovation (RI) AREA Framework and the ISPMA SPM Framework, for analysis of the research findings.

Chapter 5 explains the research methodology, research design, and research methods for the literature review, data collection, interview guide, sampling, and data analysis. Ethical considerations and validation procedures are also discussed within this chapter.

Chapter 6 presents the findings from the interviews, followed by a validation of the findings. Chapter 7 discussed the findings in light of the literature and the selected frameworks. Limitations of the study are also mentioned in this chapter.

Chapter 8 provides a conclusion of the research. Chapter 9 provides a list of references in APA style. Chapter 10 contains all the appendices referenced throughout the report.

3. Literature Review

3.1. Introduction

This review focuses on the key literature relevant to the research problem and the research questions mentioned in the previous section. This will help explore the already available body of knowledge and identify the key areas for further contributions. This review covers topics including artificial intelligence (AI), state-of-the-art AI products, ethical issues of AI, responsible AI principles and frameworks, product management, integration of responsible AI, challenges and implications, and a conclusion in the end.

3.2. Artificial Intelligence (AI)

Before discussing responsible AI and its integration into product management, it is important to understand what artificial intelligence is. This section provides an overview of definitions, classification, and ethical issues of AI and a brief list of widely used AI products.

3.2.1. Definitions of Artificial Intelligence

The term Artificial intelligence, AI in short, was coined by John McCarthy in a summer workshop at Dartmouth in 1956 (Tecuci, 2012). McCarthy (2004) defines AI as "the science and engineering of making intelligent machines, especially intelligent computer programs." However, this is not the only definition of AI. The term "Artificial Intelligence (AI)" lacks a universally accepted definition, leading to varied interpretations within and outside the field (P. Wang, 2008, 2019). Numerous attempts have been made to define artificial intelligence, however, each definition has faced criticism and has failed to achieve consensus within individual disciplines, let alone universal agreement (Abbass, 2021). Definitions of AI are inherently challenging and they cannot completely achieve universal applicability where each definition may serve different contexts more effectively (Abbass, 2021). In their 2019 study, Bawack et al. reviewed how industry experts define AI, finding that practitioners use various terms interchangeably to describe AI technologies such as machine learning, deep learning, NLP, computer vision, robotics, speech recognition, and neural networks.

Nevertheless, it is important to define artificial intelligence for various reasons. Martinez (2019) discusses the challenges stemming from the undefined nature of artificial intelligence (AI), emphasizing the significant legal and policy issues that arise without a clear conceptual framework. The evolving nature of AI complicates efforts to establish a definitive definition, which impacts both technological development and regulatory frameworks (Martinez, 2019). Abbass (2021) also highlights the organizational struggles in defining AI, stressing the implications such definitions carry for technological assessments, legal considerations, and strategic planning within businesses.

As mentioned earlier, different researchers, practitioners, and organizations propose different definitions of artificial intelligence. Samoili et al. (2020) compiled a comprehensive summary of AI definitions spanning from 1955 to 2019.

In a broader sense, there are two aspects of AI i.e. AI as a scientific discipline and AI as a technological system.

AI as a Scientific Discipline

Tecuci (2012) defines Artificial intelligence (AI) as the field focused on creating systems with human-like intelligence, such as perception, natural language processing, problem-solving, learning, and adaptation. It is a broad interdisciplinary field intersecting with computing, mathematics, linguistics, psychology, neuroscience, engineering, statistics, economics, control theory, cybernetics, and philosophy. Tecuci (2012) also identifies key research areas in AI i.e. knowledge representation, problem-solving, planning, knowledge acquisition, learning, natural language, speech, vision, and robotics.

According to (EU AI HLEG, 2019b) AI, "as a scientific field, encompasses various methods and techniques. These include machine learning (with specific examples being deep learning and reinforcement learning), machine reasoning (encompassing planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which involves control, perception, sensors, actuators, and integrating these methods into cyber-physical systems)".

Wang (1995, 2008), described five typical ways to define AI, corresponding to establishing the similarity between intelligent computer and the human mind by structure, behavior, capability, function, and principle, respectively. Abbass (2021) defines AI as the automation of cognition and the ability of machines to socially integrate, perform complex tasks, and communicate effectively with high-information content messages.

AI as a Technological System

AI as a technological system is also defined by various organizations, each highlighting different aspects. The US Department of Defense focuses on AI's ability to perform tasks requiring human intelligence. The EU and OECD emphasize AI's capacity to process data, make decisions, and adapt based on its environment. Some notable definitions include:

- **The US Department of Defense AI Strategy:** "the ability of machines to perform tasks that normally require human intelligence" (Allen, 2020).
- EU AI High-Level Expert Group: "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behavior by analyzing how the environment is affected by their previous actions" (EU AI HLEG, 2019a).
- **European Parliament:** "'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments" (European Parliament, 2024).
- **OECD:** "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment" (OECD, 2024).

Considering this it can be said that each researcher can decide how to employ the term "AI," but clarity is crucial when discussing findings, ensuring a thorough understanding of its implications (P. Wang, 2019). For this research, AI as a technological system will be focused instead of AI as a scientific discipline. Also, definitions by the EU and OECD will be used as a reference for this research.

3.2.2. Classification of Artificial Intelligence

AI can be classified in various ways. For instance, Strelkova (2017) categorizes artificial intelligence into three levels: ANI, AGI, and ASI. ANI, or Artificial Narrow Intelligence, excels in a specific domain, such as defeating a chess champion but lacks broader capabilities. AGI, or Artificial General Intelligence, surpasses human intelligence, possessing abilities like reasoning, problem-solving, and abstract thinking. ASI, or Artificial Super Intelligence, represents intellect far superior to humans across all domains, including scientific innovation and social skills (Strelkova, 2017).

According to Morandín-Ahuerma (2022), AI can be classified in two main ways: based on its cognitive abilities and based on its autonomy. In terms of cognitive capacity, AI systems are categorized as weak or limited AI, general AI, and superintelligence. Meanwhile, in terms of autonomy, AI can be reactive, deliberative, cognitive, or autonomous.

AI can also be divided into the following subfields:

Machine Learning: It enables machines to perform tasks autonomously by learning from data, through a process known as training (OECD, 2024), rather than being explicitly programmed by humans (Allen, 2020; Morandín-Ahuerma, 2022). ML is grounded in various disciplines, including statistics and computer science, encompassing four main algorithm families: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning, each differing in data usage and learning methods (Manning, 2020).

Deep Learning: It is a highly effective machine learning technique (Manning, 2020). It involves utilizing large, multi-layered artificial neural networks that operate with continuous numerical representations, similar to the hierarchical structure of neurons in the human brain. It is capable of generalizing well from limited data and scaling efficiently to accommodate large datasets and computational demands (Manning, 2020).

Generative AI: A subfield of AI, that focuses on creating new data by leveraging deep learning and neural networks to identify and replicate patterns from existing data (Ray, 2023). Popular generative AI models include GANs, VAEs, and Transformer-based models (Aydın & Karaarslan, 2023) which enable systems to generate diverse content such as music, images, and text without specific training (Mondal et al., 2023).

3.2.3. State of the Art of AI Products

AI is widely used in modern vehicles for functions like antilock brakes and fuel injection as well as in Google's search engine, Facebook's Newsfeed, and email spam filters (Strelkova, 2017). AI is also crucial in aviation, military, manufacturing, and finance, especially in high-frequency trading (Strelkova, 2017). According to (Martinez, 2019), initially, applications of AI were limited, such as serving as a chess opponent, but have since expanded to outperform humans in various complex tasks such as consumer behavior prediction, medical diagnostics, and creative endeavors. Moreover, emerging AI applications include mind reading, death prediction, and advanced surveillance, indicating a trajectory toward AI systems surpassing human abilities (Martinez, 2019).

There are various AI products in the market offering different features to their users across various verticals (Product Hunt, 2024; Top Startups, 2024; Y Combinator, 2024). These AI products have attracted investors in the past couple of years. In Q1 2024, venture funding for AI startups increased to $12.2 billion across 1,166 deals, a 4% rise from Q4 2023, according to Crunchbase data (Metinko, 2024). Following is a list of some of the AI products that gained traction in recent years.

1. **ChatGPT:** AI chatbot by OpenAI that utilizes GPT-3.5 and GPT-4 to engage in human-like conversations, making it useful for various tasks such as customer support, content creation, and interactive assistance (Toolify, 2023).

2. **DALL-E 2:** AI system that generates realistic images from text, offering features like outpainting, inpainting, and high-resolution image creation (OpenAI, 2024).

3. **GitHub Copilot:** An AI-powered tool integrated with popular code editors enhances developers' productivity by providing contextual code completions, chat assistance, and code explanations within their development environment (GitHub, 2024).

4. **Adobe Firefly:** A standalone web app using generative AI to enhance creative workflows, offering text-to-image, text-based editing, and media generation, integrated with Adobe's flagship apps and Adobe Stock (Adobe, 2024).

5. **Claude:** Advanced AI developed by Anthropic that prioritizes safety, accuracy, and security, excelling in text and code generation, idea facilitation, image transcription, and real-time multilingual translation for enhanced productivity (Claude, 2024).

3.3. Ethical Issues of AI

Ethical issues of AI are extensively examined in the literature (Wirtz et al., 2020) with an increasing interest in the ethical implications and social impacts of artificial intelligence (AI) across various sectors, including industry, academia, and the public (Hagerty & Rubinov, 2019). AI ethics are also covered by media which has an influence on public opinion and technology adoption (Ouchchy et al., 2020). However, the sophistication of discussions on AI ethics in media is limited, lacking depth in both technical understanding and ethical frameworks (Ouchchy et al., 2020).

In a study conducted by the SHERPA project funded by the EU, Stahl (2021) highlights numerous ethical issues associated with AI. It reflects a long-standing tradition of ethics discussions in AI literature which is now emerging from policy perspectives as well. Stahl (2021) empirically investigated people's perceptions of AI ethics identifying and defining 39 distinct ethical issues related to AI. The identified ethical issues span a wide range, including concerns such as lack of trust, privacy issues, biases, job displacement, and potential military applications. These issues encompass impacts on innovation, public services, health, democracy, and human rights, reflecting diverse challenges posed by AI technologies.

In another study, Khan et al. (2022) conducted a systematic literature review (SLR) identifying 22 ethical principles and 15 challenging factors critical for AI system designers including 4 main principles i.e. transparency, privacy, accountability, and fairness. According to Khan et al. (2022) the challenges in AI ethics, such as lack of ethical knowledge and vague principles, hinder effective implementation of ethical principles. Mikalef et al. (2022) also outline similar ethical challenges across different principles of responsible AI including Fairness (biased outcomes, consumer bias exploitation, price homogenization, design neglect), Transparency (lack of explainable AI procedures, insufficient risk communication, traceability issues), Accountability (unclear responsibility, inadequate response mechanisms, "no-fault" policies), Robustness and Safety (AI endangering human life, data poisoning, malicious AI use), and Data Governance (privacy breaches, biased data, unauthorized access). Wach et al. (2023) highlight the dangers of Generative AI (GAI), particularly ChatGPT, citing potential biases, privacy and security risks, ethical concerns, and the possibility of creating harmful or deceptive content.

Wirtz et al. (2020) discuss 4 key challenges of AI i.e. AI rulemaking for humans, AI discrimination, Moral dilemmas, and Compatibility of machine and human value judgment. Wirtz et al. (2020) elaborates that AI's decision-making in regulating human behavior is rational and driven by its programmed rules, lacking emotional or conscious factors, which can lead to unintended consequences. AI discrimination, highlighted by researchers and governments, addresses biases in AI systems stemming from inaccurate training data, potentially leading to unjust outcomes in human-related decisions like hiring or loan approvals. Moral dilemmas arise when AI systems confront conflicting ethical choices, where implemented rules may change during learning processes unless AI is constrained by rigid guidelines, risking autonomy. Ensuring AI aligns with human values without developing divergent behavior poses a challenge, requiring governance to mitigate risks and ensure beneficial AI applications (Wirtz et al., 2020). According to Whittlestone et al. (2019), ethical ideals for AI often clash in practice, where prioritizing one value may necessitate compromising another. For instance, Advancing complex algorithms for better predictions could reduce transparency, while data-driven technologies might undermine full assurance of data privacy. Despite these risks, communities may accept them for substantial benefits like innovative cancer treatments. Whittlestone et al. (2019) further provide hypothetical illustrations for ethical dilemmas:

1. **Balancing algorithm accuracy in decisions with fairness:** A jurisdiction adopts a recidivism risk algorithm for bail and parole decisions, which is highly accurate on average but disproportionately discriminates against black defendants due to higher false positive rates.
2. **Balancing personalization benefits with societal solidarity:** A company offers personalized insurance using algorithms trained on detailed datasets, leading to tailored services but potentially undermining support for publicly funded services.
3. **Balancing service efficiency with privacy:** A public hospital shares patient data with a private company to implement a diagnostic algorithm, enhancing healthcare efficiency but raising concerns about patient privacy and consent.
4. **Balancing automation's convenience with human self-actualization:** AI enables an all-purpose personal assistant that enhances convenience but reduces opportunities for personal skill development and creativity, potentially homogenizing human experiences.

Meek et al. (2016) mentioned the ethical issues of AI including its interaction with humans and other life forms, potential existential risks posed by AI systems, the need for regulatory frameworks similar to pharmaceuticals, and the establishment of national and international ethics committees to oversee AI development and ensuring ethical standards are upheld. Mikalef et al. (2022) also argue that a comprehensive understanding of AI should include scrutiny of its potential unintended consequences because analyzing these effects during AI development and deployment can enhance our understanding of the value derived from such technological investments. Meek et al. (2016) highlight the necessity of addressing ethical and risk issues in managing emerging technologies like AI, ensuring AI is not intrinsically harmful is important for its safe integration into society.

Healthcare is an important sector where ethical issues of AI are critical. Safdar et al. (2020) discussed four key ethical issues of AI in the context of healthcare. First, AI algorithms often exhibit selection bias, leading to reduced accuracy for underrepresented groups and potential over-reliance on automated systems, risking erroneous results. Second, AI applications in medicine struggle with rare conditions and may fail to generalize across different practice settings, necessitating careful integration and local data augmentation. Third, ownership and usage rights of patient-derived data vary by jurisdiction, raising complex issues in a global AI marketplace, especially under regulations like the GDPR. Fourth, large datasets essential for ML models challenge privacy protections, with regulations like the GDPR requiring strict data use and transparency, though true de-identification remains difficult (Safdar et al., 2020).

Focusing on the social impacts of AI across five global regions, Hagerty & Rubinov (2019) indicate that AI's social impacts vary significantly depending on the geographical context, with local cultural and social settings profoundly shaping perceptions and understandings of AI. Research in U.S. settings shows that AI-driven technologies often entrench social divides and exacerbate social inequality, particularly among historically marginalized groups. This pattern of entrenching social divides is evident globally, with low- and middle-income countries being more vulnerable to AI's negative impacts and less likely to benefit from its advancements. Hagerty & Rubinov (2019) mentioned that the current analyses of AI are predominantly biased toward U.S. perspectives and lack comprehensive research, especially from regions outside the U.S. and Western Europe.

Nevertheless, the ethical issues of AI are central to discussions on the future governance and regulation of AI technology (Wirtz et al., 2020).

3.3.1. Real-Life Cases of Ethical Issues of AI

This section provides a brief list of real-life cases of ethical issues of some of the AI products. For this purpose, AIAAIC's open repository is used which catalogs incidents and controversies related to AI, algorithms, and automation, offering an objective view to help users analyze the ethical and risk implications of these technologies (AIAAIC, 2024).

1. **EU Election and False Information:** AI chatbots from major tech companies spread misinformation about the European election by providing incorrect dates and ballot information, highlighting AI's tendency to hallucinate (Politico, 2024).
2. **Autopilot Car Crash:** A Tesla driver in autopilot mode faces vehicular homicide charges after his vehicle struck and killed a motorcyclist in Washington, highlighting safety concerns over AI-driven auto-drive features (New York Post, 2024)**.**
3. **AI Song Generators and Copyright Infringement:** Major music labels are suing AI song generators Suno and Udio for copyright infringement, alleging they mimic existing songs without permission (The Guardian, 2024).
4. **Robot Arm Crushed Factory Worker:** A worker at a factory in Thailand died after being crushed by a robot arm while laying out metal sheets, raising safety concerns in manufacturing despite factory claims of proper robot function (Jolly, 2024).
5. **Google's AI Search Tool's Misleading Responses:** Google's AI-generated search summaries, known as "AI Overviews", generated misinformation and bias, including false historical claims and dangerous advice (Euronews, 2024).
6. **WHO's AI Health Chatbot:** The WHO's AI chatbot SARAH sometimes gives outdated or incorrect answers to questions related to basic health information due to its training on older data. This raises ethical concerns about misinformation in public health, highlighting the limitations and risks of AI in healthcare (Bloomberg, 2024).
7. **YouTube's captions insert explicit language in kids' videos:** YouTube's AI-generated captions on kids' videos often misinterpret words into explicit language, raising ethical concerns about content suitability for children. Researchers highlight the issue of "inappropriate content hallucination," where AI algorithms inaccurately transcribe audio, leading to potentially harmful outcomes (Simonite, 2022).

3.4. Responsible AI (RAI)

3.4.1. Definitions of Responsible AI

Similar to Artificial Intelligence, Responsible AI is subject to various definitions and interpretations. These identify moral and ethical considerations, empirical approaches, ethical behavior, and societal consequences (Deshpande & Sharp, 2022). Dignum (2019) highlights that Responsible AI means different things to different people and contexts, for instance, governance policies for AI research and deployment, the role of developers, issues of inclusion and diversity, and reflections on the benefits and risks of AI.

