Scandinavian Journal of Information Systems

Volume 34 | Issue 2 Article 4

12-31-2022

Responsible Artificial Intelligence Systems Critical considerations for business model design

Markus Philipp Zimmer
University of Turku, maphzi@icloud.com

Matti Minkkinen
University of Turku, matmin@utu.fi

Matti Mäntymäki
University of Turku, matti.mantymaki@utu.fi

© Scandinavian Journal of Information Systems, 2022 34(2), 113-162

Responsible Artificial Intelligence Systems

Critical considerations for business model design

Markus Philipp Zimmer
Leuphana University Lüneburg, Germany, and
University of Turku, Finland
markus.zimmer@leuphana.de

Matti Minkkinen
University of Turku, Finland
matti.minkkinen@utu.fi

Matti Mäntymäki
University of Turku, Finland
matti.mantymaki@utu.fi

Abstract. Commercializing responsible artificial intelligence (RAI) involves translating ethical principles for developing, deploying, and using AI into business models. However, prior studies have reported tensions between commercial interests (e.g., development speed or accuracy) and societal interests (e.g., privacy or human rights) that can undermine RAI's value proposition. Conceptually, we distinguish two business model development perspectives on AI and responsibility: innovating responsible business models leveraging AI and designing RAI business models. Taking the second perspective, we investigate the value proposition of RAI through business model design by employing a two-stage research approach consisting of focus groups and member checking. Empirically, we present the learnings from identifying the design elements for RAI business models. These include two themes that can underlie such business models: providing vs. enabling RAI systems and the observation that the tensions in RAI's value proposition are paradoxical, not dilemmas. With our conceptual groundwork and empirical insights, we make three contributions that offer critical considerations for RAI business model design. First, we conceptualize two pathways for designing RAI business models: a corner path

Accepting editor: Arisa Shollo

© Scandinavian Journal of Information Systems, 2022 34(2), 113-162

to commercialized RAI systems vs. direct path to commercialized RAI systems. We argue that these paths have distinct implications for the responsible in RAI. Second, we reflect the sociotechnical nature of RAI systems by emphasizing the criticality of the social for responsibility. Third, we outline a research agenda for developing RAI business models.

Key words: Artificial Intelligence, AI, Responsible AI, Ethical AI, Business Models, Business Model Design.

1 Introduction

Artificial intelligence (AI) poses both opportunities and risks to human welfare that require managing AI (Accenture, 2018; Berente et al., 2021; Dignum, 2019; European Commission, 2020). Research into AI dates back to the 1950s, although initial enthusiasm was followed by an AI winter with decreased funding, at least until the 1990s (Sun & Medaglia, 2019). However, AI research and development has intensified in recent years because of increases in computing power and the availability of data. While strong, human-level AI remains a hypothetical prospect, AI currently powers advanced algorithmic systems that promise to enhance business processes and decision making across societal sectors from policing to banking (Joshi et al., 2021; Tarafdar et al., 2019). However, capturing the potential of AI through systems' learning and adaptation capabilities (Kaplan & Haenlein, 2019) will require organizations to rethink their operations, structures, and business models (Fountaine et al., 2019).

In addition to rethinking business models, AI presents a broader challenge at the societal level: ensuring that AI systems respect human values and ethical boundaries. When discussing the management of AI risks, scholars have suggested concepts such as ethical, trustworthy, and responsible AI (RAI) (Dignum, 2019; Jobin et al., 2019; Thiebes et al., 2021), arguing that AI should aspire to moral values when interpreting and learning from external data (Dignum, 2017). This notion has also found its way into businesses and policymaking. In particular, the European Union (EU) has adopted an active role in defining AI approaches and principles to reap the benefits of the technology while ensuring its responsible development and use (European Commission, 2020; Minkkinen, Zimmer, et al., 2022). The EU's goal is to promote the development of AI that respects human rights and ensures that RAI is "ethical, secure and cutting-edge AI made in Europe" (European Commission, 2018). This will involve developing the value proposition of RAI, meaning identifying and leveraging its commercial strengths. Without a respective value proposition, RAI could simply remain a set of principles, rather than becoming a potential source of competitive advantage (Mittelstadt et al., 2016; Morley et al., 2020; Seppälä et al., 2021).

In existing studies, this quest for RAI's value proposition has been discussed in terms of the commercialization challenges of implementing ethical principles in the design and operation of commercially viable AI (Eitel-Porter, 2021; Gasser & Almeida, 2017; Morley et al., 2020; Whittlestone et al., 2019). Danaher et al. (2017) report tensions between commercial interests and privacy, especially when sourcing external data. Similarly, Whittlestone et al. (2019) suggest that using data to improve efficiency and quality would conflict with individuals' privacy and autonomy. Further, they illustrate the tension between accuracy in predictions and decisions and fair and equal treatment. Morley et al. (2020)132(3429 contend that applying ethics requires significant effort from AI developers, while current AI ethics methods lack usability. Collectively, these studies suggest that compared to AI, RAI offers ethical advantages but has a weaker commercial value proposition. However, a technology's value proposition is a key aspect of its commercialization.

The commercialization of technology focuses on turning an innovation into a product or service (Markman et al., 2008). Applied to RAI, this raises questions about the value proposition of RAI, the partnerships, the key activities required for its delivery and potential customers who may use or purchase the technology (Osterwalder et al., 2005). These aspects are central to the business model concept, and scholars have suggested that business models are a powerful device for strategizing the nature and delivery of a technology's value proposition (Doganova & Eyquem-Renault, 2009). Extending these notions to RAI, we argue that we need to study the design elements of RAI business models (value proposition, potential customers, key partnerships, and key activities) to turn RAI into a competitive advantage. However, when seeking to commercialize RAI, we consider it pivotal to adopt a critical approach when considering the innate commercialization challenges, especially the trade-offs we may face between commercialization and responsibility. Thus, we aim to study the design elements and development approaches for RAI business models to identify critical considerations for addressing RAI's commercialization challenges and related research issues. Accordingly, we develop a conceptualization of RAI business models and elaborate on this concept through an explorative empirical study.

Next, we outline the concepts of RAI and business model development. Conceptually, our study introduces a definition of RAI as sociotechnical systems and two perspectives on business model development, AI and responsibility (*innovating responsible business models leveraging AI* vs. *designing RAI business models*). With regard to the latter, we refer to commercializing RAI. Subsequently, we describe our research approach, which involves focus groups and member checking (Birt et al., 2016; Stewart et al., 2007). We then present our empirical findings on the key design elements for RAI business models and two underlying themes (*providing* vs. *enabling*). Based on these conceptual and empirical insights, we offer three contributions to RAI research. First, we conceptualize two pathways for designing RAI business models (*corner* vs. *direct*) and discuss the implications of these two pathways for the responsible[1] in RAI. Second, our findings indicate that the responsible in RAI stems from responsible technical and social activities. While the former present necessary criteria, the latter are sufficient criteria to render AI responsible. This emphasizes the criticality of the social[2]—in sociotechnical—for RAI. Third, we present a research agenda for developing RAI business models.

2 Examining the value proposition of responsible artificial intelligence

2.1 What are responsible artificial intelligence systems?

The increase in computer power and the availability of data have rekindled practitioners' and researchers' interest in AI (Ulnicane et al., 2021). While this interest has generated studies on the technology's capabilities, it has also sparked discussions on related risks because of the complex and often inscrutable AI systems (Berente et al., 2021; Clarke, 2019). This resulted in increasing attention being afforded to RAI and the emergence of AI-related policy initiatives on national (Schiff et al., 2020) and cross-national levels (European Commission, 2020). Despite this attention and significant research (e.g., Dignum, 2019), the concept of RAI lacks a concise and established definition. However, to define RAI—more precisely, RAI systems—we must first define AI.

In their literature review, Samoili et al. (2020, 8) identify four common features of AI definitions: perception of the environment, information processing, decision making (including reasoning and learning), and achievement of specific goals. The EU's High-Level Expert Group on AI's (2019a) definition adds a further two features: AI systems are designed by humans, and they act "in the physical or digital dimension by perceiving their environment through data acquisition" (p. 6). Accordingly, we can consider an information system (IS) as a type of AI when it is capable of learning and adapting in a partly autonomous, bottom-up manner. Moreover, this would be based on data rather than using formalized rules, such as if-then statements (Dignum, 2019, p. 13; Kaplan & Haenlein, 2019). According to Kaplan and Haenlein (2019) "AI [is] a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation." This means that although AI systems can learn and adapt to solve complex tasks, their learning and adaptation can cause them to evolve along unforeseen trajectories. More importantly, their abilities offer opportunities and pose risks to human welfare (Berente et al., 2021).

Recognizing these risks, researchers have suggested different mitigation approaches, including explainable AI (Barredo Arrieta et al., 2020; Laato, Tiainen, et al., 2022; Meske et al., 2022) ethical AI (Eitel-Porter, 2021; Floridi et al., 2018; Jobin et al., 2019), human-centric AI (Shneiderman, 2020), trustworthy AI (European Commission, 2020; High-Level Expert Group on Artificial Intelligence, 2019b; Thiebes et al., 2021), AI governance (Laato, Birkstedt, et al., 2022; Mäntymäki et al., 2022a, 2022b; Seppälä et al., 2021) and RAI (Dignum, 2019, 2020; Minkkinen, Niukkanen, et al., 2022). In this study, although we acknowledge the breadth of concepts addressing AI risks, we use the term RAI for three reasons. First, its inclusiveness allows for explorative research without adopting a narrow perspective (e.g., the concept of explainable AI presumes a technical perspective). Second, it is well established among researchers (Dignum, 2020) and practitioners (Accenture, 2018; Eitel-Porter, 2021; PwC, 2019) making it a shared conceptual device. Third, it addresses both the design of AI and its use (i.e., RAI is also a matter of responsibly using AI). While we view this inclusiveness and widespread use as strengths of the RAI concept, this means that RAI can take different meanings.

