AI in Product Management

Prof. Zillur Rahman
Department of Management Studies
Indian Institute of Technology, Roorkee

Lecture - 59
Challenges & Ethical Considerations

. Welcome to this NPTEL online certification course on artificial intelligence in product management. Now we are discussing module 59, and we will be talking about challenges and ethical considerations. So this is what we are discussing. That is part 13: ethical considerations and future trends in AI for product management.

And this is module 59. Now, to give you an overview of this module, we will start with understanding human rights in AI and UNESCO's ethics policies. Then we will talk about key concepts of ethical product management and its importance in AI development, identifying the characteristics of ethical products in AI, learning from AI case studies, misapplications, and their consequences. And then we will discuss exploring AI privacy protection tools and their roles in ethical product deployment. To start with the introduction, AI presents product managers with new ways to enhance development, optimize performance, and improve user experience.

However, it also introduces challenges like algorithmic bias, data security, and privacy concerns. As AI continues to evolve, product managers must be agile, continuously adapting to emerging trends while ensuring their products meet user expectations and ethical standards. Balancing personalization with privacy is crucial as consumers grow more concerned about data usage and transparency. Now, let us look at the human rights approach to AI.

The rapid rise in artificial intelligence has created many opportunities globally, from facilitating healthcare diagnosis to enabling human connections through social media and creating labor efficiencies through automated tasks. However, these rapid changes also raise profound ethical concerns. These arises from the potential AI systems have to embed bias, contribute to climate degradation, threaten human rights and more. Recognizing the urgency of this challenge, UNESCO published the first-ever Global Standards on AI Ethics, with 10 core principles lay out a human rights-centered approach to the ethics of AI. The first is proportionality and do no harm.

The use of AI systems must not go beyond what is necessary to achieve a legitimate aim. Risk assessment should be used to prevent harms which may result from such users. The second is safety and security. Unwanted harms, safety risk as well as vulnerabilities to attack security risk should be avoided and addressed by AI actors. The third is right to privacy and data protection.

Privacy must be protected and promoted throughout the AI lifecycle. Adequate data protection frameworks should also be established. The fourth is multitasking and adaptive governance and collaborations. International laws and national sovereignty must be respected in the use of data. Additionally, participation of diverse stakeholders is necessary for inclusive approaches to AI governance.

The fifth is responsibility and accountability. AI systems should be auditable and traceable. There should be oversight, impact assessment, audit, and due diligence mechanisms in place to avoid conflicts with human rights, norms, and threats to environmental well-being. The sixth is transparency and explainability.

The ethical deployment of AI systems depends on their transparency and explainability, that is, T&E. The level of T&E should be appropriate to the context, as there may be tensions between T&E and other principles such as privacy, security, and safety. The seventh is human oversight and determination. Member states should ensure that AI systems do not displace ultimate human responsibility and accountability. The eighth is sustainability.

AI technology should be assessed against its impact on sustainability, understood as a set of constantly evolving goals, including those set out in the UN Sustainable Development Goals. The ninth is awareness and literacy. Public understanding of AI and data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media, and information literacy. The tenth is fairness and non-discrimination.

AI actors should promote social justice, fairness, and non-discrimination while taking an inclusive approach to ensure AI's benefits are accessible to all. Now, let us look at the ethical policies of artificial intelligence by UNESCO. UNESCO's ethical policies on artificial intelligence aim to provide a comprehensive and global framework to guide the development, deployment, and governance of AI technologies in a way that ensures these systems align with fundamental human rights, social justice, and long-term sustainable development. The rapid growth of AI technologies has the potential to bring significant

benefits to societies, but it also raises ethical concerns such as biases, inequality, and potential misuse.

As such, UNESCO's policies focus on promoting principles that safeguard the well-being of all individuals, particularly vulnerable or marginalized groups. Now, the policy area 1: ethical impact assessment. Member states should introduce frameworks for impact assessment, such as ethical impact assessment, to identify and assess benefits, concerns, and risks of AI systems, as well as appropriate risk prevention, mitigation, and monitoring measures, among other assurance mechanisms.