In a report published by the Global Index on Responsible AI, R. Adams et al. (2024) referred to Responsible AI as the creation, development, implementation, and governance of AI systems in a manner that honors and safeguards all human rights while adhering to ethical principles throughout the AI lifecycle and value chain. According to R. Adams et al. (2024), responsible AI requires that everyone involved in the national AI ecosystem is responsible for the human, social, and environmental consequences of their actions.

Lu et al. (2022) refer to responsible AI as the ethical development of AI systems to benefit humans, society, and the environment. (Y. Wang et al., 2020) describe RAI as a governance framework that emphasizes designing and implementing ethical, transparent, and accountable AI solutions while maintaining individual trust, minimizing privacy invasion, meeting stakeholder expectations and regulatory requirements, and placing end-users at the center of AI design and implementation. Dignum (2019) defines RAI as focusing on human accountability in creating intelligent systems that enhance human well-being and sustainability. It emphasizes designing AI to respect societal values, ensure transparency, and address ethical considerations in diverse multicultural contexts.

Despite the diversity in definitions, there is a common approach where Responsible AI is often defined using certain AI principles. For instance, according to Barredo Arrieta et al. (2020), RAI is "systematic adoption of several AI principles for AI models to be of practical use." Barredo Arrieta et al. (2020) suggest that these principles, which include explainability, fairness, accountability, and privacy, are essential for AI models to be effectively utilized in real-world settings. Similarly, de Laat (2021) denotes responsible AI as "AI that is fair and non-biased, transparent and explainable, secure and safe, privacy-proof, accountable, and to the benefit of mankind." For the purpose of this research, these two definitions will be used.

3.4.2. Frameworks for Responsible AI

The increasing use of AI and the importance of addressing ethical concerns related to AI have resulted in the creation of numerous guidelines and frameworks aimed at ensuring that AI technologies are developed and deployed in a "responsible" manner. These frameworks use different terminologies such as Ethical AI or Trustworthy AI along the term "responsible AI".

Jobin et al (2019) conducted a scoping review of 84 non-legal documents issued by various organizations to map existing principles and guidelines for ethical AI. The majority of these documents were published by private companies, followed by governmental agencies but also other contributors including academic institutions, intergovernmental organizations, and non-profits. Most of these documents originated from economically developed countries, such as the USA, the UK, and the European Union while African and South American countries had no independent representation. Many of these guidelines were intended for multiple stakeholders while others targeted specific groups within the issuer's sphere, such as employees. Some documents focused on the public sector, private sector, developers or designers, organizations, and researchers. According to Jobin et al. (2019), these frameworks include 11 key ethical principles including transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity. Transparency, justice and fairness, non-maleficence, responsibility, and privacy were frequently referenced, appearing in over half the sources while no single principle was common across all documents. Another important point is that despite convergence around certain principles, significant differences exist in their interpretation and the specific recommendations derived from them (Jobin et al. 2019).

Barletta et al. (2023) classified responsible AI frameworks w.r.t. the type of proposing institution using three categories: COMPANIES, UNIVERSITIES, and NO-PROFIT ORG / COMMUNITIES / PUBLIC ENTITIES (NPG/COMM/PE). Barletta et al. (2023) further classified these frameworks as Principle (for abstract ethical principles), Guideline (for concrete guidelines), and Tool (for verifying compliance and supporting the implementation of principles or guidelines). According to Barletta et al. (2023) NPG/COMM/PE contributed the most frameworks, followed by COMPANIES and then UNIVERSITIES. Also, COMPANIES and NPG/COMM/PE focused more on Principles, whereas UNIVERSITIES predominantly proposed Guidelines; COMPANIES and UNIVERSITIES also introduced a higher proportion of Tools compared to NPG/COMM/PE.

Barletta et al. (2023) also identify funding disparities affecting framework prevalence, notably between companies and non-profits, with a trend of frameworks lacking actionable tools, hindering the effective implementation of principles. Barletta et al. (2023) find that most frameworks for Responsible AI address the four main ethical principles (i.e. Transparency, Fairness, Technical Robustness & Safety, and Privacy & Data Governance) to some extent, though often partially. However, some frameworks neglect one to three principles, possibly due to the lack of global consensus on RAI definitions and principles, and ethical concerns like Ethics Bluewashing and Digital ethics shopping among companies.

Barletta et al. (2023) find that most Responsible AI frameworks focus predominantly on early phases of the SDLC, such as Requirements Elicitation, neglecting crucial phases like Design and Deployment. This lack of comprehensive guidance poses challenges for implementing and validating RAI principles across the entire SDLC. Additionally, there is a notable scarcity of practical tools accompanying theoretical frameworks across all entity types, with companies leading in offering tools that cater to both technical and non-technical stakeholders, aiming to provide comprehensive support throughout all SDLC phases. Barletta et al. (2023) highlight the fragmented nature of current Responsible AI recommendations, advocating for a unified framework to seamlessly integrate knowledge across organizational contexts and foster widespread adoption in public and private sectors.

Another study was conducted by Narayanan & Schoeberl (2023) which reviewed over 40 responsible AI frameworks revealing a diversity of approaches from various entities. According to Narayanan & Schoeberl (2023), the literature highlights the evolution of tools for implementing responsible AI systems, emphasizing the challenge of tool selection and application. Process-based frameworks offer a systematic approach to address these needs by guiding organizations in prioritizing design aspects, integrating accountability into teams, and engaging impacted communities. Frameworks vary in specificity, often balancing flexibility with the need for clear implementation guidance. Challenges include determining target audiences and aligning framework capabilities with organizational needs, which can be time-consuming and complex. Narayanan & Schoeberl (2023) also developed a matrix to aid organizations in selecting and implementing frameworks tailored to their operational contexts. This matrix categorizes frameworks by user relevance and utility, facilitating informed decisions about framework adoption based on specific AI system components, lifecycle stages, or related characteristics.

3.5. Responsible AI Principles and Practices

As mentioned earlier one common approach to define responsible AI is to use certain AI principles. Each framework provides its own set of principles. Here are some of the most frequently discussed principles across various frameworks:

3.5.1. Fairness

Fairness is a key principle of AI Ethics which emphasizes that AI systems should be inclusive and accessible throughout their life cycle, avoiding unfair discrimination (Zhu et al., 2022). It emphasizes the avoidance of discriminatory decisions by the AI system that impact values such as dignity and justice, aiming to foster social fairness (Khan et al., 2022). Fairness issues include biased outcomes, exploitation of consumer biases, price homogenization and market collusion, and inaccessibility due to design negligence (Mikalef et al., 2022).

According to Whittlestone et al. (2019), fairness remains a concept without a universal definition in philosophy where definitions may vary depending on language, culture, and political context. Different approaches include: Utilitarianism seeks to maximize overall benefit in distributions of outcomes, egalitarianism strives for as much equality as possible, and minimax theory prioritizes the maximum benefit for the worst-off. Alternatively, some theories focus on the fairness of outcome determinants, considering whether outcomes stem from individual choices or external factors like historical injustices (Whittlestone et al., 2019). Dignum (2019) also mentioned various normative interpretations of fairness leading to different actions. Thus, it is crucial to clarify the interpretation used and explicitly state how norms are implemented in a system, as this depends on the context and the designers' perspectives (Dignum, 2019). According to Laato et al. (2022), fairness in AI involves the subjective perception of whether decisions or recommendations feel just. Studies suggest that users may distrust or perceive an AI model as unfair if they believe the task it performs is unsuitable for algorithmic decision-making, regardless of improvements made to the system.

De Laat (2021) argues that numerous interpretations of fairness led to the development of specific, precise fairness metrics integrated into open-source software tools, enabling customization of machine learning models to align with desired fairness criteria. However, selecting or designing an appropriate fairness function for specific applications ex-ante is complex and context-dependent (Prem, 2023). Moreover, defining precise mathematical definitions of fairness for machine learning systems is also challenging because optimizing for various intuitive fairness dimensions simultaneously can be mathematically impossible. This ambiguity complicates efforts to ensure that AI, data, and algorithms consistently adhere to principles of fairness in practical applications (Whittlestone et al., 2019).

The EU AI HLEG (2019a) identified two dimensions of fairness in AI i.e. substantive and procedural dimensions. Substantively, it aims to ensure equitable distribution of benefits and costs, prevent bias and discrimination, and promote equal access to opportunities. Procedurally, fairness requires transparent and contestable decision-making processes, accountability for decisions, and mechanisms for effective redress against AI-driven decisions. This approach seeks to balance societal interests while respecting individual freedoms and ensuring ethical AI deployment.

De Laat (2021) highlights that bias in datasets can lead to biased machine-learning models, resulting in discriminatory predictions against specific groups. Ethical concerns arising from unfair decisions based on factors like race, age, or gender necessitate detecting biases related to sensitive data affecting protected groups to ensure fairness in AI (Barredo Arrieta et al., 2020). This requires early consideration in the data processing pipeline, with best practices and bias mitigation mechanisms integrated at various stages (Zhu et al., 2022).

Benjamins et al. (2019) discuss various definitions: “unawareness” removes sensitive variables but may overlook proxy features; “group fairness” considers fairness for all individuals; “individual fairness” models differences within the population; and counterfactual fairness explores bias causes through causal graphs. There are two main actions that can be considered for fairness in ML i.e. evaluation and mitigation. Evaluation involves measuring bias against one or more criteria such as statistical parity difference and equal opportunity difference, while mitigation aims to reduce bias in sensitive attributes through techniques like pre-processing, in-processing, and post-processing (Benjamins et al., 2019).

Prem (2023) argues that when fairness is seen as a property of AI classifiers, engineers focus on algorithmic improvements. However, examining fairness at the application and societal levels raises questions about whether algorithmic tuning alone can adequately address ethical concerns, shifting the discussion to broader social and political considerations of fairness (Prem, 2023). Vesnic-Alujevic et al. (2020) highlight that research often addresses the ethical challenge of biases and discrimination in AI, however, policy documents lack detailed elaboration on this issue, which could increase societal inequalities and harm certain groups.

3.5.2. Accountability

Accountability in AI systems involves identifying and assigning responsibility to individuals across the lifecycle of AI development and ensuring human oversight to oversee AI outcomes (Zhu et al., 2022). It aims to address liability issues by assigning responsibility for system decisions and actions to stakeholders, thus mitigating potential culpability concerns (Khan et al., 2022). Accountability demands that those legally or politically responsible for harm must justify or compensate for their actions (Vesnic-Alujevic et al., 2020), emphasizing the need for both technical and social accountability throughout the AI system's development, implementation, and operation (Khan et al., 2022). Accountability issues include unclear responsibility for AI outcomes, lack of mechanisms to address unintended AI consequences, and the prevalence of "no-fault" policies when AI causes harm (Mikalef et al., 2022).

The accountability principle is linked to other AI principles. For example, fairness as it requires mechanisms to ensure responsibility for AI systems and their outcomes, both pre- and post-development, deployment, and in operations (EU AI HLEG, 2019a). Transparency because the system must be understood prior to making liability decisions (Khan et al., 2022). Safety because ensuring AI applications meet safety thresholds through rigorous testing is critical for accountability (Mikalef et al., 2022).

According to the EU AI HLEG (2019a), accountability involves auditability, minimization and reporting of negative impact, trade-offs and redress. Auditability allows for the assessment of algorithms and design processes, contributing to trust through internal and external evaluations without necessarily revealing business models or intellectual property. Minimizing and reporting negative impacts involves identifying, assessing, documenting, and responding to potential harms, with protections for whistle-blowers and entities raising legitimate concerns, and using proportionate impact assessments. Trade-offs should be addressed methodically, with clear documentation, ensuring that AI systems do not proceed without ethically acceptable solutions. Lastly, mechanisms for redress must be accessible to address unjust adverse impacts, particularly for vulnerable groups, ensuring trust through the availability of adequate redress options (EU AI HLEG, 2019a).

Zhu et al. (2022) suggest that incorporating product features and management processes that ensure the provenance and traceability of AI artifacts and decisions can significantly enhance the accountability of AI systems. According to Vesnic-Alujevic et al. (2020), implementing the right to explanation can also contribute to heightened responsibility and accountability. Furthermore, de Laat (2021) argues that AI providers must be capable of providing a comprehensive account of every step in the production process of their solutions. However, while many companies superficially acknowledge it, only three out of the 24 firms surveyed by de Laat (2021) have actually developed accountability tools specifically tailored for AI.

Mikalef et al. (2022) mentioned that the concept of accountability has been widely explored in literature, especially in fields like healthcare and autonomous vehicles where outcomes are highly significant. There is considerable debate regarding who should be held accountable, considering factors such as intentions, motives, and reasoning behind AI actions (Mikalef et al., 2022). Vesnic-Alujevic et al. (2020) argue that as autonomous systems proliferate and decision-making is increasingly delegated to them, assigning responsibility and accountability becomes increasingly challenging.

3.5.3. Transparency & Explainability

Transparency in AI involves the ability to clearly describe, examine, and replicate how AI systems make decisions and adjust to their surroundings, including the origin and evolution of the data they use (Dignum, 2019). This openness fosters trust by ensuring clarity in all aspects related to the system, such as data sources, development processes, and stakeholder involvement in decisions impacting humans or having significant ethical implications (Dignum, 2019). The terms 'transparency', 'explainability', 'interpretability', and 'intelligibility' are frequently used interchangeably to describe what 'black-box' algorithms are perceived to lack (Whittlestone et al., 2019).

Transparency and responsible disclosure are essential so that individuals are aware when an AI system is significantly affecting them and can identify when they are interacting with an AI system (Zhu et al., 2022). Transparency issues include a lack of formalized procedures for explainable AI, inadequate communication of potential risks and limitations of AI systems, and limited traceability of erroneous AI decisions (Mikalef et al., 2022). Dignum (2019) provided a checklist for transparency with four major areas i.e. openness about data, openness about design processes, openness about algorithms, and openness about actors and stakeholders.

According to the EU AI HLEG (2019a), transparency involves traceability, explainability, and communication in AI systems. Traceability involves documenting data sets, data processes, and algorithms to ensure decisions can be traced and errors identified. Explainability requires that AI decisions and the processes behind them can be understood by humans, with the trade-off between explainability and accuracy carefully managed. Significant AI decisions impacting people's lives must be explainable in a way suitable for various stakeholders. AI systems must clearly identify themselves as non-human, allow for human interaction options, and disclose their capabilities and limitations appropriately to users (EU AI HLEG, 2019a).

Explainability refers to the capacity to clarify both the technical operations of an AI system and the associated human decisions (Prem, 2023). Transparency involves revealing AI artifacts like source code, data, algorithms, models, and documentation, which aids in explainability but doesn't fully satisfy stakeholders who seek insights into AI behavior and decision-making causes, posing challenges in earning their trust through detailed explanations (Zhu et al., 2022).

According to Dignum (2019), opacity in Machine Learning, often referred to as 'black-box' algorithms, poses a significant challenge to transparency in Artificial Intelligence. These algorithms, designed to enhance functional performance, comprise numerous component functions that, while individually simple (like statistical regression methods), collectively create systems too complex to comprehensively analyze or verify (Dignum, 2019). Despite their ability to optimize outputs for tasks such as image recognition or text classification, they achieve this by adjusting outputs to inputs without revealing the underlying structure of the function being approximated (Dignum, 2019). Explainability in deep learning algorithms is challenging because they focus on enhancing performance and fine-tuning outputs to specific inputs without revealing the underlying function's structure (Dignum, 2019).

While algorithmic transparency is often seen as universally important, there is little evidence on what kinds of explanations are preferred by different people in different contexts (Whittlestone et al., 2019). Benjamins et al. (2019) argue that the relevance and type of explainability in AI systems vary by domain, with greater importance in areas like medical diagnosis compared to movie recommendations. However, most domains still require some level of explainability. The explanations should be tailored to the transparency requirements and specific profiles of each domain, ideally incorporating domain-specific knowledge to provide deeper insights beyond simple feature rankings in supervised machine learning (Benjamins et al., 2019).

3.5.4. Privacy and Data Protection

This principle emphasizes that AI systems must uphold privacy rights, ensure data protection, and maintain data security throughout their entire life cycle (Zhu et al., 2022). This fundamentally involves individuals' rights to manage and control their personal information (Khan et al., 2022). Benjamins et al. (2019) also emphasize that privacy and security are integral throughout the lifecycle of AI systems, which heavily rely on data. This ensures utmost respect for individuals' privacy rights and their personal data, whether the data is personal or anonymized/aggregated (Benjamins et al., 2019). These principles extend beyond AI systems, as most organizations already implement processes to safeguard privacy and security (Benjamins et al., 2019). Explainability in ML models helps assess privacy by revealing how models interpret data, preventing privacy breaches from opaque internal representations. However, excessive explainability may compromise differential privacy by exposing sensitive data origins to unauthorized parties (Barredo Arrieta et al., 2020).

According to the EU AI HLEG (2019a), privacy and data governance include respect for privacy, quality and integrity of data, and access to data. Firstly, it stresses that AI systems must ensure privacy and data protection throughout their lifecycle, encompassing both user-provided information and data generated about users during interactions. The potential for AI systems to infer sensitive personal attributes highlights the need to prevent unlawful discrimination through data collection. Secondly, the quality and integrity of data used in AI training are pivotal, requiring mitigation of biases, inaccuracies, and malicious data that could alter system behavior. Testing and documentation of data processes are essential across all stages of development and deployment. Lastly, clear protocols for data access within organizations are necessary, ensuring that only qualified personnel with legitimate reasons can access individuals' data, thereby safeguarding privacy and data security.

