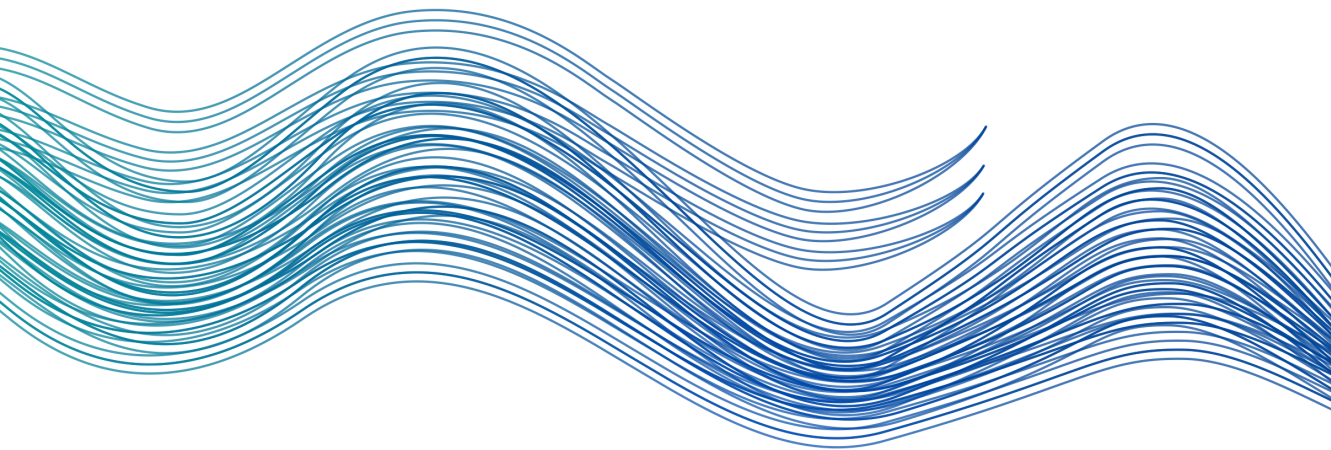**TURUN
YLIOPISTO
UNIVERSITY
OF TURKU**

ROADMAP TO COMPETITIVE AND SOCIALLY RESPONSIBLE ARTIFICIAL INTELLIGENCE

Matti Minkkinen & the AIGA project consortium

UNIVERSITY OF TURKU
Turku School of Economics
Department of Management and Entrepreneurship
Information Systems Science
ROADMAP TO COMPETITIVE AND SOCIALLY RESPONSIBLE
ARTIFICIAL INTELLIGENCE
February 2023

EXECUTIVE SUMMARY

The roadmap to competitive and socially responsible artificial intelligence (AI) offers an overview of AI governance drivers and tasks. It is intended for organizations using or planning to use information systems that include AI functionalities, such as machine learning, natural language processing, and computer vision. Responsible AI is still an emerging topic, but legal and stakeholder requirements for AI systems to comply with societally agreed standards are growing. In particular, the European Union's proposed Artificial Intelligence Act is set to introduce new rules for AI systems used in high-risk application domains. However, beyond binding legislation, soft governance, such as guidelines and ethics principles, already seeks to differentiate between socially responsible and irresponsible AI development and use practices.

The roadmap report begins by laying out its target group, instructions, and structure and then moves on to definitions. Next, we introduce the institutionalization of AI as a necessary background to the consideration of AI governance. The main roadmap section includes a visual representation and explanation of the six key drivers of competitive and socially responsible AI:
1) Movement from AI ethics principles to AI governance
2) Responsible AI commercialization potential and challenges
3) AI standardization
4) Automation of AI governance
5) Responsible AI business ecosystems
6) Stakeholder pressure for responsible AI

The roadmap is followed by a future research agenda highlighting five emerging research areas: 1) operational governance mechanisms for complex AI systems, 2) connections to corporate sustainability, 3) automation of AI governance, 4) future of responsible AI ecosystems, and 5) sociotechnical activities to implement responsible AI. Researchers and research funding bodies play a key role in advancing competitive and socially responsible AI by deepening these knowledge areas.

Advancing socially responsible AI is important because the benefits of AI technologies can be reaped only if organizations and individuals can trust the technologies to operate fairly, transparently, and according to socially defined rules.

This roadmap was developed by the Artificial Intelligence Governance and Auditing (AIGA) co-innovation project funded by Business Finland during the years 2020 to 2022. The roadmap was cocreated by researchers, company practitioners, and other AIGA project stakeholders.

Acknowledgments

Table of contents

1 Introduction

1.1 About this report

Target group. This roadmap report is for *companies and public organizations that develop and use AI or intend to use it in their operations, products, or services*. The report is written from the perspective of Finnish companies and applies in particular to companies operating in Europe that must comply with the coming European Union (EU) AI Act. However, the roadmap is also relevant to companies in other parts of the world, as similar AI regulations will likely be developed in regions beyond Europe.

Instructions for use. This roadmap presents *sets of tasks* that organizations need to execute to develop and use AI in a competitive and socially responsible manner by 2027. The boxes in the roadmap illustrations refer to these tasks, which are clusters of steps and processes rather than single activities. This level of abstraction makes it possible to fit the required tasks into one readable roadmap visualization. Based on how the sets of tasks have been designed, we, the authors of this report, consider them all necessary for the implementation of AI systems in a competitive and socially responsible manner. All the tasks need to be tackled somehow. However, we do not mandate a particular way of tackling these tasks, nor do we claim that there is only one correct way to do so. For some tasks, best practices (e.g., the implementation of standards) may emerge rapidly, while others (e.g., developing the organization's ecosystem strategy) may leave more room for differences in implementation.

Roadmap development process. The roadmap was developed iteratively during the Artificial Intelligence Governance and Auditing (AIGA) project funded by Business Finland. The planning started in January 2021, and the roadmap work began in August 2021. The development followed a design science approach (Hevner et al., 2004) with a dual emphasis on rigor and relevance. The process included an in-person AIGA consortium workshop on November 12, 2021, several virtual workshops and feedback sessions, and more focused dialogues.

Structure of the report. This roadmap report first defines the key concepts of competitive and socially responsible AI. Then, we outline the background of the institutionalization of AI. The main roadmap section (Roadmap to Competitive and Socially Responsible Artificial Intelligence) presents six roadmap drivers and the managerial and technical tasks under each driver. Together, they paint a comprehensive picture of developments in competitive and socially responsible AI over the next five years. After the roadmap section, we present a future research agenda with five areas that define the coming research landscape in AI governance and auditing.

1.2   Definitions of terms

In the following section, we establish some basic definitions. The aim is to provide working, rather than exhaustive, definitions for this roadmap document.

Artificial intelligence (AI). AI is an umbrella term that refers to various technologies and techniques united by the overall focus on developing humanlike capabilities. AI often refers to a system's ability to interpret and learn data and then use these learnings to achieve goals and tasks through adaptation (Kaplan & Haenlein, 2019). AI can also be understood as a moving frontier of computational advancements that addresses increasingly complex decision-making problems (Berente et al., 2021).

AI ethics. AI ethics refers to moral principles (e.g., fairness), ethical guidelines, codes, and frameworks that inform and guide the design, development, and deployment of AI (Koniakou, 2022).

AI system. An AI system is an information technology (IT) artifact that includes AI technologies and is surrounded by people, organizations, work systems, and institutions (Dignum, 2020; March & Smith, 1995). In other words, this roadmap document considers AI systems to be parts of sociotechnical environments that include human actors. In practice, the core of an AI system is an IT system that uses techniques such as machine learning and natural language processing.

Responsible AI system. A responsible AI system is a sociotechnical system wherein an AI agent and humans interpret and learn from data and use these learnings to achieve specific goals and tasks through flexible adaptation in a manner that is judged responsible according to rule- and consequence-based criteria (Zimmer et al., 2022; cf. Kaplan & Haenlein, 2019).

AI governance. Organizational AI governance is a system of rules, practices, processes, and technological tools that are employed to ensure that an

organization's use of AI technologies aligns with its strategies, objectives, and values; fulfills legal requirements; and meets the principles of ethical AI followed by the organization (Mäntymäki et al., 2022).