Such impact assessments should identify impacts on human rights and fundamental freedoms, in particular, but not limited to the rights of marginalized and vulnerable people or people in vulnerable situations, labor rights, the environment and ecosystems, and ethical and social implications. Governments should adopt a regulatory framework that sets out a procedure, particularly for public authorities, to carry out ethical impact assessments on AI systems to predict consequences, mitigate risks, avoid harmful consequences, facilitate citizen participation, and address societal challenges. The assessment should also establish appropriate oversight mechanisms, including auditability, traceability, and explainability, which enable the assessment of algorithms, data, and design processes, as well as include external reviews of AI systems.

Ethical impact assessments should be transparent and open to the public where appropriate. Such assessments should also be multidisciplinary, multistakeholder, multicultural, pluralistic, and inclusive. Public authorities should be required to monitor the AI systems implemented and/or deployed by those authorities by introducing appropriate mechanisms and tools. Policy area 2. Ethical governance and stewardship.

Member states should assess AI systems for suitability, methods, and potential human rights impact, prohibiting use if violations are likely. States should encourage the public and private sectors to involve stakeholders in AI governance and consider appointing independent AI ethics officers to oversee ethical assessment and monitoring. States should develop a digital ecosystem to support ethical AI, promoting access to AI resources, infrastructure, and knowledge sharing for inclusive development. States should create mechanisms with support for low- and middle-income countries, particularly least developed countries, landlocked developing countries, and

small island developing states to ensure global involvement in AI governance, including facilitating travel for representatives to participate in international discussions. States

should use AI in sensitive areas. For example, law enforcement, welfare, media, and healthcare should implement safeguards to protect human rights. Digital capacity, judicial capacity in AI. States should focus on enhancing judicial skills.

To make AI-related decisions, uphold human oversight and protect human rights in AI use. The next is AI safety standards. Governments and multilateral organizations should lead in establishing international safety standards for AI, ensuring transparency and funding research on AI risks. Then comes human rights compliance. Policies should ensure AI aligns with international human rights, respecting cultural diversity without compromising universal rights.

Policy area 3 is on data policy. Member states should work to develop data governance strategies that ensure the continual evaluation of the quality of training data for AI systems, including the adequacy of data collection and selection processes, proper data security and protection measures, as well as feedback mechanisms to learn from mistakes and share best practices among all AI actors. Member states should ensure privacy safeguards, including laws against surveillance. AI actors should

follow international standards, conduct privacy assessments, and use privacy by design throughout AI systems. Member states should protect individuals' rights over personal data through frameworks ensuring transparency, safeguards for sensitive data, data protection, accountability, the right to access and erase data, and independent oversight, especially for commercial use and cross-border transfers. Member states should establish or strengthen data policies to ensure the security of personal and sensitive data, especially information that could cause harm if disclosed, such as criminal records, income, genetic and health data, and data on identity or social attributes. The fourth policy area is development and international cooperation. Member states and transnational corporations should

Prioritize AI ethics by including discussions of AI-related ethical issues in relevant international, intergovernmental, and multi-stakeholder forums. A number of states should ensure that the use of AI in areas of development, such as education, science, culture, communication and information, healthcare, agriculture and food supply, environment, natural resources, and infrastructure management, economic planning and growth, among others, adheres to the values and principles set forth in the recommendations. Member states should work through international organizations to provide a platform for international cooperation on AI for development, including by contributing expertise, funding, data, domain knowledge, infrastructure, and facilitating multi-stakeholder

collaborations to tackle challenging development problems, especially for LMICs, and particularly LDCs, LLDCs, and SIDS.

Member states should promote international collaborations on AI research, fostering inclusive research centers and networks that support greater participation and leadership from researchers in LMICs, LDCs, LLDCs, and SIDS. Member states should support AI ethics research by engaging international organizations, research institutions, and corporations to promote ethical use across public and private sectors, adopting ethical frameworks for different cultures and contexts, and advancing feasible technology-driven solutions. Member states should foster international cooperation in AI to bridge geotechnological divides, encouraging technological exchanges and consultations across countries, sectors, and technological levels, in line with international law

to promote inclusive innovation. Policy Area 5: Environment and Ecosystem. Member states and business enterprises should assess the direct and indirect environmental impact throughout the AI systems' lifecycle, including but not limited to its carbon footprint, energy consumption, and the environmental impact of raw material extraction for supporting the manufacturing of AI technologies, and reduce the environmental impact of AI systems and data infrastructure. Member states should ensure compliance of all AI actors with environmental laws, policies, and practices. Member states should introduce incentives when needed and appropriate to ensure the development and adoption of rights-based and ethical AI-powered solutions for disaster risk resilience, the monitoring, protection, and regeneration of environmental ecosystems, and the preservation of the planet.