Privacy is intricately tied to preventing harm, especially critical due to AI systems' impact. Effective data governance is essential, ensuring data quality, integrity, relevance, access protocols, and secure processing to safeguard privacy (EU AI HLEG, 2019a). Khan et al. (2022) mention that regulatory bodies are tasked with creating laws for data privacy, particularly challenging in AI environments where systems handle user data through processes like cleaning and interpretation. Self-governing AI systems raise concerns about data access, privacy, security, and transparency, highlighting complex challenges that necessitate further research (Khan et al., 2022).

3.5.5. Robustness and Safety

Robustness and Safety principles emphasize that AI systems should be designed with a proactive strategy to address risks, ensuring they function as expected while reducing the likelihood of accidental or unforeseen damage (Mikalef et al., 2022). According to Dignum (2019), safety involves ensuring that the system performs its intended functions without causing harm to users, resources, or the environment and that its purpose aligns with human rights and values. Robustness and safety issues include AI decisions placing human life at risk (e.g., autonomous vehicle dilemmas), data poisoning or model leakage, malicious use of AI (e.g., targeted disinformation campaigns, non-state weaponized consumer drones, AI terrorism), irreproducible outcomes/decisions (Mikalef et al., 2022).

According to the EU AI HLEG (2019a) technical robustness includes resilience to attacks, fallback plans, accuracy, reliability, and reproducibility. AI systems must be safeguarded against vulnerabilities and adversarial attacks to prevent data poisoning, model leakage, or infrastructure damage, and should be designed to mitigate dual-use and abuse by malicious actors. Safety mechanisms, including fallback plans and human intervention, are essential to ensure systems operate correctly without causing harm, and risks must be assessed and managed proactively. Accurate AI systems require meticulous development and evaluation processes, especially in critical applications affecting human lives. Lastly, reliability and reproducibility are vital for ensuring consistent performance across various conditions and for enabling thorough scrutiny and validation of AI systems' behaviors (EU AI HLEG, 2019a).

Mikalef et al. (2022) emphasize the significance of creating robust and safe AI applications to address both intentional and unintentional issues. As AI becomes more widespread, it is crucial to understand how these concerns translate into functional requirements from domain experts to algorithms (Mikalef et al., 2022). Unintended consequences of AI systems can include programming errors, biased results, privacy breaches, and incorrect decisions (Dignum, 2019). To mitigate these risks, a structured, open, and value-centered design process is essential, along with formal mechanisms for measuring adaptability and robust fallback plans, such as switching to rule-based procedures or involving human operators. Intended malicious uses of AI, from email spam to cyber warfare, pose significant threats; addressing these requires a combination of regulation, responsibility, and developing AI techniques to assess and deflect malicious objectives (Dignum, 2019).

3.6. Challenges of Responsible AI

The literature identifies a range of challenges related to responsible AI, spanning technical, organizational, policy, and social dimensions. Khan et al. (2022) conducted a systematic literature review identifying the primary challenges in the areas of knowledge and expertise, organizational management, and available tools and technologies. The most frequently mentioned problem is insufficient ethical knowledge, with vague principles being the next most common issue. Other challenges include highly general principles, conflict in practice, interpreting principles differently, and lack of technical understanding. Addressing these challenges requires ongoing ethical education and adherence to relevant policies and regulations (Khan et al., 2022).

In a survey conducted by Morley et al. (2023), it was revealed that there is a lack of clarity over roles and responsibilities, a disconnect between availability and demand for pro-ethical design resources, and the challenge of justifying the costs of pro-ethical design without clear immediate returns. Schiff et al. (2020) identified five challenges in translating AI principles into effective practices including the complex and unpredictable nature of AI's social and ethical impacts, unclear accountability for ethical outcomes, conflicting disciplinary focuses, difficulty accessing and applying responsible AI methodologies, and organizational silos that impede effective interdisciplinary collaboration.

In a multivocal literature review Lu, Zhu, Xu, Whittle, et al. (2022) highlighted that the lack of explainability can erode trust and is recognized as one of the most pressing issues in responsible AI that needs to be addressed. Another challenge is that varying and slowly enacted AI regulations across jurisdictions create interoperability issues for organizations. Also, the challenge of insufficient training data for AI systems persists due to escalating concerns about data privacy (Lu, Zhu, Xu, Whittle, et al., 2022).

Deshpande & Sharp (2022) identified challenges in building responsible AI systems including the difficulty in considering the broader impact on all stakeholders beyond immediate users, a lack of clarity and consistency in ethical AI guidelines, and the varying legal and regulatory obligations across different jurisdictions. Additionally, conflicting priorities at individual, organizational, and national levels complicate the integration of ethical considerations into AI system development and maintenance (Deshpande & Sharp, 2022).

3.7. Product Management (PM)

The term "Product Management" is used for hardware products as well but for the purpose of this research it will be used interchangeably with "Software Product Management or SPM".

There is no universally accepted definition of product management, but it is generally agreed that SPM merges technical and business perspectives in software product development Maglyas et al. (2017). Ebert & Brinkkemper (2014) define product management as "the discipline and business process which governs a product from its inception to the market or customer delivery and service in order to generate biggest possible value to the business". The goal of SPM is to extend the lifespan of products as much as possible while keeping expenses low, maximizing profits, and maintaining market leadership (Springer & Miler, 2022).

According to Maglyas et al. (2017), software product management (SPM) is a multidisciplinary field that includes product and release planning, strategy, and development, supporting a product from inception to maintenance. It emphasizes the product's value to customers, aligning it with value-based software engineering. SPM is a complex socio-technical phenomenon, involving both social interactions within an organization and technical activities like development, architecture, and testing (Maglyas et al., 2017).

According to Springer & Miler (2022), SPM involves processes for defining, launching, developing, growing, maintaining, and retiring a software product. It is interconnected with strategy development, requirements engineering, project management, agile development, product marketing, and business analysis. Unlike project management, it prioritizes customers, sales, user feedback, and ongoing product growth (Springer & Miler, 2022). In a broader scope, SPM is defined as business management at the product, product line, or product portfolio level within a software organization (Maglyas & Fricker, 2014). It serves as a framework for strategizing, conceptualizing, developing, launching, managing, and marketing new products to the market (Maglyas & Fricker, 2014). Hyrynsalmi et al. (2021) highlight that the relationship between Software Product Management and general product management has been debated since SPM's emergence. Initially linked to Software Configuration Management and borrowing theories from mature fields like product management, requirements engineering, and marketing, recent work shows a decreased emphasis on direct connections to general product management (Hyrynsalmi et al., 2021).

4. Conceptual Framework

To analyze and discuss the research findings, two frameworks have been selected. The first is the Responsible Innovation (RI) Framework, which outlines key dimensions for conducting innovation responsibly: anticipation, reflexivity, inclusion, and responsiveness. It ensures that innovation processes are aligned with societal needs and ethical considerations. The second is the ISPMA SPM Framework, which offers an overview of software product management practices, including strategic management, product strategy, and product planning. It is designed to help organizations effectively manage the lifecycle of software products from conception to delivery, ensuring alignment with business goals and market demands.

4.1. Responsible Innovation - AREA Framework

The concept of "Responsible Innovation (RI)" or the broader concept of "Responsible Research and Innovation (RRI)" is defined in various ways across policy and academic discourses (Herrmann, 2023). Stilgoe et al. (2013) define responsible innovation as "taking care of the future through collective stewardship of science and innovation in the present". Von Schomberg (2012) proposed a working definition for RRI i.e. "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)".

For the operationalization of responsible innovation, one prominent framework is proposed by Owen et al. (2013) that includes four dimensions of responsible innovation i.e. Anticipatory, Reflective, Deliberative and Responsive. These dimensions are adopted by Engineering and Physical Sciences Research Council (EPSRC) who preferred using the terms Anticipate, Reflect, Engage, and Act (AREA), but these terms align with the substance and meaning of the dimensions as defined by Owen et al. (2013). For this research, EPSRC's terms are used because of their mnemonic advantage.

The four dimensions of RI are described by Owen et al. (2013) as follows:

1. **Anticipate** (Anticipatory): Analyzing both intended and unintended impacts—economic, social, environmental, etc.—using techniques like foresight,

technology assessment, and scenario development. These methods help generate expectations and explore alternative outcomes by prompting questions like "what if..." and "what else might it do?" While not aimed at precise predictions, they provide a valuable space for identifying and discussing potential issues and impacts of innovations, fostering reflection on their purposes and promises.

2. **Reflect** (Reflective): Examining the fundamental purposes, motivations, and possible effects involves considering what is understood such as existing regulations, ethical reviews, or other governance mechanisms as well as what remains unknown, including associated uncertainties, risks, gaps in knowledge, assumptions, questions, and dilemmas.

3. **Engage** (Deliberative/Inclusion)**:** Opening up visions, purposes, questions, and dilemmas to extensive, collective discussion through dialogue and engagement with diverse stakeholders allows for a wide range of perspectives to reshape issues and identify potential conflicts. The aim of such deliberation is both normative, promoting democracy, equity, and justice, and substantive, ensuring that decisions about innovation reflect and incorporate diverse social knowledge, values, and meanings.

4. **Act** (Responsive)**:** Employing this collective process of reflection to guide and shape the direction, pace, and trajectory of innovation through effective participatory and anticipatory governance mechanisms. This should be an iterative, inclusive, and adaptable process with dynamic learning capabilities.

According to Owen et al. (2013), these combined dimensions aim to achieve two main goals: first, to build "reflexive capital" by iteratively and inclusively examining the purposes, processes, and outcomes of innovation; second, to use this reflexive capital to guide decisions on innovation goals and adjust its trajectory in response to uncertainty and change.

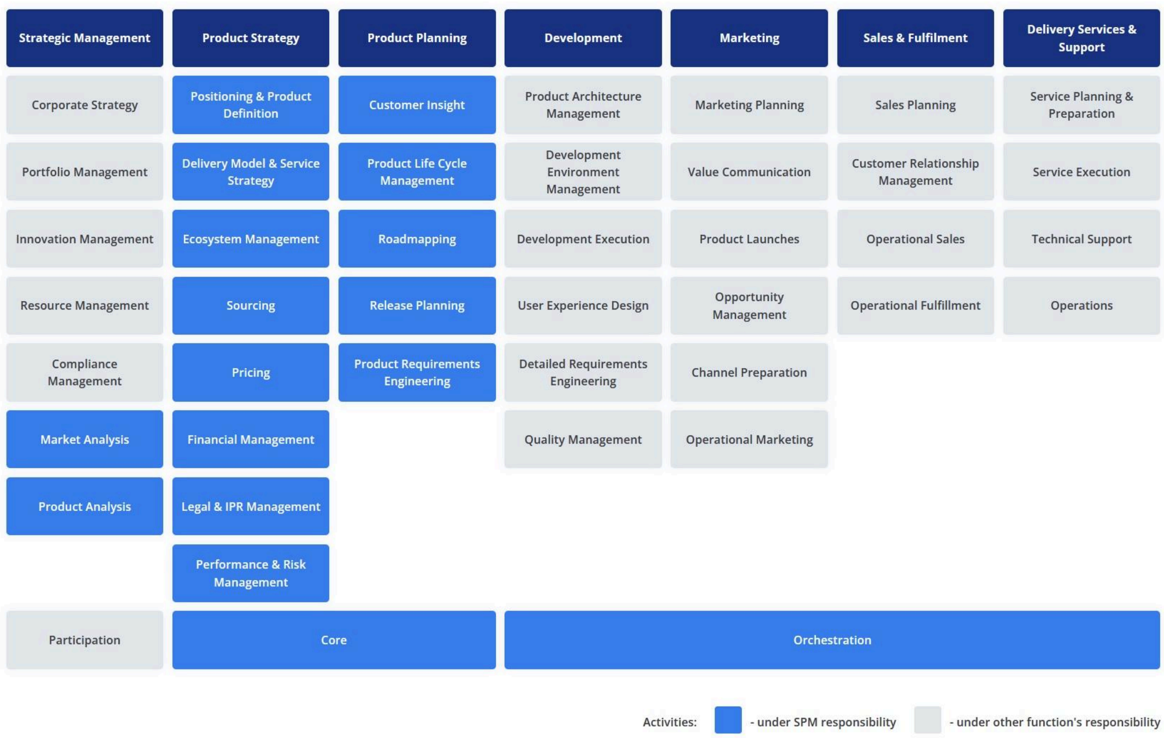
Dreyer et al. (2017) highlighted that RRI lacks clear consensus or connection to established disciplines like technology assessment and business ethics. It has also overlooked related advancements in areas like CSV and CSR discussions, sustainable finance, and ethical leadership. The business community views responsible innovation as reductionist and disconnected from current business practices in innovation and product development management, market analysis, and compliance (Dreyer et al., 2017). To address this gap, the ISPMA SPM framework has been chosen in addition to the RI AREA framework.

4.2. ISPMA SPM Framework

The ISPMA (International Software Product Management Association) SPM (Software Product Management) Framework is a comprehensive model designed to guide organizations in the effective management of software products throughout their lifecycle (Kittlaus, 2022). This framework encompasses various critical aspects of software product management, offering a structured approach to ensure alignment with strategic objectives, market needs, and operational efficiency (Kittlaus, 2022). The ISPMA SPM framework is divided into several main sections, each focusing on distinct areas of product management. These sections include Strategic Management, Product Strategy, Product Planning, Development, Marketing, Sales & Fulfillment, and Delivery Services & Support. Based on the level of involvement of product managers, these sections are categorized into three groups: participation, core, and orchestration. Each group encompasses various practices, outlining the specific responsibilities and focus areas within different aspects of product management.

Figure 1

ISPMA SPM Framework V.2.0



Note. From The International Software Product Management Association, by ISPMA, 2024, https://ispma.org/bok/

According to Kittlaus (2022), **Participation** includes practices where product managers represent their products at the corporate level. **Orchestration** includes practices that are managed by respective functions within the company, but orchestrated by the product manager to ensure alignment with the product strategy and plan. **Core** includes practices that product managers are usually directly responsible especially those related to "Product Strategy" and "Product Planning". For this research, core practices will be focused which are defined by Kittlaus (2022) as follows:

1. **Market Analysis:** Analysing current and future aspects of a market, including market structure, competitors, market shares, and customer preferences and behavior.
2. **Product Analysis:** Analysing current and future aspects of a business, including KPIs like revenue, revenue distribution, footprints, and market shares.
3. **Positioning & Product Definition:** Positioning defines how to communicate a product to customers, while product definition clarifies its purpose, scope, and evolving focus, including use, functionality, quality, design, compatibility, customization, and delivery.
4. **Delivery Model & Service Strategy:** The delivery model outlines how a product is made available, such as through licensing or SaaS, while the service strategy defines the services included and their providers.
5. **Ecosystem Management:** Coordinating and enhancing cooperation among stakeholders within a software ecosystem to boost business success, with responsibilities varying by company size and role.
6. **Sourcing:** Ensuring all required resources are available when they are needed.
7. **Pricing:** Set, communicate, and negotiate prices in a convincing way.
8. **Financial Management:** Planning, tracking, and influencing financial aspects.
9. **Legal & IPR Management:** Take care of all legal product aspects
10. **Performance & Risk Management:** Using KPIs to continuously monitor and improve product performance and identify and manage risks throughout the product lifecycle.
11. **Customer Insight:** Knowing and understanding of the problems and the environment in which customers operate.
12. **Product Life Cycle Management:** The management of the business and technical aspects of a software product with regard to its position in its life cycle.

13. **Roadmapping:** A flexible strategic tool for planning the long-term evolution of a software product, outlining intended deliverables and goals over time.
14. **Release Planning:** The process of selecting the requirements for the next release.
15. **Product Requirements Engineering:** A systematic approach to eliciting, documenting, analyzing, and managing requirements, while addressing market, technical, and economic goals.

While these practices are relevant to AI products as well, the ISPMA SPM framework lacks specific guidance on managing AI products except for a brief discussion under the "Data-Input-Driven Approach" section. Similarly, ethical issues related to AI are touched upon in the "Compliance Management" section, but it does not explore responsible AI principles, frameworks, or their relevance to AI product management. For an AI product manager, the ISPMA SPM framework alone may not adequately address the ethical challenges associated with AI, particularly in implementing responsible AI practices. For this, the RI AREA framework has been selected to complement the ISPMA SPM framework.

4.3. Integrating the Two Frameworks

This conceptual framework provides a structured and comprehensive approach to analyzing responsible AI principles in product management practices, ensuring that the research findings are grounded in both ethical considerations and practical management strategies.

The four dimensions of the RI AREA framework are used as the overarching structure where under each dimension relevant findings from the interviews will be discussed. The ISPMA framework will be used as a reference to relate those findings with various practices within product management. This will also help identify relevant practices from the ISPMA SPM framework and how they are related to specific dimensions of the RI AREA framework in the context of responsible AI integration and challenges that an AI product manager faces.

By integrating the RI AREA framework's ethical and societal focus with the ISPMA SPM framework's practical product management guidance, the findings will be analyzed to address how responsible AI principles are embedded in product management practices. This integrative approach ensures that AI products are not only technically and commercially viable but also ethically sound and socially responsible.

5. Methodology

This chapter details the research methodology and research methods employed in this study including literature review, pilot study, data collection, interview guide, sampling, data analysis, ethical considerations, and validation of research findings. By outlining the comprehensive process involved in this research, this chapter aims to provide a clear framework for researchers to replicate the study or leverage its methodology for future investigations.

5.1. Research Design

This study is grounded in inductive reasoning, pragmatic paradigm, and qualitative research approach with semi-structured interviews as the primary data collection method.

Inductive reasoning is suitable for this study because of the explorative nature of this study. Inductive reasoning involves open-minded data collection to form theories and draw generalizable conclusions from the gathered data. It involves an iterative process of moving back and forth between data collection and theoretical reflection (Clark et al., 2021).