Competitive and socially responsible AI. By competitive and socially responsible AI, we mean AI that fulfills societal, legal, and business requirements. This means that the AI is developed and used sustainably and is, at the same time, competitive in the global market. These two aspects support each other in the long run because only legally and ethically sustainable AI can survive in the market once AI regulation is enforced. However, some companies may reap profits from unsustainable and ethically questionable practices and business models using AI. The competitiveness element means that AI is economically viable, and its use can be justified from a business perspective. Social acceptability, in turn, means that AI capabilities are used in a manner that is judged responsible according to applicable rules (e.g., legislation and ethical principles) and considers the foreseeable impacts of the AI system (Zimmer et al., 2022).

1.3 Background: Institutionalization of AI

The use of AI technologies is spreading rapidly in different sectors and use cases, such as healthcare, transport, and marketing. In particular, big data analytics and machine learning technologies have become many organizations' operations. More recently, generative AI technologies, such as the ChaptGPT chatbot and powerful image generation algorithms, are spreading to new application areas and potentially amplifying the disruptions caused by AI technologies. At the same time, the societal and legal governance of AI is actively debated worldwide. This shows that we are currently undergoing an institutionalization of AI both in the sense of AI diffusion and increasing demands to institutionalize effective AI governance. In parallel with the societal debate, AI regulation is proactively developed, most visibly in the EU with the upcoming EU AI Act (European Commission, 2021). In addition, AI ethics principles have proliferated, converging around principles such as transparency, fairness, and accountability (Jobin et al., 2019). This development has led to a situation in which organizations that use AI systems need to follow legal, societal, and ethical requirements, but these requirements are still evolving and are poorly understood. In this rapidly moving situation, organizations and political decision-makers need to look ahead to consider a roadmap of competitive and socially responsible AI for the coming years.

Over the past five years, responsible AI, AI governance, and AI auditing have been topics of increasing interest among practitioners and academics (Dignum,

2020; Mäntymäki et al., 2022; Minkkinen, Niukkanen, et al., 2022; Mökander et al., 2021). There is reason to believe that they will continue to be hot topics in the coming years, with the EU's proposed AI Act likely to become enforceable over the next five years.

Organizations are currently at different stages in their awareness and implementation of AI governance, and they are likely to take different development paths. While heavily regulated sectors, such as finance and healthcare, are up to speed on relevant regulations and required organizational processes, countless small and medium-sized companies are only now waking up to AI governance requirements. Nevertheless, we can perceive an overall institutionalization of AI governance, most visibly over the past five years and within the EU.

As a precursor to AI governance, the General Data Protection Regulation (GDPR) already regulates data collection and processing for organizations operating within the EU. Because AI systems learn based on data, regulations on data protection are highly pertinent to AI systems. In addition, there is sectoral legislation (e.g., in healthcare and transport) that already regulates AI systems. However, at present, this already-existing AI legislation is generally not well known (Viljanen & Parviainen, 2022).

The EU AI Act, proposed in 2021, is a recent important step forward in AI regulation (European Commission, 2021). The act is the product of legislative developments in the EU that commenced in 2018 with the Declaration of Cooperation on Artificial Intelligence. The AI Act outlines requirements related to risk management, transparency, and quality systems, among other topics (European Commission, 2021). By means of the ensuing policy process, the EU aspires to be a key player in defining rules related to digitalized societies.

In addition to binding legislation, soft law instruments, such as standards and certificates, can play an important role. Standards can complement binding legislation by providing voluntary and concrete guidance to help organizations comply with legislation. More broadly, internationally recognized AI standards can disseminate best practices, foster trust among stakeholders, and promote the beneficial development of AI systems (Cihon, 2019). AI standardization is generally still in its early stages, with most of the standards under development and expected to be published within the next five years (European Commission Joint Research Centre, 2021).

There are two notable international standards-developing organizations relevant to AI standards. The first is ISO/IEC JTC 1, a joint technical committee for standardization in the information and communication technology field, with a subcommittee focusing on AI standardization—ISO/IEC JTC 1/SC 42. The

second key AI standardization body is the Institute of Electrical and Electronics Engineers (IEEE) Standards Association, which has worked on, for example, Wi-Fi and Ethernet standards. There are only a handful of published standards, but research in this area is gaining momentum.

Certification systems for AI governance are being developed in parallel with standardization work. The proposed EU AI Act envisions a network of authorities with the power to certify organizations that introduce AI systems into the European market (European Commission, 2021). The IEEE has also launched the Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS), which aims to create specifications for certification and marking processes that advance transparency, accountability, and reduction in algorithmic bias in autonomous and intelligent systems.

Going forward from standards and certifications, the auditing of AI is currently promoted as a way to ensure that organizations govern their AI systems appropriately and as a potential source of economic growth due to emerging auditing services (Koshiyama et al., 2021). Before the EU AI Act is finalized, AI auditing services continue to emerge, and there is significant uncertainty about the final form of the act. It is important to note that AI auditing takes different forms, from consultancy and advisory services to certifications. Before the AI Act is passed, organizations can conduct *ethics-based auditing* (Mökander et al., 2021), either internally or using an external service provider, to ensure that their AI systems adhere to ethical AI principles, such as transparency. While the ethics-based auditing of AI systems can alleviate stakeholder pressure and concerns, it does not offer assurance regarding legal compliance.

Over time, the development of the auditing and oversight of AI systems is likely to shift, at least in part, from "softer" ethics-based auditing to certifications and legal auditing services. Compliance with AI Act (AIA) requirements is set to become a requirement for companies to operate at least in the European market. In turn, ethics-based auditing could cater to more demanding sectors in which customers or investors require ethical business practices beyond the minimum legal requirements. This means that organizations must remain attuned to the AI Act, sectoral legislation, and changing stakeholder requirements.

Why is this important for organizations that develop and use AI? Competitive and socially responsible AI is important because binding legislation sets only the bare minimum requirements for responsible AI. Organizations have much to gain from going above and beyond this minimum level. Not breaking the law is an important first step, but successful organizations consider the responsible development and use of AI more broadly than simply a legal compliance requirement. For example, it is self-evident that a bank operates within legal

boundaries, but customers increasingly expect banks to also consider sustainability aspects when contemplating investment and lending opportunities.[1] In this respect, responsible AI could be compared to environmental sustainability, for which stakeholder expectations have already grown and become institutionalized in corporate sustainability reporting.

---

[1] For example, https://www.unep.org/news-and-stories/story/why-financial-institutions-are-banking-sustainability

2 Roadmap to competitive and socially responsible artificial intelligence

This chapter presents the roadmap to competitive and socially responsible AI, which is structured into six drivers that contain tasks for organizations that develop and use AI systems. The drivers are "From AI ethics principles to AI governance," "Responsible AI commercialization potential and challenges," "AI standardization," "Automation of AI governance," "Responsible AI business ecosystems," and "Stakeholder pressure for responsible AI."

Fig. 1 shows an overview of the roadmap to competitive and socially responsible AI. The six main drivers are visualized as lanes pointing toward the same desired end goal. Within each lane, several tasks need to be executed to help reach the end goal of competitive and socially responsible AI. Links are also drawn between the tasks to indicate which tasks most likely need to be executed before others. However, the drivers are not placed in priority order, and their importance for different organizations will vary.

In summary, the first driver outlines the organizational steps from AI ethics principles to practicable AI governance. The second outlines the potential and challenges of commercializing responsible AI through appropriate business models. The third outlines organizational measures related to AI standardization. The fourth outlines the automation of AI governance, which enables governance by design and through automated tools. The fifth outlines responsible AI ecosystems that are emerging and will provide opportunities to procure and provide responsible AI services. Finally, stakeholder pressure for responsible AI is rising, leading to the need to report on responsible AI performance.

Fig. 1 Roadmap to competitive and socially responsible artificial intelligence

The following sections present the six sets of tasks in the roadmap. The section titles are the drivers that enable the managerial and technical tasks indicated by the colored notes in the illustration. Under the section for each driver, we explain the tasks that need to be conducted by organizations that develop and use AI systems. The roadmap is intended as an overview rather than a detailed action plan because different types of private and public organizations in various industries will execute the tasks differently. For example, banks will have priorities and sectoral requirements that differ from those of clothing retailers or universities. Therefore, the tasks should be understood as a checklist of things to consider and execute in a contextually appropriate manner in the coming years. The desired endpoint, competitive and socially responsible AI, is the same for all organizations. However, how this looks and the specific path to get there will differ for each organization.