Involve the participation of local and indigenous communities throughout the lifecycle of AI systems and should support circular economy-type approaches and sustainable consumption and production patterns. When choosing AI methods, given the potential data-intensive or resource-intensive character of some of them and the respective impact on the environment, member states should ensure that AI actors, in line with the principles of proportionality, favor data-, energy-, and resource-efficient AI methods. Requirements should be developed to ensure that appropriate evidence is available to show that an AI application will have the intended effect or that the safeguards accompanying an AI application can support the justification of its use. The sixth policy area is gender. Member states should ensure that the potential for digital technologies and artificial intelligence to contribute to achieving gender equality.

Is fully maximized and must ensure that the human rights and fundamental freedoms of girls and women and their safety and integrity are not violated at any stage of the AI systems lifecycle. Moreover, ethical impact assessments should include a transversal gender perspective. Member states should ensure that the potential of AI systems to advance the achievement of gender equality is They should ensure that these technologies do not exacerbate the already wide gender gap existing in several fields in the analogous world and instead eliminate those gaps. These gaps include the gender wage gap, unequal representation in certain professions,

Limited presence in top management boards and AI research teams, educational disparities, inequitable digital and AI access, and an unequal distribution of unpaid work and caregiving responsibilities. Policy area 7 is culture. Member states are encouraged to incorporate AI systems, where appropriate, in the preservation, enrichment, understanding, promotion, management, and accessibility of tangible Documentary and intangible cultural heritage, including endangered languages, as well as indigenous languages and knowledges. For example, by introducing and updating education programs related to the application of AI systems in these areas where appropriate and by ensuring a participatory approach targeted at institutions and the public. Member states are encouraged to examine and address the cultural impact of AI systems, especially natural language processing applications such as automated translation and voice assistance, On the nuances of human language and expression.

These assessments should inform strategies to maximize benefits, bridge cultural gaps, enhance understanding, and address negative impacts such as reduced use leading to the loss of endangered languages, dialects, and cultural expressions. The state should promote AI education and digital training for artists and creative professionals to assess the suitability of AI technologies for use in their profession and contribute to the design and implementation of suitable AI technologies, as AI technologies are being used to create, produce, distribute, broadcast, and uphold artistic freedom. Member states should promote awareness and evaluation of AI tools among local cultural industries and small and medium enterprises working in the field of culture to avoid the risk of concentration in cultural markets.

Policy area 8 is education and research. Member states should work with international organizations, educational institutions, and private and non-government entities to provide adequate AI literacy education to the public at all levels and in all countries, in order to empower people and reduce the digital divide and excessive inequalities resulting from the

widespread adoption of AI systems. Member states should promote the acquisition of essential skills for AI education, including literacy, numeracy, coding, digital skills, media literacy, critical thinking, teamwork, communication, and AI ethics, particularly in regions with educational gaps.

Member states should promote awareness programs on AI development covering data, opportunities, challenges, and the impact of AI on human rights, including children's rights. These programs should be accessible to both technical and non-technical groups. Member states should ensure that AI researchers are trained in research ethics and require them to include ethical considerations in the design, production, and publication, especially in the analysis of the datasets they use, how they are annotated, and the quality and scope of the results with possible applications. Member states should promote the participation and leadership of girls and women, diverse ethnicities and cultures, persons with disabilities, marginalized and vulnerable people or those in vulnerable situations, minorities, and all persons not enjoying the full benefits of digital inclusion.

In AI educational programs at all levels, as well as monitoring and sharing the best practices in regard to other member states. Member states should encourage private sector companies to facilitate the access of the scientific community to their data for research, especially in LMICs, in particular LDCs, LLDCs, and SIDs. Policy area 9 is communication and information. Member states should ensure that AI actors respect and promote freedom of expression, as well as access to information, with regard to automated content generation, moderation, and curation.