Scholars have discussed at least five definitions of RAI. We initially outline these definitions as the groundwork for providing our working definition of RAI systems. Researchers have conceptualized RAI as follows: (1) a governance framework, (2) organizational choices, (3) practices of using AI, (4) legal and ethical investigations, and (5) designing and implementing AI in alignment with human values (Benjamins, 2021; Dignum, 2019; Eitel-Porter, 2021; Taylor et al., 2018; Wang et al., 2020). Wang et al. (2020) define RAI as a governance framework used to "harness, deploy, evaluate, and monitor AI machines," with a focus on designing and implementing ethical AI systems. Benjamins (2021) considers RAI to be a set of organizational choices about AI principles and their technical articulation. Eitel-Porter (2021) understands RAI as the practice of using AI to empower employees and businesses and to ensure a fair impact on customers and society. Taylor et al. (2018) understand RAI as an "umbrella term for investigations into legal, ethical and moral standpoints of autonomous algorithms or applications of AI." Finally, Dignum (2019, p. 119) suggests that RAI refers to the design and implementation of AI in alignment with human values. She states that "[r]esponsible AI means that AI systems should be designed and implemented in ways that recognise and are sensitive to human interaction contexts without infringing on core values and human rights." Subsequently, she elaborates her definition by positing interactivity, autonomy, and adaptation as core features of AI systems, which should be complemented with accountability, responsibility, and transparency (ART) to ensure RAI (2020). She also argues that "[a]ddressing ART will require a sociotechnical approach to design, deployment, and use of systems, interweaving software solutions with governance and regulation" (Dignum, 2020, pp. 217-218). In this study, we present all five meanings, since each illuminates certain aspects that characterize RAI, although individually they fall short of defining its full meaning. To provide a working definition of RAI, we particularly draw on Dignum's definition of RAI and Kaplan and Haenlein's (2019) definition of AI systems. Accordingly, by combining Dignum's notion of the sociotechnical nature of RAI systems and her criteria for when we can deem AI systems responsible with Kaplan and Haenlein's AI definition, we propose the following working definition of RAI systems:

> *A responsible AI system is a sociotechnical system in which an AI agent and social entities jointly interpret external data, learn from such data, and use this learning to achieve specific goals and tasks through flexible adaptation in a manner that is judged responsible according to deontological and consequentialist criteria.*

Having defined RAI systems, we now elaborate on the key terms in our working definition. We conceptualize an RAI system as a sociotechnical system because AI agents and social entities perform the activities that render AI systems responsible. These AI agents can be software components, algorithms, or machine learning models, whereas social entities can be individuals, organizational units, or entire organizations. While the former can interpret, learn, and adapt from external data, the latter provides data, designs, and interpretations and uses the AI agent for certain purposes. Jointly, their activities should be aligned with human values. Thus, we do not suggest that AI agents and social entities have the same role. Rather, we posit that AI systems require both technical and social means to be responsible.

In our definition, a comprehensive understanding of RAI systems involves both deontological and consequentialist criteria. Deontology is the predominant approach in AI ethics and is based on duties and rules (Hagendorff, 2020). Deontological criteria are fundamental values, ethical principles, social norms, and laws (Jobin et al., 2019). We argue that legal compliance is the bare minimum for RAI systems. Complementing deontological criteria, consequentialist criteria pertain to the outcomes of actions (i.e., consequences), such as the adverse impacts of automated decisions on individuals (Hagendorff, 2020; Trocin et al., 2021).

In essence, we define RAI as AI that exercises its capabilities in a manner that is considered responsible. Different from AI, the concept of RAI rests on the assumption that there can be irresponsible AI systems. If every AI system were responsible by default, there would be little need for RAI as a separate concept. Moreover, responsibility would not be a differentiating factor between AI and RAI systems. The AI ethics literature has extensively discussed the risks of biases, algorithmic opacity, and ethically reprehensible AI uses such as propaganda (e.g., Dignum, 2020; Floridi et al., 2018; Hagendorff, 2020). Thus, AI capabilities are not always exercised responsibly.

The literature supports the assumption that RAI involves managing tensions and trade-offs. The predominant framing of RAI encompasses risk mitigation, sacrificing efficiency, accuracy, and commercial interests (Clarke, 2019; Danaher et al., 2017; Whittlestone et al., 2019). According to this framing, RAI involves trade-offs between efficiency and privacy (including individual autonomy) and between accuracy and fair and equal treatment (Whittlestone et al., 2019). Further, Danaher et al. (2017) report tensions between commercial interests, which favor development speed and profit, vs. societal interests, which favor individuals' rights, privacy, and dignity. Morley et al. (2021) suggest that RAI requires compromises between being too flexible and too strict as well as between devolved and centralized responsibility. In our view, the framing of these tensions in the literature suggests that they pose either/or questions, which demand that actors choose one option over the other (e.g., commercial or societal interests) for their resolution (Smith et al., 2010; Smith & Lewis, 2011; Whittlestone et al., 2019). This indicates that these tensions pose a challenge for designing a value proposition for RAI systems that features both ethical principles and commercial viability.

We argue that these tensions inevitably point toward theoretical and practical problems in and for the design of RAI systems' value proposition. This raises two questions: how do these tensions play out in specifying RAI systems' value proposition, and how can we turn RAI systems into a commercially viable product or service when RAI means sacrificing the efficiency and other benefits that AI offers. To the best of our knowledge, except for a set of consultancy reports (e.g., Accenture, 2018; PwC, 2019), previous work has not addressed this problem of either/or tensions in the value proposition of RAI systems.

2.2 Sketching the value proposition of responsible artificial intelligence through business model development

The value proposition of technological innovations is crucial for the commercialization of these innovations. Commercialization is a process that aims to turn technological innovations into viable products or services (Chebo & Wubatie, 2021; Eldred & McGrath, 1997). However, this process often fails because of a lack of understanding of a technology's value proposition, its potential customers, and the key partners for delivering this value proposition (Laird & Sjoblom, 2004). These three aspects (value proposition, potential customers, and key partners) are at the core of the business model concept (Osterwalder et al., 2005). Therefore, in the following paragraphs, we consult existing work on business model development to identify concepts that are suitable for sketching the value proposition and related key elements (i.e., potential customers and key partners) of RAI systems.

We understand business models as "a simplified and aggregated representation of the relevant activities of a company […] [which] describes how marketable information, products and/or services are generated by means of a company's value-added component" (Wirtz et al., 2016, p. 41). Hence, they are an abstraction of how a company seeks a competitive advantage through activities of creating, exchanging, and capturing value (Keen & Qureshi, 2006; Teece, 2010). This definition implies that business models concentrate on a single firm. However, they also include the activities of partners (e.g., vendors, suppliers, and policymakers) and customers (Al-Debei & Avison, 2010; Zott & Amit, 2010), which are interdependent of a firm's activities when operating their business model. Thus, business models represent a firm's reality (Doganova & Eyquem-Renault, 2009).

Besides representing reality, companies can use business models to sketch a potential reality (Doganova & Eyquem-Renault, 2009). From this perspective, we conceive business models as market devices or instruments that enable managing tensions in the commercialization process. Thus, beyond explaining a firm's reality, business models can serve as blueprints for orchestrating organizational activities around an information system (IS) (i.e., the IS renders the value proposition) and involving an IS (i.e., the IS supports the value proposition) such that they contribute to organizational performance (Baden-Fuller & Morgan, 2010). This renders business model development a crucial activity in the commercialization process of a technological innovation such as RAI (Chesbrough & Rosenbloom, 2002; Dmitriev et al., 2014; Markman et al., 2008). However, the question then becomes how to develop business models for a potential reality.

In this study, we differentiate between two approaches to business model development: business model innovation and business model design (Baden-Fuller & Morgan, 2010; Chesbrough, 2010; Wirtz et al., 2016). The first refers to innovation within an existing business model (e.g., leveraging new technologies), while the second involves designing a business model from scratch (Chesbrough, 2010; Wirtz et al., 2016). We consider business model design a key activity for understanding and sketching the value proposition of RAI systems for two reasons. First, the value proposition of RAI systems remains elusive (Mittelstadt et al., 2016; Morley et al., 2020). Second, existing studies on RAI have reported tensions within RAI systems' value proposition (Clarke, 2019; Danaher et al., 2017; Whittlestone et al., 2019). Taking a business model design approach, we examine how companies can address these tensions to create and deliver the value propositions of RAI systems (Wirtz et al., 2016; Zott & Amit, 2010). To accomplish this aim, we next consider the elements that constitute a business model.

Various business model frameworks propose a number of design elements that jointly represent a business model (Zott & Amit, 2010). We subsume these elements to five, which are shared among these frameworks: value proposition, potential customers, key partners, key activities, and finances (Al-Debei & Avison, 2010; Ballon, 2007; Nenonen & Storbacka, 2010). Value proposition refers to determining and positioning a company's offerings (Hedman & Kalling, 2003; Nenonen & Storbacka, 2010). The element of potential customers is closely connected to value proposition, since the offering determines potential customers and vice versa (i.e., to whom it poses value) (Nenonen & Storbacka, 2010; Osterwalder et al., 2010; Teece, 2010). The third commonly shared element is value networks or key partners, who assist in delivering the value proposition (Al-Debei & Avison, 2010; Ballon, 2007; Osterwalder et al., 2010). The fourth element of key activities refers to how a company organizes its value creation (Hedman & Kalling, 2003; Nenonen & Storbacka, 2010; Osterwalder et al., 2010). The final shared element includes finances, the revenue model, or earnings logic (Al-Debei & Avison, 2010; Nenonen & Storbacka, 2010; Osterwalder et al., 2010). Together, these elements constitute the business model concept.

Considering that our research purpose is to examine the design elements and development approaches for RAI business models, we focus on the elements of value proposition, potential customers, key partners, and key activities. Value proposition is at the core of our research aim and is closely related to potential customers (value for who?), key activities (how is value produced?), and key partners (with who is value produced?) (Osterwalder et al., 2010). Further, extending our view with respect to the commercialization process of technological innovation, next to value proposition, potential customers and key partners are key issues that can determine why commercialization fails (Laird & Sjoblom, 2004). As a form of business model representation, we adopted the business model canvas presented by Osterwalder et al. (2010), since it is widely used among practitioners. Before describing the research approach, in the following section, we conceptualize two perspectives on business model development, AI and responsibility.

2.3 Two perspectives on business model development, artificial intelligence and responsibility

Drawing on the two conceptions of business model innovation and business model design (Wirtz et al., 2016), we distinguish between two perspectives on business model development, AI and responsibility. First, from the perspective of business model innovation, organizations can use AI to innovate an existing responsible business model toward a business model that leverages the value proposition of AI to achieve some societal good (e.g., leveraging AI to offer personalized advertisements and then using the profits to plant trees to combat climate change). Second, by adopting a business model design perspective, organizations can create and implement operations that leverage the value proposition of RAI, constituting an RAI business model. In this paper, we focus on the second perspective of how organizations can design business models for RAI systems. To clarify our focus, in the following section, we describe both perspectives and ground them in prior work on AI and RAI.