2.1 From AI ethics principles to AI governance

Fig. 2 shows an overview of the advancement from AI ethics principles to AI governance. Each item in the figure is an organizational task or a set of tasks, and the arrows indicate the sequential order of the tasks. The tasks are further explained in the following sections.
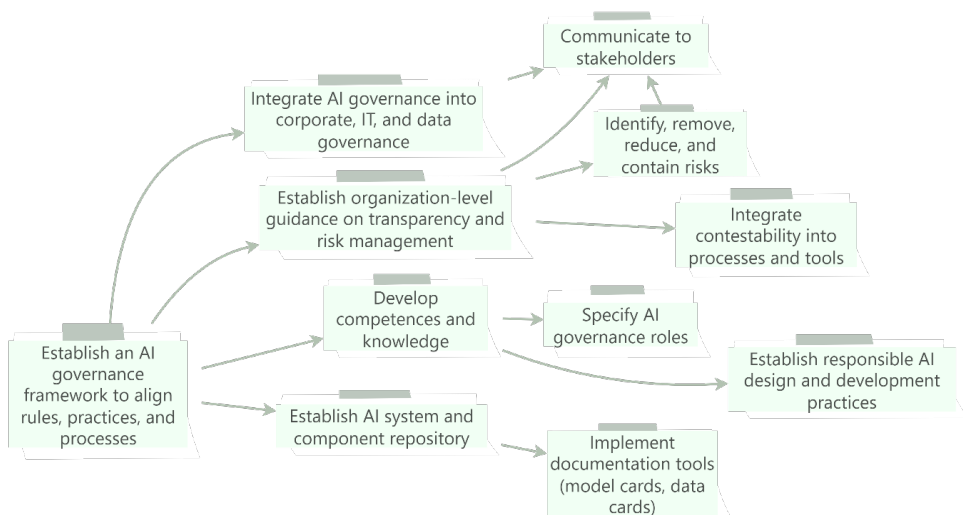

Fig. 2 From AI ethics principles to AI governance

Establish an AI governance framework to align rules, practices, and processes. The emerging AI regulatory framework requires organizations to put in place guidance on AI development and use. The guidance at this level intends to establish common practices among AI development teams and across the organization. On the roadmap to competitive and socially responsible AI, organization-level guidance must be in place early, before the enforcement of upcoming AI regulation, because it lays the foundation for subsequent organizational AI governance efforts.

Integrate AI governance into corporate, IT, and data governance. AI governance does not take place in a vacuum and needs to be part of an organization's overall governance system, including integration with corporate, IT, and data governance. In some organizations, if AI provides only small added functionalities to existing IT systems, AI governance can be primarily covered by existing IT governance and data governance processes. However, as organizations' AI portfolios grow and AI functionalities become more critical to their operations, AI governance and issues such as AI system explainability and accountability need to be handled appropriately. Data governance is central to AI governance because AI systems rely on data to learn and operate. Therefore, organizations should consider establishing a repository to document data sources, access conditions, processing workflows, and data quality controls. Organizations must also identify what information to provide to users and AI subjects and how to deliver it. In most cases, users will be informed in a digital space. This will likely require the developer to design and implement various interfaces.

Establish organization-level guidance on transparency and risk management. Organizations need to understand their AI systems' purposes, benefits, and risks to enable responsible AI development. These should be in the documentation (e.g., model cards; Mitchell et al., 2019) of the respective AI systems and included in the organizations' AI system repositories. However, before this is possible, organizations must create common guidelines that indicate which aspects of AI systems need to be documented. This work begins by raising general management-level awareness of AI governance requirements and potential AI risks if governance is not in place. If there is no awareness of the issue, AI governance is unlikely to achieve the necessary management commitment, and thus, there is little motivation to draft the required guidelines.

Organizational guidelines define the processes for producing more specific documentation (e.g., in model cards and an organizational AI system repository). They also indicate metrics for measuring AI system performance, algorithm performance, and data quality. Metrics must be defined at the organizational level to ensure their uniformity and comparability vis-à-vis different AI systems.

In their guidelines, organizations need to produce descriptions of the workflows, processes, and tools used to ensure sustainable AI. These include automated monitoring tools and mapping the regulatory environment early in an AI system's life cycle. The regulatory mapping process provides the organization with basic information on the regulatory environment of the AI system. The process should ensure that regulatory instruments are reviewed and that the primary regulatory instruments and constraints affecting the AI system are known to the development team when the development work begins. Once a tentative understanding of the future intended purpose and users of an AI system is reached, the organization's legal function should conduct an in-depth legal analysis to identify critical regulatory risks, constraints, and design parameters. Managers should communicate these regulatory focal points to the development team.

Integrate contestability into processes and tools. Organizations need to develop their contestability capabilities comprehensively. Some AI systems will face contestability requirements. For example, public authorities may be subject to legislative rules that require them to direct their customers to appeal bodies. Firms may wish to open escalation channels for complaints or appeals against decisions. The developers should identify and document these requirements and pressures to facilitate the development work. Some decisions may also be subjected to contestation. Developers should review whether the decision-making targets should have an avenue to trigger a review of the decision that affects them. If regulation or customer needs require contestation possibility, the developer should build a system to implement contestability. This will likely require the developer to design and implement various interfaces and organizational processes.

Develop competencies and knowledge. Implementing AI governance is a significant competence development challenge and opportunity for organizations. Employee education and training are required to promote AI knowledge and awareness of ethics and governance themes, such as responsibility and privacy. Education helps ground internal AI discussions in realistic expectations rather than in misconceptions, fears, and AI hype. Training sessions may include seminars, webinars, workshops, and online courses. Moreover, organizations can develop their knowledge base by participating in research activities, such as following the latest studies and participating in AI research projects and initiatives, such as the IEEE ECPAIS. Organizations also need to develop a solid understanding of their own data and algorithms through, for example, mathematical model validation and documentation to avoid using black-box systems that organizational users do not understand.

Establish responsible AI design and development practices. Achieving socially responsible AI systems requires responsible design and development practices. This means a set of practices and tools that make it possible to practically implement organizational ethical AI policies. Responsible AI design and development constitutes a broad domain, ranging from software development life cycles to the gender and ethnic diversity of development teams. Some key practices are human oversight (retaining human control over the decision-making process); the search for simple solutions to balance accuracy, complexity, and interpretability; the implementation of responsibility by design (incorporating responsibility from the start), understandable explainability targeting the right audience, and stakeholder engagement in design and development. On a more technical level, responsible design and development practices refer to, for example, machine learning operations (MLOps) practices throughout the development pipeline, including bias mitigation, model validation, and continuous monitoring. Before deploying AI systems and algorithmic components, the developer should develop key metrics to assess whether the AI system and components achieve a sufficient level of performance and are safe to deploy.

Specify AI governance roles. Effective AI governance requires clear roles and responsibilities for governing particular AI systems. Currently, roles are still relatively unclear for many, and they are assigned differently in organizations, with CEOs responsible for AI systems in some organizations and development teams and users responsible in others. There is a tendency for responsibility to be shifted downstream to the deployer organizations and end users of AI systems. However, this may change with the introduction of specific requirements for AI system providers in the EU AI Act. Especially in larger organizations, a dedicated AI oversight unit can help by providing a precise organizational location for the overall coordination of AI governance.

Establish an AI system repository. The AI system developer and users should have an AI system repository that contains information on all AI systems operated or used by the organization. Once the organization commences development work or deploys an AI system, it should assign a unique identifier *(system ID)* and enter it into the AI repository. The AI system repository should hold information on the system's intended purpose, possible unintended uses, and descriptions of the operational environment and business model.

This documentation needs to be in place at the latest when the EU AI Act becomes enforceable. While documentation needs stemming from regulation may differ according to the risk level of the system, the application of the same documentation requirements to all AI systems is likely to be most efficient.

In addition, the AI system repository should include information on risks and impacts, accountability, and transparency. AI systems may impose various risks on users and third parties. A thorough understanding of the risk and impact of the AI system should guide development activities. Accountability requirements mean that the AI system repository should indicate the AI system owner and the roles, tasks, and decision rights related to the respective AI system. The development team should also identify and document transparency and explainability requirements. In practice, the AI system repository can be automatically populated as far as possible from the model cards or similar documentation methods linked to the respective AI systems.