Pragmatic paradigm is appropriate for this study as it emphasizes action, practical outcomes, and solutions to problems over antecedent conditions. It focuses on the research problem and utilizes diverse approaches to understand it, with methodological freedom. This will help address research objectives i.e. to examine approaches for integration of responsible AI in product management and to identify the practical challenges of it (Creswell, 2018).

Qualitative research approach is selected as it seeks deep insights by engaging people, communities, or organizations and interprets the social world through participant perspectives (Clark et al., 2021). Considering the explorative nature of this study, the qualitative approach is suitable for gathering insights and analyzing using an appropriate theoretical framework.

Semi-structured interviews (SSIs) were chosen as the data collection method because they offer a balance between the structured nature of close-ended surveys and the flexibility of open-ended sessions (W. C. Adams, 2015). SSIs are particularly suitable for gathering the independent thoughts of the interviewees (Clark et al., 2021). Given that this research topic is relatively uncharted, semi-structured interviews enable an exploratory conversation, allowing for the identification of useful leads for further discussion and analysis.

5.2. Literature Review

A semi-systematic literature review was conducted to identify and explore already existing literature on the subject. The following steps were taken to conduct this review:

1. **Reference Searching:** Keywords such as "Responsible AI", "Responsible AI principles", "Responsible AI frameworks", "Ethical Issues of AI", "Responsible Innovation", "Product Management", "Product Management Practices", "Product Management Frameworks", "AI Products", "Artificial Intelligence", "AI", and "Definition of AI" were used in various combinations to identify relevant studies.
2. **Reference Management:** References were shortlisted based on relevance, number of citations, and publication date.
3. Reading and Notes Taking: A three-pass approach (Keshav, 2007) was employed for further shortlisting and noting key points.
4. **Notes Organization:** Notes were organized under different themes both deductively and inductively.
5. **Paraphrasing and Summarizing:** Lengthy quotations were paraphrased and summarized to condense information.
6. **Synthesizing and Writing:** The final draft was prepared with appropriate headings and subheadings.

Step	Tool
Reference Searching	Google Scholar, Elicit, Litmap, Google Search
Reference Management Reading and Notes Taking	Zotero
Notes Organization	Microsoft Excel
Paraphrasing and Summarizing	ChatGPT
Synthesizing and Writing	Microsoft Word

Table 1: Tools Used for Literature Review

Note that Google Scholar, Elicit, and Litmap were used to search for scholarly literature. However, other material such as news reports, industry reports, policy guidelines, websites, and blog posts were also used for the review. Google search was used to find such material.

5.3. Data Collection

The primary objective of data collection was to gain insights from experienced professionals in the field and to understand their diverse perspectives. For this, data were collected through semi-structured interviews (SSIs) with industry practitioners, researchers, and policymakers working in product management and responsible AI spaces. Interviews were conducted online with the selected participants, lasting 20-54 minutes with an average of 30 minutes per interview. Microsoft Teams was used as the main tool for conducting these interviews. This platform facilitated audio and video recording, as well as the auto-generation of transcripts, ensuring an accurate and efficient data collection process. An interview guide was developed to ensure consistency across interviews. All interviews were recorded with participants' consent and transcribed verbatim for analysis.

5.3.1. Sampling

Purposive sampling was employed to select participants for this study, dividing them into two major groups: Product Managers working on AI products, and Consultants, Researchers, and Policymakers working in the responsible AI space (see Appendix A). This approach ensured that participants possessed sufficient and relevant experience to provide valuable insights on the subject (Martinson & O'Brien, 2015). A total of 20 participants were interviewed, comprising 12 AI Product Managers, 4 Responsible AI Researchers/Policymakers, and 4 Responsible AI Consultants from 10 countries, including the US, UK, Denmark, Germany, India, and China, and were affiliated with 20 organizations including unicorns, top-tier consulting firms, Ivy League university, as well as small to medium-sized enterprises.

LinkedIn was the primary source for identifying potential participants. Connection requests were sent to 145 individuals, of which 81 were accepted. Out of these, 74 were contacted for interviews, and 27 agreed to participate. Finally, 20 individuals were interviewed. 14 people declined due to various reasons, including company policy, workload, medical leave, paternity leave, and other personal reasons. Those unable to participate were asked to refer someone else who might be relevant and interested, which helped secure a few additional interviewees.

To identify potential participants, LinkedIn searches were conducted using keywords such as "AI Product Managers" and "Responsible AI." Another effective strategy for connecting with researchers and practitioners in the responsible AI field was sharing a summary of a YouTube lecture by a renowned professor in the field as a LinkedIn post. This post was reposted by the professor and 19 other individuals and organizations, resulting in over 9,500 impressions which include some of the interviewees.

5.3.2. Pilot Study

Considering the early and evolving nature of the research problem, this research study adopts an exploratory approach. To facilitate this process, a pilot study was conducted to benefit the initial phase and provide preliminary guidance. Clark et al. (2021) emphasize that pilot interviews are instrumental for testing interview flow, refining questions, and gaining experience to identify and address any preparatory issues. For the pilot study, a product management consultant, who is also a fellow at a product management association, was chosen through purposive sampling. A semi-structured interview was conducted with this participant via online meeting, allowing for an open exploration of the topics related to product management and responsible AI. The insights gained from the pilot study informed the subsequent research steps and shaped the direction of the study.

5.3.3. Interview Guide

Two interview guides were developed (see Appendix A) for conducting semi-structured interviews: one for AI product managers and one for RAI consultants, researchers, and policymakers. While both guides covered similar topics, they differed in focus. The first guide emphasized AI product management and the integration of responsible AI at the product level, whereas the second guide concentrated on responsible AI frameworks and their integration at the organizational level. These interview guides were developed based on the research questions, literature review, and pilot study. These interview guides help structure the conversation while allowing for flexibility to pursue emerging themes. The guide includes both overarching questions and probes to dive deeper into specific aspects of the participants' experiences (Clark et al., 2021).

5.4. Data Analysis

A thematic analysis was conducted to systematically analyze the interview data. This process follows six stages as proposed by Braun & Clarke (2006). However, some stages were adapted for instance themes were deductively identified instead of inductive identification.

1. **Familiarization:** Interview recordings were auto-transcribed verbatim, and converted into semi-verbatim using ChatGPT as well as manually, and initial ideas were noted.
2. **Initial coding:** Key phrases and sentences were coded and highlighted for quotations.
3. **Identifying themes:** Themes were identified based on the sections in interview guide.
4. **Reviewing themes:** Themes were reviewed to ensure alignment with coded data.
5. **Defining themes:** Themes were refined and renamed for consistency and coherence.
6. **Producing the report:** The final draft for the findings chapter was prepared.

Tool	Used For
MS Teams	Auto-Generating Transcripts
ChatGPT + Manual	Cleaning Transcripts
Microsoft Word	Coding, Highlighting, Writing and Editing
Microsoft Excel	Organizing Quotations under Themes

Table 2: Tools Used for Thematic Analysis

5.5. Ethical Consideration

Consent was obtained from all interview participants, ensuring they understood the purpose of the study and the measures in place to protect their confidentiality. Some of the interviewees were not concerned about their confidentiality while other interviewees explicitly asked not to disclose their identity. In either case, any identifying information was removed from the transcripts to maintain the anonymity of all the participants. However, interviewees were informed that their prior consent would be taken in case their identity is disclosed as part of a dissemination strategy to share the findings of this research via LinkedIn posts, video clips, articles, presentations, or any other appropriate channel.

5.6. Validation

Validation of the findings of this research was important to establish the credibility of the research. For this purpose, a summary of key research findings (see Appendix B) was sent to all the participants via email with a request to review and share their feedback. Some of the interviewees responded with their feedback and comments which can be found under the Feedback and Validation section in the Findings chapter of this report.

6. Findings

6.1. Pilot Study

In a pilot study, the interviewee emphasized the role of a product manager in upholding both strategic guidelines and ethical standards. It was noted that while compliance management is determined at the executive level, its implementation falls to product managers, who must address conflicts between ethical guidelines and business interests.

> *"which guidelines the company wants to follow in these areas [regulatory compliance, sustainability and ethics], are made on the level of the executive management. But once the decisions are made, the product managers need to make sure that these decisions are followed and implemented on the product side." - Interviewee #1*

The pilot study also highlighted product management practices relevant to ethics and compliance, including product strategy and planning, delivery models, service strategies, product lifecycle management, customer insight, and road mapping.

In addition to the established practices, the study identified two key areas requiring further empirical research to enhance the body of knowledge on product management: the management of AI products and the use of AI tools in product management.

> *"What are the special tasks of the product manager when he manages an AI-based product. In that area we need more input regarding best practices. The other area is the question of how can product managers make use of AI-based tools to improve the productivity and quality of their work." - Interviewee #1*

> *"those are the two areas, related to AI that are most relevant where we need to extend our body of knowledge over time. Based on the current situation that there is no real best practice yet in these areas, we could only add elements to our body of knowledge so far on a more abstract level that's what we have done." - Interviewee #1*

Given this research gap and the relevance of these topics to the research problem of this study, both of these topics have been covered during the interviews with other participants, particularly AI product managers.

6.2. Product Management Practices for AI Products

The core practices of product management remain the same whether dealing with AI products or traditional software products. However, there are some additional pracitces that are performed by the product managers in the case of AI products.

> *"The key principles remain the same. I feel it doesn't really change as much if it's an AI product or simple automation or some other thing." - Interviewee #6*

> *"There are certain things that are added probably when you are working with an AI heavy product. The outline still remains the same." - Interviewee #4*

A typical AI product management life cycle includes practices such as product discovery, product visions, road mapping, pricing, and release planning.

> *"The typical life cycle that we follow for AI product is we do discovery [prove the value of use case], then alpha [initial experiments], beta [scale experiments to larger users, tech ecosystem] and then we go live i.e. productionization.." - Interviewee #4*

> *"Making product visions, making sure it aligns with the company vision, product road mapping, and product planning. I do all of that." - Interviewee #7*

When asked about practices within the ISPMA SPM framework, most of the interviewees said all of them are relevant for AI products in some sense. However, some practices become more important. These include *Market and Product Analysis* to understand the needs for different ethical values and different responsible practices, *Positioning and Product Definition* as it has ethical implications, *Delivery Model & Service Strategy* which is often overlooked, *Ecosystem Management* and *Sourcing* since a variety of toolkits, frameworks, data are used in AI model building and training, *Legal and IPR Management* and *Performance and Risk Management* because responsible AI is rooted in risk analysis, management and mitigation, *Product Life Cycle Management* because ML/AI Ops is less mature and require scrutiny.

Nevertheless, the first step for building an AI product is to evaluate whether AI is the most appropriate solution for the problem at hand or not, as it may not always be the best fit.

> *"One of the key questions is also always, is AI the right solution for the problem? Because you could decide to use AI every single time if you wanted to, but it's not always the right solution to the problem." – Interviewee #5*

Therefore, it is important to focus on the user's problem rather than technology. Also, it can start with building with traditional logical structure before employing AI model in a product.

> *"We always start with the problem and then the users rather than have a technology in mind and trying to think about potential use cases." - Interviewee #9*

> *"the MVP [Minimum Viable Product] looks entirely different for AI. You want to see can I use that for data collection potentially so that I can actually in the future build something more sophisticated [using AI]?" - Interviewee #5*

One major difference that AI product managers have to deal with is that traditional software products are deterministic i.e. there is a degree of certainty in their outputs. In contrast, AI products are probabilistic i.e. their outputs cannot be predicted with 100% certainty. It requires additional safety nets and experiments as compared to traditional software products.

> *"a lack of predictability, that's why you need additional safety nets and safeguards around product to make sure it is properly implemented." – Interviewee #14*

> *"to increase the quality and the consistency of the responses, one needs to conduct a lot more experiments. Constantly conducting experiments is a big difference compared to traditional software development." – Interviewee #14*

Also in the case of AI products, product managers have to spend a significant amount of time on research than on traditional development-related tasks.

> *"Things becomes a bit more complicated because you also have the whole feasibility factor and a lot of more unknown. You are in research space more than you are in a classical development space." – Interviewee #5*

On the product development side, product managers have to deal with AI-specific coding such as consuming of foundation model APIs, building AI models, and data processing. This makes the management of AI products more technical than a typical software product.

> *"You have to deal with APIs on the back end and also a whole bunch of AI specific coding which requires processes such as data sourcing, collection and then making the model and/or finding existing models." - Interviewee #7*

> *"When it comes to product development it is much more technical and if you're not really good at it, good luck because you will not be able to optimize the product because you don't really understand the technicalities." - Interviewee #7*

On the people side, it is important to factor in the readiness and openness of the people to adopt AI systems. Customers are more concerned about their data in the case of AI products which requires product managers to communicate and educate customers about the product.

> *"Your customer base is gonna have questions. You need to understand what the solution is doing so that you can answer those questions." - Interviewee #3*

> *"The most important capability is communication. You will be interacting with customers that have no idea about what data science and machine learning is. You also have to be able to communicate with your team." - Interviewee #5*

> *"No one likes headcount reduction so are people open to share what they currently do or are they first educated enough to maybe adopt to start using a AI system. That angle also needs to be explored. - Interviewee #4*

Last but not least, because of potential ethical issues of AI products, product manager have to navigate various regulatory requirements and make sure that the product team is following necessary procedures where needed.

> *"Other differences would be the legal implications of AI, since unfortunately, AI is not particularly well. There's not a lot of regulation around AI today, and the lack of regulation creates ambiguity around its use." - Interviewee #14*

> *"Generally a product manager would get all the business requirements and then feed that into the technical team. But in AI side of it, responsible component, ethical considerations and regulatory requirements come into play. - Interviewee #10*

> *"You need to take into consideration, the upcoming EU AI act, then GDPR regulations, then industry specific data protection requirements." - Interviwee #15*

6.3. Practices for Integrating Responsible AI

The integration of responsible AI varies widely in maturity and approach across organizations. These strategies often span the entire organization rather than focusing solely on product management. However, product management remains a key component of it.

In most organizations, some dedicated people and teams are responsible for navigating regulations and implementing standard operating procedures.

> *"We have a Governing Council [a strategic initiative with representations from product management, tech, infrastructure team, architecture team, legal, etc.] who has laid out a standard operating procedure, and based on that we have tweaked our model development lifecycle with checkpoints for responsible AI." – Interviewee #4*

> *"the companies that have responsible AI teams do tend to be further ahead of their maturity. Because those responsible AI teams are responsible for being able to get some of that standard practice in place." - Interviewee #19*

AI model releases are guided from the start by structured templates, comprehensive documentation, and extensive testing to ensure adherence to best practices and safeguards.

> *"From the very beginning, we have templates, like a structured process that governs all AI large language model based releases. There's a lot of documentation that goes behind it to make sure that you're documenting your practices, safeguards, design principles. Then there's a lot of testing that goes behind it as well." – Interviewee #14*

Some organizations integrate validation and risk assessments throughout the AI development lifecycle, addressing high-risk applications with specific risk management strategies while incorporating ethical considerations from data collection through to deployment.

> *"We do validation and risk assessment multiple times during the model development lifecycle. If it is low then we continue development and it will eventually go to production. But if it's high risk application which impacts regulated process then we need to find out the ways of managing and mitigating the risks." - Interviewee #4*

One of the interviewees shared that they inform customers about the use of their data for transparency purposes. However, it is difficult to balance intellectual property rights with the level of transparency and explainability in a competitive market with similar products.

"We disclose how we use their [customer] data and that we do generate a derivative products based on that data. Users want a lot of insights and explainability into how the model works and how the model is built. But we also in a very competitive space where a lot of people are trying to build similar models." - Interviewee #5

One of the approaches for integrating responsible AI is incorporating a human in the loop to validate and make corrections to the predictions made by the AI system, particularly in high-stakes domains like finance, which helps build trust and ensures accuracy.

"we had a human in the loop, there was a person who was responsible for validating result or editing in case it was wrong and we used that kind of feedback to train the system back. Human in the loop helped build trust." – Interviewee #6

Integrating RAI is a collaborative effort where all the team members raise concerns. However, most of the concerns are raised by engineers who sometimes refuse to proceed. It helps draw the attention of the product manager to brainstorm for possible solutions.

"some engineers are like we are not gonna go ahead with doing this because this is wrong. That's when you're like, OK, let's actually focus on this problem that you just flagged and let's get to the bottom of this." - Interviewee #7

"They need to have that responsible AI culture. Product managers have to make sure that they lead the entire team focusing on whether they have taken into all these responsible AI practices." - Interviewee #10

Since it is a rapidly evolving field, product managers tend to learn from other people in the industry. However, it is not always helpful which leads them to continuously reference the latest academic work and expert insights to guide their approach.

"a lot of people don't have a lot of experience in dealing with it. So we can't always find a reference point in the industry. We have to look up research papers and see if there was any kind of publications in this area and who is the expert in this area and try to read their work pipeline. What kind of thing they suggested." - Interviewee #7

For the integration of responsible AI principles, some product managers adopt both open-source toolkits and specialized software solutions, with a focus on testing these frameworks in controlled environments before broader deployment.

> *"We are lately giving much importance to responsible AI processes and principles and toolkits that help you. There are some of the open source toolkit that will clearly help you to identify the decision making up that AI model." - Interviewee #11*

> *"there's some software solutions like Credo AI and Fiddler AI and Mission Control that are built to scale tested frameworks.." - Interviewee #19*

However, integrating AI vendor products requires careful consideration of ethical factors, including the source of the technology, intellectual property rights, and financial arrangements, to ensure responsible adoption and integration.