When studying AI in business, scholars tend to view AI as instrumental, meaning something to be infused into existing business models for added value (Ransbotham et al., 2017). We discovered studies on how AI drives change and innovation in business models (Armour & Sako, 2020; Burström et al., 2021; Lee et al., 2019; Sjödin et al., 2021; Zaki, 2019) and how AI can provide business value (Garbuio & Lin, 2019; Wamba-Taguimdje et al., 2020). Some researchers have also examined how AI can support business models that promote sustainability (Di Vaio, Boccia, et al., 2020; Di Vaio, Palladino, et al., 2020; Toniolo et al., 2020). In these studies, the focus is on the responsibility of business models as a whole, linking them to research on corporate social responsibility (van Marrewijk, 2003) and organizational value logics (Laasch & Pinkse, 2020). In this paper, we ascribe the notion underlying this research (using AI as a component of responsible business models) to business model innovation (Wirtz et al., 2016). Thus, we refer to this first perspective as "innovating responsible business models that leverage AI." We argue that this perspective does not address the responsibility of AI, the value proposition of RAI systems, or the respective commercialization challenges.

It could be argued that the EU's intent to establish RAI as a competitive advantage (Antonov & Kerikmäe, 2020) requires a different perspective: establishing RAI business models instead of infusing AI in responsible business models. Accordingly, existing work on translating AI ethics into practice has focused on providing actionable mechanisms for implementing RAI systems (Morley et al., 2020, 2021). Breidbach and Maglio (2020) investigated the ethical implications of data-driven business models and suggest that the design of ethical data-driven value propositions would be a novel area of inquiry. Further, Kumar et al. (2021) examined the role of RAI in value formation and market performance in a healthcare context in India. Both studies share the concept of placing RAI systems at the center of business model development. In this case, responsibility emanates from the RAI system, not from the business model as a whole. We argue that this perspective requires the creation of business models for RAI systems, rather than incorporating AI systems into (responsible) business models. Thus, we refer to this perspective as designing RAI business models. This second perspective demands the design of business models that leverage the value proposition of RAI systems and address their commercialization challenges.

We conceptualize the two outlined perspectives in Table 1. We differentiate these perspectives based on their underlying business model development approach and ground them in related literature streams on AI. Further, we situate our study as taking the perspective of designing RAI business models. Hence, we focus on designing RAI business models based on the value proposition of RAI systems.

	Innovating responsible business models leveraging AI	*RAI business models*
Underlying business model development approach	• Business models using AI as one part of the value proposition for responsible business. • Responsible business models comprise activities that infuse value creation with elements of ethics and sustainability to make the business model responsible (e.g., using AI in sustainability projects or ethically and sustainably investing revenue created using AI). • Takes a business model innovation approach. • The responsibility focuses on the overall business model and not on AI itself.	• Business models commercializing the value proposition of RAI systems. • Respective business models comprise activities that create value with RAI systems; i.e., the value creation through AI itself is responsible. • Takes a business model design approach. • The main responsibility rests on the RAI system.

Related AI literature streams	• AI supporting business models that promote sustainability (Di Vaio, Boccia, et al., 2020; Di Vaio, Palladino, et al., 2020; Toniolo et al., 2020) • AI and digital technologies driving business model innovation (Armour & Sako, 2020; Burström et al., 2021; Lee et al., 2019; Sjödin et al., 2021; Zaki, 2019)	• Translating AI ethics principles into practice (Morley et al., 2020, 2021)132(3429. • Ethical data-driven value propositions (Breidbach & Maglio, 2020). • RAI in value formation and market performance (Kumar et al., 2021).

Table 1. Two perspectives on business model development, AI and responsibility

3 Research approach

Following the business model development literature, we investigated design elements that could partially inform potential RAI business models. To accomplish this, we conducted focus group discussions (Stewart et al., 2007) and an elaboration workshop involving member checking (Birt et al., 2016). Member checking is a qualitative research technique for exploring the credibility of results and their resonance with participants through interviews, surveys, or in our case, a workshop (Birt et al., 2016). Hence, we followed a two-stage research process, which interlaced the activities of data collection and data analysis.

3.1 Data collection

We collected qualitative data in the first stage by employing focus groups (Stewart et al., 2007), while in the second stage, we used member checking (Birt et al. 2016). Table 2 provides an overview and details of the data collection.

Stage	*Objective*	*Details on data collection*

#1	Five concurrent focus groups (March 2021)	Gain initial insights on practitioners' understanding of the selected design elements for RAI business models.	• No. of participants: 19 (including 5 facilitators). • Means of data collection: Written notes on focus groups and their final presentations, notes on debriefing among focus group facilitators, and completed business model canvases.
#2	Elaboration workshop (using member checking) (April, 2021)	Present preliminary findings from the focus groups to gain additional insights on practitioners' framing of the selected design elements for RAI business models.	• No. of participants: 18 (including 2 facilitators). • Means of data collection: Recording workshop sessions, written answers to posed questions, written notes.

Table 2. Data collection in the two-staged research process

We conducted both stages with industry experts from five different companies participating in a joint research project on Artificial Intelligence Governance and Auditing (AIGA), which has the purpose of examining the value proposition of RAI systems. Before outlining the data collection, we provide an overview of the involved companies and informants in Table 3.

Company	*Description*	*Job roles*
Company A	Large consulting company (<1,500 employees) offering strategic consulting, service design, software development, AI, analytics, and cloud and cloud-integration services.	Head of Research, Head of Sustainable AI, Business Lead (Data-Driven Business), Insight Lead, Data Scientist, Data Business Designer.
Company B	Small/medium-sized consulting company (<100 employees) offering digital solution design.	Executive Advisor, Sales Director, Principal Consultant.

Company C	Large consulting company (<1,000 employees) offering digital strategy, software engineering, and data and intelligence services.	Head of AI and Data Works, Competence Lead, Design Researcher, Service and UX Designer.
Company D	Small/medium-sized company (<50 employees) offering data and AI strategy, data science, and data architecture services.	Co-founder, Analytics Executive, Chief Data and AI Officer.
Company E	Small/medium-sized company (<50 employees) offering an AI-based cloud service.	Founder, CEO.

Table 3. Description of companies participating in the data collection process

During the first stage of data collection (March 2021), we conducted a virtual workshop on potential RAI business models with experts from the participating companies and the AIGA project. This workshop comprised five concurrent focus groups tasked with discussing the selected design elements (value proposition, potential customers, key partners, and key activities). Researchers have previously used focus groups and workshops for business model design (Blaschke et al., 2016; Turetken & Grefen, 2017). The focus group approach allowed us to collect data efficiently from busy experts and to generate new insights through synergistic interactions between the participants (Stewart et al., 2007, pp. 42-43). The virtual focus groups were used for practical reasons: the COVID-19 pandemic did not allow physical meetings. While the virtual setting could limit the interaction, it made participation possible for the time-constrained experts. In total, 19 people participated in the workshop. Each of the five groups consisted of three to five industry experts from the participating companies (see Table 3). The participants encompassed different areas of expertise, including data management, privacy, digital strategy, data science, AI consulting, and AI product development. The uniting factor was that each participant had at least five years of experience in their area of expertise. In many cases, they had more than 10 years of experience across several areas connected to digital strategy. Most participants were also in senior management roles. Further, five researchers from the AIGA project moderated the focus groups. They employed a nondirective approach to allow space for individual views and spontaneous interactions among participants (Stewart et al., 2007, p. 91). We provided each focus group (and its moderators) with a simplified business model canvas that consisted of the selected design elements and a set of guiding questions (Figure 1). The intention was for them to capture the key points of the discussion using our simplified business model canvas. We randomly allocated the participants to the focus groups and ensured that each group included experts from different companies. The focus groups were allocated approximately 30 min for discussions.

WANTS TABLE 4 ABOUT HEREIn the second stage of data collection (April 2021), we elaborated on our findings in a second virtual workshop that involved member checking (Birt et al., 2016). However, the objective of this workshop was not to validate our preliminary analysis. Rather, it was to collect further data to complement our focus group data, allowing us to elaborate on our initial insights. On this occasion, we adopted a structured and directive approach (Stewart et al., 2007, p. 91) and invited the same industry partners to the second workshop. While some participants from the focus groups joined, we also had new participants. We started the second workshop with a recap of the first, including a presentation of our preliminary findings. Subsequently, we invited all the participants to comment on and discuss our findings. For this, we used an online tool to pose questions to the audience. Each participant could then respond individually without seeing the other participants' remarks. We decided to use this online tool and its anonymity to offer each industry expert the opportunity to participate, comment, complement, or even object to the statements from the focus groups. Moreover, by not being able to see other experts' remarks and comments, this mitigated the risk that participants would assimilate their responses or refrain from sharing their views due to a fear of rebuttal. Three consecutive questions (Table 4) were posed, and we shared the participants' remarks in an open discussion during the virtual workshop.

Key activities	**Value Proposition**	**Customers**
What are your organizations' key activities to deliver these benefits?	What are the benefits that your organization offers with responsible AI?	Who are your organization's recipients of responsible AI benefits?
Key partners		
Who are the key partners in these key activities?		

Figure 1. Simplified business model canvas with guiding questions adapted from Osterwalder et al. (2010).

© Scandinavian Journal of Information Systems, 2022 34(2), 113-162

3.2 Data analysis

In our two-stage research process, the data collection and analysis stages were interlaced. Figure 2 illustrates the research process and the interlacing of data collection and FIGURE 2 ABOUT HEREanalysis. First, we consulted our notes from the five focus groups, as well as the focus groups' business model canvases. We tabulated the business model canvases in a spreadsheet, listing the selected design elements (value proposition, customers, key activities, and key partners) as rows and the five focus groups as columns. We then transferred and merged the data collected in the notes and focus groups' business models to the respective cells in the spreadsheet. This resulted in a schema for initially analyzing the focus group data individually (i.e., within each focus group itself) and then cross-comparing the five focus groups' data for each design element. An excerpt of the preliminary findings is presented in tabular form in the Appendix (see Table 10).