AI system developers and users must also consider that complex AI systems consist of algorithmic components. The organization should define and document the intended purpose of all algorithmic components in an understandable way at the beginning of the development process. Understanding the AI system, its components, and their interactions requires the developer and the user to define and document the environment in which the algorithmic component is embedded, the function it performs, and the processes it involves and interacts with.

Identify, remove, reduce, and contain risks. The proposed EU AI Act requires a risk management system in organizations that deal with high-risk AI applications. Risk management can be seen as a pipeline for dealing with risks, which needs to be prepared before the enforcement of the AI Act. The first step is the identification of risks. The second is removing risks that need to be removed and reducing those that the organization cannot remove. The remaining risks must then be contained and communicated internally and externally to stakeholders. Risk deliberations and measures should be documented. When the AI Act is in place, regulators will enforce and monitor the risk management system, and ultimately, the noncomplying organizations will face sanctions. The highest penalty in the AI Act proposal (Art. 71) is 30 million euros or 6% of worldwide annual turnover, whichever is higher. Apart from sanctions, companies will face potentially significant reputation damage. In particular, risks to health, safety, and fundamental rights should be identified, analyzed, and documented. Discrimination risk, in turn, always arises when the AI system treats a person differently from another and the AI system user cannot demonstrate that the different treatment was not based on the person's age, origin, nationality, language, religion, belief, opinion, political activity, trade union activity, family relationships, health status, disability, sexual orientation, or other personal characteristics. Discrimination-related risks should be carefully identified, analyzed, and documented. The developer should build a system that allows for the monitoring of AI system risk and impact. The EU regulatory framework will

likely require a post-market surveillance system to be built to track AI system performance.

Implement documentation tools (model cards and data cards). Model cards are recommended to keep track of AI systems' requirements and competitive advantages. The information should be embedded in the systems' source code, and the model cards are one way of embedding these requirements. AI systems are embedded in complex technical environments. Infrastructure, system architecture, and interfaces fundamentally affect their impact. The infrastructure, architecture, interface-related risks, and impacts should be documented to ensure AI system sustainability. Model cards can also document transparency, explainability, and contestability-related requirements and their monitoring. Data cards, in turn, can be used to document data quality and monitoring issues.

Communicate with stakeholders. In addition to robust documentation tools, organizations need solid communication practices for information to reach stakeholders and to build trust in the organization's use of AI systems. Communication practices include providing information about the organization's data and algorithms and informing involved parties about human–AI interaction and automated decision-making. Forward-looking organizations offer additional information on top of the bare minimum set by the GDPR. For example, AI register platforms have been piloted and can be used to classify an AI portfolio and share the information with stakeholders. It is also important to acknowledge in communications that AI systems may be imperfect and may have specific weaknesses compared to human processing. Finally, it is crucial to inform stakeholders when they are dealing with an AI system—for example, in the case of chatbots communicating with customers.

Further reading

AIGA ⚑ Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Putting AI ethics into practice: The hourglass model of organizational AI governance (arXiv:2206.00335). arXiv. https://doi.org/10.48550/arXiv.2206.00335

AIGA ⚑ Seppälä, A., Birkstedt, T., & Mäntymäki, M. (2021). From ethical AI principles to governed AI. Proceedings of the 42nd International Conference on Information Systems (ICIS2021). International Conference on Information Systems (ICIS), Austin, Texas. https://aisel.aisnet.org/icis2021/ai_business/ai_business/10/

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

AIGA Laato, S., Birkstedt, T., Mäntymäki, M., Minkkinen, M., & Mikkonen, T. (2022). AI governance in the system development life cycle: Insights on responsible machine learning engineering. Proceedings of the 1st Conference on AI Engineering - Software Engineering for AI. CAIN, New York, N.Y. https://doi.org/10.1145/3522664.3528598

AIGA Laato, S., Mäntymäki, M., Minkkinen, M., Birkstedt, T., Islam, A. K. M. N., & Dennehy, D. (2022). Integrating machine learning with software development lifecycles: Insights from experts. ECIS 2022 Proceedings. ECIS, Timișoara, Romania.

AIGA Muiruri, D., Lwakatare, L. E., Nurminen, J. K., & Mikkonen, T. (2022). Practices and infrastructures for machine learning systems: An interview study in Finnish organizations. *Computer*, *55*(6), 18–29. https://doi.org/10.1109/MC.2022.3161161

AIGA Seppälä, A. (2021). Implementing Ethical AI: From Principles to AI Governance [Master's thesis, University of Turku]. https://www.utupub.fi/handle/10024/151387

AIGA Niemenmaa, A. (2021). ART of AI governance: Pro-ethical conditions driving ethical AI governance [Master's thesis, University of Turku]. https://www.utupub.fi/handle/10024/152484

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. ArXiv:1810.03993 [Cs]. https://doi.org/10.1145/3287560.3287596

Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data cards: Purposeful and transparent dataset documentation for responsible AI. ArXiv:2204.01075 [Cs]. http://arxiv.org/abs/2204.01075

Tournas, L. N., & Bowman, D. M. (2021). AI insurance: Risk management 2.0. *IEEE Technology and Society Magazine*, *40*(4), 52–56. https://doi.org/10.1109/MTS.2021.3123750

EU AI Act, Article 11 (Technical documentation)

EU AI Act Article 17 (Quality management system)

2.2 Responsible AI commercialization potential and challenges

Fig. 3 provides an overview of responsible AI commercialization tasks and their sequential order. The tasks are explained further in the following sections.
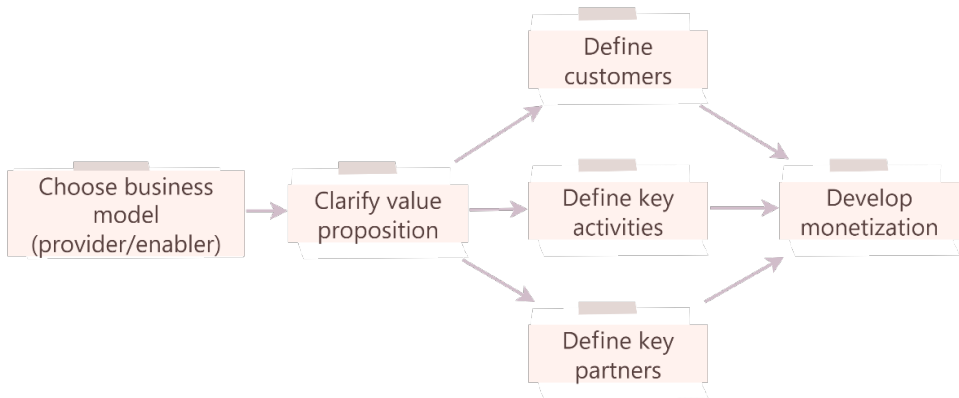
Fig. 3 Responsible AI commercialization potential and challenges

Choose business model (providing/enabling). Implementing ethical principles in designing and operating commercially viable AI presents commercialization challenges. Studies suggest that compared to other kinds of AI, responsible AI offers ethical advantages but a weaker commercial value proposition. Organizations need not select either ethical or competitive AI, but they do need to carefully consider the business model of their (responsible) AI systems. Two basic types of business models leverage the value proposition of responsible AI: providing and enabling responsible AI. The providing type of business model creates value by developing and offering an actual responsible AI system. The second business model type enables responsible AI by supporting, facilitating, and assessing the development and operation of responsible AI systems, for example, by conducting auditing. The choice of the business model (providing or enabling responsible AI) defines the content of the following steps: value proposition, customers, activities, partners, and monetization.

    Clarify value proposition. Organizations need to clarify the value proposition of responsible AI in their contexts. Generally, the potential value proposition of responsible AI is twofold. First, AI creates value by automating business processes and increasing operational efficiency. Second, responsible AI extends this value proposition by providing the additional value of explainable, transparent, and societally responsible AI solutions. For example, by increasing users' trust and complying with regulations and stakeholder requirements, responsible AI can promote more sustainable and future-proof commercialization of AI.

Define customers. Following the value proposition, organizations must define the customer groups for their responsible AI solutions. Two generic customer groups are organizational customers (business-to-business) and consumers (business-to-consumer). Each of these requires different approaches to considering customer requirements and commercializing AI. Organizations operating within the "providing" type of business model offer AI systems to other organizations or consumers. In contrast, for the "enabling" type of organizations (e.g., consultancies and auditing firms), potential customers are the organizations developing and offering responsible AI. From a broad perspective, there is a long chain of customers of responsible AI, ranging from businesses to the entire society.