Appropriate frameworks, including regulation, should enable transparency for online communication and information operators and ensure that users have access to a diversity of viewpoints, as well as processes for prompt notification to users on the reasons for removal or other treatment of content. and appeal mechanisms that allow users to seek redress. Member states should use AI systems to improve access to information and knowledge. This can include support to researchers, academia, journalists, the general public, and developers to enhance freedom of expression, academic and scientific freedom, access to information, and increase proactive disclosure of official data and information. Member states should also create an enabling environment for media to have the rights and resources to efficiently

effectively report on the benefits and harms of AI systems and also encourage media to make ethical use of AI systems in their operations. Ethical Policy Area 10 – Economy and

Labour: Member states should assess and address the impact of AI systems on labour markets and the implications for education requirements in all countries, with special emphasis on countries where the economy is labour-intensive. This includes introducing core and interdisciplinary skills at all education levels and helping workers adapt to a changing job market and understand AI ethics. Skills like critical thinking, communication frameworks, empathy, and the ability to transfer knowledge should be taught alongside technical and low-skilled tasks, with curricula regularly updated to reflect in-demand skills. Memberships should support collaboration agreements among governments, academic institutions, vocational education,

And training institutions, industry worker organizations, and civil society to bridge the gap of skill set requirements to align training programs and strategies with the implications of the future of work and the needs of industry, including small and medium enterprises. Project-based teaching and learning approaches for AI should be promoted, allowing for partnerships between public institutions, private sector companies, universities, and research centers. Member states should ensure competitive markets and consumer protection by preventing abuses of dominant market positions, including monopolies in AI systems across their life cycles. AI developers should respect ethical AI standards when exploring or applying systems in countries with less developed regulations while complying with international laws and local standards.

Additionally, member states should collaborate internationally to strengthen regulatory frameworks, improve enforcement mechanisms, and ensure equitable access to AI technologies for all nations. Policy area 11: health and social well-being. Member states should endeavor to employ effective AI systems for improving human health and protecting the right to life, including mitigating disease outbreaks while building and maintaining international solidarity to tackle global health risks and uncertainties, and ensure that the deployment of AI systems in healthcare is consistent with international law and their human rights law obligations. Member states should ensure that actors involved in healthcare

AI systems take into consideration the importance of the patient's relationships with their family and with healthcare staff. Member states should develop guidelines for human-robot interactions and their impact on human-human relationships based on research and directed toward the future development of robots, with special attention to the mental and physical health of human beings. Attention should be given to the use of robot care for older people and persons with disabilities in education, as well as toy robots, chatbots, and companion

robots for children and adults. Member states should encourage and promote collaborative research into the effects of long-term interaction of people with AI systems, paying particular attention to the psychological and cognitive impact that these systems can have on children and young people. This should be done using multiple norms, principles, protocols, approaches, and assessments of the modification of behavior and habits, as well as careful evaluation of the downstream cultural and societal impacts.

Furthermore, Member States should encourage research on the effect of AI technologies on health system performance and health outcomes. Now, let us dive deep into ethical product management. With the pace of technology moving Like no other time in history, product managers find themselves at the intersection of innovation, business objectives, and ethical responsibilities. As digital products become increasingly integrated into our daily lives, shaping behaviors and influencing societies, the ethical implications of these creations have never been more profound or far-reaching.

From social media platforms, that can sway public opinion to AI systems that make decisions affecting millions. The products we build can transform the world, for better or for worse. The growing importance of ethics in product management stems from recognizing that greater responsibility comes with greater power. Key ethical challenges in tech today span a wide spectrum.

Data privacy and user consent have become critical concerns in an era of big data and personalized experiences. Algorithmic biases threaten to perpetuate and amplify societal inequalities. The addictive nature of some digital products raises questions about digital well-being and mental health. Environmental sustainability, accessibility, and the implications of AI-driven decision-making add further complexity to the ethical landscape. Now, let us look at what is an ethical product. An ethical product is one that is designed, developed, and marketed in a way that promotes fairness, transparency, social responsibility, and environmental sustainability. Ethical products consider the well-being of users, societies, and the environment throughout their life cycle, from sourcing materials to production, usage, and disposal. Ethical products consider the well-being of users, society, and the environment throughout their life cycle, from sourcing materials to production, usage, and disposal.

