University of Reading

Scratching the surface of responsible AI in financial services: a qualitative study on non-technical challenges and the role of corporate digital responsibility

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Skouloudis, A. ORCID: https://orcid.org/0000-0002-3363-6692 and Venkatraman, A. (2025) Scratching the surface of responsible AI in financial services: a qualitative study on non-technical challenges and the role of corporate digital responsibility. AI, 6 (8). 169. ISSN 2673-2688 doi: 10.3390/ai6080169 Available at https://centaur.reading.ac.uk/123858/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.3390/ai6080169

Publisher: MDPI AG

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

AI

MDPI

Article

Scratching the Surface of Responsible AI in Financial Services: A Qualitative Study on Non-Technical Challenges and the Role of Corporate Digital Responsibility

Antonis Skouloudis [1,*] and Archana Venkatraman [2,3]

[1] Department of Environment, University of the Aegean, 81132 Lesvos, Greece
[2] Henley Business School, University of Reading, Henley-on-Thames RG9 3AU, UK
[3] IDC Europe, London W1T 2RE, UK
* Correspondence: skouloudis@env.aegean.gr

Abstract

Artificial Intelligence (AI) and Generative AI are transformative yet double-edged technologies with evolving risks. While research emphasises trustworthy, fair, and responsible AI by focusing on its "what" and "why," it overlooks practical "how." To bridge this gap in financial services, an industry at the forefront of AI adoption, this study employs a qualitative approach grounded in existing Responsible AI and Corporate Digital Responsibility (CDR) frameworks. Through thematic analysis of 15 semi-structured interviews conducted with professionals working in finance, we illuminate nine non-technical barriers that practitioners face, such as sustainability challenges, trade-off balancing, stakeholder management, and human interaction, noting that GenAI concerns now eclipse general AI issues. CDR practitioners adopt a more human-centric stance, emphasising consensus-building and "no margin for error." Our findings offer actionable guidance for more responsible AI strategies and enrich academic debates on Responsible AI and AI-CDR symbiosis.

Keywords: artificial intelligence (AI); generative AI (GenAI); corporate digital responsibility (CDR); AI and ethics; GenAI in financial services; responsible AI; qualitative study

check for updates

Academic Editor: Isidoros Perikos

Received: 19 May 2025
Revised: 9 July 2025
Accepted: 21 July 2025
Published: 28 July 2025

Citation: Skouloudis, A.; Venkatraman, A. Scratching the Surface of Responsible AI in Financial Services: A Qualitative Study on Non-Technical Challenges and the Role of Corporate Digital Responsibility. *AI* **2025**, *6*, 169. https://doi.org/10.3390/ai6080169

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial Intelligence (AI) and Generative AI (GenAI) promise transformative benefits by accelerating decision-making, automating routine tasks, and unlocking novel products and services. Yet, their rapid, pervasive adoption has outpaced our understanding of how to govern them responsibly. High-profile mishaps and regulatory scrutiny have underscored a widening "theory-to-practice" gap in Responsible AI (RAI): while scholars and industry bodies articulate principles of fairness, transparency, and accountability [1,2], detailed guidance on operationalising such ideals remains thin on the ground, particularly in highly regulated sectors such as that of financial services.

The paper seeks to address such a gap by examining two intertwined research questions: (a) What non-technical barriers inhibit the practical translation of RAI principles into routine governance and operations within financial services organisations? and (b) How can Corporate Digital Responsibility (CDR) frameworks be leveraged to bridge the distance between high-level RAI guidelines and the real-world challenges faced by practitioners in financial institutions? To address these questions, the paper's specific objectives are (a) to conduct semi-structured interviews with AI/GenAI stakeholders across European financial institutions to elicit and map non-technical obstacles that reflect areas of alignment, tension,

or omission between theoretical prescriptions and practitioners' lived experiences, (b) to explore whether CDR orientation (as a set of core values, expected behaviours and ethical guidelines) can inform actionable governance mechanisms, decision-making processes, and accountability structures for AI/GenAI deployment, and (c) add to academic discourse and practitioner debates by laying the ground for further research on the identified barriers, cross-industry comparisons, and longitudinal assessments of AI governance efficacy along with targeted strategies to mitigate barriers and accelerate responsible adoption of AI/GenAI.

Drawing from 15 in-depth interviews with AI executives, risk officers, and CDR practitioners, this is the first study that attempts to systematically identify categories of organisational, cultural, and human-centric barriers that impede RAI implementation. We then examine how a CDR orientation—rooted in ethical data stewardship, stakeholder engagement, and sustainability—can help translate high-level RAI principles into actionable governance practices. In this respect, our analysis reveals three core findings: (1) practitioners perceive GenAI risks as more urgent than traditional AI challenges; (2) nine distinct non-technical barriers—from role ambiguity and legacy processes to human-factor and budget constraints—undermine RAI adoption; and (3) CDR practitioners adopt a more human-centric, consensus-driven approach to RAI, emphasising shared values such as "no margin for error" and placing trust and purpose at the centre of governance.

By integrating such insights, our contributions are threefold. First, we extend the Responsible AI literature [3,4] by empirically validating and richly describing organisational and cultural impediments that transcend technical concerns. Second, we advance CDR scholarship [5,6] by demonstrating its relevance to AI governance and outlining how CDR dimensions could be mapped onto RAI mechanisms. Third, we offer a practical taxonomy of nine non-technical barriers, equipping practitioners with a diagnostic lens for designing holistic governance protocols and laying the groundwork for future quantitative validation and cross-industry comparison.

The remainder of the paper is structured as follows. Section 2 reviews the literature on Responsible AI and Corporate Digital Responsibility, highlighting the distance between high-level principles and practitioner needs. Section 3 outlines the qualitative methodology and sampling approach. Section 4 presents our thematic analysis findings on non-technical barriers and the mediating role of CDR. Section 5 discusses implications for theory and practice and concludes with limitations and directions for future research.

2. Background

2.1. AI/GenAI: New Directions for the Financial Sector

AI and GenAI are rapidly adopted by financial services, transforming them like never before [7]. Within the financial services industry alone, AI is used across multiple business streams and use cases, including improved fraud detection, more accurate credit scoring, efficient investment management decisions, personalised customer service experiences, and effective risk mitigation with improved, insight-driven risk management. For instance, HSBC, one of the world's largest banks, has embraced AI solutions for common use cases such as processing time minimisation or improving customer experience and is doubling down on employing AI to combat financial crime and money laundering [8]. It has now developed and deployed an AI algorithm that can detect "suspicious activity on its own, without us telling it what to look for", as Richard D. May (HSBC's Group Head of Financial Crime, Global Banking & Markets and Commercial Banking) indicated in November 2023. With AI, HSBC is checking 1.35 billion transactions for signs of financial crime every month across 40 million customer accounts with dramatic positive impact and reports 60% fewer false-positive cases, i.e., an industry term for when transactions are incorrectly flagged as

suspicious by AI. False positives pose a burden on resources and also reduce the confidence in the system's quality of output because every time there is an alert of suspicious behaviour, investigators have to manually review it. Thus, improving the predictive power in risk detection with AI has become a game-changer for HSBC to address financial crime.

The market size of AI in finance is estimated to record a 25.7% CAGR, growing from USD 7.3 billion in 2021 to over USD 22.6 billion by 2026. Within this sector, the use of AI in the banking market alone is projected to quadruple from USD 6.8 billion in 2022 to over USD 27 billion by 2027 (Bhatnagar and Mahant [9]: 446). Such growth is not restricted to large financial institutions but all entities across the financial sector, including capital markets, retail banking, fintechs, and insurance providers, which will be using AI even more in the future, indicating "the pervasive nature of AI" (Bhatnagar and Mahant [9]: 441). At the same time, GenAI (i.e., AI that can auto-generate responses by analysing vast datasets) can bring further improvements and innovation opportunities [10]. GenAI can generate content in the form of text, images, videos, or audio material by learning from existing data to accelerate knowledge management, decision-making, and the creation of new content [11]. Within the banking sector, GenAI is "*paving the way for previously inconceivable innovations*" (Botunac et al. [12]: 4). Supporting arguments for this claim can also be found in the McKinsey Global Institute's report, where the prediction that GenAI can yield up to USD 4.4 trillion/year in value across all industries is made, with financial services expected to have "one of the largest opportunities" [13].