Define key activities. In parallel with specifying customers, organizations need to identify the key activities in their responsible AI business model. Generally, there are at least four types of activities: technologically developing responsible AI, understanding the market needs for responsible AI, auditing responsible AI systems, and raising awareness and lobbying. In general, activities can be divided into more technical (e.g., AI development) and social categories (e.g., raising awareness).

Define key partners. In addition to customers, networked organizations should consider the partners with whom they can develop or enable responsible AI solutions. Three basic types of partners may be identified. First, customers (both organizational customers and consumers) can also be seen as partners in designing and successfully commercializing responsible AI. Second, enablers, such as consultancy and auditing firms, technology and research institutions, and investors, provide resources and assistance to make responsible AI possible. Third, regulators set rules and incentives for responsible AI, and organizations should monitor regulatory developments closely.

Develop monetization. Finances, revenue models, and earnings logic are crucial parts of business models according to established business model frameworks. The reason monetization is placed last in this roadmap theme is that, in our view, responsible AI business model design should start from responsible AI rather than from highly commercialized but ethically problematic AI, and organizations should seriously consider the other elements in addition to monetization. In this way, commercialization and responsibility are treated as equally important, and organizations have fewer path dependency problems compared to when they try to convert lucrative but ethically problematic AI systems into responsible ones.

Further reading

AIGA ⬛ Zimmer, M. P., Minkkinen, M., & Mäntymäki, M. (2022). Responsible artificial intelligence systems: Critical considerations for business model design. *Scandinavian Journal of Information Systems*.

AIGA ⬛ Eneh, L. (2022). *Business model for sustainable AI consulting*. [Master's thesis, Tampere University].

Kumar, P., Dwivedi, Y. K., & Anand, A. (2021). Responsible artificial intelligence (AI) for value formation and market performance in healthcare: The mediating role of patient's cognitive engagement. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10136-6

2.3   AI standardization

Fig. 4 shows an overview of organizational tasks related to AI standardization, and the arrows indicate their sequential order. The tasks are outlined further in the following sections.
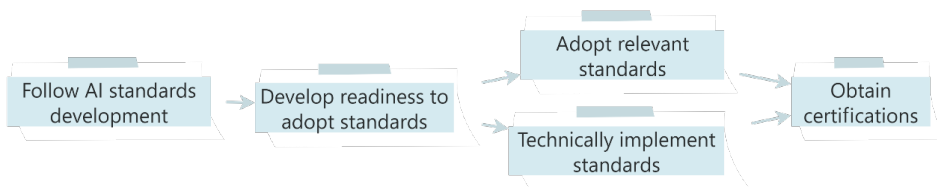

Fig. 4 AI standardization

Follow standards development. Regarding the implementation of AI standards, the first step for organizations is to follow the standardization landscape and, when relevant, participate in standardization work. AI-related standards are being developed in multiple streams. At the time of writing, the development of standards is at an early stage, providing opportunities to influence the work. The two main actors are the IEEE and the International Organization for Standardization (ISO), while the European Committee for Standardization develops standardization based on ISO standards at the European level. AI standards, such as data quality standards and AI system life cycle standards, will

provide guidance on many aspects of AI systems, including transparency, explainability, and risk management.

Develop readiness to adopt standards. Before implementing standards, organizations need to develop the readiness to adopt them. This is achieved by building awareness of standards in relevant roles and teams within the organization, assessing gaps in relevant AI governance processes compared to the requirements set out by standards, and assessing which standards are relevant.

Adopt and technically implement relevant standards. Organizational standards adoption is likely to follow similar adoption processes as other comparable technology standards, such as cybersecurity standards. Alongside the formal organizational adoption of standards, they need to be technically implemented in the relevant AI systems and architectures.

Obtain certifications. The proposed EU AI Act sets up an emerging ecosystem of *notified bodies* and *notifying authorities*, which means, in practice, a network of auditors and regulatory authorities. The introduction of the EU AI Act into force will move the auditing and oversight landscape toward certifications of compliance with the new regulation. After the AI Act is in place, organizations developing AI systems for European markets need to ensure that their systems comply with the regulation's requirements. AI Act certifications will address this need.

Further reading

AIGA Tamminen, P. (2022). *Operationalizing transparency and explainability in artificial intelligence through standardization* [Master's thesis, University of Turku]. https://urn.fi/URN:NBN:fi-fe2022053039454

Cihon, P. (2019). Standards for AI governance: International standards to enable global coordination in AI research & development. Future of Humanity Institute.

ISO/IEC JTC 1/SC 42 Artificial intelligence. https://www.iso.org/committee/6794475.html

IEEE P7000™ Standards. https://ethicsinaction.ieee.org/p7000/

2.4 Automation of AI governance

Fig. 5 shows an overview of organizational tasks related to automation, and the arrows indicate their sequential order. The tasks are further explained in the following sections.
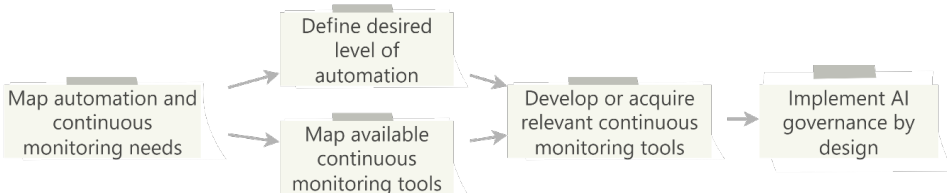

Fig. 5 Automation of AI governance

Map automation and continuous monitoring needs. Effective, continuous performance monitoring is crucial to sustainable AI use. Likewise, algorithmic component performance can be ascertained solely by deploying metrics to evaluate performance. All components should be subject to constant performance monitoring. Data quality should also be monitored. Monitoring requires that developers design and document data quality metrics to enable the consistent monitoring of data. Automation is a natural fit for AI governance because AI systems learn and adapt based on data, thus adopting patterns that are not hard-coded by designers. Monitoring the compliance and responsible operation of such systems is challenging because they operate at high speeds and can produce surprising outcomes. Therefore, automated and software-based monitoring systems can ease the burden on human operators. The first step in implementing such automation is mapping the automation and monitoring needs related to specific AI systems and their features. The needs can depend on regulatory requirements (e.g., specific regulations in sectors such as healthcare) and the risk profile of the use case (low-risk/high-risk). In the first instance, organizations need to ensure that automated solutions are permitted by law and then consider their business cases compared to human oversight.

Define the desired level of automation. A crucial question in implementing automated AI governance is the desired level of automation. With a comparatively low level of automation, automated governance can assist developers and auditors and provide additional information on the fairness of algorithms, for instance. This can be implemented, for example, via dashboards that display key indicators. With a high level of automation, the automated system can make decisions, and the human oversight agent can intervene in the case of incorrect decisions but

normally takes a backseat position. Higher levels of automation can raise the efficiency of AI governance, similar to other process automation. However, there is a risk of introducing a second-order problem, whereby opaque algorithmic systems monitor other opaque algorithmic systems.

Map available continuous monitoring tools. Tools for continuously monitoring and auditing AI systems are still in early development stage, but tools for monitoring individual metrics, such as fairness, already exist. Therefore, organizations should map the existing state of the art and identify existing components before developing their own solutions. Techniques for continuous monitoring already exist (Myllyaho et al., 2022), so it is essentially a matter of automating them and finding a suitable combination of tools.

Develop or acquire relevant continuous monitoring tools. Organizations can develop or acquire continuous monitoring tools depending on their needs, the automation level, and the mapping results. In the coming years, off-the-shelf solutions for individual components may be introduced, and their integration into a functioning continuous monitoring system will then be the crucial step. Such solutions may include, for example, tools for data monitoring, bias monitoring, and mitigation, as well as for conducting periodic health checks of AI systems.

Implement AI governance by design. Once the AI governance tool chain is in place, organizations can implement AI governance in their AI systems by design and by default. The goal is to reach an AI governance system that "just works" as far as possible, enabling calm governance with minimal heavy compliance processes that would disrupt AI design and development work. The extent to which AI governance can be automated depends on the context and the regulatory requirements. Organizational AI governance roles and processes are unlikely to become completely obsolete due to automation, especially with the requirements stemming from the coming EU AI Act.