Such a product goes beyond Mere functionality or profit motives; it is carefully designed, developed, and marketed with a sense of accountability to people and the planet. Now, what are the key characteristics of ethical products? One is user safety and well-being.

Ethical products do not harm users physically or psychologically. They should avoid promoting addictive behaviors or disseminating harmful information.

The next is environmental responsibility. These products are developed with consideration for their environmental impact. This includes using sustainable materials and processes that do not contribute to climate change or resource depletion. Ethical products aim to reduce their carbon footprints and promote ecological balance.

The next is fair labor practice. Fair labor practices in ethical products are often associated with ethical sourcing practices, meaning that the materials used are obtained under fair labor conditions. This includes ensuring a safe working environment, fair wages, and respect for workers' rights throughout the supply chain. The next characteristic of ethical products is inclusivity and accessibility. Ethical products strive to be inclusive.

Catering to diverse users' needs and ensuring that no group is unfairly excluded from benefiting from the product. This involves considering various user experiences during the design and development phases. So now let us talk about ethical considerations in AI product management. With the rise of AI comes a new set of ethical challenges that product managers must address. As AI systems become more integrated,

into products, issues such as user privacy, algorithmic biases, and ethical concerns about the use of data come to the forefront. It is the responsibility of product managers to ensure that their AI-powered products are not only effective but also ethical and transparent. One of the most pressing ethical concerns in AI is the potential for algorithmic biases to affect decision-making processes. AI systems rely on data, and if the data is biased, it can lead to unfair or discriminatory outcomes.

Product managers must be vigilant in ensuring that their AI systems are trained on diverse datasets and that they actively monitor the output for signs of bias. Addressing these issues early in the product development process can help prevent ethical pitfalls down the line. User privacy is a critical concern as these systems require extensive personal information to operate effectively. With these needs comes a significant responsibility to protect and secure users' privacy.

Product managers must ensure that their products comply with data protection regulations and that users are fully informed about how their data is being used. Transparent communication with users is key to building trust and maintaining the integrity of the product. The rise of AI in machine learning has introduced a new layer of complexity to

product ethics. Algorithmic bias is a significant concern, as AI systems can perpetuate or amplify existing societal biases.

For instance, AI-driven hiring tools have been found to discriminate against certain demographic groups, while facial recognition systems have shown lower accuracy rates for women and people of color. Product managers must be vigilant in identifying and mitigating these biases, which often requires looking beyond the code to the underlying data and assumptions. that inform AI models. So, let us start by looking at the first one: the accountability and responsibility challenge. Determining accountability for AI-driven decisions can be challenging, especially when errors or unintended consequences occur.

Since AI decisions can impact users' lives significantly, there must be clear accountability structures. For example, if an autonomous vehicle causes an accident, it may be unclear whether responsibility lies with the software, the hardware provider, or the product team. Product managers should establish clear accountability frameworks within the product lifecycle and ensure that human oversight is embedded into AI systems. The next is the social and psychological impact AI-powered products can have on users' mental well-being, behavior, and social interactions.

The use of AI in social media, for instance, can contribute to issues like addiction, misinformation, and social polarization. For example, social media algorithms that prioritize engagement can lead to echo chambers or the spread of sensationalized content, impacting users' mental health and perspective. Product managers should design AI systems that prioritize users' well-being and avoid manipulative techniques. Next is the ethical use of data for AI training.

Many AI systems rely on data collected from online sources, sometimes without clear user consent, raising ethical questions about data ownership and usage rights. For example, facial recognition models trained on images scraped from social media without user consent raise serious ethical concerns about privacy. Product managers should ensure that training data is sourced responsibly with proper consent and usage rights. Next are the long-term societal implications. Widespread adoption of AI has potential long-term impacts on employment, economic inequality, and societal norms.