Indeed, the future of finance infused with AI and GenAI is not far, and the underlying negative implications they may have are topics of continuous debate and investigation by industry members, policymakers, and academia. While McCarthy's AI definition from 2007 has a positive connotation to capture the imagination of users to employ AI for organisational benefits [14], today's extensive use of intelligent systems as well as the early negative consequences (e.g., see Twitter's chatbot Tay, Microsoft's Zo, or Google's Gemini AI failures of offensive content) have sparked AI practitioners and enthusiasts to better understand its 'sinister' side and unintended consequences [2,15].

As we stand at the brink of entering an AI-infused world, ensuring a clear practical approach for organisations to adopt AI and GenAI responsibly becomes a huge imperative. Almost every discourse on AI and GenAI highlights the challenges, risks, and implications, urging caution and for users to adopt AI responsibly to ensure they capitalise on the benefits that AI can bring in while doing so without undermining data privacy, ethics, transparency, or accountability and also by addressing the potential bias in the system. Thus enters the discourse on RAI.

2.2. The Notion of Responsible AI—From Theory to Practice

Responsible AI, in its essence, is a formal approach to guide the implementation of AI techniques at scale by focusing on systems' fairness, models' explainability, and organisations' accountability [4]. It is an encompassing concept that "*imposes the systematic adoption of several AI principles for AI models to be of practical use*" (Arrieta et al. [4]: 84). RAI guidelines not only focus on the explainability of AI models (i.e., technical capability where it is possible to trace and explain how models make predictions and decisions) but also include considerations around fairness, accountability, and privacy. While RAI has become a commonly accepted direction to guide AI deployment, the challenge is that there is an overwhelming amount of principles, guidelines, and emerging regulations (such as the EU AI Act) to warn adopters of the potential challenges, but limited practical guidance is available. Take RAI principles alone; so far, 38 academic studies explore RAI constituents [3] that also apply to GenAI [1,16]. Fjeld et al. [1] identify 47 principles such as accountability, responsibility, transparency, security, privacy, and fairness, without "unanimity" among adopters to

give a clear pathway for new organisations to espouse. Moreover, academics observe that although there is a recent wave of studies on RAI, practical guidance or relevant research at the organisational, business, or individual levels remains scarce (Mikalef et al. [2]: 259). Scholars call for studies into making RAI actionable, indicating that growing AI adoption is only resulting in growing confusion and complexity on how to actually implement RAI practices [3]. The literature highlights a "*thorny gap*" [1] between high-level concepts and real-world application. There is a strong argument that while the detailed discourse on identifying and defining RAI principles is essential, much of the emphasis and focus is "*quite high-level and abstract and do not provide much guidance for practitioners regarding the realities of deploying AI in practice*" (Mikalef et al. [2]: 266). This clearly underlines the importance of thinking practically about RAI frameworks. Relevant examples of gaps between theory and practice are well-articulated in recent studies, including how RAI principles can be vague, open to multiple interpretations, and hard to apply to everyday practices [17]; how practitioners face challenges to turn "*theoretical understanding of potential inequities into concrete action*" [18]; and the challenging task of precisely articulating how organisations frame actions vis-à-vis governance (Elliot et al., 2021 [5]). In this respect, Mikalef et al. [2], among others, also call for more research and evidence-based insights to improve our knowledge on the barriers to RAI.

2.3. Beyond the Techno-Legal Domains: Barriers in the Practical Implementation of Responsible AI

Responsible AI is multi-faceted, covering different aspects, as those highlighted by Merhi [3] (Figure 1). When many of these dimensions are likely to be affected by an individual organisation's approaches, culture, processes, skills, motives, maturity, access to resources, industry regulations, and more, how can a theoretical-technical focus alone suffice in successful practice of RAI? Highlighting the future of financial institutions, researchers maintain that the significance of AI, LLMs, and GenAI is undeniable but requires strong collaboration between technology expertise and human experience to shape the sustainable transformation of the industry [19]. Yet, the literature highlights a chasm between academic research priority when it comes to AI governance and what the real-world practitioners actually require. Managers and practitioners want "*organizational tactics and stakeholder management rather than technical methods alone*" (Rakova et al. [18]: 2). Rakova et al. [18], in fact, highlight an underlying tension between the academic and practitioners' domains where theory is disapproved by practitioners working in corporate terrains. Thus, there is an emerging need for organisation-level frameworks with aligned decision-making, coordination, and communication—all of which are a far cry from technical frameworks. In a similar vein, in their study, Schiff et al. [17] argue that to truly understand RAI in greater depth and scrutiny, experts need to expand their scope of attention to include a wider spectrum of issues influencing both social and human welfare. They call for practitioners, policymakers, and academics to go beyond the "narrow set of topics such as bias, transparency, privacy, or safety", avoid treating them as non-interdependent, siloed concepts, and "instead, the full range of topics and their complex interdependencies needs to be understood" even though "*(. . .) such a task can be enormously difficult*" (Schiff et al. [17]: 3).

The primary focus for a wide range of key publications on RAI is on technical challenges (tools, security, interpretability, explainability, and/or ways to address model bias, etc.), privacy risks and regulatory implications, or recommendations to address technical issues and legal compliance. In contrast, limited information exists on non-technical barriers even when researchers argue that the ease-of-implementation, culture, and organisational 'macro-motivators' drive RAI [20]. Under such a persistent lack of a comprehensive view or practical guidance and real-world contexts, RAI principles will risk amounting to nothing more than just being "claims" (Schiff et al. [17]: 2), while relevant practical barriers

will persist partially due to the limited exploration of respective non-technical barriers in materialising RAI.

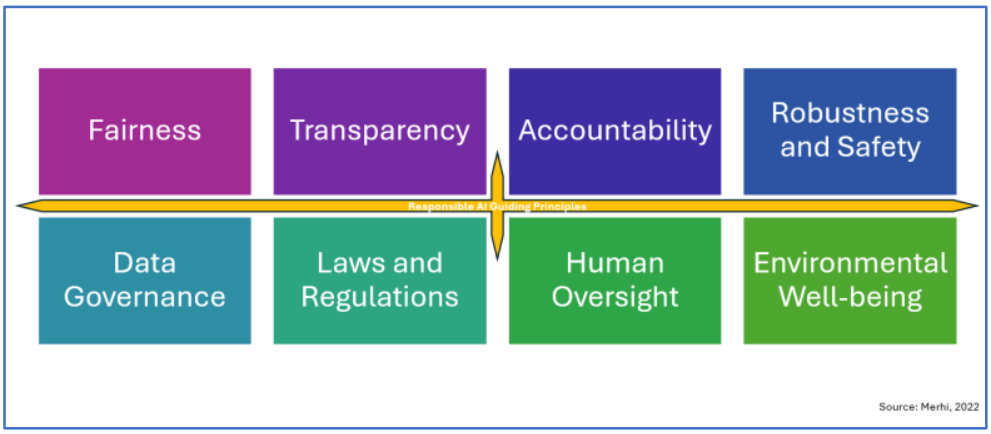

Figure 1. Principles of Responsible AI [3].