Further reading

AIGA ⚑ Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T., & Nurminen, J. K. (2021). Systematic literature review of validation methods for AI systems. *Journal of Systems and Software*, *181*, 111050. https://doi.org/10.1016/j.jss.2021.111050

AIGA ⚑ Myllyaho, L., Raatikainen, M., Männistö, T., Nurminen, J. K., & Mikkonen, T. (2022). On misbehaviour and fault tolerance in machine learning systems. *Journal of Systems and Software*, *183*, 111096.

AIGA ⚑ Minkkinen, M., Laine, J., & Mäntymäki, M. (2022). Continuous auditing of artificial intelligence: A conceptualization and assessment of tools and frameworks. *Digital Society*, *1*(3), 21. https://doi.org/10.1007/s44206-022-00022-2

AIGA ⚑ Mäkinen, S. (2021). Designing an open-source cloud-native MLOps pipeline [Master's thesis, University of Helsinki]. http://urn.fi/URN:NBN:fi:hulib-202103301763

Eulerich, M., & Kalinichenko, A. (2018). The current state and future directions of continuous auditing research: An analysis of the existing literature. *Journal of Information Systems*, *32*(3), 31–51. https://doi.org/10.2308/isys-51813

2.5   Responsible AI business ecosystems

Figure 6 provides an overview of organizational tasks related to responsible AI business ecosystems, and the arrows indicate their sequential order. These tasks are further explained below.
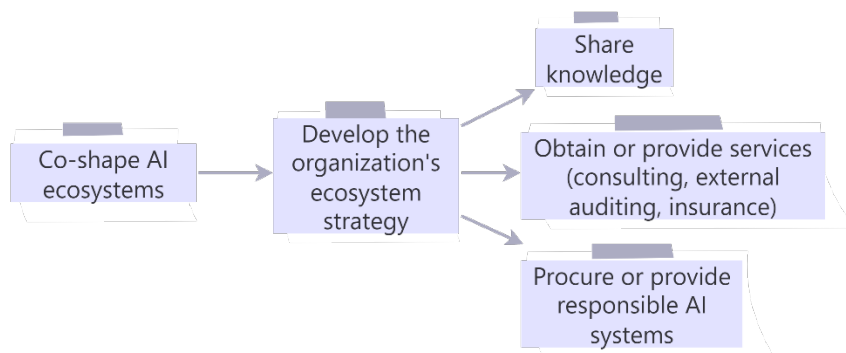

Fig. 6 Responsible AI business ecosystems

Co-shape AI ecosystems. Europe takes an ecosystem approach to trustworthy AI, focusing on an ecosystem of excellence and one of trust. The motivating vision is that a regulatory and business ecosystem will ensure the enforcement and practical implementation of the EU AI Act and responsible AI more broadly. *Ecosystems* generally refer to an evolving set of actors, activities, artifacts, and related institutions and relations that are important for actors' innovative performance (Granstrand & Holgersson, 2020). Ecosystems include different activity clusters, activities, and actor roles. The purpose of an ecosystem is to produce a central

value proposition through the work of ecosystem members (Adner, 2017). Responsible AI ecosystems are shaped in the present before the EU AI Act is in place. Among the potential participants, there needs to be a consensus of vision—that is, an understanding of the overall value proposition provided by the ecosystem. Further, the ecosystem is likely to require one or more *orchestrators* that drive the development and operation of the ecosystem. At present, it is uncertain who the orchestrator could be, whether it is a central entity, such as an EU institution; another public actor (e.g., an innovation agency or regulatory body); a hybrid set of public–private actors; or a set of private companies (e.g., the so-called Big Five technology companies). Organizations can benefit from early involvement in the responsible AI ecosystem in three ways. First, they can gain reputational benefits by going further than the minimum level required by binding legislation. Second, they can forge ecosystem partnerships early, save costs, and tailor their AI systems with help from ecosystem members. Third, they can gain organizational learning by being early with regard to experimenting with and using responsible AI systems.

Develop the organization's ecosystem strategy. How should companies attach themselves to this emerging ecosystem? This is a critical question in the coming years. Companies can adopt a proactive ecosystem strategy when the ecosystem is still emerging and place themselves in a more central position than later entrants. After the AI Act is in place, the ecosystem will be more fixed, and new entrants will have less leverage. Forging a productive AI ecosystem strategy in companies will require significant management attention in the coming years. This will also likely create new business opportunities for consultants to advise organizations on ecosystem strategy.

Share knowledge. An essential function of the responsible AI ecosystem is to act as a forum for knowledge exchange. For example, sector-specific responsible AI events for network members can provide vital information on recent and upcoming developments in AI regulation, standardization, self-regulatory guidelines, tools, and best practices.

Obtain or provide services. After developing an ecosystem strategy, organizations can obtain services, such as consulting, external auditing, and insurance, from the ecosystem. Alternatively, they can provide such services to the ecosystem, depending on their position. The needs for various responsible AI services differ depending on industry and company size, for example. Still, the overall demand for responsible AI services, such as legal services and auditing, is set to rise in the coming years. Given the importance of risk management in responsible AI, insurance companies are likely to play a central role in the future ecosystem.

Procure or provide responsible AI systems. Organizations can procure or provide responsible AI systems, depending on their ecosystem position. This is a more concrete form of "selling responsible AI" than the services mentioned above. In this case, what is exchanged is an information system, or part of an information system, that responsibly utilizes AI, thus easing the compliance pressures of the buying organization. Certifications for responsible AI (see the section on standardization) are likely to appear over the next few years, and they will further institutionalize the selling of responsible AI systems.

Further reading

AIGA ⚑ Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2022). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10269-2

AIGA ⚑ Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2021). Towards ecosystems for responsible AI: Expectations on sociotechnical systems, agendas, and networks in EU documents. In D. Dennehy, A. Griva, N. Pouloudi, Y. K. Dwivedi, I. Pappas, & M. Mäntymäki (Eds.), *Responsible AI and analytics for an ethical and inclusive digitized society* (pp. 220–232). Springer International Publishing. https://doi.org/10.1007/978-3-030-85447-8_20

AIGA ⚑ Laine, J. (2021). Ethics-based AI auditing core drivers and dimensions: A systematic literature review [Master's thesis, University of Turku]. https://www.utupub.fi/handle/10024/152795

Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, *62*, 102441. https://doi.org/10.1016/j.ijinfomgt.2021.102441

Wirtz, B. W., Langer, P. F., & Weyerer, J. C. (2022). An ecosystem framework of AI governance. In J. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. Young, & B. Zhang (Eds.), *The Oxford handbook of AI governance*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.24

European Commission. (2020). *WHITE PAPER on artificial intelligence—A European approach to excellence and trust*.

2.6   Stakeholder pressure for responsible AI

Fig. 7 shows an overview of organizational tasks related to stakeholder pressure for responsible AI, and the arrows indicate their sequential order. These tasks are further explained in the following sections.
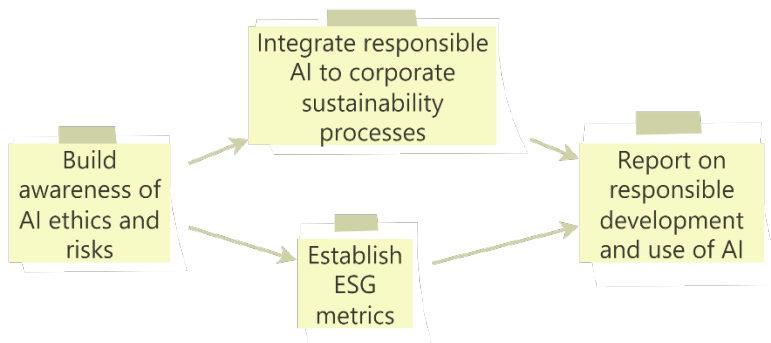
Fig. 7 Stakeholder pressure for responsible AI

Build awareness of AI ethics and risks. Awareness of AI as an ethical and sustainability issue still needs to be improved among organizations and stakeholders, such as investors. However, ethically aware consumers, sustainable investing professionals, and frontrunner organizations are already driving the change toward mainstreaming AI as a corporate sustainability issue. Therefore, organizations should proactively build awareness internally of AI ethics issues and risks. This may include, for example, internal training sessions and workshops and the integration of AI ethics into onboarding materials and processes.