> *"Where you are sourcing from? What are the ethical considerations you have applied in order to bring that? Is it your IP [Intellectual Property] or their IP? Is there a certain percentage that they get? All those discussions come into play when you're adopting somebody's code or somebody's work in your product." - Interviewee #10*

One of the interviewees highlighted the importance of meticulous data management in model production, emphasizing the need for careful handling of training data to ensure quality and privacy. They stress the significance of preprocessing and post-processing steps to prevent issues like data contamination and leakage of sensitive information while maintaining rigorous standards for data sanitization and privacy throughout the model's lifecycle.

Some interviewees mentioned that their organizations are not integrating responsible AI principles into product management.

> *"My last two company actually did not employ responsible AI principle, and many of the companies do not. Once you are MVP where your product is working. That's the time it takes a focus but most of the time pushed at the backbench." - Interviewee #6*

> *"In product management, not yet. Because, we're still in the process of creating the guidelines on how to responsibly and ethically use AI tools." - Interviewee #9*

6.4. Frameworks of Responsible AI

The EU AI Act is seen as more stringent and comprehensive compared to frameworks from companies like Microsoft and Google, which are viewed as less impactful and motivated by self-preservation; larger institutional frameworks are considered more stable and authoritative than individual company proposals.

> *"The demands of the EU AI Act is much harder than the Microsoft and also Google's. It is very low compared to EU AI Act. OECD is much better and the EU AI Act has looked into OECD and covering some of the same areas." - Interviewee #2*

> *"Anything that is not governmental is just a source of inspiration. But what's really important is the regulator. It's good that we have AI act coming up." - Interviwee #15*

> *"bigger institutional frameworks that capture some degree of consensus are probably more likely to be stable than small one off proposals that come out from the people in AI research team at Google [for example]" – Interviewee #13*

Organizations are increasingly focusing on developing customized frameworks grounded in their own principles and specific use cases, integrating regulatory standards like the EU AI Act and OECD guidelines to create comprehensive and practical solutions that align with their core values and address conflicting regulations.

> *"Instead of picking straight out of someone's playbook, why don't we go to the source and build from the ground up ourselves like understand all those principles are and how specifically do those apply in our own specific use cases?" - Interviewee #9*

> *"We follow these principles [RAI principles] as well, but all of these have to tie up with our core values and core principles. And from there on bit draw how these principles will be brought into practice." - Interviewee #4*

> *"We developed a lot of tools and methods which we put together to our own framework. It's running quite successful. We are aligned with OECD. But for example, we [organization] have our own principles as well, which are a little bit broader than most of the principles which companies have." - Interviewee #12*

> *"those [EU AI Act, OECD] get integrated into the our framework, to make sure that our framework is as close to a superset of all the regulation as possible. Obviously, to*

> *be a true superset with all of the competing different frameworks, it's nearly impossible just because sometimes it's conflicting regulation." - Interviewee #14*

Organizations emphasize aligning with specific regulations like the EU AI Act, focusing on practical harm at the use-case level and integrating frameworks such as NIST for actionable guidelines, rather than solely relying on abstract principles.

> *"We personally aligned using the EU AI act, because we thought of harm at the use case level, in the operationalization or in the use of the system on the ground and the impact on end users. Not principles [like] OECD." - Interviewee #16*

> *"Principles are not enough. Frameworks should get into development processes. Something like the NIST risk management framework, you don't even necessarily need to elaborate the principles to begin implementing" – Interviewee #20*

> *"The framework that we use tends to be implementation agnostic. It gives a set of requirements and guidelines. It's up to the engineer to decide how they meet those requirements. It turns out that, organizationally, we've found a set of tools, services, libraries that we tend to trust. We use those by default." - Interviewee #14*

Existing frameworks often lack specificity and integration into product management workflows, requiring adaptation to fit roles and responsibilities; a comprehensive, product-life-cycle-spanning framework is needed for effective application across various aspects like risk, legal, financial, and sourcing.

> *"A lot of the frameworks does not really fall easily in the workflow because the frameworks are so loose, it's so hard to use them." – Interviewee #7*

> *"A lot of them are really not on product management. They're more like a social technical approach to how do we build solutions in a very generic way. You just have to retrofit it to you own roles and responsibilities." - Interviewee #9*

> *"I would say there's good frameworks built for a couple of these core cells, but if I were to say here's a RAI framework for product strategy, I don't think there's a comprehensive one that exists." - Interviewee #19*

Selecting an appropriate framework depends on the specific context of a business, including its size and industry, with a preference for globally recognized frameworks; companies should choose and justify their framework based on their unique needs.

> *"I would say choose your framework, back it up, have evidence, explain why, and then you should be OK. The framework that you choose is ultimately the discretion of each company. But as long as it's recognized by the EU AI act." - Interviewee #8*

> *"It's important that there's a community-driven direction that people follow. But it's just easiest to select one set that is globally accepted and adopted." - Interviewee #17*

> *"the kinds of interventions or frameworks that might work at Google are probably not going to work at a startup. Companies find or develop practices and interventions that are from organizations that are more similar to their own." - Interviewee #13*

Organizations frequently reference respected standards like OECD and NIST RMF for guidance, though these frameworks often require customization to effectively address specific industry needs and organizational contexts.

> *"OECD seems to be the largest and most respected community that people reference for standards more regularly than others. But we also consider NIST AI RMF, the UN, the Hiroshima Code of Conduct, and various others, including the Bletchley Declaration and Canada's standards" - Interviewee #17*

> *"the framework that is probably the most influential on things that we are doing at the [organization name] is the NIST RMF in part because that encompasses considerations that have high applicability in industry." – Interviewee #13*

Implementing clear frameworks accelerates development, boosts team confidence and creativity, and reduces the need to pull models from production.

> *"It speeds up development pace by just making it a clear path forward. I see people a lot more confident in their work. Increasing creative problem solving because these frameworks enable teams to understand the constraints and oftentimes you do need another layer of creativity and you need to push your thinking further. The last one is a reduction in having to pull a model from production." - Interviewee #19*

6.5. Challenges of Integrating Responsible AI

Interviews identified various challenges in regard to the integration of responsible AI. These challenges can be grouped into the following categories:

Increased Overheads

Emphasis on documentation and compliance in model development imposes a significant burden on teams, requiring extensive reporting, testing, and transparency.

> *"It's an additional overhead on the development team. You need to create reports, order trail, respond to multiple questions, make sure that the kind of data that you're using that is documented, the process that you're using for algorithm selection that is documented, explainability, transparency of models is documented." - Interviewee #4*

Unrealistic Customer Expectation

Users often expect flawless outputs from generative AI, despite the inherent limitations and the reality that the AI technology cannot achieve perfect accuracy.

> *"There is quickly a user expectation that the output is perfect because now it's a computer doing it's not a human so you expect the output to be perfect but in reality and AI is rarely capable of doing something to perfection." - Interviewee #5*

Complexity of the Whole Affair

Addressing the complexity of generative AI remains a significant challenge, with ongoing efforts to tackle fundamental issues like hallucination and the rapid evolution of the field.

> *"The complexity of the whole thing. That definitely is the biggest issue in terms of the product being something that can be relied upon and is trustworthy." - Interviewee #3*

> *"It's fundamentally hard problem and a lot of things are being worked in that space. For example, hallucination, there's no one solution to it." - Interviewee #6*

> *"Everybody is trying to get their head around. We have data scientists working on it every day. They say they have a hard time keeping up." – Interviewee #2*

> *"We're never gonna be able to predict all the challenges that AI throws at us just with the pace of how fast it develops." - Interviewee #19*

Productivity-Responsibility Tradeoff

Product development teams often face the dilemma of choosing between improving model performance or ensuring explainability and fairness, with limited resources and a focus on meeting business requirements driving their decisions.

> *"Sometimes there's X amount of resources and a limited amount of time and you have to choose. Do you make a better model? Or do you stop development and say we need to be able to explain [the model]?" - Interviewee #15*

Keeping up with Compliance

Navigating the evolving regulatory landscape requires significant effort to stay informed, ensure compliance, and effectively communicate complex regulations, because of a lack of clear standards and varying interpretations within the industry.

> *"Staying up to date with emerging regulation requires a lot of vigilance as well as work to then go back and respond back to it. Then helping our customers even understand what the regulation means and what it is because we're definitely in a moment where the overall landscape is changing" - Interviewee #14*

> *"If compliance frameworks aren't distilled and applied correctly, they can become too complex or overly simplistic for businesses to understand and use."- Interviewee #8*

> *"The lack of clear standards for specific types of systems and then the context in which those systems are being used." - Interviewee #17*

Team Coordination

Product managers often grapple with the complexity of integrating diverse and distributed teams, each with different focuses and interpretations of transparency, while also managing logistical and organizational challenges in a non-hierarchical structure.

> *"Integrating different and distributed teams together is one of the challenges typically project managers face here." - Interviewee #11*

> *"If you have a diverse teams, everyone is understanding transparency differently, so there is a lot of communication and a lot of what should we do with deciding what." - Interviewee #12*

> *"One challenge we face is simply trying to get a handle on all of the client work because we're not a hierarchically structured organization. That's a challenge just of logistics, communication and organization, not of like ethics or something of that nature." - Interviewee #16*

Skill Shortages

Product managers and development teams often struggle with limited AI expertise and resources, facing challenges in understanding AI principles and retaining skilled professionals because of rapid technological advancements and high turnover rates.

> *"Most of the product managers deal with the fact that they know very little. They don't really know AI so much themselves. The techniques require you to dig really deep to make the optimal decision." - Interviewee #7*

> *"lack of resources that are very conversant with AI principles or AI ethics and our ways of working if it's in PM. We find that even our legal team are maybe not as well equipped on AI strategies." - Interviewee #9*

> *"Retaining knowledge, the fact that data science and data engineering are positions that make or break your team. They have very high turnover because these skills are very much in demand." - Interviewee #15*

Negligence in the Early Stages

Product teams often overlook ethical considerations until the final stages, risking legal repercussions and project delays, as a lack of early focus on responsibility and compliance can lead to significant setbacks and reactive fixes.

> *"The challenge is whether they [product team] have taken into those ethical considerations is very important, else they succumb to those ethical pitfalls at the end when it comes back to bite them, when the customers throw in a lawsuit or something." - Interviewee #10*

> *"Ethical or being responsible, it sounds weak and it doesn't sound like there is business behind it. It is often neglected until the end of the launch and then just before the launch there is like we need to have the regulation in place." - Interviewee #12*

6.6. Strategies to Deal with Challenges

To address the challenges, AI product managers employ the following strategies:

Automated Risk Assessment and Metrics

One strategy is automating risk assessments to minimize overhead and implementing metrics to prevent hallucination. Also conducting extensive fairness and bias checks while avoiding synthetic data to ensure reliable and consistent performance.

> *"We're to create automated workflows where they [developers] just need to select 6-7 checkboxes and the risk assessment is done automatically. We're just trying to streamline that process so that the overhead is minimal." - Interviewee #4*

> *"We try to introduce metrics to test for hallucination for instance. We do the RAG [Retrieval-Augmented Generation] part to making sure that it will not answer in all kinds of areas that it should not." - Interviewee #2*

Customer Onboarding and Communication

Interactive onboarding sessions ensure new customers understand the product, while ongoing efforts focus on transparent communication of AI model outputs, accuracy, and data quality to align expectations and showcase capabilities.

> *"Every time we get a new customer on board, we make sure to do an onboarding session so that they have the fundamental knowledge [of the product]. They can ask questions and make sure that we are aligned on the capabilities." - Interviewee #5*

Customer Feedback Loop

Customer feedback mechanisms are used to identify and address issues like bad language, PII, and vulnerable code in AI outputs, which then inform updates to frameworks and guidelines to enhance quality and best practices.

> *"We also have mechanisms for our customers to provide feedback to us [like identifying] bad language, PII, potentially vulnerable code [generated by AI]. We take all of that input and we look for ways that we can improve the quality of the output. Then we feed that back to the folks who are defining the frameworks and the*

> *guidelines to say we think that there's a gap in our guidelines here. We should update this to use this new best practice." - Interviewee #14*

Building the Right Team

Hiring product managers with expertise in generative models and business is crucial, though training those without a math background is time-consuming; dedicated teams ensure adherence to responsible AI practices by collaborating across engineering and governance.

> *"You hire a product manager who have training in generative models and product management. If you don't have any, train them but training them takes a whole lot of time if they don't have mathematics background. So the best way is to just hire people with good understanding of generative models and good understanding of business and stuff." - Interviewee #7*

> *"We have multiple people dedicated to ensuring that we are following responsible AI best practices. They're product folks, they're not just a policy wonk. They work with engineering teams, with the governance teams back at [parent organization name] in order to make sure that we're both staying up to date with new guidelines, new policies, all of the tests that were implementing, the safeguards appropriately" - Interviewee #14*

Internal and External Collaboration

Organizations seek consultation from subject matter experts from both industry and academia. Also stay updated on regulations via social media and news, while collaborative team cultures and engaging workshops are essential for developing effective solutions through inclusive discussions.

> *"they [organizations] are asking for consultation advice. They also closely following the regulation and updates from people on LinkedIn and other social media platforms and news outlets. There are toolkits being produced and I think each company has their own toolkit." - Interviewee #8*

> *"I think if the culture of the team is collaborative, then you find solution quite easily. If you do like engaging workshops also you can create amazing solutions for these problems. But that is the work as a consultant, bring in everyone on the desk and then they talk. You need to translate for all of the voices." - Interviewee #12*

6.7. Implications

Impact on Product Management

One of the implications is that with AI products, product lifecycle management must continuously address ethical considerations and the evolving impact of AI models on users.

> *"performance [of AI models] can change over the course of time depending upon how you train, fine tune. The ethical questions will no longer be static so product lifecycle management has to constantly revisit for ethical alignment." – Interviewee #20*

One interviewee highlighted that compliance and documentation requirements for RAI implementation will slow down the process, while another stressed that product teams must embrace this, as it is essential for effective communication between different teams.

> *"productionization of solutions would become slower. Companies are trying to document experiments, that will also slow down the entire process." - Interviewee #4*

> *"The documentation is a hindrance to innovation, that mindset now for AI product development teams has to change." – Interviewee #20*

Product managers must possess technical skills and a solid understanding of technology to effectively lead and manage their teams, ensuring they are not overshadowed by developers.

> *"It's going to be difficult to be PM and not have tech exposure and tech skills and technology. The fundamentals should be in place. Otherwise it is not possible to be in control and you end up where the team is led by the developers." - Interviwee #15*

Improved Product Safety

Some interviewees believe as responsible AI practices advance, product safety will improve, leading to greater trust, and reduced risk for end users.

> *"Increased trust, increased understanding. Also weeding out of bad actors in the market, ideally. Better product safety that's what I hope." - Interviewee #17*

> *"As RAI matures, the safety of the products will mature with it and the end users, the public will be susceptible to fewer harms. That's my guess." - Interviewee #16*

Positive Influence of Regulations

Regulatory requirements are anticipated to encourage the adoption of responsible practices, ideally integrating safety considerations throughout the product development cycle.

> *"I'm hoping with some of the regulatory requirements that it will help unblock a little bit enough to get companies out of this inertia point and into more actual responsible practices and be able to grow from there." – Interviewee #19*

> *"In an ideal scenario, they [product teams] would be influenced in the sense that these kinds of considerations would be happening throughout the product development life cycle and that there would be a clear process to follow that improves the safety of the products." - Interviewee #13*

Automation and Skill Shift

As automation advances, it will reduce the need for human labor in routine checks. It will create the need for new skills and roles to adapt to the changing landscape of work.

> *"There will likely be more automation that helps to perform a lot of these checks without as much human intervention. I don't think human interventions ever gonna go away. But it won't require as much human labor." - Interviewee #14*

> *"We will have a lot of revenues generated by machines and the need of human labor will decrease. Part of RAI will also be the transitioning out to build new skills to help people to get a new skills and be employed for other task." – Interviewee #18*

Shift in the Nature of Products

The future will see a shift towards utilizing data and specialized models as central elements of innovation and business strategy, with an increasing focus on integrating data through language interfaces rather than traditional user interfaces.

> *"Increasingly, language will be the user interface and then we will be bringing data from multiple kind of sources to a single threaded conversation. I suspect that a lot more product managers will be thinking about how they create viable businesses and viable products based upon the value of their data, rather than just on the value of their user experience as defined by pixels on the screen." - Interviewee #14*

6.8. Recommendations

Recommendations varies due to the diverse structures of product management teams from single individual to centralized functions. However, there are some general recommendations that can be adapted by each organization according to their specific organizational setup.

Familiarize with the Regulations and Frameworks

Product managers should familiarize themselves with various regulation and frameworks, and utilize available resources and standards to ensure compliance and effective integration.

> *"Get introduced to the EU AI Act. I like the way EU AI Act thinks about the citizens and don't go for the lowest bar that you meet." - Interviewee #2*

> *"There is a whole suite of resources like RAI Institute or training that we provided or the NIST AI RMF. If organizations are overwhelmed by the EU AI act to start with, I think looking at a standard like ISO 42001 is a good way to start to get some of those controls in place at an organizational level." - Interviewee #17*

Think beyond Regulations

Although regulations are important and even binding in certain cases, interviewees encouraged to think beyond regulations and focus on the intent behind regulations and ensure transparency throughout the AI development lifecycle.

> *"Go beyond the letter of the law and think about intent. Because the legislation is always gonna be behind when we talk about this kind of fast moving IT technology. Have a guilty conscience. Be mindful of what could go wrong." - Interviewee #3*

> *"We should not focus only on the compliance aspects, but we should try to figure out a way to recheck the entire AI development life cycle." - Interviewee #4*

Adopt a Holistic Approach

Organizations should adopt a comprehensive, phased strategy for implementing responsible AI, starting with defining core values and operationalizing them across the organization.