The emphasis on relevant organisational-level/micro-level challenges is highlighted in the literature; for instance, Arrietta et al. [4] indicate that implementing RAI in organisational settings pertains to a balancing act between the "*major cultural and organisational changes needed to enforce such principles over processes endowed with AI functionalities*", and the readiness of "*IT assets, policies and resources*" to deploy it (Arrieta et al. [4]: 108). Arrieta et al. go on to characterise it as "*(. . .) the gradual process of rising corporate awareness around the principles and values of Responsible AI*" (p. 108) where success rests. In their book on the 'AI Dilemma', Powell and Kleiner [21] poignantly argue that the challenge is hardly about the ability of the machines to learn and self-generate content, but it is actually about the human ability to manage the growing abilities of these systems. They emphasise the value of non-technical capabilities such as raising one's own awareness and abilities and taking the responsibility seriously to have robust control, not 'illusory' control. Similarly, Merhi [3] identifies eleven critical barriers to RAI, categorises them in three major groups—Technological, Organisational, and Environmental (TOE)—and examines them under the lens of the TOE framework (first introduced by Tornatzky et al. [22]), highlighting "*how the context of an organization can play a determinant role*" in the adoption of a new technology. Merhi's research and analysis assigned maximum weight to the technological barriers, but this study includes caveats that the barriers identified are by no means exhaustive given the dearth of studies on RAI barriers. The author further highlights that many of the systematic studies focus on the healthcare sector and call for further "*(. . .) quantitative and qualitative papers that collect primary data on barriers" that "impact Responsible AI in other industries*" (Merhi [3]: 1157).

Similarly, the World Economic Forum [23] highlights how prioritising responsible design and use of GenAI systems from an early point is crucial to ensure a "positive future" and that "AI serves as a force for good" (WEF [23]: 2). To this effect, it formed an AI Governance Alliance (AIGA), whose members met at a summit to assess and provide guidance for AI use. Post this 'Responsible AI Leadership: A Global Summit on Generative AI' in April 2023 in Presidio, USA, WEF's AIGA released "30 action-oriented recommendations aimed at guiding generative AI towards meaningful human progress" (WEF [23]: 2). Rooted in WEF's human-centric AI vision, the recommendations were categorised into 3 themes, spanning the full GenAI life cycle (Figure 2).

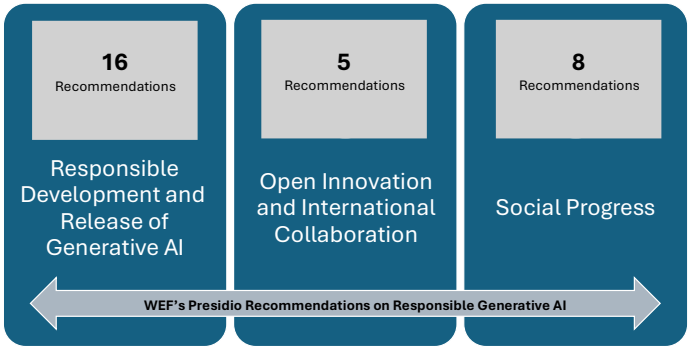

Figure 2. WEF's Responsible Gen AI Recommendations based on the Presidio Summit and in partnership with AI Commons [23].

In June 2024, WEF also introduced a PRISM framework that builds on the Presidio framework. Its purpose is to guide organisations and investors through the "nuanced landscape" of AI (WEF [24]: 4). The PRISM framework emphasises the focus on non-technical barriers and "*stresses the importance of organizational readiness over mere technological capability*" (WEF [24]: 3) and calls for "*active engagement*" between AI stakeholders and social innovators to "*jointly enable the ethical adoption of AI for positive impact*" (WEF [24]: 3). WEF's framework, based on insights from numerous social innovators and experts, stresses non-technical capabilities in implementing AI successfully since "*internal preparedness often outweighs technological and data considerations*" (WEF [24]: 8).

It is evident that when a fundamentally transformational technology (compared to anything before it), such as (Gen)AI, is deployed, conventional approaches such as external enforcements through command-and-control regulations or fines can become inefficient instruments as organisations are likely to engage in mere check-box exercises. Some refer to this as the temptation of solutionism, i.e., the "*belief that complex socio-technical and political problems can be 'solved' (or avoided) by the introduction of new techniques*" when exploring responsibility gaps in AI deployment (Santoni de Sio and Mecacci [25]: 1072). Thus, what is required is a holistic, non-technical-orientated view beyond the realms of technology and/or legality and more into ethics, morals, and philosophy perspectives in embracing AI responsibly at a practical level. In sum, experts call for new modes of thinking; for instance, such profound perspectives can be found in Bietti [26], who investigates tech ethics and, using evidence-based arguments, indicates a tectonic shift in technology, with technology's carbon footprint, industry movements and whistleblowing, flurry of regulations, and geopolitical tensions around technological behemoths, among others, necessitating a much more deliberate, comprehensive, and rich investigation in order to establish what "*the tech companies and stakeholders owe to humans, to animals and to the planet*" (Bietti [26]: 218). The so far limited resolution of non-technical barriers as well as recommended actions on exploring the pivotal role of organisations, culture, and corporate awareness make it a worthwhile research endeavour to explore them through the lens of Corporate Digital Responsibility (CDR), which is already adopted by some financial institutions and can potentially assist in translating RAI's theory into practice.

2.4. Corporate Digital Responsibility as a Facilitator of Responsible AI

Scholars have examined whether Corporate Digital Responsibility can serve as a "*mechanism to demystify governance complexity*" [5] but call for further research to disambiguate its critical role and contribution. As a broader set of ethical principles and actions, CDR encompasses active human oversight and responsible technology use, including AI systems and techniques [27]. While there are several attempts to frame CDR, one commonly adopted definition is to view it as a "*set of shared values and norms guiding an organisation's*

operations" when it comes to the four key processes related to data and digital technologies (Lobschat et al. [27]: 875): development of technology and data capture, operations and decision-making, inspection-governance and impact assessment, and continuous refinement of technology and data. In sum, CDR offers a valuable framework to guide business leaders' judgement and options when it comes to digital technologies, drawing from the notion of business ethics while being distinct from corporate responsibility, which is now becoming obsolete and is less focused on digital technology per se.

Being a relatively modern corporate concept, it is fast gaining traction, especially among businesses in Europe [28], partially due to the EC's recent directives and emerging policy directions as well as European businesses' tendency to take a governance-, trust-, and ethics-orientated approach, especially to enterprise technology. Back in 2018, the German Federal Ministry of Justice and Consumer Protection launched the CDR initiative in cooperation with large European organisations, including leading financial services entities, such as ING, DKB, and Barmer, among others. The participants sought to actively engage in shaping the digital future with society's best interests at heart, going beyond legal compliance to taking voluntary actions on material issues describing digital responsibility. Currently, the Office of CDR Initiative is operated by the Consumer Policy and is focused on clearly understanding the risks and opportunities of digital technology, including AI, contributing to hard-to-answer but pressing questions such as "*how does artificial intelligence actually works and when might the use of AI lead to disadvantages for us?*" [29].

Recent research explores a symbiosis between CDR and AI ethics [6]. In their study, Olatoye et al. explore such a symbiotic relationship between ethical governance (driven by CDR) and technology, particularly around RAI, and call such a connection inseparable. They focus on CDR for developing an ethical framework required for RAI. Such ethics-orientated frameworks can take RAI beyond technical and legal perspectives to identify emerging barriers and encourage individual (i.e., corporate-level) accountability and responsibility. In this respect, Elliot et al. [5] find more than 160 ethical AI principles and stress that these tend to confuse rather than guide organisations given the lack of harmonisation and/or alignment in the implementation approaches. According to these researchers, trust and purpose—which are essential tenets for financial services organisations for appropriately using data and digital technologies—must be codified to make them implementable across scales. Elliot et al. attempt to harmonise the approaches while at the same time, raising awareness on the role and importance of CDR as a mechanism to "*demystify governance complexity and to establish an equitable digital society*" (Elliot et al. [5]: 179).