   Integrate responsible AI into corporate sustainability processes. Corporate sustainability has gathered steam in recent years as a combination of corporate social responsibility, stakeholder theory, and sustainable development. Alongside building awareness of AI ethics issues, organizations need to integrate responsible AI into corporate sustainability processes, such as by developing corporate sustainability strategies, creating working groups, engaging with senior management, and engaging in stakeholder communication activities. In the early stages, this will likely require well-positioned responsible AI champions who keep the topic on the agenda. Integration into sustainability processes ensures that

32

awareness of responsible AI is diffused more broadly into the organization rather than remaining an isolated topic.

Establish environmental, social, and governance metrics for responsible AI. Organizations need to adopt or develop metrics to credibly demonstrate their performance in responsible AI. Environmental, social, and governance (ESG) metrics are increasingly important to investors. Still, there are no standardized measures for companies or investors to track responsible AI dimensions, such as fairness and transparency, and each organization may define these terms differently. However, AI governance is increasingly institutionalized through the EU AI Act and other legislation. Investors, customers, regulators, and other stakeholders are likely to increasingly expect adequate performance in the future. The development of metrics begins with the identification of material issues—that is, issues relevant to future financial performance—and the exploration of how these can be measured. In the future, environmental and social impact metrics can be adapted to measure AI impacts.

Report on responsible development and use of AI. Companies are increasingly expected to address ESG issues as part of their sustainability reports, and a corporate sustainability reporting directive is in preparation in the EU. The responsible development and use of AI is still an emerging issue, but in the coming years, it will likely become part of ESG reporting in some way. Therefore, forward-looking organizations should seek ways to integrate the responsible development and use of AI, as well as realistic consideration of potential risks and unintended impacts, into their sustainability reporting.

Further reading

AIGA ⚑ Minkkinen, M., Niukkanen, A., & Mäntymäki, M. (2022). What about investors? ESG analyses as tools for ethics-based AI auditing. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01415-0

AIGA ⚑ Niukkanen, A. (2021). Taking the responsible use of AI into account in ESG analyses [Master's thesis, University of Turku]. https://www.utupub.fi/handle/10024/152622

Sætra, H. S. (2021). A Framework for evaluating and disclosing the ESG related impacts of AI with the SDGs. *Sustainability*, *13*(15), 8503. https://doi.org/10.3390/su13158503

Brusseau, J. (2021). AI human impact: Toward a model for ethical investing in AI-intensive companies. *Journal of Sustainable Finance & Investment*, 1–28. https://doi.org/10.1080/20430795.2021.1874212
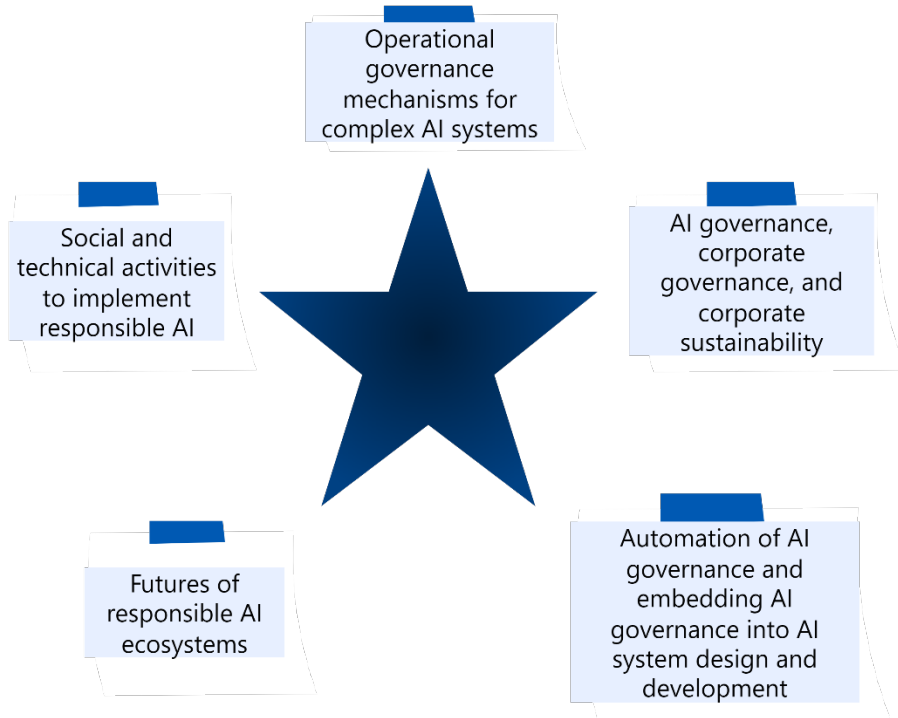
3   Future research agenda

Fig. 8 Future research agenda for competitive and socially responsible AI

AI governance and auditing present numerous potential topics for interdisciplinary and international research in the coming years (Fig. 8). As a synthesis of the issues in the roadmap presented in this report, we highlight five central directions for future research.

1. Operational governance mechanisms for complex AI systems. The translation of AI ethics into practice has been repeatedly raised as a crucial issue in taking responsible AI scholarship forward (Morley et al., 2020). There is still a significant gap between abstract ethical principles, such as fairness, on the one hand and technical research on AI on the other. Operational AI governance mechanisms have the potential to fill this gap, but such tools need to be better

understood. For example, by studying frontrunner organizations, researchers can learn best practices for implementing AI governance in real-life AI systems in production use.

2. AI governance, corporate governance, and corporate sustainability. Organizational AI governance belongs under the umbrella of corporate governance, and AI ethics is also linked to corporate sustainability activities. While initial scoping has been done (Mäntymäki et al., 2022), the links between these governance areas are poorly understood. Thus, research is needed to conceptualize AI governance within corporate governance, IT governance, and data governance and to clarify the unique aspects of AI governance compared to these more established domains. This research stream will help incorporate AI governance into organizations' daily activities and senior management tasks.

3. Automation of AI governance and embedding governance into AI system design and development. In the future, AI governance will increasingly take place in an automated fashion and by design, operating closer to the speed of the software development processes and the coding workflows. To enable this, AI governance concerns and issues need to be integrated into the tools used by software development teams, such as version control systems (Stirbu et al., 2022). While AI governance tools have been mapped, the research landscape on the practices and tools needed for testing and monitoring AI systems is still emerging (Muiruri et al., 2022). In particular, the continuous auditing and validation methods provide significant ground for future research (Minkkinen, Laine, et al., 2022; Myllyaho et al., 2021).

4. Future of responsible AI ecosystems. Responsible AI ecosystems are currently emerging in the EU and beyond, driven by advancing AI strategy and legislation, as well as the business potential of AI governance (Minkkinen, Zimmer, et al., 2022). However, at present, the form of such ecosystems remains unclear. Several potential structures exist, such as centralized ecosystems around single platforms or more horizontal networks organized around a central value proposition. Therefore, researchers should investigate stakeholders' expectations and visions of responsible AI ecosystems and the methods of transforming these expectations into plans, strategies, and institutional structures.

5. Social and technical activities for implementing responsible AI. Responsible AI systems are *sociotechnical systems* in which AI agents and social entities together process data in a responsible manner (Dignum, 2020; Zimmer et al.,

2022). Advancing and implementing responsible AI systems require various activities, such as defining requirements, developing AI systems, and auditing and oversight. These activities can be characterized as primarily technical (e.g., developing AI) or social (e.g., understanding market needs for responsible AI). Technical activities are a necessary condition for responsible AI systems because the underlying system design needs to make responsible use possible. However, social activities (e.g., defining requirements and auditing) are a sufficient condition for responsible AI because technical implementation alone cannot ensure the responsible use of AI (Zimmer et al., 2022). Therefore, research is needed on the interplay of social and technical activities to implement responsible AI. Furthering this research agenda will provide foundational knowledge on responsible AI as a basis for AI governance research and practice.