> *"Look up best practices, build out the right policy into the software development life cycle in collaboration with legal, data privacy, infosec." – Interviewee #16*

> *"Start with a holistic full picture strategy, not a single strategy on a single product or a single ethical value and forgetting to look at the ecosystem that it exists in. It doesn't have to be done all at once. It's better to build over time." - Interviewee #19*

> *"Determine what their responsible AI values are depending on their specific use case. Then figuring out some of the operationalizations. How to translate from the higher order company level stuff all the way down to product teams." – Interviewee #13*

Engage Top Management

RAI should be a strategic priority involving top management to effectively allocate resources and mitigate risks. PMs should make a compelling case to stakeholders for investing in RAI, emphasizing its strategic value, risk mitigation, and potential to build user trust.

> *"It's a very strategic topic for companies in general and then in particular for product management. They [top management] should really understand why it is important as a strategic decision to invest time and resources in responsible AI." - Interviewee #18*

> *"Make a case to your stakeholders on how it would be beneficial if they adopt RAI practices and the kind of case you would make is it would mitigate a lot of risk, and it could build trust with the users." - Interviewee #6*

Build a Responsible Team and Culture

Organizations should prioritize hiring and cultivating a skilled, interdisciplinary team with a deep understanding of RAI principles. It is crucial to foster a cohesive team culture, ensure ongoing education and to communicate about these responsibilities.

> *"They cannot really be very responsible if they never know the details of what actually you have to be responsible for. So talent management, hire people who are amazing in generative models that's the first step in responsible AI." - Interviewee #7*

> *"Hire an interdisciplinary team. You need people able to understand the different components and they should be like an orchestra to approach this topic in cooperative manners with an open mindset and constructive mindset." - Interviewee #18*

> *"Responsible AI starts with the people and the company culture. Are your people trained and educated on these concepts and frameworks?" - Interviewee #19*

6.9. Other Insights

6.9.1. Use of AI Tools in Product Management

There was a mixed response to the use of AI tools, mainly LLMs. Some product managers find them useful for getting inspiration in certain use cases such as generating acceptance criteria, user stories, and user personas.

> *"I have been having great success for it [in-house LLM] to write acceptance criteria. Also using it to be creative around generating some input data. I'm not taking it in 100%, but using it as inspiration. […] also generated to do a little bit of API development but it's just more for the fun of it." - Interviewee #2*

> *"we use a couple of AI tools like the premium versions of ChatGPT and Claude for refining the ideas we already have, creating user stories, and suggest different user personas for specific use cases. Just a bit of suggestions, it's like a brainstorming partner." - Interviewee #9*

> *"[from a competitor perspective] I've had extensive looks at the ones that have a similar type of customer base. On a personal side, I only have used chatGPT [when] I was looking for a job, just a co-writer in my personal efficiency." - Interviewee #3*

On the other hand, some product managers do not use it instead they rely on their domain knowledge, experience, and templates that they have created.

> *"Typically no. With so many years [of experience] we kind of know what needs to be done and we have some piece of work already done in similar space. I typically rely on the sources that I already have. I have a nearly laid out the life cycle of a product and what needs to be created or what needs to be drawn out I can refer back to some of the content I already have". - Interviewee #4*

> *"I've tried using ChatGPT but I found it to be absolutely terrible. I have tried to use it a bit for inspiration in terms of communication, but I've rarely ended up using anything other than maybe the structure. But it's also because I am very deep in the industry, which means you have to obtain some kind of context knowledge which ChatGPT doesn't necessarily support at the same level." - Interviewee #5*

> *"the product development teams use it [GPT] if it's an LLM leveraged AI application. But if it is not, we don't use ChatGPT as a tool for the product management. I don't use it, but obviously people do leverage it." - Interviewee #10*

Some interviewees shared concerns regarding the use of AI tools because of data protection, intellectual property, and hallucination issues with possible solutions such as deploying in-house LLM and careful evaluation of the output generated by AI tools.

> *"You do not want to also compromise any IP related work. […] You may need to run it against your knowledge repository and not to the entire part of the ChatGPT because if you do that then it gets exports outside as well." - Interviewee #10*

> *"There needs to be an AI culture. If you're using a tool, that's great [but] you need to be mindful where you can't. You need to make sure it still goes through an internal QA process. If it's a question of transferring blame, you're the one using the tool. You're responsible for the output." - Interviewee #8*

6.9.2. Technical Workflow for Building AI Model

One of the interviewees explained the overall workflow of building an AI model as a product, which is converted into microservices and APIs. This process starts with understanding the problem, thinking about the solution approach, and then working with the data.

> *"We think about the data collection from multiple sources involving data preprocessing, data cleaning, data masking, data wrangling, data translation, and removing unnecessary white spaces, filling up unfilled fields in the data set. Then we check any outlier in the data set." - Interviewee #11*

> *"We spent much time on cleaning the data because when we have a right set and right amount of data, the AI model prediction will be on the right side. Data cleaning occupies a very important phase in our AI model engineering." - Interviewee #11*

After processing data, the next two stages are feature engineering and algorithm selection for which various manual and automated tools and techniques are used.

> *"From the data we create features, feature generation, feature optimization, feature selection. Features decide the quality of the prediction. We have to focus more on*

identifying the relevant features, irrelevant features can be removed. There are so many techniques for feature identification, removal, and choosing." - Interviewee #11

"There are several automated platforms and frameworks that help to create AI model based on the algorithm chosen. Also manually we can check with multiple machine learning algorithms like decision tree, random forest, support vector machine, logistic regression, linear regression. Then Boosting algorithms, bagging algorithms, calibrate algorithms, ensemble algorithms are there." - Interviewee #11

Another interviewee mentioned that most corporate ML models use custom implementations of open-source algorithms, such as those from scikit-learn, tailored to specific business cases. The focus is on data integration and responsible usage, whether its a simple regressions or advanced models like LLMs.

"In 80% of the cases we take from something like scikit-learn or a public Python library and then make a custom implementation and a custom business case around it on premise or on cloud for the corporate. But then you build a pipeline around it. It's all about the data that goes in the predictions it prints and how you integrate that with your in-house software." - Interviewee #15

Last stages include evaluation and optimization to meet performance standards. Once optimized, the model is deployed in a cloud environment using ML platforms, with continuous monitoring and updates to maintain performance and address issues like data quality decline and concept drift.

"Once a model is arrived then we evaluate whether the AI model is able to give the required result required performance required throughput. If they are not up to the mark, then we go back using back propagation to do the correction within the parameters. Then we will again check it. AI model engineering. AI model evaluation. AI model optimization." - Interviewee #11

"We use ML apps platform to deploy that ML model in a cloud environment. Then there are observability platform, security platform, monitoring framework that help to continuously update the model to keep up its performance because the data quality may go down, the concept drift can occur." - Interviewee #11

6.9.3. Regulatory Compliance

Overall, the response to AI regulations was positive. Regulations are viewed as drivers of innovation, with responsible AI practices offering a competitive advantage and enhancing product and company value. This approach is seen as fostering a virtuous cycle and increasing market share, despite ongoing challenges and areas for improvement.

> *"There is this old school view of compliance being the business prevention department. In this new world, I think compliance should be the innovation driver, the business facilitator." – Interviewee #8*

> *"I don't think in that way [regulations hinder innovation]. I think it brings value to the product, also to the company and then it will create a virtuous cycle and increase the market shares of the companies investing in responsible AI." – Interviewee #18*

> *"We are moving in the right direction. It's not perfect, there are a lot of open question. But at least it is the first step and this is an important first step. It's important to see that there is at least something and we can improve it with time." - Interviewee #18*

However, financial and other penalties for non-compliance remain a crucial aspect of regulation. The shifting landscape, including actions by venture capitalists and executive orders, is pushing companies to prioritize ethical frameworks and compliance to access lucrative markets like the EU and government contracts.

> *"Right now, the incentives are more into hey, you'll pay a lot of money if you don't comply with the acts [regulation]." – Interviewee #6*

> *"EU market is really important and they [other countries] just want to make sure their product based on AI will be marketable in the EU." - Interviewee #18*

> *"you [startup company] just pushed the ethics to the side. Then venture capitalists were starting to pull back their money, and they were starting to say, I ain't gonna fund a company that's not thinking about potential ramifications of their product. So the ground has been shifting in any particular place." - Interviewee #20*

> *"One of the things we've seen is that executive orders are pushing our government towards requiring the framework itself. Here everybody wants to sell to the*

> *government like everybody. So as they're rolling out AI products, they have to be thinking about the framework, and that's probably a better approach than trying to implement evaluations or standards of care." - Interviewee #20*

While regulation is important, the focus should be on deploying responsible AI solutions aligned with business objectives to prove their value. Demonstrating the business case for responsible AI fosters self-regulation, ensuring companies meet legal requirements and remain competitive in the market.

> *"Although regulation is important, regulation will not save us. I would say actually the effort should be channeled into deploying responsible AI solutions that are connected to business impact and business objectives so that we can prove the business case for responsible AI and ethics." - Interviewee #19*

6.9.4. Relevance of the Research

Most interviewees expressed their interest in reading the research outcomes, highlighting the lack of documented practices and a strong need for more information on the topic.

> *"I'm looking forward to your research. I've seen AI applications in product management, but not on how to do that in a responsible way. And if there's been that, there's no documented research around those practices." - Interviewee #9*

> *"We need more people like you going out there and performing this kind of survey. The entire industry is trying to make sense of this stuff and everyone, myself included, is just hungry for information and to understand what the consensus is. Thank you for going out there and helping to advance all of our work." - Interviewee #14*

One interviewee mentioned the issues of academic solutions in a practical setting. This highlighted the need for research focusing on industry practices.

> *"a lot of the solutions proposed in academia were not tractable for actual practice, either because it would take so much time to get situated in that literature or because they wouldn't integrate with workflows, weren't practical, relied on assumptions about data or capacities we would have that we didn't have." - Interviewee #13*

6.10. Feedback and Validation

Key findings were shared with the interviewees for feedback and validation. One interviewee noted a shift in the industry, with companies increasingly prioritizing AI effectiveness over RAI, privacy, and security concerns to achieve promised productivity gains.

> *"a lot of companies are starting to move past RAI, privacy, and security concerns and become much more interested in making AI effective in their workplace. The industry has collectively made a lot of promises about productivity gains, and now customers expect to realize those gains." - Interviewee #14*

This could mean ethical issues are overlooked. However, according to the interview, this shift is because of better understanding of responsible AI challenges and improved processes.

> *"Folks aren't overlooking RAI, but the problems/processes seem to be better understood, and therefore require less effort. I certainly don't want to imply that RAI is a "solved problem". Just like any engineering fundamental (e.g. security, privacy, accessibility, reliability), RAI will always require attention. But it's not as big, ambiguous, and unpracticed as it once was." - Interviewee #14*

Also productivity is not the only incentive to adopt AI, it has other use cases as well. For instance, LLMs are expected to be applied to security and vulnerability remediation.

> *"While "productivity" is still one of the most tempting benefits, I suspect that we'll see LLMs applied to more use cases than simply increased productivity over the next 1-2 years. Personally, I'm starting to look at more use cases related to security and code quality where LLMs can assist with vulnerability remediation." - Interviewee #14*

For AI product managers, this means they have to ensure customer success which requires working with solution engineers and communicating with customers.

> *"Consequently, an increasingly important part of Product Manager's job is to ensure customer success. This means working directly with solution engineers (i.e. the engineers+consultants typically responsible for working directly with customers) to ensure they also understand how the technology works, what patterns+processes result in good outcomes, and how to perform the change management necessary to adopt new tools." - Interviewee #14*

7. Discussion

7.1. Product Management Practices for AI Products

To effectively integrate responsible AI principles into product management, it is important to first understand the specific practices involved in managing AI products. The findings of this study highlighted that product management practices for AI products are built upon traditional practices but include additional considerations unique to the nature of AI technologies. Core practices such as product discovery, vision development, road mapping, pricing, and release planning remain central to managing AI products. However, these practices are adapted to address the specific challenges posed by AI's probabilistic nature.

Unlike traditional deterministic software, AI products require a more iterative and experimental approach due to their inherent unpredictability. This necessitates additional safety nets, extensive experimentation, and a greater emphasis on research. Also, AI product managers must navigate complex technical tasks, such as integrating APIs and building AI models. Building an AI model involves data handling, feature engineering, algorithm selection, optimization, cloud deployment, and ongoing monitoring to address data quality and performance issues. Understanding this technical workflow is important for product managers to integrate responsible AI throughout the AI model development lifecycle.

Other key practices include assessing whether AI is the right solution for a given problem, managing the AI-specific coding and data processes, and ensuring responsible and transparent practices to comply with regulations like the EU AI Act and GDPR. As an AI product manager, effective communication with both customers and development teams is also crucial to manage customer's expectations and help them understand the functionality of AI products.

This study also explored the use of AI tools in product management, especially large language models (LLMs), which received mixed responses. Some product managers appreciate LLMs like ChatGPT and Claude for tasks such as generating acceptance criteria and user stories, finding them useful for inspiration and brainstorming. However, others prefer relying on their domain expertise and established templates, arguing that AI tools may not provide the contextual accuracy needed for their specific needs. Some interviewees use in-house LLMs because of data protection, intellectual property, and the reliability of AI outputs concerns.

7.2. Integration of Responsible AI into Product Management

Integrating Responsible AI (RAI) into product management is essential for the ethical development and deployment of AI products. The majority of the interviewees said that their organization integrates responsible AI. However, this process varies widely in maturity and approach among organizations but typically involves several key practices. Structured templates, comprehensive documentation, and rigorous testing are some of the common practices to guide AI product releases. Some organizations integrated risk assessment throughout the AI development lifecycle, with particular attention to high-risk applications.

Effective RAI integration also requires collaboration across teams. Some organizations have dedicated RAI teams that create and enforce guidelines and procedures. These teams work with product management, legal, and other departments to ensure the alignment of various teams. Product managers often need to address concerns raised by engineers and other stakeholders, to build a culture of responsibility. Since the field of RAI is rapidly evolving, product managers must stay informed through academic research and industry best practices.

When it comes to responsible AI frameworks, most organizations customize them to fit specific needs and workflows. AI product managers must ensure that their chosen framework aligns with their organization's values and regulatory requirements while remaining adaptable to industry changes. These custom-built frameworks typically incorporate regulatory standards such as the EU AI Act and OECD guidelines. Frameworks from private organizations are often used as inspiration for technical setups but not necessarily for values and principles. This is partly due to product managers' skepticism about the intentions of private organizations and the fact that private company frameworks are non-binding and more prone to change compared to the EU AI Act.

Although most of the interviewees said their organizations integrate responsible AI, not all organizations prioritize this integration in their AI products. Some neglect RAI entirely, while others postpone it until the final stages of the product lifecycle, just before launch. This delay is often due to resistance from product teams, driven by limited resources and the eagerness to ship the product. Additionally, top management may fail to recognize the business value of integrating responsible AI, further hindering its implementation. This raises concerns about the potential harm these AI products might cause.

Based on the research findings, practices within the ISPMA SPM framework are mapped on each dimension of the Responsible Innovation framework, as presented in Table 3.

Responsible Innovation Dimensions	**ISPMA SPM Framework Practices**
Anticipate	**Market and Product Analysis:** Understanding market needs and ethical values through market and product analysis helps anticipate potential impacts and risks associated with AI products. This foresight ensures that AI systems are designed with future ethical considerations in mind and align with market demands and regulatory requirements. **Roadmapping:** Incorporates responsible AI practices by setting strategic milestones and checkmarks. This helps anticipate and address future ethical and operational challenges throughout the product's lifecycle. Structured templates and comprehensive documentation are used to foresee potential issues and guide the development process.
Reflect	**Product Life Cycle Management**: Managing the product throughout its lifecycle, including monitoring for model drift and tracking, requires ongoing reflection on the ethical implications and performance of the AI system. This continuous review helps adapt practices based on insights and challenges. **Performance and Risk Management**: Regular risk analysis and management practices involve reflecting on and mitigating risks associated with AI products. This reflexivity ensures that responsible AI practices are continuously evaluated and refined based on performance metrics and risk assessments.
Engage	**Customer Insight**: Gathering feedback from users and domain experts ensures that diverse perspectives are included in the development process. This practice helps align AI products with user needs and ethical standards by incorporating their

	insights into the product's design and functionality. **Ecosystem Management**: Coordinating with various stakeholders, including technology providers and partners, enhances the inclusivity of AI development. This collaboration ensures that different viewpoints and expertise are considered, contributing to a more responsible and comprehensive approach to AI.
Act	**Positioning and Product Definition**: Responsively defining and positioning the product involves adapting to ethical standards and regulatory requirements. Providing public-facing documentation and clear product definitions ensures that stakeholders are informed about the ethical aspects and functionality of the AI product. **Delivery Model & Service Strategy**: Adapting the delivery model and service strategy to include responsible AI practices demonstrates responsiveness to evolving ethical and regulatory requirements. Ensuring that documentation and processes are in place for different delivery teams helps maintain alignment with responsible AI principles throughout the product's deployment. **Legal and IPR Management**: Addressing legal and intellectual property issues responsively ensures compliance with regulations and ethical standards. This involves adapting to legal requirements and managing IP considerations to support the responsible use and development of AI technologies. **Sourcing**: Ensuring transparency and ethical practices in sourcing data and resources is a responsive approach to managing the inputs and impacts of AI products. This involves adapting sourcing practices to meet ethical and regulatory standards.

Table 3: Responsible Innovation Dimensions and ISPMA SPM Practices

7.3. Challenges and Strategies for Integrating RAI in PM

This study identifies several key challenges encountered by AI product managers during the integration of responsible AI and strategies for addressing these issues. A concise overview of these challenges and corresponding strategies is presented in Table 4.