Second, we analyzed the data from the elaboration workshop. The data included a recording of the workshop, participants' written responses to our questions on the preliminary findings (Table 4) and our notes taken during the workshop. We viewed the participants' responses to validate our preliminary findings and searched for any additional material, in line with the member check focus group approach (Birt et al., 2016). Accordingly, we analyzed these responses to identify further or complementary design elements, linking them to the design elements they addressed. We then assessed whether they refined, extended, or defied our preliminary findings and incorporated them accordingly. Refining responses provided a better understanding of the practitioners' framing of the business model design elements, while extending responses built on our preliminary findings and provided additional insights into the design of RAI

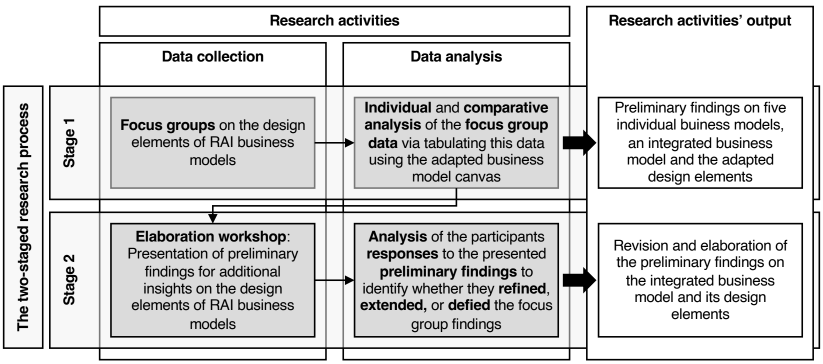

Figure 2. The two-stage research approach, including data collection, data analysis, and each stage's outputs.

Q#	*Questions posed to workshop participants*	*Underlying notion*
#1	How do the key observations resonate with your own views on RAI business models?	• The question focuses on the preliminary findings, and whether they resonate with the respondents. • Objective of gaining complementary material closely related to existing material.
#2	Considering your own views on RAI business models, what do you think are the key questions and issues in these business models?	• The questions focus on the respondents' interpretations of the first workshop's findings. • Objective of broadening the analysis to identify further design elements of respective RAI business models.
#3	Can you name important aspects of RAI business models that are missing in the analysis?	• This question focuses explicitly on what is missing from the preliminary findings. • Objective of inviting respondents to specifically complement what is missing, rather than building on the preliminary findings.

Table 4. Questions posed to participants in the second workshop to elaborate on our preliminary findings

business models. Defying responses expressed caution about the preliminary findings, suggesting a different framing.

4 Findings

In this section, we present our findings on each design element in the following sequence: value proposition, customers, key activities, and key partners. We then outline the learnings from integrating these design elements into one business model.

4.1 Potential value proposition

By analyzing the focus group data, we identified two tenets of RAI's value proposition. The first was based on AI itself, and the second was based on activities underlying the notion of RAI. For example, the participants stressed that AI creates value through "automat[ing] business processes" (Focus group #2), entailing increased "operational efficiency" (Focus group #5). In addition, the focus groups discussed how RAI extends this value proposition. In other words, they positioned RAI to offer the value proposition of AI (i.e., increased operational efficiency) plus additional value. They related this additional value to RAI as being "value-driven, explainable, and transparent" (Focus group #1) and "societally acceptable" (Focus group #2).

Being value-driven, RAI "considers the customers of AI solutions to increase their life quality" (Focus group #4). This renders RAI systems acceptable within society. This is because they do not harvest data solely for profit and also deliver a service that benefits the end user (i.e., the person who the RAI solution's "decisions" ultimately affect). This notion suggests the consequentialist ethical stance (i.e., focus on outcomes), which was included in our RAI system definition. Explainability and transparency support this value proposition, as end users can grasp and evaluate how and why an RAI solution has arrived at one decision (or recommendation) over another (cf., Barredo Arrieta et al., 2020). Further, they can assess whether an AI solution complies with applicable regulations when handling their data. The participants in the focus groups argued that by combining these aspects, RAI intrinsically involves "risk management" (Focus group #4 and #5) (cf., Clarke, 2019).

According to the focus group discussions, AI systems pose certain ethical, legal, financial, and reputational risks. Collecting sensitive data can have ethical and legal consequences, which could entail being fined (i.e., financial risk) and (most likely worst) could engender reputational damage that might impede future business. Positioning RAI as "AI that complies with regulations and is safe to use" (Focus group #4) means managing these risks and building a sustainable and trusted brand. Hence, RAI features the value proposition of managing risks related to AI while simultaneously building a sustainable brand (cf., Eitel-Porter, 2021). This sustainability argument was also reflected in other future-oriented arguments discussed within the focus groups.

The focus group discussions generated value propositions related to the future. Focus group #2 argued for RAI "increas[ing] the lifetime of AI investments and reduc[ing] the risk of AI investments being wasted (for changing regulations)." Accordingly, RAI was "future-proof," enabling AI systems to "stand the test of time (e.g., fulfill current and anticipated regulatory requirements)" (Focus group #3). This would help companies to establish a "sustainable brand" (Focus group #4) and build trust for investors and customers. These arguments connect RAI to future orientation and sustainability, enabling an AI innovation model in which companies and societies "make haste slowly" (Floridi, 2019) rather than promoting risky and potentially destructive AI innovations.

Complementing these findings from the focus groups, the elaboration workshop provided three key observations of RAI's value proposition. First, one participant extended the focus groups, highlighting the importance of differentiating between at least "two somewhat different business models/value creation: utilization of RAI and helping organizations to build RAI." Second, another participant defied the preliminary findings from the focus groups. The focus group discussion revealed the notion that AI makes decisions, meaning that RAI could offer "responsible decisions" (Focus group #4). Defying this notion, the participant stated, "I'm wary of 'selling responsible decisions'. I would not be comfortable promising that to a customer, since there are always humans involved, and AI systems need to evolve as well. We can't control that."

This suggests that RAI may provide responsibly derived recommendations rather than responsible decisions, leaving the decision making to human actors. Moreover, this underpins the notion that RAI systems are sociotechnical systems. Finally, the elaboration workshop revealed that RAI offers a competitive advantage. Customers might prefer RAI systems, entailing an increase in data that could improve RAI services in the future. Thus, when designing an RAI business model, organizations should build on a value proposition that is itself sustainable rather than adding sustainability activities to an AI business model. Table 5 presents the remarks from the elaboration workshop and the resulting adaptations.

Focus group results addressed in the elaboration workshop	*Remark type and participants' remarks*	*Adapted result after elaboration workshop*
If the value proposition of AI is selling decisions, RAI systems sell responsible decisions.	*Defying*: "I'm wary of 'selling responsible decisions'. I would not be comfortable promising that to a customer, since there are always humans involved, and an AI system needs to evolve as well. We can't control that."	Defies the focus groups' statement that RAI systems "sell responsible decisions." Accommodating this defiant remark, we changed this value proposition to offering responsibly inferred recommendations to stress the sociotechnical nature of RAI systems.
Focus groups did not differentiate between business model themes.	*Extending*: "I see that there could be two somewhat different business models/value creation: utilization of RAI and helping organizations to build RAI."	Extends the focus group discussions, adding two business model themes: (1) providing RAI and (2) enabling organizations to develop RAI.

Focus groups stressed the value proposition of RAI along the responsibility (or ethical) dimension over commercial interests.	*Defying*: "[…] mostly external incentives [e.g., laws and regulations] are being mentioned. Actually, there are cases (not rare ones) where adopting RAI (e.g., privacy-preserving AI) is more profitable because trust gets you more data."	Defies the underlying assumption that RAI systems sacrifice the profitability of AI systems, counterarguing that RAI creates trust, resulting in an increase in data for learning and adapting the system.

Table 5. Adaptations to value propositions in RAI business models from the elaboration workshop

4.2 Potential customers

The focus groups' conversations about potential customers revealed two groups: organizational customers and consumers (or end-users). This differentiation relates to the often-used demarcation between business-to-business (B2B) and business-to-consumer (B2C) operations (cf., Laasch & Pinkse, 2020). Interestingly, when referring to consumers as end-users, the focus groups used the term "citizens" (Focus groups #1 and #2) rather than consumers. Focus group #5 stressed that RAI attracts "ethically aware and responsible customers." Extending this notion, the participants mentioned "society and democracy" (Focus groups #4 and #5) as customers of RAI. They stressed that even in the context of organizational customers, RAI business models might face a "[p]otentially long chain of customers from B2B to the entire human society" (Focus group #4).

Similar to the business model themes suggested under the value proposition, the focus groups implicitly differentiated between two different themes when discussing organizational customers. The first builds on RAI systems as systems offered to other organizations or consumers, while the second focuses on supporting organizations operating under the first business model theme. Within the first business model theme, RAI attracts organizational customers and consumers. Within the second business model theme, operated by actors such as AI consultancies and auditing firms, potential customers are the organizations developing and offering RAI. This constellation suggests a network of actors who can be both providers and customers of RAI systems within this network.

4.3 Potential key activities

We identified four key activities in potential RAI business models: 1) technologically developing RAI, 2) understanding the market needs for RAI, 3) auditing RAI systems, and 4) raising awareness and lobbying.

With regard to technologically developing RAI systems, the focus groups highlighted the importance of "XAI [explainable AI] for product development and developing RAI solutions" (Focus group #4). This activity includes developing tools and methods that would enable organizations to develop RAI, such as AI libraries and RAI open-source solutions that organizations could leverage. Hence, this could facilitate the development of RAI systems for technological advances and reusability. Moreover, the latter could reduce the investments required to develop RAI.

The second key activity is to understand the market needs for RAI. The focus groups offered several suggestions on who should be asked regarding market needs, as well as how. In terms of who, the focus groups highlighted organizations that could buy RAI systems as a service. However, they also noted that consumers (i.e., end users) should be involved, not merely organizations. The underlying goal was to understand these actors' needs and demands regarding RAI systems. When discussing how, they stressed "work[ing] with customers and ask[ing] questions" (Focus group #3). Focus group #2 specifically emphasized "researching the needs of clients, end-users, and people affected by AI solutions." In the elaboration workshop, this point was refined to not simply researching or asking about their needs; rather, it should entail co-creating using "service design methods to find out how to create value for both of them [the organization offering RAI and the customer]." Hence, the second key activity that emerged was understanding the needs of organizational customers and consumers regarding RAI systems.

One key activity mentioned by the focus groups was auditing (cf., Mökander et al., 2021). By receiving an audit of their AI systems as RAI systems, organizations could obtain a certificate in the future for their RAI. This certificate could help build trust in their solution, supporting their RAI business models. This renders auditing a key activity in RAI business models, as it can differentiate AI systems from RAI systems.