Special emphasis is placed on CDR in fintechs and financial institutions driven by AI, where it can act as a potential framework to navigate through complexities of RAI and as a means of codifying trust, which then becomes practically valuable; CDR is seen to have the potential to assist financial services in embedding "*purpose and trust*" and in driving RAI adoption (Elliot et al. [5]: 185). This is particularly critical as, according to an Edelman survey of trusted sectors between 2015 and 2019, the financial services sector emerges as the least trusted sector [30]. If financial services are bracing to reap the potential rewards of the AI and GenAI capabilities, then they need to do it in a trustworthy manner or face risks of alienation from customers. Trust, transparency, ethics, purpose, and a culture of accountability cannot be embedded merely through externally dictated technical or legal enforcements but can be better addressed by understanding non-technical (employee-related, procedural-operational, organisational, and environmental) challenges and putting into practice robust AI governance mechanisms. In this respect, CDR is more holistic at an organisational level, as illustrated in Figure 3, encompassing three dimensions with trust and purpose firmly sitting in the middle, making it ever more crucial in the AI-infused

era: managing the environmental impact of technology, ensuring societal well-being, and promoting economic transparency.

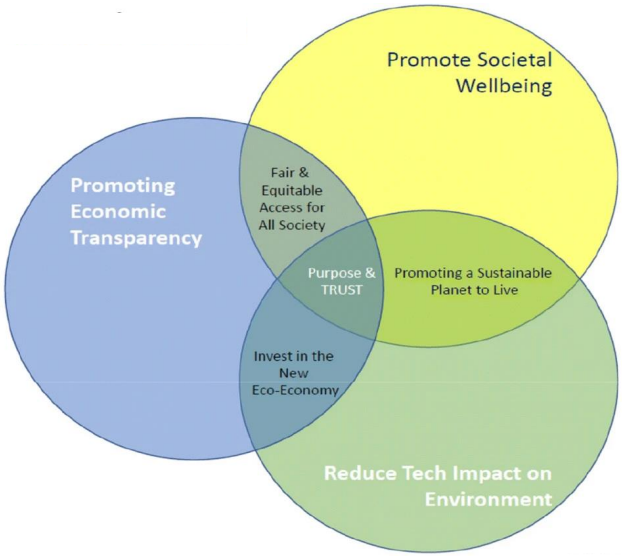

Figure 3. Central tenets of CDR [5].

Another critical point that indicates the synergies and value of CDR in making RAI practical is that CDR necessitates robust data management and data responsibility, as data is at the core of AI and GenAI systems. As Dhake et al. [19] relevantly point out, "*(. . .) the efficacy of GenAI and LLMs in the banking sector hinges on the accessibility of high-quality and consistent data*" (p. 255). Much of the CDR literature highlights its role in ensuring data responsibility by emphasising multiple data responsibility pillars that pertain to CDR. According to Cheng and Zhang [31], these include six key components requisite both in the digitalisation (unbiased data acquisition, data protection, and data maintenance) and digitisation (appropriate data interpretation, objective predicted results, and tackling value conflicts in data-driven decision-making). The three digitalisation components of data responsibility appear to be at the centre of policymakers and regulators' approach in drafting the EU AI Act, indicating the essential role of CDR in embedding data responsibility principles when putting RAI into practice. The data responsibility components are arguably less about technical requirements and more about developing an appropriate culture, attitude, organisation, and supporting processes to uphold them.

The literature also points to CDR's role in stimulating organisations to go beyond their legal obligations and highlights how such frameworks are less rigid and can be tweaked to suit specific industry requirements, lending themselves to AI applications vis-à-vis financial services. Rapid advancements in AI and GenAI, eagerness among financial institutions to use it, CDR tenets, the multi-faceted nature of RAI, practical challenges despite the abundance of RAI frameworks, and the urgency to turn theory to practice all direct to the urgent need to explore the non-technical barriers in practical approaches toward RAI. Identifying non-technical barriers in financial services and exploring the value of CDR in making RAI actionable could take us closer to a holistic, unanimous blueprint for implementation; build the case for CDR investment; and encourage follow-up research on non-technical dimensions. Without practical use, RAI frameworks connote the illusion of control, fictitious control.

Exploratory studies such as ours can offer insights on professionals' perceptions and attitudes to disambiguate the respective frameworks and 'humanise' them to allow for actionable insights for practitioners or a blueprint for implementing RAI, which can increase confidence in investing in AI and boost business performance while building the

required and much-desired trust and clear data strategies. Such research complements the World Economic Forum's recently introduced PRISM framework that outlines layers of AI implementation, including adoption pathways, where it highlights business-level non-technical risks (such as change management or stakeholder engagement) as high risks [24]. Indeed, bridging the glaring gap between ideal frameworks and ground realities in this hugely complex discipline can indeed be a hard and cumbersome task. Yet, the identification of non-technical and organisational barriers around making RAI actionable can help demystify the theory and, ultimately, set a clear roadmap for progress without jeopardising societal values and principles.

In this regard, to investigate the non-technical challenges and CDR's facilitative role, we now detail our qualitative methodology.

3. Material and Methods

The literature review clearly highlighted the profound and far-reaching impact of (Gen)AI, making responsible use of AI a sheer necessity. Yet, "*there are many difficulties to implementing Responsible AI in practice*" (Wang et al. [20]: 3), with scholars highlighting how real-world needs and expectations for using AI or machine learning fairly are "*neglected in the literature so far*" (Holstein et al. [32]: 2). The few studies on the 'principles-to-practice' gaps retain a sharp focus on technical aspects surrounding RAI. Still, academics stress that translating RAI to practical adoption requires "*an expansive scope of attention to the full set of issues influencing human well-being*" (Schiff et al. [17]: 3) that falls well beyond technical specifications and focuses on resolving "*various (. . .) organizational barriers*" that practitioners face (Holstein et al. [32]: 12). Exploring the full range of barriers is tempting, but "*such a task can be enormously difficult*" (Schiff et al. [17]: 3) but indicates the need to examine perspectives previously overlooked in order to enrich current knowledge (and practice).

Drawing from such arguments, we employed a qualitative research strategy to gather experiences, perceptions, and/or attitudes of professionals in finance regarding non-technical/organisational challenges to practicing RAI under the scope of recommending practical steps and adapting organisational practice. We focus down on financial services, as AI is to generate more than USD 140 BN annually by 2025 for this industry [9], denoting the increased penetration of intelligent systems into this highly regulated sector of pivotal importance to the economy. In this respect, through the narratives, the emerging role of the CDR conceptual framework is examined in order to understand whether an alignment between RAI and CDR is reconciled in financial institutions and gauge CDR's potential to address the barriers identified in this study, given its potential to "demystify governance complexity" around AI's deployment (Elliot et al. [5]: 179).

Following an inductive approach to explore the research objectives, this study can be considered exploratory with contributions to both theory and practice. We opted for semi-structured interviews since "qualitative research entails generating theories inductively rather than testing theories that are specified at the outset" (Bell et al. [33]: 366) and attempted to identify patterns in barriers that may be hard to resolve or those deemed to be more pressing to resolve (thereby providing insights to practitioners on priority issues and imperatives required for realising Responsible (Gen)AI in financial services). To achieve this, an interview protocol was devised and pilot-tested prior to the main data-gathering tasks.