The challenges identified in this study resonate with the challenges discussed in the available literature. For example, Morley et al. (2023) highlighted the challenge of justifying the extra time and costs of 'pro-ethical' design without a clear ROI. This aligns with the "increased overheads" and "productivity-responsibility tradeoff" challenge. Khan et al. (2022) mentioned management and technical staff lack awareness of the moral and ethical complexities involved in AI systems, this resonates with a "skills shortage". Regulations vary across jurisdictions (Lu, Zhu, Xu, Whittle, et al., 2022) which makes it difficult to "keep up with compliance". "team coordination" challenge is discussed by Schiff et al. (2020) who mentioned the "many hands" and "division of labor" problems in AI development which arise from the separation between technical and non-technical teams, leading to dispersed accountability and ineffective communication, which impedes addressing AI's broader social impacts. Schiff et al. (2020) also discussed the "complexity of the whole affair", highlighting AI's impacts on human well-being are complex and multifaceted, yet discussions often focus too narrowly on issues like bias and transparency, neglecting broader, interconnected risks.

The strategies in this study also align with the existing literature including engaging top management, forming a dedicated team for RAI (Lu, Zhu, Xu, Whittle, et al., 2022), customer communication, adopting a holistic approach (Akbarighatar, 2024) and external collaboration (Shneiderman, 2021). It important to note that these strategies are not universally applicable for various reasons. For instance, automated risk assessment might necessitate a custom-built tool which is a significant undertaking. Interactive sessions may be feasible for high-ticket B2B customers but not for low-ticket B2C customers. Forming a dedicated RAI team with highly skilled resources may be impractical for small startups with limited staff. Engaging top management in a large organization might be challenging for AI product managers.

Nevertheless, AI product managers can use the findings as a reference to understand the challenges they might face and to select, combine, and tailor strategies to address the specific challenges they encounter during the development and management of an AI product.

Challenges	Strategies
Increased Overheads: Documentation and compliance add significant burden on teams, requiring extensive reporting, testing.	**Automated Risk Assessments:** Create automated workflows for risk assessments and testing with minimal manual input.
Unrealistic Customer Expectations: Users often expect flawless outputs from AI, despite the technology's inherent limitations.	**Customer Communication:** Help new customers understand the product via onboarding sessions and in-app information.
Complexity of the Whole Affair: The complexity of ethical issues and the rapid pace of technological evolution of AI.	**Build a Right Team:** Hire product managers with expertise in both AI models and responsible AI challenges.
Productivity-Responsibility Tradeoff: Reluctance to RAI implementation from product teams due to limited resources.	**Think Beyond Regulations:** Communicate the potential risks to product teams beyond mere compliance including business risks.
Keeping up with Compliance: Navigating the evolving regulatory landscape to ensure regulatory compliance.	**Form an RAI Team:** To continuously monitor and adapt to the evolving regulatory landscape, ensuring ongoing compliance.
Team Coordination: Integrating diverse and distributed teams with varying focuses and interpretations of RAI principles.	**Adopt a Holistic Approach:** Build a comprehensive strategy and policy with clear responsibilities for each team member.
Skill Shortages: Limited AI expertise and high turnover rates of skilled professionals.	**External Collaboration:** Engage with subject matter experts and arrange team workshops to develop necessary skills.
Negligence in the Early Stages: Ethical considerations are often overlooked until the final stages of product development.	**Engage Top Management:** Build case for strategic value of RAI, its potential to build user trust and risk of regulatory panelities.

Table 4: Challenges and Strategies for Integrating RAI in PM

7.4. Limitations & Future Research

This research provides insights into the management of AI products and the integration of responsible AI principles. However, there are several limitations to this study that present opportunities for future research. Specifically, this study has three notable limitations:

1. **Scope**: This research offers a high-level understanding of the research problem but it does not cover detailed implementation of specific responsible AI principles. For example, mapping each Responsible AI (RAI) principle to specific practices within product management. Future research could look into implementation details for specific RAI principles to provide a technical guide for AI product managers.
2. **Sample**: This study primarily reflects the perspective of Product Managers. To gain a more holistic understanding, it is crucial to include the viewpoints of other stakeholders within the organization. Future studies could investigate the perspectives of the management team responsible for governance policies and the engineering teams involved in developing AI products and implementing RAI principles.
3. **Methodology**: This study employed interviews as the primary data collection method. Alternative methodologies, such as case studies or focus groups, could provide deeper insights. For example, conducting focus groups with diverse product teams—including product managers, designers, engineers, and experts in responsible AI policy and research—could facilitate a more focused dialogue on specific problems and solutions.

Future researchers can address any or all of these limitations to expand this research and contribute to the field of responsible AI, particularly regarding its integration into the product management of AI products. For this purpose, targeting "responsible AI products" directly could be a useful strategy, as these products deal with customers lacking responsible practices and can share their own implementation experiences and challenges.

8. Conclusion

AI products are increasingly embedded in our daily lives, despite the well-documented ethical concerns of AI. Massive investments are pouring in from venture capitalists and tech giants alike, signaling a future dominated by automation. But what kind of future are we heading towards with such automation?

> *"The rise of powerful AI will be either the best, or the worst thing, ever to happen to humanity. We do not yet know which." ("Stephen Hawking Warns of Dangerous AI," 2016)*

While we cannot predict the future with certainty, we do hold the power to shape it through our choices today, particularly those influencing the course of technological innovation. This research aims to guide, those who influence AI's trajectory, toward responsible innovation by offering insights into integrating responsible AI within product management. It explored current practices, challenges, and strategies employed by product managers.

Although integrating responsible AI is an organization-wide effort, product managers play a crucial role in orchestrating it. This involves engaging top management by presenting the business case for responsible AI, addressing productivity and responsibility concerns raised by product teams, and communicating with customers to build trust while navigating the evolving technological and regulatory landscape.

This study may not resolve the overarching ethical challenges associated with AI. However, it provides valuable insights for industry professionals, academics, and policymakers who are invested in addressing these concerns. Building a better future is our collective responsibility because this is what we pass on to the next generation. The responsibility does not lie on AI but on ourselves.

> *"We are building now what will remain for the future generation. We will be the ancestor of our kids when we will die, so how can we become better ancestor for them? I think this is a question in the context of AI." - Interviewee #18*

9. Bibliography

Abbass, H. (2021). Editorial: What is Artificial Intelligence? *IEEE Transactions on Artificial Intelligence*, *2*(2), 94–95. IEEE Transactions on Artificial Intelligence. https://doi.org/10.1109/TAI.2021.3096243

Adams, R., Adeleke, F., Florido, A., Santos, L. G. de M., Grossman, N., Junck, L., & Stone, K. (2024). *Global Index on Responsible AI 2024*. The Global Index on Responsible AI. https://coral-trista-52.tiiny.site/

Adams, W. C. (2015). Chapter Nineteen: Conducting Semi-Structured Interviews. In *Handbook of Practical Program Evaluation* (pp. 492–505). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119171386.ch19

Adobe. (2024). *Adobe Firefly—Free Generative AI for creatives*. https://www.adobe.com/products/firefly.html

AIAAIC. (2024). *AIAAIC*. https://www.aiaaic.org/home

Akbarighatar, P. (2024). Operationalizing responsible AI principles through responsible AI capabilities. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00524-4

Allen, G. (2020). *Understanding AI Technology*.

Aydın, Ö., & Karaarslan, E. (2023). Is ChatGPT Leading Generative AI? What is Beyond Expectations? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4341500

Barletta, V. S., Caivano, D., Gigante, D., & Ragone, A. (2023). A Rapid Review of Responsible AI frameworks: How to guide the development of ethical AI. *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*, 358–367. https://doi.org/10.1145/3593434.3593478

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2019). *Artificial Intelligence in Practice: Implications for Information Systems Research*.

Benjamins, R., Barbado, A., & Sierra, D. (2019). *Responsible AI by Design in Practice* (arXiv:1909.12838). arXiv. https://doi.org/10.48550/arXiv.1909.12838

Bloomberg. (2024). *WHO's New AI Health Chatbot SARAH Gets Many Medical Questions Wrong—Bloomberg*. https://www.bloomberg.com/news/articles/2024-04-18/who-s-new-ai-health-chatbot-sarah-gets-many-medical-questions-wrong?utm_source=website&utm_medium=share&utm_campaign=twitter&sref=E9Urfma4&mc_cid=775f4d56eb&mc_eid=c1a573459d

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. https://doi.org/10.1191/1478088706qp063oa

Clark, T., Foster, L., Sloan, L., Bryman, A., Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). *Bryman's Social Research Methods* (Sixth Edition, Sixth Edition). Oxford University Press. https://global.oup.com/academic/product/brymans-social-research-methods-9780198796053

Claude. (2024). *Claude*. https://claude.ai/

Creswell, J. W. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.).

de Laat, P. B. (2021). Companies Committed to Responsible AI: From Principles towards Implementation and Regulation? *Philosophy & Technology*, *34*(4), 1135–1193. https://doi.org/10.1007/s13347-021-00474-3

Deshpande, A., & Sharp, H. (2022). Responsible AI Systems: Who are the Stakeholders? *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 227–236. https://doi.org/10.1145/3514094.3534187

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6

Dreyer, M., Chefneux, L., Goldberg, A., Von Heimburg, J., Patrignani, N., Schofield, M., & Shilling, C. (2017). Responsible Innovation: A Complementary View from Industry with Proposals for Bridging Different Perspectives. *Sustainability*, *9*(10), Article 10. https://doi.org/10.3390/su9101719

Ebert, C., & Brinkkemper, S. (2014). Software product management – An industry evaluation. *Journal of Systems and Software*, *95*, 10–18. https://doi.org/10.1016/j.jss.2013.12.042

EU AI HLEG. (2019a). *Ethics Guidelines for Trustworthy AI High-Level Expert Group on Artificial Intelligence*.

EU AI HLEG. (2019b, April 8). *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Euronews. (2024, May 28). *Google's new AI search tool causes concern due to misleading responses*. Euronews. https://www.euronews.com/next/2024/05/27/googles-new-ai-summaries-tool-causes-concern-after-producing-misleading-responses

European Parliament. (2024). *Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (ec) No 300/2008, (eu) No 167/2013, (eu) No 168/2013, (eu) 2018/858, (eu)*

2018/1139 and (eu) 2019/2144 and Directives 2014/90/Eu, (eu) 2016/797 and (eu) 2020/1828 (artificial Intelligence Act). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CONSIL%3APE_24_2024_REV_1&qid=1718701204804

GitHub. (2024). *GitHub Copilot · Your AI pair programmer*. GitHub. https://github.com/features/copilot

Hagerty, A., & Rubinov, I. (2019). Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence. *arXiv.Org*. https://arxiv.org/abs/1907.07892v1

Herrmann, H. (2023). What's next for responsible artificial intelligence: A way forward through responsible innovation. *Heliyon*, *9*(3), e14379. https://doi.org/10.1016/j.heliyon.2023.e14379

Hyrynsalmi, S., Suominen, A., & Seppänen, M. (2021). A Bibliographical Study of Software Product Management Research. *2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 1–8. https://doi.org/10.1109/ICE/ITMC52061.2021.9570214

ISPMA. (2024). *The International Software Product Management Association*. https://ispma.org/bok/

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), Article 9. https://doi.org/10.1038/s42256-019-0088-2

Jolly, B. (2024, March 28). *Alarming moment robot arm crushes worker to death in factory*. The Mirror. https://www.mirror.co.uk/news/world-news/shocking-moment-robot-arm-crushes-32459232

Keshav, S. (2007). How to read a paper. *ACM SIGCOMM Computer Communication Review*, *37*(3), 83–84. https://doi.org/10.1145/1273445.1273458

Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., & Akbar, M. A. (2022). Ethics of AI: A Systematic Literature Review of Principles and Challenges. *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering*, 383–392. https://doi.org/10.1145/3530019.3531329

Kittlaus, H.-B. (2022). *Software Product Management: The ISPMA®-Compliant Study Guide and Handbook*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-65116-2

Laato, S., Tiainen, M., Najmul Islam, A. K. M., & Mäntymäki, M. (2022). How to explain AI systems to end users: A systematic literature review and research agenda. *Internet Research*, *32*(7), 1–31. https://doi.org/10.1108/INTR-08-2021-0600

Lu, Q., Zhu, L., Xu, X., & Whittle, J. (2022). *Responsible-AI-by-Design: A Pattern Collection for Designing Responsible AI Systems* (arXiv:2203.00905). arXiv. https://doi.org/10.48550/arXiv.2203.00905

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2022). *Responsible AI Pattern Catalogue: A Multivocal Literature Review* (arXiv:2209.04963). arXiv. https://doi.org/10.48550/arXiv.2209.04963

Maglyas, A., & Fricker, S. A. (2014). The Preliminary Results from the Software Product Management State-of-Practice Survey. In C. Lassenius & K. Smolander (Eds.), *Software Business. Towards Continuous Value Delivery* (pp. 295–300). Springer International Publishing. https://doi.org/10.1007/978-3-319-08738-2_21

Maglyas, A., Nikula, U., Smolander, K., & Fricker, S. A. (2017). Core software product management activities. *Journal of Advances in Management Research*, *14*(1), 23–45.

https://doi.org/10.1108/JAMR-03-2016-0022

Manning, C. (2020, September). *Artificial Intelligence Definitions*. https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

Martinez, R. (2019). Artificial Intelligence: Distinguishing Between Types & Definitions. *ARTIFICIAL INTELLIGENCE*, *19*.

Martinson, K., & O'Brien, C. (2015). Chapter Eight: Conducting Case Studies. In *Handbook of Practical Program Evaluation* (pp. 492–505). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119171386.ch19

McCarthy, J. (2004). *What is Artificial Intelligence?*

Meek, T., Barham, H., Beltaif, N., Kaadoor, A., & Akhter, T. (2016). Managing the ethical and risk implications of rapid advances in artificial intelligence: A literature review. *2016 Portland International Conference on Management of Engineering and Technology (PICMET)*, 682–693. https://doi.org/10.1109/PICMET.2016.7806752

Metinko, C. (2024, April 29). *AI Funding Stays Strong Despite Few Billion-Dollar Rounds*. Crunchbase News. https://news.crunchbase.com/ai/robotics-venture-funding-strong-q1-2024/

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems*, *31*(3), 257–268. https://doi.org/10.1080/0960085X.2022.2026621

Mondal, S., Das, S., & Vrana, V. G. (2023). How to Bell the Cat? A Theoretical Review of Generative Artificial Intelligence towards Digital Disruption in All Walks of Life. *Technologies*, *11*(2), Article 2. https://doi.org/10.3390/technologies11020044

Morandín-Ahuerma, F. (2022). *What is Artificial Intelligence*. *3*(12).

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023). Operationalising AI ethics: Barriers, enablers and next steps. *AI & SOCIETY*, *38*(1), 411–423. https://doi.org/10.1007/s00146-021-01308-8

Narayanan, M., & Schoeberl, C. (2023). *A Matrix for Selecting Responsible AI Frameworks*. Center for Security and Emerging Technology. https://doi.org/10.51593/20220029

New York Post. (2024). *Tesla driver using autopilot mode charged with vehicular homicide in death of motorcyclist*. https://nypost.com/2024/04/24/us-news/tesla-driver-using-autopilot-mode-charged-with-vehicular-homicide-in-death-of-motorcyclist/

OECD. (2024). *Explanatory memorandum on the updated OECD definition of an AI system*. OECD. https://doi.org/10.1787/623da898-en

OpenAI. (2024). *DALL·E 2*. https://openai.com/index/dall-e-2/

Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media. *AI & SOCIETY*, *35*(4), 927–936. https://doi.org/10.1007/s00146-020-00965-5

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A Framework for Responsible Innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible Innovation* (1st ed., pp. 27–50). Wiley. https://doi.org/10.1002/9781118551424.ch2

Politico. (2024, April 11). *AI chatbots spread falsehoods about the EU election, report finds*. POLITICO. https://www.politico.eu/article/ai-chatbots-spread-falsehoods-about-the-eu-elections-report-finds/

Prem, E. (2023). From ethical AI frameworks to tools: A review of approaches. *AI and Ethics*, *3*(3), 699–716. https://doi.org/10.1007/s43681-023-00258-9

Product Hunt. (2024). *The Best Artificial Intelligence Apps and Products of 2024*. Product Hunt. http://www.producthunt.com/topics/artificial-intelligence

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Safdar, N. M., Banja, J. D., & Meltzer, C. C. (2020). Ethical considerations in artificial intelligence. *European Journal of Radiology*, *122*, 108768. https://doi.org/10.1016/j.ejrad.2019.108768

Samoili, S., Cobo, M. L., Gomez, E., De Prato, G., Martinez-Plumed, F., & Delipetrev, B. (2020). *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence* [Monograph]. Joint Research Centre (Seville site). https://publications.jrc.ec.europa.eu/repository/handle/JRC118163

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2020). *Principles to Practices for Responsible AI: Closing the Gap* (arXiv:2006.04707). arXiv. https://doi.org/10.48550/arXiv.2006.04707

Shneiderman, B. (2021). Responsible AI: Bridging from ethics to practice. *Communications of the ACM*, *64*(8), 32–35. https://doi.org/10.1145/3445973

Simonite, T. (2022). YouTube's Captions Insert Explicit Language in Kids' Videos. *Wired*. https://www.wired.com/story/youtubes-captions-insert-explicit-language-kids-videos/

Springer, O., & Miler, J. (2022). A comprehensive overview of software product management challenges. *Empirical Software Engineering*, *27*(5), 106. https://doi.org/10.1007/s10664-022-10134-5

Stahl, B. C. (2021). Ethical Issues of AI. In B. C. Stahl (Ed.), *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies* (pp. 35–53). Springer International Publishing. https://doi.org/10.1007/978-3-030-69978-9_4

Stephen Hawking warns of dangerous AI. (2016). *BBC News*.

https://www.bbc.com/news/av/technology-37713942

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Strelkova, O. (2017). *Three Types of Artificial Intelligence*. http://eztuir.ztu.edu.ua/jspui//handle/123456789/6479

Tecuci, G. (2012). Artificial intelligence. *WIREs Computational Statistics*, *4*(2), 168–180. https://doi.org/10.1002/wics.200

The Guardian. (2024, June 25). Music labels sue AI song generators Suno and Udio for copyright infringement. *The Guardian*. https://www.theguardian.com/music/article/2024/jun/25/record-labels-sue-ai-song-generator-apps-copyright-infringement-lawsuit

Toolify. (2023, November 28). *Unveiling the Hidden Dangers of Chat GPT*. https://www.toolify.ai/gpts/unveiling-the-hidden-dangers-of-chat-gpt-139610

Top Startups. (2024). *Top Startups—Funded by Sequoia, A16Z, YC*. https://topstartups.io/?industries=Artificial Intelligence

Vesnic-Alujevic, L., Nascimento, S., & Pólvora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, *44*(6), 101961. https://doi.org/10.1016/j.telpol.2020.101961

von Schomberg, R. (2012). Prospects for technology assessment in a framework of responsible research and innovation. In M. Dusseldorp & R. Beecroft (Eds.), *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden* (pp. 39–61). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93468-6_2

Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, *11*(2), 7–30. https://doi.org/10.15678/EBER.2023.110201

Wang, P. (1995). *On the Working Definition of Intelligence*.