Finally, the focus groups stressed "growing awareness, knowledge, and skillsets" (Focus group #1). This involves "[l]obbying/selling RAI solutions (do not purchase unreliable black box systems)" (Focus group #4) to regulators as well as to customers (both organizational customers and consumers). Interestingly, the participants in the elaboration workshop were particularly responsive to this third key activity. One participant highlighted that "there might be a need to make us [customers] ethical too! Activities could be transforming society with ethical business models and services." Another participant highlighted the importance of creating external incentives through regulations as a key future issue. Similarly, a third participant argued for the ability "to credibly highlight incentives for firms to use RAI. For example, for privacy-preserving AI, intuitively it seems that introducing privacy will cause a decrease in accuracy, but it can actually increase it." In summation, the fourth key activity that emerged was creating and building awareness of the value proposition of RAI, especially its ability to compete with AI systems in terms of accuracy. Table 6 presents the remarks from the elaboration workshop and the resulting adaptations.

Focus group results addressed in the elaboration workshop	*Remark type and participants' remarks*	*Adapted result after elaboration workshop*
Understanding customer needs through market research.	*Extending*: "Consider all stakeholders' point-of-views and use Service Design methods to find out how to create value for all of them."	Extends the focus group results on understanding customer needs, stating that this activity should address customers and all stakeholders and reach beyond market research (i.e., also include service design methods).
Key activities focused on developing, implementing, and operating RAI systems.	*Refining*: "There might be a need to make us [consumers] ethical too! Activities could be transforming society with ethical business models and services."	Refines the underlying assumption that consumers demand RAI systems, and that the systems need to be ethical to RAI systems may transform society to become more ethical.
Regulating as a key activity, as RAI systems sacrifice commercial interests.	*Defying*: "Credibly highlight incentives for firms to use RAI. For example, for privacy-preserving AI, intuitively it seems that introducing privacy will cause a decrease in accuracy, but it can actually increase it."	Defies the assumption that RAI requires sacrifices in commercial interests (e.g., accuracy, efficiency in AI systems). RAI can also improve accuracy.

Table 6. Adaptations to key activities in RAI business models from the elaboration workshop

4.4 Potential key partners

When discussing potential key partners, the focus groups noted that different actors are crucial in the design of RAI business models. In particular, three types of key partner emerged: customers, regulators, and enablers. Customers include organizational customers (e.g., private companies and public organizations) as well as consumers (i.e., end-users), who might be "customers' customers (in certain specific occasions—indirect market creation)" (Focus group #4). This stresses that when designing an RAI business model, organizations might be required to extend customer relations beyond their direct customers. These end users may demand and prefer RAI systems over other AI systems, rendering them key partners in designing RAI business models for successfully commercializing RAI.

With regard to enablers, the focus groups referred to consultancy and auditing firms, technology and research institutions, and investors. The first of these offers advisory and development services for building RAI and for "establishing respective AI governance frameworks" (Focus group #2). Auditing firms could audit these frameworks and the built RAI systems "creat[ing] trust in an RAI solution [providing] certification" (Focus group #2). Technology and research institutions are important key partners for advancing RAI organizationally (e.g., governance frameworks) and technologically (e.g., machine learning models) (Focus group #4). Extending this observation, a participant in the elaboration workshop stated "AI is being built with pre-existing tools and APIs provided by cloud services. The key issue here is how to integrate RAI into the work that is done with these service providers. Can we responsibly do anything without their help?" These technological partners (e.g., providers of cloud services, AI libraries, and data providers) play a crucial role in training AI. Finally, investors are critical key partners as they decide whether organizations receive resources for developing RAI systems.

The third type of key partner was regulators, who were afforded a prominent role by the focus groups. The discussions circulated around governmental bodies issuing laws and regulations on AI. Moreover, they set the rules and incentives for RAI. One participant extended this issue in the elaboration workshop, arguing that "regulation and incentive systems are the future key issue. […] will there be enough strong regulatory and citizen activism to REALLY turn incentives towards responsible use?" Since non-governmental organizations influence regulations, they were named by the focus groups, particularly "human rights organizations" (Focus group #5), as key partners for "lobbying and raising awareness," thus influencing regulation (Focus group #4). Table 7 presents the remarks from the elaboration workshop and the resulting adaptations.

Focus group results addressed in the elaboration workshop	*Remark type and participants' remarks*	*Adapted result after elaboration workshop*
Technology and research institutions are key partners for advancing RAI organizationally and technologically.	*Extending:* "AI is being built with pre-existing tools and APIs provided by cloud services, the key issue here is how to integrate RAI into work that is done with these service providers. Can we responsibly do anything without their help?"	Extends the focus groups' views on technology providers as key partners. This remark highlights the importance of technology partners, framing them as gatekeepers (e.g., what if they do not develop technology enabling RAI?).
Regulators (e.g., governmental institutions) are key partners.	*Extending:* "Ecosystem-level regulation and incentive systems are the future key issue. Can companies continue breaking the law and benefit like now, or will there be enough strong regulatory and citizen activism to REALLY turn incentives towards responsible use?"	Extends external incentives (e.g., laws and regulations) to be the key issue by specifying regulators' roles as key partners in creating external incentives for RAI.

Table 7. Adaptations to key partners in RAI business models from the elaboration workshop

4.5 Learnings from integrating the design elements into one business model

Thus far, we have presented the design elements individually. However, for these to constitute RAI business models, we must consider them in an integrated business model (see Figure 3). Based on this integration, we highlight the key elements and interdependencies. For example, the technological development of RAI is a key element because without developing the technology for RAI, organizations cannot offer RAI systems. This relates to identifying key partners who can assist in developing RAI. However, while key partners are a versatile group, organizations also face a multifaceted customer landscape. The most obvious customers are organizational customers and consumers. However, the focus groups stressed concepts such as "ethically aware consumers" and talked about "citizens, society, and democracy" as customers. This observation emphasizes the importance of developing a better understanding of how organizations can raise awareness (as a key activity) as well as how they can scale their RAI business models when starting with ethically aware customers. These key elements and their relations indicate that while the value proposition of RAI differs from AI, it still seems elusive. More succinctly, the pertinent questions become what value can organizations capture through RAI systems (not through AI) and what value do customers perceive in RAI systems compared to AI.

The focus group responses highlighted an underlying assumption regarding RAI's value proposition: its monetization falls short of that of AI. This assumption has also been expressed in prior literature (Danaher et al., 2017; Whittlestone et al., 2019). In the elaboration workshop, the participants argued to the contrary, stating that there are cases of RAI contributing to monetary interests, since it can help organizations acquire more data because consumers trust their RAI system over other systems. This indicates that the tensions within RAI's value proposition do not present either/or dilemmas. Rather, they are paradoxical tensions of both/and (Smith & Lewis, 2011).

Prior studies have identified tensions within RAI's value proposition as commercialization challenges. Specifically, there are tensions between commercial interests (e.g., the accuracy and efficiency of AI systems) and societal interests (e.g., privacy, individual autonomy, and dignity) (See 2.1). The framing of these tensions suggests that they pose either/or questions regarding RAI's value proposition. This means that when developing RAI business models, organizations must choose between a commercially viable (and competitive) business model or a responsible yet less (if at all) commercially viable business model. Opposed to this view, our findings suggest that these tensions are paradoxical (Smith & Lewis, 2011). For example, the experts argued that RAI creates trust, suggesting that customers may choose an RAI system over an AI system. As a result, RAI systems accumulate more data than AI systems, which can have positive effects on their accuracy. Shifting our conception of these tensions to paradoxes rather than dilemmas alters the prior assumption that RAI's value proposition has less commercial potential in favor of ethical principles. Instead, the paradoxical nature suggests that organizations operating RAI systems must continuously balance these paradoxical interests. In other words, as paradoxical tensions cannot be resolved (Lewis & Smith, 2014), they circulate between being latent or salient. Hence, if at some point an organization established an RAI business model rendering these tensions latent, they may require adjustments when the public conception of RAI alters, legal requirements change, or new technological advancements offer new possibilities that re-render the tensions salient.

Key activities
- Technologically developing RAI
- Understanding market needs for RAI
- Auditing RAI solutions
- Raising awareness and lobbying

Value proposition
- Value proposition of *AI itself* (i.e., improved operational efficiency)
- Value propositions *beyond AI (i.e., RAI specific value proposition)*:
  - Intrinsic risk management (e.g., ethical, financial, and legal risks)
  - Positively affect consumer' life quality
  - Offering responsibly derived recommendations
  - Future-oriented: Increased lifetime of AI investments
  - Building a sustainable AI brand

Customers
- Organizational customers (e.g., B2B RAI solutions or auditing RAI)
- (Ethically aware) consumers (i.e., end-users)
- Citizens (society and democracy at large)

Key partners
- Customers (i.e., direct customers but also customers' customers; indirect market creation)
- Enablers (e.g., technology and research institutes, AI library providers, consultancies, auditing firms, investors)
- Regulators (e.g., governmental bodies)

Figure 3. Integrated business model

Thus, we argue that the tensions within RAI's value proposition are paradoxical (not dilemmas) and require the continuous balancing of competing interests.

Finally, the integrated business model suggests two business model themes (Nenonen & Storbacka, 2010; Zott & Amit, 2010): business models that leverage the value proposition of RAI itself and those that emerge at the periphery of RAI. While the first create value by offering an actual RAI system, the second create value through enabling, meaning supporting, facilitating, or assessing the development and operation of RAI systems. The integrated business model portrays the potential content of the design elements of an RAI business model for developing and offering RAI. It takes the perspective of an organization operating the first business model theme of providing RAI. Simultaneously, the integrated business model contains a list of key partners that create value, enabling the operation of RAI. These key partners thus operate the second business model theme of enabling RAI. While both business model themes appear promising, the enabling theme relies on organizations successfully commercializing RAI within the theme of providing RAI. Accordingly, the providing RAI business model theme has a primary position in the value network as a prerequisite for the enabling business model theme.