If not carefully managed, AI can exacerbate job displacement or deepen existing inequalities. For example, automation in industries such as retail or manufacturing can lead to significant job losses, affecting workers' livelihoods and exacerbating economic divides. Product managers should consider AI's broader societal impact, supporting upskilling

initiatives and promoting policies. The future of product management is inextricably linked with ethical considerations as technologies continue to advance and our products become increasingly integrated into people's lives.

The responsibility of product managers will only grow. We must stay vigilant. Continuously educating ourselves about emerging ethical issues and adapting our practices accordingly. But knowledge alone is not enough. The time for action is now.

As product managers and leaders, we have the power to shape not just our products but the future of technology and its impact on humanity. This is our call to action. Commit to embedding ethical considerations into every stage of your product development process. Foster a culture of ethical awareness within your teams and organizations. Develop and implement ethical metrics to measure the true impact of your products.

Stay informed about emerging technologies and their ethical implications. Engage in broader discussions about tech ethics and contribute to shaping ethical standards in our industry. Now let us look at AI case studies for misapplication and consequences. So we start with McDonald's and its AI experiment with drive-thru ordering blunders. After working with IBM for three years to leverage AI for drive-thru orders, McDonald's called the whole thing off in June 2024.

The reason? A slew of social media videos showed confused and frustrated customers trying to get the AI to understand their orders. One TikTok video In particular, it featured two people repeatedly pleading with the AI to stop as it kept adding more chicken nuggets to their order, eventually reaching 260. On June 13, 2024, an internal memo obtained by trade publication Restaurant Business revealed McDonald's would end the partnership with IBM and shut down the test.

The next example is Air Canada paying damages for chatbot lies. In February 2024, Air Canada was ordered to pay damages to passengers after its virtual assistant gave them incorrect information during a particularly difficult time. Jake Moffatt consulted Air Canada's virtual assistant about a bereavement fare following the death of his grandmother in November 2023. The chatbot told him he could buy a regular-price ticket from Vancouver to Toronto and apply for a bereavement discount within 90 days of purchase. Following that advice, Moffatt purchased a one-way ticket

—a CA$794.98 ticket to Toronto and a CA$845.38 return flight to Vancouver. But when Moffatt submitted his refund claim, the airline turned him down, saying bereavement fares

cannot be claimed after tickets have been purchased. Moffatt took Air Canada to a tribunal in Canada, claiming the airline was negligent and misrepresented information via its virtual assistant. According to tribunal member Christopher Rivers, Air Canada argued, claiming it could not be held reliable for the information provided by its chatbot.

Rivers denied that argument, saying that the airline don't take reasonable care to ensure its chatbots are accurate. So, he ordered the airline to pay Moffat Canadian dollar 812.02, including Canadian dollar 650.88 in damages. The third is NYC's AI chatbot encourages business owners to break the law. In March 2024, the Markup reported that Microsoft-powered chatbot MyCity was giving entrepreneurs incorrect information that would lead to them breaking the law. Unveiled in October 2024, MyCity was intended to help provide New Yorkers with information on starting and operating businesses in the city as well as housing policy and workers' rights.

The only problem was that the Markup found MyCity falsely claimed that business owners could take a cut of their workers' tips Fireworkers who complain of sexual harassment and serve food that has been nibbled by rodents. Another case study, healthcare algorithm failed to flag black patient. In 2019, a study published in Science revealed that a healthcare prediction algorithm used by hospitals and insurance companies throughout the US to identify patients in need of high-risk care management program was far less likely to flag black patients. High-risk care management program provided trained nursing staff and primary care monitoring to chronically ill patients in an effort to prevent serious complications.

But the algorithm was much more likely to recommend white patients for these programs than the black patients. The study found that the algorithm used healthcare spending as a proxy for determining an individual's healthcare need. But according to Scientific American, the healthcare cost of sicker black patients were on par with the cost of healthier white patients, white people, which means they received lower risk scores even when their need was greater. The study's researchers suggested that a few factors may have contributed.

People of color are more likely to have lower incomes, which, even when insured, may make them less likely to access medical care. Implicit bias may also cause people of color to receive lower-quality care. While the study did not name the algorithm or the developer, the researchers told Scientific American they were working with the developer to address the situation. Now, let us look at the tools for AI privacy protection. To protect privacy

from the potential risks associated with AI technologies, a range of tools and best practices have been developed to address concerns around data security, unauthorized access, and misuse.