Wang, P. (2008). *What Do You Mean by "AI"?*

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Wang, Y., Xiong, M., & Olya, H. (2020). Toward an understanding of responsible artificial intelligence practices. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 4962–4971. http://hdl.handle.net/10125/64352

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). *Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for research*.

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration. *International Journal of Public Administration*, *43*(9), 818–829. https://doi.org/10.1080/01900692.2020.1749851

Y Combinator. (2024). *AI (Artificial Intelligence) Startups funded by Y Combinator (YC) 2024*. Y Combinator. https://www.ycombinator.com/companies/industry/ai

Zhu, L., Xu, X., Lu, Q., Governatori, G., & Whittle, J. (2022). AI and Ethics—Operationalizing Responsible AI. In F. Chen & J. Zhou (Eds.), *Humanity Driven AI: Productivity, Well-being, Sustainability and Partnership* (pp. 15–33). Springer International Publishing. https://doi.org/10.1007/978-3-030-72188-6_2

10. Appendices

10.1. Appendix A - Interviews

10.1.1. Interview Guides

Sections	**Interview Questions**	**RQ**
Introduction	1. Could you please introduce yourself and share your experience with AI products?	General
AI Product Management	2. What kind of practices or activities do you perform as a product manager working with AI products? a. Do you use AI tools in your product management? Which ones and how? 3. How do you compare these practices with traditional product management practices?	General
RAI Integration	4. Does your organization integrate responsible AI principles into product management? a. If yes, how? If not, why not? 5. Do you use any specific RAI framework? Which one and why? If not, why not?	RQ 2
Challenges & Implications	6. What challenges do you face when you integrate RAI in product management? Examples? 7. How does your organization address the challenges of integrating RAI principles in PM? 8. What kind of implications do you foresee for integrating RAI principles in PM practices?	RQ 3
Recommendations	9. What would you recommend for organizations seeking to integrate RAI into their PM practices?	RQ 2
Concluding Remarks	10. Is there anything else you would like to add or if you have any questions?	General

Table 5: Interview Guide for AI Product Managers

Sections	**Interview Questions**	**RQ**
Introduction	1. Could you please introduce yourself and share your experience with responsible AI and AI products?	General
RAI Frameworks	2. How do you see different frameworks for RAI? a. Which ones are more relevant for product teams? 3. What are your thoughts on RAI framework by private companies [Google, Microsoft]? a. How would you respond to the concern regarding the credibility of RAI framework by private companies?	RQ 2
RAI Integration	4. How do organizations currently approach the integration of responsibility AI, particularly in their product management?	RQ 2
Challenges & Implications	5. What challenges do you see organizations face when they integrate responsible AI principles? 6. How are organizations currently dealing with these challenges? 7. What kind of implications do you foresee for integrating RAI principles in PM practices?	RQ 3
Recommendations	8. What would you recommend for organizations seeking to integrate RAI into their PM practices?	RQ 2
Concluding Remarks	9. Is there anything else you would like to add or if you have any questions?	General

Table 6: Interview Guide for RAI Consultants, Researchers and Policy Makers

10.1.2. Description of Interviewees

#	Designation	Category	Duration
1	Product Management Consultant	Product Management	48 minutes
2	Senior Product Manager	Product Management	40 minutes
3	Product Manager	Product Management	30 minutes
4	Associate Director Product Management	Product Management	33 minutes
5	Senior Product Manager AI	Product Management	42 minutes
6	Product Manager	Product Management	26 minutes
7	Senior Product Manager AI	Product Management	40 minutes
8	Director	Consultant RAI	20 minutes
9	Senior Product Manager	Product Management	30 minutes
10	Associate Director AI	Product Management	30 minutes
11	Chief Architect & VP AI Division	Product Management	23 minutes
12	Managing Consultant	Consultant RAI	20 minutes
13	AI Governance Fellow	Research & Policy	28 minutes
14	VP Product	Product Management	36 minutes
15	Product Manager AI	Product Management	32 minutes
16	Director	Consultant RAI	54 minutes
17	Managing Director	Research & Policy	39 minutes
18	Independent Ethics Expert	Research & Policy	32 minutes
19	Author & AI Ethicist	Consultant RAI	29 minutes
20	Professor & Director AI Ethics	Research & Policy	27 minutes

Table 7: Description of Interview Respondents

10.1.3. Codebook for Thematic Analysis of Interviews

#	Code	Description
Theme: Product Management Practices for AI Products		
1	Key Practices	Core activities for managing AI products.
2	Product Life cycle	Stages from discovery to deployment of AI products.
3	Traditional SPM	SPM practices for standard software like web, mobile app.
4	People Aspect	SPM practices that involve people (customers, teams).
5	Compliance	SPM practices that involve legal and ethical issues.
Theme: Practices for Integrating Responsible AI		
6	RAI People	Individuals and teams dedicated to RAI implementation.
7	Templates	Predefined formats for documenting AI development.
8	Risk Assessment	Evaluating potential risks throughout the AI lifecycle.
9	Transparency	Clear and accessible information about AI products.
10	External Resources	Utilizing industry research and expert insights for RAI.
11	Toolkits	Tools and resources for implementing RAI.
Theme: Frameworks of Responsible AI		
12	Private Frameworks	RAI frameworks proposed by private organizations.
13	Overlapping Principles	Shared principles across various RAI frameworks.
14	Custom Framework	RAI frameworks tailored for organizational needs.
15	Framework Issues	Challenges and limitations within existing RAI frameworks.
16	Framework Selection	Choosing the most appropriate RAI framework.

17	Influential Frameworks	Widely adopted RAI frameworks.
Theme: Challenges of Integrating Responsible AI		
18	Increased Overheads	Additional resources required for implementing RAI.
19	Customer Expectations	User expectations for the AI products.
20	Complexity	The intricate challenges involved in integrating RAI.
21	Productivity Tradeoff	Balancing productivity with RAI implementation.
22	Compliance Challenge	Complexities of adhering to RAI regulations and standards.
23	Team Coordination	Collaboration among teams for RAI implementation.
24	Lack of Knowledge	Insufficient expertise and understanding of AI and RAI.
25	Negligence	Lack of integration of RAI in the product lifecycle.
Theme: Strategies to Deal with Challenges		
26	Automated Workflows	Streamlining RAI processes through automated systems.
27	Checks and Metrics	Benchmarks and monitoring systems for AI performance.
28	Communication	Informing customers about AI product and use of their data.
30	External Resources	Utilizing industry research and expert insights for RAI.
31	Collaborative Culture	Teamwork and shared responsibility in AI products.
32	Feedback Loop	Customer insight identifying ethical issues in AI product.
Theme: Implications		
33	Impact on PM	Influence of RAI on product management practices.
34	Better Products	Improvement in quality and safety of AI products.
35	Improved Processes	Improvement in RAI implementation processes.

36	Further Automation	Automation of RAI operations that are manual currently.
37	Nature of Product	Change in characteristics and features of AI products.
Theme: Recommendations		
38	Regulations	Legal standards and guidelines governing AI development.
39	Beyond Regulation	Exceeding basic compliance to ensure RAI implementation.
40	Engage Management	Involving leadership in supporting RAI implementation.
41	Team and Culture	A collaborative environment for RAI implementation.
42	Holistic Approach	A comprehensive strategy for integrating RAI.
Theme: Use of AI Tools in Product Management		
43	Useful	Positive response to the use of AI tools
44	Not Useful	Negative response to the use of AI tools
45	Why Not Using	Reasons for not using AI tools
Theme: Technical Workflow for Building AI Model		
46	AI Model Building	Process of building an AI/ML Model
Theme: Regulatory Compliance		
47	Innovation	Regulations facilitating innovation
48	Penalties	Consequences of failing to meet regulatory standards in AI
Theme: Relevance of the Research		
49	Research Gap	Areas in AI PM lacking sufficient study or evidence

Table 8: Codebook for Thematic Analysis of Interviews

10.2. Appendix B - Key Findings for Feedback and Validation

Here are the key findings of the study after analyzing the data generated by 20 interviews with professionals in AI product management and responsible AI space.

AI Product Management Practices

- The core principles of product management remain the same for AI and traditional software products, though AI products require additional practices.
- The first step in AI product development is to evaluate if AI is the most appropriate solution for the problem, focusing on user needs rather than technology.
- AI product managers must handle the probabilistic nature of AI products, which requires additional safety nets, experiments, and safeguards compared to traditional software.
- Managing AI products involves more research than traditional software products due to feasibility factors and unknowns, placing product managers more in a research space.
- AI product management is more technical, involving AI-specific coding, consuming foundation model APIs, building AI models, and data processing.
- Effective communication and customer education are crucial as customers are often more concerned about data in AI products and may need guidance to adopt AI solutions.
- Navigating regulatory requirements and ensuring adherence to ethical considerations and responsible AI practices is essential, involving legal implications, data protection regulations, and industry-specific requirements.

Integration of Responsible AI in Product Management

Currently Status and Approaches

- Integration of responsible AI practices varies widely across organizations, often encompassing the entire organization rather than being limited to product management.
- Some organizations have dedicated teams and governing bodies for navigating regulations and implementing standard procedures for responsible AI.

- Structured templates, comprehensive documentation, and extensive testing are used to ensure adherence to best practices and safeguards for AI models.
- Risk assessments are integrated throughout the AI development lifecycle, with specific risk management strategies for high-risk applications.
- Balancing transparency with intellectual property rights is challenging, but informing customers about data usage remains a priority for transparency.

Frameworks of Responsible AI

- The EU AI Act is more stringent and comprehensive compared to company-specific frameworks, which are seen as less impactful and driven by self-preservation.
- Organizations are increasingly developing their own frameworks, grounded in their principles and use cases, and integrating regulatory standards like the EU AI Act and OECD guidelines.
- Emphasis is placed on aligning with specific regulations and integrating frameworks like EU AI Act, NIST for actionable guidelines to address practical harm at the use-case level.
- Existing frameworks often lack specificity and do not integrate well into product management workflows, highlighting the need for a comprehensive framework covering the entire product lifecycle.
- Companies are encouraged to select globally recognized frameworks based on their specific context, such as business size and industry, and customize these standards to effectively address industry-specific needs and organizational contexts.

Challenges of Integrating Responsible AI

- Increased Overheads: The emphasis on documentation and compliance in model development adds a significant burden on teams, requiring extensive reporting, testing, and transparency efforts.
- Unrealistic Customer Expectations: Users often expect flawless outputs from generative AI, despite the technology's inherent limitations and the reality that perfect accuracy is rarely achievable.
- Complexity of the Whole Affair: The complexity of generative AI, including issues like hallucination and the rapid pace of technological evolution, remains a significant challenge for ensuring reliability and trustworthiness.

- Productivity-Responsibility Tradeoff: Teams face a dilemma between enhancing model performance and ensuring explainability and fairness, often constrained by limited resources and business requirements.
- Keeping up with Compliance: Navigating the evolving regulatory landscape demands significant effort to stay informed, ensure compliance, and interpret complex regulations, especially due to the lack of clear standards.
- Team Coordination: Integrating diverse and distributed teams with varying focuses and interpretations of transparency poses logistical and organizational challenges, particularly in non-hierarchical structures.
- ·Skill Shortages: Limited AI expertise and high turnover rates create difficulties for product managers and development teams in understanding AI principles and retaining skilled professionals.
- Negligence in the Early Stages: Ethical considerations are often overlooked until the final stages of product development, leading to potential legal issues and project delays due to insufficient early focus on responsibility and compliance.

Strategies to deal with Challenges

- Automated Risk Assessments and Metrics: Automating risk assessments with minimal manual input and implementing metrics to test for hallucination are strategies used to ensure reliable AI performance and prevent errors.
- Customer Onboarding and Communication: Interactive onboarding sessions help new customers understand the product, while transparent communication about AI model outputs, accuracy, and data quality aligns expectations and showcases capabilities.
- Customer Feedback Loop: Mechanisms for collecting customer feedback on AI outputs, such as detecting bad language and sensitive information, are used to update frameworks and improve quality and best practices.
- Building the Right Team: Hiring product managers with expertise in generative models and business is crucial, and dedicated teams work to ensure adherence to responsible AI practices by collaborating across engineering and governance.
- Internal and External Collaboration: Organizations engage with subject matter experts, stay updated on regulations through various channels, and foster collaborative team cultures and workshops to develop effective AI solutions.

Implications

- Ethical Considerations: AI product lifecycle management must continuously address ethical considerations and the evolving impact of AI models on users.
- Compliance and Documentation: The need for rigorous compliance and documentation for responsible AI (RAI) implementation may slow down production but is essential for effective team communication.
- Product Safety and Trust: As responsible AI practices advance, product safety will improve, leading to greater trust and reduced risk for end users.
- Regulatory Influence: Regulatory requirements are anticipated to encourage the adoption of responsible practices, integrating safety considerations throughout the product development cycle.
- Increase in Automation: Automation will reduce the need for human labor in routine checks, creating new skills and roles to adapt to the changing AI-driven work landscape.
- Shift in Nature of Products: The future will see a shift towards utilizing data and specialized models as central elements of innovation and business strategy, with an increasing focus on integrating data through language interfaces rather than traditional user interfaces.

Recommendations

- Familiarize with Regulations and Frameworks: Product managers should understand regulations like the EU AI Act and utilize resources such as the NIST AI RMF and ISO 42001 standards to ensure compliance and effective integration.
- Think Beyond Regulations: Emphasis should be on the intent behind regulations, ensuring transparency throughout the AI development lifecycle, and considering potential risks beyond mere compliance.
- Adopt a Holistic Approach: Organizations should implement a phased strategy for responsible AI, starting with core values and integrating best practices across the software development lifecycle in collaboration with legal, data privacy, and information security teams.
- Engage Top Management: Responsible AI (RAI) should be a strategic priority involving top management to allocate resources effectively and mitigate risks, highlighting its strategic value and potential to build user trust.

- Build a Responsible Team and Culture: Hiring skilled, interdisciplinary teams and fostering a culture of responsibility and ongoing education are crucial for successful RAI implementation, ensuring team members understand and uphold RAI principles.

Other Insights

Use of AI Tools in Product Management

- Product managers have mixed responses to using AI tools, particularly LLMs, with some finding them useful for generating acceptance criteria, user stories, user personas, and for refining ideas, brainstorming, and creating specific use case suggestions.
- Some product managers prefer relying on their domain knowledge, experience, and pre-existing templates over AI tools.
- Concerns about data protection, intellectual property, and the accuracy of AI-generated content lead some to advocate for in-house LLMs and careful evaluation of AI outputs.
- There is a recognition that using AI tools requires a mindful approach, including adherence to internal QA processes and responsibility for the generated content.

Technical Workflow for Building AI Model

- Building an AI model as a product involves understanding the problem, devising a solution, and working with data, including extensive data preprocessing and cleaning to ensure high-quality predictions.
- The workflow includes feature engineering, where relevant features are generated, optimized, and selected to enhance prediction quality, utilizing various manual and automated tools and techniques.
- Algorithm selection involves testing multiple machine learning algorithms, such as decision tree, random forest, and support vector machine, and employing advanced techniques like boosting and ensemble methods.
- The final stages involve evaluating and optimizing the model to meet performance standards, followed by deployment in a cloud environment with continuous monitoring and updates to maintain performance and address issues like data quality decline and concept drift.

- Corporate ML models often use custom implementations of open-source algorithms tailored to specific business cases, with a strong focus on data integration.

Regulatory Compliance

- Current incentives focus on financial penalties for non-compliance, making risk and compliance crucial for AI products.
- Regulation is driving innovation, as incorporating responsible AI into products and processes provides a competitive advantage.
- Compliance should be viewed as an innovation driver and business facilitator, not a hindrance, and streamlined processes can enhance efficiency and productivity.
- Regulations add value to products and companies, creating a virtuous cycle that increases market share for those investing in responsible AI.
- There is a growing emphasis on ethical AI practices from investors and governments, with frameworks and executive orders pushing for responsible AI development to meet market and legal requirements.