5 Discussion and implications

In this paper, we examined design elements and development approaches for RAI business models. We provided a working definition of RAI systems as sociotechnical systems, in which an AI agent and social entities jointly conduct activities in a responsible manner. Moreover, we problematized tensions in RAI's value proposition as commercialization challenges. We then differentiated between two perspectives on business model development: AI and responsibility to examine RAI's value proposition. We refer to the first as innovating responsible business models leveraging AI, and the second as designing RAI business models. We adopted the second perspective and conducted five focus groups and a member-checking workshop to study RAI's value proposition (and related design elements). Next, we reflect on three contributions to research on RAI. First, we outline two pathways for designing RAI business models and critically reflect their implications for managing the tensions that are innate within RAI's value proposition. Second, we argue that RAI systems' sociotechnical nature is reflected in the integrated business model, highlighting the criticality of the social in sociotechnical for RAI. Finally, we outline a research agenda for RAI business models.

5.1 Two pathways for designing responsible artificial intelligence business models

This study contributes to RAI business model design. By investigating approaches for developing RAI business models, we found that the business model literature offers two perspectives on business model development: AI and responsibility. This can be viewed as innovating responsible business models leveraging AI vs. RAI business models (see Table 1 in Section 2.3). While we consider both important, we focused on the second perspective (i.e., business models building on the value proposition of RAI). From this perspective, we proposed disentangling AI responsibility (see Section 2.1) and AI commercialization (see Section 2.2) into separate axes and investigating different paths in the resulting matrix to understand how organizations can design RAI business models. This positioning of responsibility and commercialization as two separate axes has two benefits. First, it allows us to combine the two problem areas discussed in the conceptual framework (RAI and the commercialization of technologies) into a single problem of commercializing RAI. Second, it allows a more elaborate view of the paradoxical tensions between commercial interests and responsibility, which were evident in our findings and in the existing literature (Asatiani et al., 2021; Eitel-Porter, 2021; Gasser & Almeida, 2017; Whittlestone et al., 2019) and within the focus group discussions. Figure 4 presents the resulting two-by-two matrix.

We suggest two pathways for designing RAI business models based on insights into the potential value proposition (see Figure 4). While the corner pathway (1) leverages the value proposition of AI itself to arrive at RAI business models, the direct pathway

(2) starts by developing RAI and then proceeds to commercialization. Both pathways start in the matrix's bottom half at a low level of AI commercialization. We argue that the pathways offer two approaches for addressing the paradoxical tensions in RAI systems' value proposition.

The corner pathway (1) takes a business model design approach in which a viable business model for AI is primary (upward arrow to quadrant IV). Subsequently, RAI requirements are introduced (rightward arrow to quadrant III). This pathway addresses the paradoxical nature of RAI's commercialization challenges by first emphasizing commercialization and then introducing responsibility. While this approach reduces the complexity of balancing the paradoxical tensions at the start, it poses the long-term challenge of integrating RAI into already commercialized business model designs without negatively affecting the level of commercialization. In a well-aligned business model (Solaimani et al., 2015), integrating the responsible use of AI could lead to cascading changes in requirements, or inertia could be encountered due to path dependencies. Thus, in this pathway, companies postpone managing the paradoxical tensions in RAI's value proposition (as discussed in Section 2.1) to a later path point to reap quick returns from commercializing AI. However, path dependencies may render these tensions salient or create new ones when companies integrate responsibility into their established business models.

In contrast, the direct pathway takes a business model design approach in which RAI is primary (arrow originating in quadrant II). In other words, it directly seeks commercialization of RAI through respective business models (upward arrow to quadrant III). This renders the design elements identified within this study key components in the early stages of business model design. While this pathway requires organizations to balance the paradoxical tensions in RAI systems' value propositions directly, it avoids path dependencies that could emerge on the corner path. However, the key challenges in this pathway are finding a convincing revenue model connected to RAI activities and continuously managing the paradoxical tensions. In fact, while organizations might find a balance that renders the tensions latent, new legislation, public conceptions of RAI, or technological advancements could re-render them salient.

Conceptually, the two pathways are equifinal, meaning that they both reach the same general outcome: an RAI business model. However, we argue that the two pathways present a critical consideration for managing the paradoxical tensions in RAI business model design. The corner pathway (1) prioritizes commercialization over responsibility. In contrast, the direct pathway (2) treats the two dimensions of commercialization and responsibility as equally important. This highlights that while the pathways may conceptually arrive at the same outcomes, they consider different priorities of design

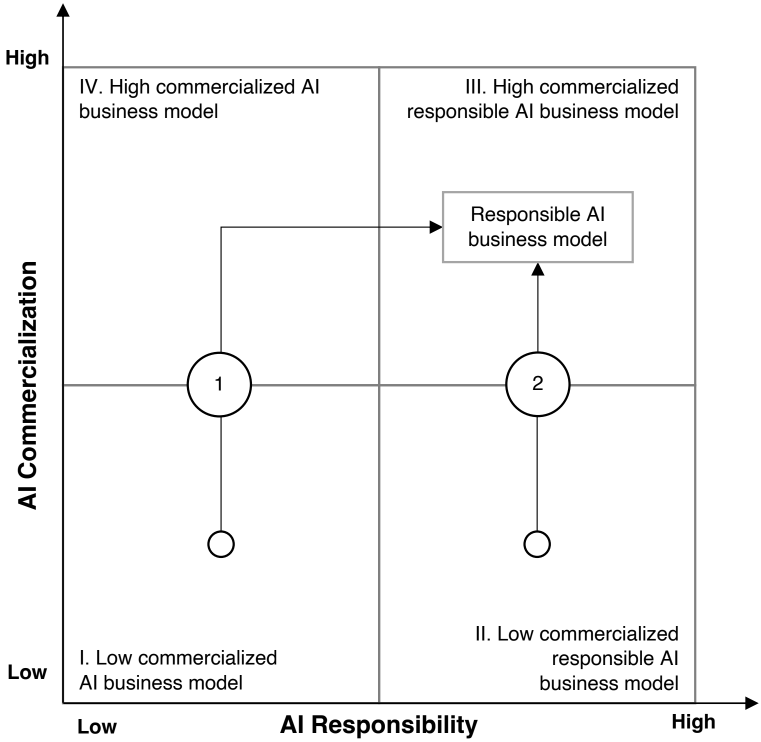

Figure 4. Two pathways for designing RAI business models

objectives in practice: monetary gains against ethical principles. We posit that regardless of the equifinal outcome (i.e., an RAI business model), these priorities introduce different design assumptions that underlie the design decisions along the two pathways. This means that whenever the paradoxical tensions in RAI business models become salient, these assumptions drive the decisions and actions that render them latent. Hence, we argue that the two pathways highlight the critical consideration of *which design objectives do we prioritize*. A consideration that qualitatively matters for how we manage the paradoxical tensions and the ensuing process and outcomes of RAI business model design.

5.2 The sociotechnical nature of responsible artificial intelligence

When defining RAI systems, we characterized them as sociotechnical systems (see Section 2.1). Our findings offer grounds for specifying this sociotechnical nature of RAI.

Hence, the study presents a contribution to the emerging literature on RAI and the technically oriented discussion on explainable AI. In Table 8, we relate the identified key activities to the value proposition of RAI and define the primary nature of these relations (i.e., technical or social). We focus on the design element of key activities since these express interactions between technical and social actors that constitute the value proposition of RAI. These interactions provide insights into which actions and relations render RAI systems responsible. We excluded the other design elements (i.e., potential customers and key partners), since these capture actors, not actions. The last column in Table 8 presents reflections on how these activities specify the sociotechnical nature of RAI systems.

Table 8 illustrates that the key activities for creating an RAI systems' value proposition are sociotechnical in nature. Apart from "technologically developing RAI," which is mainly technical, all activities are primarily social or feature social and technical aspects. This implies that although the value proposition of RAI is founded on a technological basis, it rests on social activities. In other words, the responsibility in RAI stems from technical implementations (e.g., explainable AI and transparent AI), but involves respective social implementations. These comprise crafting regulations that specify both technical and social responsibility criteria for AI. We argue for technical and social criteria since an AI system that fulfills the technical responsibility criteria can still be used irresponsibly (e.g., employee or work monitoring systems). Our findings suggest that responsibility in RAI stems primarily from social and primarily technical activities, such that technical responsibility is a necessary criterion for RAI and social responsibility is a sufficient criterion. Hence, while technical responsibility criteria are necessary for RAI, they are not sufficient for rendering AI systems RAI systems. For this, we also require RAI systems to fulfill the social responsibility criteria. We argue that this foregrounds the criticality of the social for the responsible in RAI.

The social involves activities that define responsibility *ex ante* (e.g., understanding the market needs for RAI), the development of RAI systems, and render it *ex post* (e.g., auditing RAI systems and raising awareness). Before developing RAI systems, we set the criteria that determine when AI is responsible. Subsequently, we draw on these criteria to examine whether AI is responsible in accordance with our technical and social criteria (i.e., whether we use AI responsibly). Hence, our findings suggest that the responsible in RAI rests on our *ex ante* socially constructed criteria, our *ex post* examination of these criteria, and our *ex post* use of RAI systems. All these activities are primarily social, highlighting the criticality of the social in sociotechnical. In other words, it highlights the importance of critically considering when and how we make AI responsible.

5.3 Research agenda on responsible artificial intelligence business model design

This study presents a starting point for future research. While we provide answers to our research aim, our learnings also point to new research issues (see Table 9). Hence, we contribute a research agenda on RAI business models and their design to the emerging literature on commercializing RAI. The learnings from the integrated business model (such as the need for future research on business model themes) and, more importantly, from the value creation network (or ecosystem) are implicitly shown in the integrated business model (cf., Minkkinen et al., 2021; Minkkinen, Zimmer, et al., 2022). Researchers could either extend the identified business model themes of providing and enabling RAI or reveal additional business model themes by studying specific industries or sectors. Such insights can help to abstract ideal types or archetypes (Baden-Fuller & Morgan, 2010) of RAI business models. While the findings on the RAI business model design elements offer a starting point, they also reveal potential research issues.

Identified key activities	*Related elements of RAI's value proposition*	*Primary sociotechnical nature of the activity*	*Reflections on the RAI definition*
Technologically developing RAI	• Implementing RAI systems in organizations or society. • Making AI explainable and transparent. • Developing tools and methods for RAI.	Technical	Social entities design technical entities to fulfill technical responsibility requirements.
Understanding the market needs for RAI	• Ensuring AI systems' societal acceptability. • Establishing a sustainable AI brand. • Managing AI risks. • Supporting organizations. that provide RAI • AI solutions increase customers' and stakeholders' quality of life.	Social	Social entities define both social and technical responsibility criteria for AI and ensure its responsible use.
Auditing RAI systems	• Monitoring AI systems' explainability and transparency. • Testifying the legal and regulatory compliance of AI systems. • Managing AI risks.	Technical Technical Social	Social entities assess technical entities and their use against, both social and technical responsibility criteria.
Raising awareness and lobbying	• Influencing AI regulations toward RAI. • Defining responsibility criteria for AI.	Social Sociotechnical	Social entities define and examine both technical and social responsibility criteria.