3.1. Interview Protocol

In order to capture rich, experience-driven insights into the organisational and non-technical challenges associated with implementing RAI and the potential role of CDR, a

semi-structured interview protocol was developed. This qualitative approach was chosen to align with this study's exploratory research aims and to facilitate in-depth narratives from senior professionals in the European financial services sector. Semi-structured interviews are well-suited for such exploratory studies where the objective is to identify themes inductively, rather than test pre-existing hypotheses [33,34]. The structure and content of the interview protocol drew directly from recognised calls in the literature to bridge the gap between high-level RAI principles and their real-world application [1–3]. The guide comprised 31 primary questions, grouped into seven thematic sections, with follow-up prompts for elaboration. A pilot test of the instrument ensured clarity, thematic relevance, and alignment with current debates in the RAI and CDR literature. The sections of the research instrument were as follows (with Interview sections 4 and 5 being the longest, i.e., half of the interview time was dedicated to these sections):

Interview section 1/Introductory Context: This section included three opening questions to gather background on the participant's role and the organisation's AI/GenAI strategy. These foundational questions helped establish the institutional and functional context of each respondent, in line with qualitative, exploratory research norms for contextualising practitioner insights [34,35].

Interview section 2/Organisational Vision and Use Cases: Three questions addressed the organisation's AI vision, use cases, and principal concerns to trace practitioner priorities and technological orientations. These were informed by the literature identifying the rising application of AI/GenAI in finance and the lack of clarity on how organisations are responding to emergent ethical challenges [2,9,19,24].

Interview section 3/Understanding and Applying RAI: Four questions explored participants' conceptual understanding of RAI, adopted frameworks (e.g., EU AI Act, OECD, US NIST), key stakeholders involved, and internal governance journeys. These questions reflect the "principles-to-practice" challenge widely identified in the literature [1,4,5] and aim to determine how RAI is 'translated' into organisational mechanisms (i.e., operationalised in practice).

Interview section 4/Identifying Non-Technical Barriers: This section contained six questions specifically designed to elicit descriptions of practical, non-technical challenges (barriers) faced during AI/GenAI adoption, including examples of such barriers and their perceived impact. The questions also distinguished between technical and organisational barriers, following recent calls to move beyond technology-centric approaches in AI ethics [3,17,18].

Interview section 5/Strategies for Overcoming Barriers: Eight questions investigated approaches to addressing the previously identified barriers, including which are hardest to resolve, prioritisation logic, and any changes made to the frameworks as barriers became visible. The items—questions comprising this section—were devised to capture adaptive strategies and map the barrier-resolution loop and inspired by studies urging better understanding of how AI governance evolves in practice [21,32].

Interview section 6/Corporate Digital Responsibility: Four questions focused on whether CDR was adopted within the organisation and its potential or actual contribution to overcoming non-technical barriers. These questions were grounded in emerging calls positioning CDR as a bridge between ethical intent and operational practice in digital governance [5,6,27].

Interview section 7/Open Reflection and Emerging Topics: Three closing questions provided participants with space to share additional reflections, future expectations, or concerns (on AI/GenAI governance trajectories) not covered in previous sections. This open-ended set of questions is in line with best practices in qualitative research to allow for

generating unanticipated themes, further emergent insights, and enhancing the interpretive depth of qualitative inquiry [35,36].

The full list of interview items is provided in Appendix A; question codes were used during the analysis to maintain traceability between participant responses and protocol structure. This design allowed for thematic consistency while preserving flexibility for contextual probing, in line with established methodological guidance.

3.2. Sample Identification

This study's participants are 15 professionals working in European financial institutions who were selected through purposive-convenience sampling and satisfied the following criteria: (i) their organisation uses AI and GenAI, (ii) they have previous experience and a comprehensive understanding of RAI (albeit from different domains and departments in their organisations), and (iii) they play a key role in implementing RAI in their organisation. The interviewees represent organisations from the United Kingdom, the Netherlands, Germany, France, and Switzerland, i.e., economies with well-established and robust financial institutions. The sample includes leaders and practitioners from data management, data security, IT, digital transformation, business development, developer teams, and CDR and AI leads. Among the interviewees, three have engaged in CDR strategy formulation, and their respective experience proved to be essential to explore relevant nuances in translating RAI. The participants work in retail banking, investment banking, and insurance services, and one works in an organisation offering fintech applications; they were drawn from the first author's contact list and from professional networking platforms using profile criteria that matched this study's purpose. The convenience sampling approach we followed does risk self-selection bias but was mitigated through role stratification (varied functions) and cross-country inclusion (capturing national nuances). Likewise, the gender ratio of interviewees (4 female and 11 male) reflects known demographic skews in the digital and (fin)tech sectors [37]. While this imbalance may narrow perspective diversity, (a) all female participants held high-influence positions (role seniority), and (b) thematic convergence across genders both reduced this concern.

The subtopics examined during the interviews (such as CDR's role or types of frameworks used) were drawn from the academic and grey literature exploring these emerging concepts [5] (Elliot et al., 2021) with an aim "*to provide a rich descriptive, exploratory and explanatory study*" (Azungah [35]: 384) on aspects and issues where the current literature is nascent, making the selected research instrument appropriate for the scope of this study. An administration approach that mitigates ethical risk was followed, in line with well-established recommendations indicated in the business research literature [33]. This includes explaining upfront the purpose and research objectives with an information sheet, maintaining the confidentiality and anonymity of the participants, securing each participant's informed consent, and offering a choice of interview implementation (recorded, unrecorded, and face-to-face). Nine of the semi-structured interviews were video recorded (Microsoft Teams environment), and two were conducted face-to-face and audio recorded. In four cases, following the participant's request not to record the interview, detailed/verbatim notes were taken. Each interview lasted between 45 and 90 min. Ethical approval for this study was obtained from the Research Ethics Committee of the University of Reading (reference code: SREC-HBS-20240928-ARVE1689; date of approval: 30 September 2024). Key information describing the 15 interviewees along with identifying codes is presented in Table 1.

Table 1. Participants' details.

Interviewee's id Code	Gender	Position/Executive Role	Country	Organisation	CDR Experience
P1	M	Strategy Officer	The Netherlands	Fintech	-
P2	F	Chief Technology Officer	The United Kingdom	Insurance	-
P3	F	Director, App Delivery Officer	The United Kingdom	Insurance	-
P4	M	Security Officer	The United Kingdom	Stock Exchange	-
P5	M	Data Management; Sustainability Officer	France	Insurance	✓
P6	F	Chief Information Security Officer	Switzerland	Investment Bank	-
P7	M	DevOps Engineer	The Netherlands	Pension Funds	-
P8	M	Security Officer	The United Kingdom	Retail Bank	-
P9	M	AI Governance Officer	The United Kingdom	Investment Bank	-
P10	M	Risk Management and Compliance Officer	The Netherlands	Pension Funds	-
P11	M	Strategy Officer	Germany	Ethical bank	✓
P12	M	Cybersecurity Officer	The United Kingdom	General Bank	-
P13	M	Data Management Officer	The United Kingdom	Insurance	-
P14	M	AI Strategy Officer	The United Kingdom	Multi-Agent AI Services	✓
P15	F	Head of AI and Data	The United Kingdom/Germany	Trading Platform	-

3.3. Data Analysis

Reflexive thematic analysis [38] was selected due to its compatibility with our research aims: while grounded theory is well-suited for building substantive theory from data, our approach sought to extend established constructs (e.g., Fjeld et al.'s 'thorny gap') and produce a fruitful, practice-orientated taxonomy of barriers. Reflexive thematic analysis supports this by enabling us to identify and make sense of patterns across the data in a transparent, flexible manner, while still attending to researcher reflexivity and the co-construction of meanings.