These tools and strategies play a crucial role in maintaining trust, compliance, and the ethical use of AI systems. Here are some key tools that can help safeguard personal and organizational data. One is the Microsoft Responsible AI Toolbox. The Microsoft Responsible AI Toolbox is a suite of integrated tools and functionalities designed to help operationalize responsible AI in practice.

It provides a collection of models, data exploration and assessment tools, user interfaces, and libraries that enable developers and stakeholders of AI systems to develop and monitor AI more responsibly and take better data-driven actions. The toolbox includes four visualization widgets. The first is the Responsible AI Dashboard, a unified interface that consolidates mature Responsible AI tools, allowing comprehensive model assessment, debugging, and informed decision-making. The second is the Error Analysis Dashboard for identifying model errors and discovering cohorts of data in which the model underperforms.

Fairness dashboards. For understanding models' fairness issues, using various group fairness metrics across sensitive features and cohorts. Another is Google Cloud's AI Explanation and What-If tool. Google Cloud's AI Explanation and What-If tools are designed to increase fairness, responsibility, and trust in AI model decisions. What-If tool.

The What-If tool is an open-source visualization for inspecting any machine learning model. It allows users to explore how models behave under different scenarios, providing a deeper understanding of model decisions. The tool is compatible with various platforms, including Jupyter Notebooks, TensorBoard, and Cloud AI Platform Notebooks. AI explanations provide feature attribution for models deployed on AI platforms. It offers built-in visualization capabilities for image data and works on tabular data.

When requesting explanations, Users receive predictions along with feature attribution information, which shows how each feature contributes to the predictions. This feature helps identify the most important features and understand how they impact model decisions. Then there is Granica. Granica is an AI infrastructure platform.

for building traditional and generative AI that is safe, effective, and low-cost. Granica Screens offers real-time sensitive data discovery, classification, and masking for both data

lakes and end-user LLM prompts. Granica Screen uses high-efficiency ML-powered scanning algorithms to process and safely unlock data for training and prompting at high accuracy and without driving up compute costs. Nightfall AI Nightfall AI is an enterprise data leak prevention platform for software as a service, generative AI, email, and endpoints.

It provides sensitive data discovery, encryption, and exfiltration protection. It also serves as a unified platform for sensitive data mapping and management across SaaS applications. Arthur AI offers a suite of AI observability tools to help monitor and fix issues with AI models. Products include an LLM evaluation service and LLM firewall to validate user prompts and model responses and ML monitoring and optimization platform and AtherChat, which is a turnkey plug and play AI chat platform with an integrated firewall.

CrowdStrike Falcon is an AI native security solution that leverages machine learning and security and behavioral AI. This AI security software consolidate next generation antivirus, endpoint detection and response, and it will be 4 by 7 managed threats hunting service into a single lightweight agent. Providing organizations with complete and protection across their endpoints. CrowdStrike use behavioral AI to detect anomalies in user endpoint behavior. This means it can monitor current activity as compared with past users action to protect the perimeter.

Then comes Zscaler. Zscaler offers cloud-based internet security and web filtering, focusing on data loss prevention. Key data loss prevention capabilities include securing data in motion, endpoint data protection, and mitigating risks from misconfiguration. It provides visibility into users' input in generative AI applications and allows isolation of apps. to manage sensitive data risks.

It offers enhanced AI-driven outcomes and capabilities, specifically for IT and security teams, to manage their security efficiently. Cylance, now part of BlackBerry, is a leading provider of AI-driven endpoint security solutions. By using machine learning models, Cylance can accurately identify and stop emerging threats before they cause harm. The platform offers proactive threat hunting, automated threat response, and incident investigation capabilities to enhance security operations. So, to conclude this module, we first discussed human rights in AI and UNESCO's ethics policies.

Then, we reviewed ethical product management. Thereafter, we explored the key characteristics of ethical products. We also examined AI case studies to understand misapplications and their consequences. Finally, we discussed various AI privacy

protection tools. And these are some of the references from which the material for this module was taken.

Thank you.