Table 8. RAI business model key activities and reflections on RAI systems' sociotechnical nature

Research topic	*Future research agenda*	*Potential research issues*
RAI business models	Investigating specific RAI business models to understand the underlying business model themes and identify RAI business model archetypes.	• How do specific RAI business model themes differ? • How does the critical consideration of design objectives reflect in business model themes? • What RAI business model archetypes emerge?
Pathways and development approaches	Exploring the suggested development approaches and the two pathways for designing RAI business models in practice.	• How do business models that leverage AI differ from RAI business models in practice? • How can organizations design RAI business models using the corner or direct pathway in practice? • What challenges emerge when managing the paradoxical tensions depending on the pathway? • How do the design objectives that underpin the pathways drive business model design? • How do organizations manage the paradoxical tensions in RAI's value proposition? • How do external influences (e.g., regulations and ethical principles of stakeholders) shape RAI business models?
Sociotechnical nature	Investigating the interplay of primarily technical and primarily social entities and actions in sociotechnical RAI systems and their respective responsibility criteria.	• How do existing technical and social responsibility criteria for AI relate? • How do audits evaluate both the technical and social responsibilities of AI systems?

Research topic	*Future research agenda*	*Potential research issues*
Value proposition	Exploring the value proposition specific to RAI. In particular, the paradoxical nature of the tension within this value proposition.	• How can organizations demarcate the value propositions of responsible AI from other AI solutions? • How do potential customers perceive the value proposition of RAI?
Potential customers	Exploring the customer landscape of RAI systems and how customers differ in terms of requirements and role.	• How does the value proposition vary depending on different customers? • How do RAI business models consider different societal groups (e.g., minorities)? • How can RAI business models scale from ethically aware customers?
Key activities	Understanding the processes and dynamics of societal awareness of RAI and RAI regulation, as well as the technological development of RAI systems.	• How can organizations raise awareness and of what among whom? • How can customers influence RAI regulation? • How can developers design flexible RAI systems that are adaptable to changing regulations and market needs? • How can policymakers reflect on the criticality of their regulations for when AI is responsible? • How can critical research contribute to raising awareness of RAI?

Research topic	*Future research agenda*	*Potential research issues*
Key partners	Shifting the focus from single firm business models to ecosystems of RAI to explore and understand the network of social and technical entities that jointly establishes the value proposition of RAI.	• How will the different key partners' business models complement each other when studied from an ecosystem perspective? • How can organizations leverage indirect market creation for RAI business models? • How do customers' multi-faceted roles (being customers and key partners simultaneously) affect RAI business model design? • How can organizations design value networks for cooperating with different enablers? • How can organizations cooperate with regulators to create incentives for RAI solutions?
Finances	Exploring the business model element of finances to identify potential earning logics that monetarize the RAI specific value proposition. Further, the identification of potential tensions between earning logics and responsibility.	• How can organizations successfully create a revenue stream from RAI systems' specific value propositions? • Which earning logics support the responsibility criteria in RAI systems? • How does monetarization differ between AI and RAI systems?

Table 9. Research agenda on RAI business models and their design

We conceptualized two perspectives on business model development, AI and responsibility: *innovating responsible business models leveraging AI* and *designing RAI business models*. These perspectives provide future scholars with a distinction between the phenomena of organizations using AI within responsible business models and the phenomena of organizations operating RAI business models. While the first presents opportunities for studying the use and benefits of AI systems in the context of responsible businesses, the latter emphasizes questions related to the practical implementation of ethical AI, transparent AI, or trustworthy AI principles. If we truly seek RAI systems, we need to study the latter. Researchers can build on these conceptualizations to investigate the different approaches. More critically, within the perspective of designing RAI business models, researchers can investigate the two pathways, meaning *corner* vs. *direct*, and determine how these pathways assist organizations in managing the paradoxical nature of the tensions within RAI's value proposition. This research avenue involves examining if and how the critical consideration of the underpinning design objectives entails qualitatively different RAI business models.

The sociotechnical nature of RAI systems poses the question of when RAI systems are responsible. Our findings suggest that the responsible in RAI primarily stems from the social. Nonetheless, we call for future research that studies whether the responsibility of RAI systems depends on primarily technical or primarily social activities. These questions outline a future research agenda for studying the relation between the responsibility criteria for the social and the technical of RAI systems. Lastly, we can infer future research agendas for the five design elements of business models. It should be noted that we include the element of "finances" despite (or because of) not addressing this aspect within the study.

This discussion and our research agenda (see Table 9) suggest that examining and designing RAI business models requires practitioners and scholars to take a critical stance toward the assumptions and activities that perform RAI systems. The criticality for managing the paradoxical tensions in RAI's value proposition and for turning RAI systems into viable products or services lies in our actions.

Notes

1. The expression 'the responsible in RAI systems' refers explicitly to the aspects that render RAI systems responsible in comparison to AI systems.
2. Defining RAI systems as sociotechnical systems (see Section 2.1), we, for analytically reasons, separate RAI systems into social and technical aspects. Emphasizing these aspects, we refer to the *socio*-technical as the social, and the socio-*technical* as the technical of RAI systems.

Acknowledgments

We thank the guest editors of the special issue and the anonymous reviewers for their guidance in shaping this research. Further, we acknowledge that this research was conducted within the Artificial Intelligence Governance and Auditing (AIGA) project funded by Business Finland. More information on the AIGA project can be found under: https://ai-governance.eu.

Bibliography

Accenture. (2018). *Realising the economic and societal potential of responsible AI in Europe*. Accenture. https://www.accenture.com/_acnmedia/PDF-74/Accenture-Realising-Economic-Societal-Potential-Responsible-Ai-Europe.pdf

Al-Debei, M. M., & Avison, D. (2010). Developing a unified framework of the business model concept. *European Journal of Information Systems*, *19*(3), 359-376. https://doi.org/10.1057/ejis.2010.21

Antonov, A., & Kerikmäe, T. (2020). Trustworthy AI as a Future Driver for Competitiveness and Social Change in the EU. In D. Ramiro Troitiño, T. Kerikmäe, R. M. de la Guardia, & G. Á. Pérez Sánchez (Eds.), *The EU in the 21st Century: Challenges and Opportunities for the European Integration Process* (pp. 135-154). Springer International Publishing. https://doi.org/10.1007/978-3-030-38399-2_9

Armour, J., & Sako, M. (2020). AI-enabled business models in legal services: From traditional law firms to next-generation law companies? *Journal of Professions and Organization*, *7*(1), 27-46. Scopus. https://doi.org/10.1093/jpo/joaa001

Asatiani, A., Malo, P., Rådberg Nagbøl, P., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, *22*(2), 325-352. https://doi.org/10.17705/1jais.00491

Baden-Fuller, C., & Morgan, M. S. (2010). Business Models as Models. *Long Range Planning*, *43*(2-3), 156-171. https://doi.org/10.1016/j.lrp.2010.02.005

Ballon, P. (2007). Business modelling revisited: The configuration of control and value. *Info*, *9*(5), 6-19. https://doi.org/10.1108/14636690710816417

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

Benjamins, R. (2021). A choices framework for the responsible use of AI. *AI and Ethics*, *1*(1), 49-53. https://doi.org/10.1007/s43681-020-00012-5

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *Management Information Systems Quarterly*, *45*(3), 1433-1450.

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research*, *26*(13), 1802-1811. https://doi.org/10.1177/1049732316654870

Blaschke, M., Cigaina, M., Riss, U. V., & Shoshan, I. (2016). Designing business models for the digital economy. In *Shaping the Digital Enterprise: Trends and Use Cases in Digital Innovation and Transformation* (pp. 121-136). Scopus. https://doi.org/10.1007/978-3-319-40967-2_6

Breidbach, C. F., & Maglio, P. (2020). Accountable algorithms? The ethical implications of data-driven business models. *Journal of Service Management*, *31*(2), 163-185. Scopus. https://doi.org/10.1108/JOSM-03-2019-0073

Burström, T., Parida, V., Lahti, T., & Wincent, J. (2021). AI-enabled business-model innovation and transformation in industrial ecosystems: A framework, model and outline for further research. *Journal of Business Research*, *127*, 85-95. https://doi.org/10.1016/j.jbusres.2021.01.016

Chebo, A. K., & Wubatie, Y. F. (2021). Commercialisation of technology through technology entrepreneurship: The role of strategic flexibility and strategic alliance. *Technology Analysis & Strategic Management*, *33*(4), 414-425. https://doi.org/10.1080/09537325.2020.1817367

Chesbrough, H. (2010). Business Model Innovation: Opportunities and Barriers. *Long Range Planning*, *43*(2), 354-363. https://doi.org/10.1016/j.lrp.2009.07.010

Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: Evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, *11*(3), 529-555. https://doi.org/10.1093/icc/11.3.529

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review*, *35*(4), 410-422. https://doi.org/10.1016/j.clsr.2019.04.007

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S.-M., Morison, J., Murphy, M. H., O'Brolchain, N., Schafer, B., & Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, *4*(2), 1-21. https://doi.org/10.1177/2053951717726554

Di Vaio, A., Boccia, F., Landriani, L., & Palladino, R. (2020). Artificial intelligence in the agri-food system: Rethinking sustainable business models in the COVID-19 scenario. *Sustainability (Switzerland)*, *12*(12). Scopus. https://doi.org/10.3390/SU12124851

Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, *121*, 283-314. Scopus. https://doi.org/10.1016/j.jbusres.2020.08.019

Dignum, V. (2017). Responsible Autonomy. *Electronic Proceedings of IJCAI 2017*, 4698-4704. https://www.ijcai.org/proceedings/2017/655

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use ai in a responsible way*. Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6

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

Dmitriev, V., Simmons, G., Truong, Y., Palmer, M., & Schneckenberg, D. (2014). An exploration of business model development in the commercialization of technology innovations: Business model development in technology innovations. *R&D Management*, *44*(3), 306-321. https://doi.org/10.1111/radm.12066

Doganova, L., & Eyquem-Renault, M. (2009). What do business models do? *Research Policy*, *38*(10), 1559-1570. https://doi.org/10.1016/j.respol.2009.08.002

Eitel-Porter, R. (2021). Beyond the promise: Implementing ethical AI. *AI and Ethics*, *1*(1), 73-80. https://doi.org/10.1007/s43681-020-00011-6

Eldred, E. W., & McGrath, M. E. (1997). Commercializing New Technology—I. *Research-Technology Management*, *40*(1), 41-47. https://doi.org/10.1080/08956308.1997.11671102

European Commission. (2018). *Coordinated Plan on Artificial Intelligence*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:795:FIN

European Commission. (2020). WHITE PAPER On Artificial Intelligence—A European approach to excellence and trust.