4   Conclusion

This roadmap report gave a comprehensive outlook on AI governance and auditing issues in the coming years. We started by defining the key concepts of competitive and socially responsible AI. Then, we outlined the background of the institutionalization of AI. In the main roadmap section, we presented an overview of six roadmap drivers and sketched the managerial and technical tasks under each driver. First, competitive and socially responsible AI requires the application of AI ethics principles to implement AI governance. Second, responsible AI commercialization needs to be tackled head-on. Third, AI standards are quickly emerging, and organizations need to stay up to speed on the standardization landscape. Fourth, AI governance will be embedded in automated solutions and software life cycles in addition to separate governance processes. Fifth, responsible AI business ecosystems are emerging and are co-shaped by regulators, companies, and civil society actors. Finally, stakeholder pressure for responsible AI is growing, manifested in increasing customer and investor requirements for responsible AI performance.

Together, these drivers and tasks paint a comprehensive picture of competitive and socially responsible AI developments over the next five years. After the roadmap section, we presented a future research agenda with five areas that will define the coming research landscape in AI governance and auditing and responsible AI.

The timeliness of AI governance is the overarching message of the roadmap. In other words, the central message of this roadmap document is that private and public organizations need to start devoting attention to governing their AI systems to align with coming regulatory requirements and stakeholder pressure. Being proactive now will save considerable effort in the future when responsible AI has become the expected way of operating.

The other important message is that implementing responsible AI requires numerous steps. However, the first tasks along the roadmap, such as building awareness of AI ethics and risks and following AI standards development, are already available to most organizations. Considering the task of implementing responsible AI as a roadmap consisting of numerous themes makes it less daunting

than viewing the end state of governed AI as a monolith. Moreover, support is already available for some tasks. It is likely to be increasingly available in the coming years, when the requirements of the future EU AI Act, for example, become more apparent.

Organizations will come to terms with making AI systems competitive and socially responsible in the coming years. During this change, researchers and research funding bodies will play a crucial role in advancing the knowledge base of governance mechanisms, connections to corporate sustainability, the embeddedness of governance in design and engineering, the future of responsible AI ecosystems, and the sociotechnical activities that are carried out to implement responsible AI. The future of responsible AI is being shaped in the present.

References

Adner, R. (2017). Ecosystem as structure: An actionable construct for strategy. *Journal of Management*, *43*(1), 39–58. https://doi.org/10.1177/0149206316678451

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450.

Brusseau, J. (2021). AI human impact: Toward a model for ethical investing in AI-intensive companies. *Journal of Sustainable Finance & Investment*, 1–28. https://doi.org/10.1080/20430795.2021.1874212

Cihon, P. (2019). *Standards for AI governance: International standards to enable global coordination in AI research & development*. Future of Humanity Institute.

Dignum, V. (2020). Responsibility and artificial intelligence. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 213–231). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.12

Eneh, L. (2022). *Business model for sustainable AI consulting*. Tampere University.

Eulerich, M., & Kalinichenko, A. (2018). The current state and future directions of continuous auditing research: An analysis of the existing literature. *Journal of Information Systems*, *32*(3), 31–51. https://doi.org/10.2308/isys-51813

European Commission. (2020). *WHITE PAPER on artificial intelligence—A European approach to excellence and trust*.

European Commission. (2021). *Proposal for Artificial Intelligence Act COM/2021/206 final*. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence

European Commission Joint Research Centre. (2021). *AI Watch, AI standardisation landscape state of play and link to the EC proposal for an AI regulatory framework.* Publications Office. https://data.europa.eu/doi/10.2760/376602

Granstrand, O., & Holgersson, M. (2020). Innovation ecosystems: A conceptual review and a new definition. *Technovation*, *90–91*, 102098. https://doi.org/10.1016/j.technovation.2019.102098

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

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

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Koniakou, V. (2022). From the "rush to ethics" to the "race for governance" in artificial intelligence. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10300-6

Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., & Lomas, E. (2021). *Towards algorithm auditing: A survey on managing legal, ethical and technological risks of AI, ML and associated*

algorithms (SSRN Scholarly Paper ID 3778998). Social Science Research Network. https://doi.org/10.2139/ssrn.3778998

Kumar, P., Dwivedi, Y. K., & Anand, A. (2021). Responsible artificial intelligence (AI) for value formation and market performance in healthcare: The mediating role of patient's cognitive engagement. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10136-6

Mäkinen, S. (2021). Designing an open-source cloud-native MLOps pipeline [Master's thesis, University of Helsinki]. http://urn.fi/URN:NBN:fi:hulib-202103301763

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*. https://doi.org/10.1007/s43681-022-00143-x

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Minkkinen, M., Laine, J., & Mäntymäki, M. (2022). Continuous auditing of artificial intelligence: A conceptualization and assessment of tools and frameworks. *Digital Society*, *1*(3), 21. https://doi.org/10.1007/s44206-022-00022-2

Minkkinen, M., Niukkanen, A., & Mäntymäki, M. (2022). What about investors? ESG analyses as tools for ethics-based AI auditing. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01415-0

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2021). Towards ecosystems for responsible AI: Expectations on sociotechnical systems, agendas, and networks in EU documents. In D. Dennehy, A. Griva, N. Pouloudi, Y. K. Dwivedi, I. Pappas, & M. Mäntymäki (Eds.), *Responsible AI and analytics for an ethical and inclusive digitized society* (pp. 220–232). Springer International Publishing. https://doi.org/10.1007/978-3-030-85447-8_20

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2022). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10269-2

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *ArXiv:1810.03993 [Cs]*. https://doi.org/10.1145/3287560.3287596

Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*, *27*(4), 44. https://doi.org/10.1007/s11948-021-00319-4

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Muiruri, D., Lwakatare, L. E., Nurminen, J. K., & Mikkonen, T. (2022). Practices and infrastructures for machine learning systems: An interview study in Finnish organizations. *Computer*, *55*(6), 18–29. https://doi.org/10.1109/MC.2022.3161161

Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T., & Nurminen, J. K. (2021). Systematic literature review of validation methods for AI systems. *Journal of Systems and Software*, *181*, 111050. https://doi.org/10.1016/j.jss.2021.111050

Myllyaho, L., Raatikainen, M., Männistö, T., Nurminen, J. K., & Mikkonen, T. (2022). On misbehaviour and fault tolerance in machine learning systems. *Journal of Systems and Software*, *183*, 111096. https://doi.org/10.1016/j.jss.2021.111096

Niukkanen, A. (2021). Taking the responsible use of AI into account in ESG analyses [Master's thesis, University of Turku]. https://www.utupub.fi/handle/10024/152622

Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data cards: Purposeful and transparent dataset documentation for responsible AI. *ArXiv:2204.01075 [Cs]*. http://arxiv.org/abs/2204.01075

Sætra, H. S. (2021). A framework for evaluating and disclosing the ESG related impacts of AI with the SDGs. *Sustainability*, *13*(15), Article 15. https://doi.org/10.3390/su13158503

Seppälä, A., Birkstedt, T., & Mäntymäki, M. (2021). From ethical AI principles to governed AI. *Proceedings of the 42nd International Conference on Information Systems (ICIS2021)*. International Conference on Information Systems (ICIS), Austin, Texas. https://aisel.aisnet.org/icis2021/ai_business/ai_business/10/

Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, *62*, 102441. https://doi.org/10.1016/j.ijinfomgt.2021.102441

Stirbu, V., Raatikainen, M., Röntynen, J., Sokolov, V., Lehtonen, T., & Mikkonen, T. (2022). Toward multiconcern software development with everything as code. *IEEE Software*, *39*(4), 27–33. https://doi.org/10.1109/MS.2022.3167481

Tamminen, P. (2022). *Operationalizing transparency and explainability in artificial intelligence through standardization*. https://www.utupub.fi/handle/10024/154054

Tournas, L. N., & Bowman, D. M. (2021). AI insurance: Risk management 2.0. *IEEE Technology and Society Magazine*, *40*(4), 52–56. https://doi.org/10.1109/MTS.2021.3123750

Viljanen, M., & Parviainen, H. (2022). AI applications and regulation: Mapping the regulatory strata. *Frontiers in Computer Science*, *3*. https://www.frontiersin.org/article/10.3389/fcomp.2021.779957

Wirtz, B. W., Langer, P. F., & Weyerer, J. C. (2022). An ecosystem framework of AI governance. In J. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. Young, & B. Zhang (Eds.), *The Oxford handbook of AI governance*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.24

Zimmer, M. P., Minkkinen, M., & Mäntymäki, M. (2022). Responsible artificial intelligence systems: Critical considerations for business model design. *Scandinavian Journal of Information Systems, 34(2)*. https://aisel.aisnet.org/sjis/vol34/iss2/4