Qualitative coding was employed using appropriate data analysis software in order to systematically categorise excerpts of the key informants' responses and subsequently identify patterns in the data. Transcripts were read carefully, with efforts to fully familiarise with the data long before all interviews were completed [33,38]. This task was followed by an initial round of tentative coding and then by assigning descriptive codes summarising excerpts from the transcripts into descriptions that encapsulated the content of the qualitative data. The interview protocol was designed to allow for key themes to be examined through the narratives and representative statements drawn from the responses; the semi-structured guide allowed for question-based codes that classified all respective responses to the respective interview question. Thematic analysis was employed for patterns across the qualitative dataset, i.e., certain excerpts across the responses pointing to the same direction, repeating certain meanings, or being largely non-homogenous. This allowed us to construct our narrative of qualitative findings based on the participants' viewpoints, attitudes, and perceptions. The display and interpretation analysis was aligned with the three-step approach suggested by Sekaran and Bougie [39] in terms of qualitative data reduction and was conducted in line with recommendations set forth by Ryan and Bernard [36] to identify critical themes: repetitions, metaphors and analogies, transitions, linguistic connectors,

and theory-related materials. Transcripts and findings were shared with participants for validation and to correct or avoid misinterpretations. While we did not conduct formal triangulation with additional data sources (e.g., documents or observations), we aimed for interpretive rigour through iterative discussions between researchers, detailed memo writing, and the maintenance of an audit trail of theme definitions and coding decisions. This ensured internal consistency in how themes were developed and applied across transcripts.

While this study did not achieve theoretical saturation (i.e., a concept more central to grounded theory), thematic saturation [35] was achieved within the qualitative design. Although our sample of fifteen practitioners was determined pragmatically (based on access and scheduling constraints), we monitored for thematic saturation throughout our analysis. After coding the majority of the interview transcripts, no new substantive themes emerged in the remaining few ones; instead, the iterative coding process confirmed stability, showed diminishing new insights in these last interviews, and largely confirmed existing patterns identified. This aligns with recommendations that saturation in thematic analysis is attained when data "offers no further insights to research questions" [40].

We therefore conclude that, within the scope of European financial services, our sample achieved practical saturation for the purposes of an exploratory, practice-orientated study around non-technical challenges and CDR-informed mindsets. We present these findings below, organised around the core research objectives.

4. Findings

4.1. Value and Practical Use of AI and GenAI in Financial Services

Providing context for RAI practices, participants highlighted the tremendous opportunity and tangible benefits of using AI and GenAI in finance. Although adopting it at varying degrees, participants were unanimous in admitting the transformative nature of AI in their organisations, and GenAI in particular. They all shared examples of applications and tasks or processes where their organisations are currently employing AI techniques (either actively or sparingly). The most common use case objective they indicated was for tangible benefits of cost reduction, productivity, and efficiency improvements, which is in line with Bhatnagar and Mahant's [9] finding that 36% of executives in finance have used AI to "reduce costs by 10%" and about 46% of financial services organisations reported "improved customer experience after implementing AI" (p. 446). P1 (Strategy Officer, Fintech, The Netherlands) indicated the use of AI/GenAI for money laundering and fraud detection, and P3 (Director/App Delivery Officer, Insurance, UK) for customer data analysis to identify business opportunities faster, while P2 (Chief Technology Officer, Insurance, UK) for underwriting (e.g., being able to summarise all of the client notes and documents), writing standard operating procedures (SOPs), rewriting legacy banking applications from PASCAL or COBOL to Python or JavaScript, detecting threats or cyber risks to retain cyber resilience, and improving code quality to reduce relevant vulnerabilities. P2 also identified several monotonous, tedious, and mundane tasks in financial services that are necessary but less value-adding. According to P2, "AI can do quite well", particularly in such repetitive tasks that are not unique or those low-risk tasks that do not use sensitive or personal information. This view is consistent with the Bank of England survey [41], echoing the value of AI in automating routine, repetitive, or undifferentiated activities that all FIs undertake, such as (financial) reporting, tax returns, client onboarding, and document recognition, retrieval, and processing.

4.2. AI and GenAI Risks: Inextricable Links to Barriers to Responsible AI

The thematic analysis revealed that barriers to practicing RAI are inextricably linked to two key aspects. First, the risks of deploying AI in financial operations, especially the

new risks introduced by GenAI, which is a vastly different AI technology; this is because participants kept referring back to the risks when describing implementation challenges their organisations face. Understanding AI risks is essential in putting the implementation barriers in the right context and highlighting the importance and approaches to address them. Second, technical and legal barriers such as security, privacy, and algorithms/large language models monitoring.

Multiple common perspectives around the risks of AI, particularly GenAI and how that affected their RAI implementation, were surfaced. P4 (Security Officer, Stock Exchange, UK) highlighted how her organisation has "robust" architectural processes and a governance framework to manage the risks of traditional AI and machine learning (ML) because of traditional AI and ML's maturity and its use within the organisation for between 10 and 15 years. Likewise, P12, P1, P3, P10, P8, and P7 all stated that their organisations have developed mature AI governance systems and guardrails for ensuring Responsible ML practices through safety systems and employee training to ensure testing of the input data for ethics and fairness.

Still, the risks of GenAI and practically managing it are top-of-mind across all interviewees:

> *"Gen AI is this new kid on the block that is more challenging because it's multi multimodal to include natural language, voice, text to voice, text to image creating new threat vectors and risks"*. (P4, Security Officer, Stock Exchange, UK)

Similar thoughts were echoed by P6, P2, P3, P11, P7, and P8 that their "biggest" RAI concern is more around GenAI than AI itself:

> *"(. . .) Largely because of the fact that obviously generative AI is a learning module (. . .) and it learns from itself"*. (P3, Director/App Delivery Officer, Insurance, UK)

But as highlighted by four participants (P2, P7, P1, and P12), similar risks exist in other AI uses, too. P2 and P12 referred to the recent AI mishap from McDonalds, where its voice recognition AI software to process orders in its drive-through restaurants proved unreliable. The system misinterpreted orders to provide "bacon-topped ice cream to hundreds of dollars' worth of chicken nuggets" in a single order [42]; a project initiated in 2019, and in 2024, McDonalds announced the end of this automated order-taking system. P2 used this example to show that AI is nascent and can miss things. But gravely pointed out that a similar mishap in the finance industry '*in, say, medical insurance use case will be. . .far less funny*'.

> *"If AI misses something in the medical history and the claim becomes invalid. This has a double impact: The insurer has to foot the entire bill versus the re-insurer as commonly practised; and re-insurance rates go up because the company was not diligent enough in the underwriting"*. (P2, Chief Technology Officer, Insurance, UK)

P2, P5, P4, and P8 also called out risks and issues around bias and security of AI. Still, GenAI remained the most recurring risk across interviews (identified through repetition and linguistic connectors used in thematic analysis). This is echoed in multiple studies backing GenAI as more challenging because traditional ML is fairly established with proven use cases in finance, predominantly in "pattern identification, classification, and prediction," whereas GenAI models are able to create 'original' output that is often indistinguishable from human-generated content" (OECD [43]: 6). With Bhatnagar and Mahant [9] predicting a 270% increase in the growth of AI adoption in financial services over the next four years, concerns expressed in the interviews about managing GenAI risks are also documented in the recent literature (e.g., Dhake et al. [19]). Through the narratives it was confirmed that, with its broad appeal and ease-of-use, GenAI appears to be a game-changer and also a disruptor of established RAI practices, with organisations having no option but to adapt and build new guardrails.

The latter requires an exploration of relevant non-technical challenges across people, process, skills, organisation, and budget in order to realise a principles-to-action transition for RAI based on the participants' viewpoints.

4.3. Non-Technical Barriers in Making Responsible AI Actionable

Eight participants used words such as "overwhelming" and "million-dollar question" or long, reflective pauses followed by "hmmm, good question. . ." when explaining their difficulty to implement RAI. This may encapsulate the challenges executives face in governing AI effectively, given the breakneck speed of AI evolution as well as constantly evolving RAI conceptual guidelines.