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, *1*(6), 261-262. https://doi.org/10.1038/s42256-019-0055-y

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Fountaine, T., McCarthy, B., & Saleh, T. (2019, July 1). Building the AI-Powered Organization. *Harvard Business Review*. https://hbr.org/2019/07/building-the-ai-powered-organization

Garbuio, M., & Lin, N. (2019). Artificial Intelligence as a Growth Engine for Health Care Startups: Emerging Business Models. *California Management Review*, *61*(2), 59-83. https://doi.org/10.1177/0008125618811931

Gasser, U., & Almeida, V. A. F. (2017). A layered model for AI governance. *IEEE Internet Computing*, *21*(6), 58-62. https://doi.org/10.1109/MIC.2017.4180835

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99-120. https://doi.org/10.1007/s11023-020-09517-8

Hedman, J., & Kalling, T. (2003). The business model concept: Theoretical underpinnings and empirical illustrations. *European Journal of Information Systems*, *12*(1), 49-59. https://doi.org/10.1057/palgrave.ejis.3000446

High-Level Expert Group on Artificial Intelligence. (2019a). *A definition of AI: Main capabilities and disciplines*.

High-Level Expert Group on Artificial Intelligence. (2019b). *Ethics Guidelines for Trustworthy AI*. European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2

Joshi, M. P., Su, N., Austin, R. D., & Sundaram, A. K. (2021). Why So Many Data Science Projects Fail to Deliver. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/why-so-many-data-science-projects-fail-to-deliver/

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

Keen, P., & Qureshi, S. (2006). Organizational transformation through business models: A framework for business model design. 8. Scopus. https://doi.org/10.1109/HICSS.2006.376

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

Laasch, O., & Pinkse, J. (2020). Explaining the leopards' spots: Responsibility-embedding in business model artefacts across spaces of institutional complexity. *Long Range Planning*, *53*(4), 101891. https://doi.org/10.1016/j.lrp.2019.101891

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.

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*. https://doi.org/10.1108/INTR-08-2021-0600

Laird, I., & Sjoblom, L. (2004). Commercializing technology: Why is it so difficult to be disciplined? *Business Horizons*, *47*(1), 65-71. https://doi.org/10.1016/j.bushor.2003.11.010

Lee, J., Suh, T., Roy, D., & Baucus, M. (2019). Emerging technology and business model innovation: The case of artificial intelligence. *Journal of Open Innovation: Technology, Market, and Complexity*, *5*(3). Scopus. https://doi.org/10.3390/joitmc5030044

Lewis, M. W., & Smith, W. K. (2014). Paradox as a Metatheoretical Perspective: Sharpening the Focus and Widening the Scope. *The Journal of Applied Behavioral Science*, *50*(2), 127-149. https://doi.org/10.1177/0021886314522322

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

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022b). *Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance* (arXiv:2206.00335). arXiv. https://doi.org/10.48550/arXiv.2206.00335

Markman, G. D., Siegel, D. S., & Wright, M. (2008). Research and Technology Commercialization. *Journal of Management Studies*, *45*(8), 1401-1423. https://doi.org/10.1111/j.1467-6486.2008.00803.x

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management*, *39*(1), 53-63. https://doi.org/10.1080/10580530.2020.1849465

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

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, *3*(2). https://doi.org/10.1177/2053951716679679

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., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds and Machines*. https://doi.org/10.1007/s11023-021-09563-w

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

Nenonen, S., & Storbacka, K. (2010). Business model design: Conceptualizing networked value co-creation. *International Journal of Quality and Service Sciences*, *2*(1), 43-59. https://doi.org/10.1108/17566691011026595

Osterwalder, A., Pigneur, Y., & Clark, T. (2010). Business model generation: A handbook for visionaries, game changers, and challengers. Wiley.

Osterwalder, A., Pigneur, Y., & Tucci, C. L. (2005). Clarifying Business Models: Origins, Present, and Future of the Concept. *Communications of the Association for Information Systems*, *16*(1). https://doi.org/10.17705/1CAIS.01601

PwC. (2019). *A practical guide to Responsible Artificial Intelligence (AI)* (p. 20). https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. *MIT Sloan Management Review*, *59*(1).

Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., Delipetrev, B., European Commission, & Joint Research Centre. (2020). *AI watch: Defining Artificial Intelligence: towards an operational definition and taxonomy of artificial intelligence.* op.europa.eu/publication/manifestation_identifier/PUB_KJNA30117ENN

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153-158. https://doi.org/10.1145/3375627.3375804

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/

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *Acm Transactions on Interactive Intelligent Systems*, *10*(4), 26. https://doi.org/10.1145/3419764

Sjödin, D., Parida, V., Palmié, M., & Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. *Journal of Business Research*, *134*, 574-587. Scopus. https://doi.org/10.1016/j.jbusres.2021.05.009

Smith, W. K., Binns, A., & Tushman, M. L. (2010). Complex Business Models: Managing Strategic Paradoxes Simultaneously. *Long Range Planning*, *43*(2-3), 448-461. https://doi.org/10.1016/j.lrp.2009.12.003

Smith, W. K., & Lewis, M. W. (2011). Toward a Theory of Paradox: A Dynamic Equilibrium Model of Organizing. *Academy of Management Review*, *36*(2), 381-403. https://doi.org/10.5465/AMR.2011.59330958

Solaimani, S., Bouwman, H., & Itälä, T. (2015). Networked enterprise business model alignment: A case study on smart living. *Information Systems Frontiers*, *17*(4), 871-887. https://doi.org/10.1007/s10796-013-9474-1

Stewart, D., Shamdasani, P., & Rook, D. (2007). *Focus Groups*. SAGE Publications, Ltd. https://doi.org/10.4135/9781412991841

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, *36*(2), 368-383. https://doi.org/10.1016/j.giq.2018.09.008

Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Using AI to Enhance Business Operations. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/using-ai-to-enhance-business-operations/

Taylor, S., Pickering, B., Boniface, M., Anderson, M., Danks, D., Følstad, A., Leese, M., Müller, V., Sorell, T., Winfield, A., & Woollard, F. (2018). *Responsible AI—Key Themes, Concerns & Recommendations For European Research And Innovation*. Zenodo. https://doi.org/10.5281/ZENODO.1303252

Teece, D. J. (2010). Business Models, Business Strategy and Innovation. *Long Range Planning*, *43*(2-3), 172-194. https://doi.org/10.1016/j.lrp.2009.07.003

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, *31*(2), 447-464. https://doi.org/10.1007/s12525-020-00441-4

Toniolo, K., Masiero, E., Massaro, M., & Bagnoli, C. (2020). Sustainable Business Models and Artificial Intelligence: Opportunities and Challenges. *Contributions to Management Science*, 103-117. Scopus. https://doi.org/10.1007/978-3-030-40390-4_8

Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. (2021). Responsible AI for digital health: A synthesis and a research agenda. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10146-4

Turetken, O., & Grefen, P. (2017). *Designing service-dominant business models*. 2218-2233. Scopus.

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2021). Framing governance for a contested emerging technology: Insights from AI policy. *Policy and Society*, *40*(2), 158-177. https://doi.org/10.1080/14494035.2020.1855800

van Marrewijk, M. (2003). Concepts and Definitions of CSR and Corporate Sustainability: Between Agency and Communion. *Journal of Business Ethics*, *44*(2), 95-105. https://doi.org/10.1023/A:1023331212247

Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, *26*(7), 1893-1924. Scopus. https://doi.org/10.1108/BPMJ-10-2019-0411

Wang, Y., Xiong, M., & Olya, H. (2020). *Toward an Understanding of Responsible Artificial Intelligence Practices*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2020.610

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195-200. https://doi.org/10.1145/3306618.3314289

Wirtz, B. W., Pistoia, A., Ullrich, S., & Göttel, V. (2016). Business Models: Origin, Development and Future Research Perspectives. *Long Range Planning*, *49*(1), 36-54. https://doi.org/10.1016/j.lrp.2015.04.001

Zaki, M. (2019). Digital transformation: Harnessing digital technologies for the next generation of services. *Journal of Services Marketing*, *33*(4), 429-435. Scopus. https://doi.org/10.1108/JSM-01-2019-0034

Zott, C., & Amit, R. (2010). Business Model Design: An Activity System Perspective. *Long Range Planning*, *43*(2), 216-226. https://doi.org/10.1016/j.lrp.2009.07.004

Appendix

This appendix contains Table 10, which presents an excerpt from the preliminary findings table produced in the first stage of the research process.

	Focus group #1	…	*Focus Group #5*	*Findings per design element (comparative analysis)*
Value proposition	• AI solutions stand the test of time; comply with current and future regulation • RAI certification creates trust …	…	• Legal and ethical risk management for AI • AI services make decisions; its about selling decisions	• Risk management • Future oriented, i.e., robust against future regulations. • RAI sells responsible decisions …
Potential customers	• Consumers, citizens, endusers • Organizations operating AI …	…	• Chain of customers • Organizations that order or build AI …	• New auditing market emerging • Multifacetted customer roles (e.g., customer & key partner) …
Key activities	• Governance of AI solutions • Developing RAI systems …	…	• Lobbying for RAI solutions • Develop audit framework …	• Focus on developing RAI • Growing or raising awareness …
Key partners	• Recipients: Customers, etc. • Auditing firms …	…	• Regulators, NGOs • Customers and end-users …	• Recipients as key partners: Novel ecosystem role …
Findings per focus group (individual analysis)	• Trust a central role in value proposition • Key activities focus on developing and auditing	…	RAI creates auditing market Key activities involve customers and partners …	

Table 10. Excerpt of the preliminary findings table.