> *"With AI, people tend to feel overwhelmed (. . .) and there are so many new terms popping up and challenges"*. (P11, Strategy Officer, Ethical bank, Germany)

4.3.1. Challenge #1: Determining Who's Responsible for What

Seven of the officers interviewed used words such as "tricky", "hardest", "big risk", "biggest risk", and "main fear" when attempting to elaborate on challenges to ensure accountability. In five of these seven interviews, accountability was mentioned upfront and unsolicited. Apart from accountability, security and safety were cited by an equal number of participants as the critical principles of RAI; interestingly, robustness was not mentioned by any of the interviewees.

P2 highlighted the challenge to determine who is accountable/responsible if algorithms go wrong; using a self-driving car analogy, [P2] illustrated:

> *"If. . ., if, for instance, there is graffiti on a stop sign and a malicious actor puts in a command that 'if graffiti go at 30 miles an hour' and accident happens, who is liable?"* (P2, Chief Technology Officer, Insurance, UK)

Practical challenges in ensuring accountability became more nuanced when four participants explicitly used the word "black box" to describe AI, e.g.,:

> *"Too many unknown unknowns to bullet-proof responsible AI use. When data goes in, it is like a black-box. You can only provide data for training, you cannot control the outcomes"*. (P6, Chief Information Security Officer, Investment Bank, Switzerland]

> *". . .As in we may not always understand how it reached a certain conclusion, right? Why a certain user is segmented in category A versus category B"*. (P1, Strategy Officer, Fintech, The Netherlands)

P1 placed emphasis on the value of having clear mechanisms to monitor and manage the data inputs into the AI system(s) and the outputs to facilitate "reverse engineering" for transparency. To make this practical, he continued to denote that periodic input–output audits and random checks can help organisations identify the issues, document them, and improve the system before deploying in critical use cases such as credit scores. Some participants' tone changed to a slower pace when mentioning the consequences of lack of traceability, explainability, or accountability in AI systems because of their black box nature. This partially highlighted their seriousness in addressing this need.

> *"Imagine relying on a report that is an AI outcome and it is not accurate, it risks reputation. . . Already, mortgage crisis eroded trust in financial services organisations. (. . .) Without control, if AI is allowed to unleash, it will be chaos"*. (P15, Head of AI & Data, Trading Platform, UK/Germany)

Moreover, one of the Dutch executives (P7) highlighted another practical challenge: service providers' approach to introducing AI features as default-on to their customers, including those in financial services. A few other participants described technology providers'

tactics of introducing AI features as default on or sneakily or too many, too quickly, making this a practical barrier to practicing AI use responsibly. These interviewees referred to external factors such as supply chain, absence of an implementable "shared responsibility model", AI providers' feature-innovation strategies, and the role of service partners as non-technical barriers.

P7 pointed out that one Dutch organisation blocked Microsoft's GenAI tools because, most times, the AI features were auto-enabled, and practitioners conducted detailed risk assessments for six months to obtain full control and confidence before enabling them. Another participant also highlighted the provider risk, saying any changes to the models can access new data repositories without authorisation; every time a provider makes a change to their AI service, the guardrails internally need to be revised.

> *"If a vendor changes their privacy model or rights structure, it is impossible for us to update all the models in 10 s".* (P15, Head of AI & Data, Trading Platform, UK/Germany)

Discussions on non-technical barriers were more fluid than moving from one principle after another. This shows that although there are different principles, the aim is to approach AI as holistically as possible to identify and mitigate risks. The practical barriers identified for ensuring accountability can be extended to transparency as well as fairness and inclusion.

4.3.2. Challenge #2: Preparing for and Managing Unintended Consequences at the Speed of AI

> *"New and evolving technology [such as AI] brings on new risks and it can be difficult to keep on track, build mitigants quickly."* (P9, AI Governance Officer, Investment Bank, UK)

One area of long discussion in most interviews was when participants started talking about the fairness and inclusion principle of RAI. For example, P3 raised an important issue about rolling out AI equitably and the importance of identifying pilot groups or user segments in a factual way based on transparent decisions rather than arbitrarily rolling it out based on seniority alone. P3 warned that this could create a digital divide at a social level within the organisation (see Capraro et al. [44]). Principles of fairness and inclusion in RAI would see organisations deciding user groups based on who would benefit the most from using AI tools and services in their jobs, P3 added.

P8 made similar points around transparency of decisions to ensure fairness.

> *"It's really, really important to make sure you define everything, make sure you define the taxonomy.*
>
> *Your AI risk description based on your risk taxonomies.*
>
> *Make sure you give the proper examples of each. . .*
>
> *Make sure you map them into the framework you prefer."* (P8, Security Officer, Retail Bank, UK)

Interestingly, country-specific nuances of how organisations approached RAI implementation were evident, despite the small sample size. For instance, all three of the Dutch executives referred to the Dutch government's algorithmic blunder from 2020 when highlighting practical challenges around the accountability principle. The issue involves the Dutch government's welfare surveillance program called SyRI that uses vast amounts of personal and sensitive data with algorithms to predict how likely a person is to commit tax fraud or (child) benefit fraud. The AI system was a "risk calculation model" that was developed "over the past decade", but the flawed system disqualified many genuine

benefit seekers and wrongly categorised many innocent people as frauds and even affected their access to resources and/or credit scoring. The issue was so serious that Amnesty International called on governments everywhere to instantly stop using sensitive data in risk-scoring and billed the project as "opaque 'black box' systems, in which the inputs and calculations of the system are not visible, resulting in "absence of accountability and oversight" [45]. P1 mentioned the mishap as a "delicate issue" and one that their financial services organisation is "cognisant about" when mapping out adoption plans. P10, however, highlighted the story as a learning opportunity. Through their narratives it was pointed out that this national experience has had an impact on how Dutch practitioners approach RAI. For instance, P7's organisation now follows a conservative approach based on an "AI is scary mindset", where such systems are prohibited in use cases such as in decision engines, while it relies on intensive training programs to make everyone take responsibility for using AI ethically and fairly and understand the implications of not doing so. Serious implications include loss of IP, non-compliance, exposure of sensitive data, and low-quality output from AI.

> *"We want to practice Responsible AI responsibly, especially to avoid such horrible situations. But it also risks slowing down"*. (P1, Strategy Officer, Fintech, The Netherlands)

> *"It ruined people's lives (. . .) Shows that there's too much at stake to even plan responsibly across multiple facets [of AI] . . .Responsibility is always with people and we are keeping it with people. If AI fails, it can break a lot"*. (P7, DevOps Engineer, Pension Funds, The Netherlands)

> *"We've seen things go wrong in the past and you want to learn from them."*
>
> (P10, Risk & Compliance Officer, Pension Funds, The Netherlands)

Another perspective around unintended consequences is the (lack of) organisational readiness to deal with future uncertainties. Participants imagined bad actors using AI to attack banks. Examples included how fraudsters can impersonate tens of thousands of genuine mortgage applicants by "hearing a recording of a couple seconds" to train an AI system. In this regard, two executives (P1 and P7) were particularly unsure on the overall value of AI to finance in the long term:

> *"We don't fully know whether AI in the financial services market is gonna have a net negative or net positive effect"* (P1, Strategy Officer, Fintech, The Netherlands)

4.3.3. Challenge #3: Difficulty in Ensuring Fairness and Inclusion

Additional practical barriers around fairness and inclusion include a lack of understanding of what fairness is, what AI risks are, and how to define thresholds and risk appetite. According to P13, explaining data governance is far easier than explaining fairness. It is tricky for a few reasons, he elaborated. One, people are impatient and want to do things "quickly" rather than "properly". This results in tactical decisions that make the environment complex and difficult to manage or justify.

> *"It is hard to tell how AI systems are using data for fair outputs given how automated AI systems are. . .With GenAI, you are compounding the problem and authorise decisions that lead to undesirable consequences"*. (P13, Data Management Officer, Insurance, UK)

P8 also alluded to similar concerns, saying the emphasis on speed of AI output or results is far higher than the emphasis on the fundamentals or input due diligence.

> *"Fairness is context-dependent and can be subjective"*. (P9, AI Governance Officer, Investment Bank, UK)

Many respondents had lots to say about fairness and inclusion and a desire to do the right thing. It was a repeated barrier, but participants highlighted aspects that are near and dear to them or their experiences. Although non-homogenised, the varied perspective further highlights the complexity of making RAI principles actionable. Highlighting disadvantages for certain communities, P14 identified GenAI's usage limitations for non-native English speakers as another point around fairness. One of the Dutch participants agreed that to improve quality, the team has to embed a translation service into their GenAI tool, which requires more time, resources, and skills compared to native English users.

> *"...(my) conclusion was that the tool needs to be trained in equal number of books and papers in Norwegian as it is in English to have the same impact in that language...How am I going to do that?"*. (P14, AI Strategy Officer, Multi-Agent AI Services, UK)

Another perspective around fairness and inclusion was pointed out by P11, who indicated the critical importance of decision criteria in place in order to abort an AI service in case of conflict when it yields a great revenue model but makes job roles redundant. In a similar vein, P3 raised another fairness question around rolling out AI tools "equitably", where members of the workforce do not feel less privileged for not having it.

> *"AI does have this ability to create social divide (...) Fair and inclusive AI is how you provide access so that if there are question around this, the decisions can be justified and explained"*. (P3, Director/App Delivery Officer, Insurance, UK)

4.3.4. Challenge #4: Traditional Business Approaches, Models, and Processes in Finance Unable to Cope with AI

Through their responses, executives suggested that, beyond business processes, legacy models and practices are barriers, too. Two interviewees pointed out how the banking sector has historically had a very mature culture of collaboration and cooperation and consensus-building.

> *"(...) This is one of the nicest thing about the industry where things usually change very slowly, with consensus. But with AI, it's the opposite and we're caught off-guard"*. (P11, Strategy Officer, Ethical bank, Germany)

Another past process that is considered a barrier was data management. P13 highlighted how their (small-sized) enterprise has not tagged the data or developed data access guidance for AI; adding this past practice makes practicing AI responsibly an impossible task according to this data management officer. Data access was also cited by P5, P6, and P3 and billed as a "bottleneck". P6 further highlighted how traditional governance mechanisms are incapable of dealing with the new kind of unstructured data used in AI. Aspects such as policies and guardrails around data integration or data leakage prevention, as P6 denoted, are sparse to make AI use safe.

> *"There are still lot of questions...And so, use of AI can lead to human error because there are no controls to alert the team of human errors"*. (P6, Chief Information Security Officer, Investment Bank, Switzerland)

4.3.5. Challenge #5: Balancing Trade-Offs and Meeting Expectations of Different Stakeholders

There were two distinct viewpoints among the executives interviewed, with one camp with 5 participants valuing an 'innovation-first' approach, while the rest supported a 'governance-first' approach to deploying AI/GenAI in finance. Six of the respondents (four with direct security backgrounds, one with data management, and one with AI governance) called out security or safety as their highest priority when turning RAI principles into practice. These practitioners placed emphasis on resolving privacy-by-design and

security loopholes. The former five participants focused on innovation included a Dutch fintech (P1) that has provided access to CoPilot to everyone in the organisation with the expectation of high usage and high productivity. In fact, in P1's words, in his organisation, managers obtain notification to nudge the team in case of low adoption.

> *"We are a born-digital company, so tech-savvy culture is ingrained in the workforce. Why bother a colleague if [GenAI tool] CoPilot can do. . . what you need"*. (P1, Strategy Officer, Fintech, The Netherlands)

P2, P11, and P6 expressed difficulty in surfacing innovator perspectives when security voices are the loudest. P2 admitted it is difficult not to "govern AI to death". P14 and P11 particularly expressed the desire for financial services organisations to adopt some of "Silicon Valley" attitude of innovation. But P7, P8, P4, and P15 emphasised the importance of governance. In fact, P15, with 20 years of experience in neural networks and deep learning, declared:

> *"(. . .) Of course, the most secure technology is one that has been turned off, battery removed and buried in the desert. Incredibly secure. . .But not usable"*. (P12, Cybersecurity Officer, General Bank, UK)

> *"(. . .) It' almost impossible to get to a responsible Generative AI. There is no way to establish a stable governance on pure stochastic systems like the ones based upon this AI paradigm. A system that gives you different answers to the same question can't be 100% reliable and can't be controlled or governed"*. (P15, Head of AI & Data, Trading Platform, UK/Germany)

Such a polarised view is itself a big challenge that practitioners have to reconcile first when implementing RAI that meets everyone's objectives or explain the trade-offs. In contrast to the Dutch fintech's CoPilot adoption, one UK FI took 12 months to roll out CoPilot once a decision was made. Similarly, P10 highlighted how getting approval for use in 8 weeks is "lucky and fast", as certain proof of concepts (POCs) take up to half a year because they have to go through layers of approval, which include the immediate business unit and multiple governing bodies, such as data teams, privacy teams, security architects, legal teams, and so on. Those supporting the 'security-first' approach admitted to the challenges of coping with the speed of AI expectations of an enthusiastic workforce. If adoption is faster than governance, then double the effort is required to bring it back on track, admitted P8. P4 also noted how certain groups in his organisation are excited to use the technology. But insisted that a thorough evaluation of appropriate solutions takes time, indicating the enormity of this non-technical barrier.

> *"Locking it down is like trying to get salt out of the ocean because there are so many ways around it. . .we will end up with shadow usage of different tools. They will start using and then ask for waiver to use saying they have invested time and money"*. (P7, DevOps Engineer, Pension Funds, The Netherlands)

Further highlighting the trickiness in balancing security with innovation, P4 highlighted that slow governance can backfire, with frustrated users finding newer and more convenient GenAI tools in traditional, low-risk AI facets such as machine learning, accentuating the problem.

> *"If governance is always playing catch-up, they may use GenAI even in areas where traditional ML is appropriate bringing in new risks"*. (P4, Security Officer, Stock Exchange, UK)

> *"(. . .) We're always in such a hurry to find answers"*. (P11, Strategy Officer, Ethical bank, Germany)

P3 expressed the 'pain' differently:

> *"We don't want to find out that our IP is being utilised in GenAI tools like ChatGPT, and that meant that we had to quite quickly react to being able to put the guardrails."* (P3, Director/App Delivery Officer, Insurance, UK)

The risks, if this challenge is not addressed, are far-reaching to compliance and costs too. For example, if GenAI tools are adopted without following appropriate licensing or usage agreements, there can be serious IP, security, and compliance risks. P5 delineated few examples of French companies having experienced this, adding a warning that the "Internet and AI never forget". The speed intensifies the governance or monitoring challenge, as P9 underlined that "(. . .) Responsible AI takes time. . .evidence gathering, filling in assessments, etc".

Overall, balancing the needs of innovation-orientated tech enthusiasts in the organisation, the murky world of AI risks, and long-established organisational structures or processes is tough, admitted by nine participants. In fact, interviewees cited that they are torn between the enthusiasm to use AI quickly and the traditional organisational aspects that are not yet ready for such an unprecedented new way of business operations, with regulations "somewhere in between", as P3 pointed out that regulations are a good starting point but do not provide complete, clear answers to all barriers, particularly the non-technical ones. Yet, most participants expressed a strong awareness and intention to be a catalyst for RAI practices but have no blueprint to follow.

> *"Trade-off between performance and/or profit and fairness (is one of the top barriers in implementing Responsible AI) and business stakeholders have to make those decisions"*. (P9, AI Governance Officer, Investment Bank, UK)

4.3.6. Challenge #6: Ensuring Sustainability

Eight participants mentioned sustainability or carbon footprint as a significant challenge. This was an area that participants discussed with a rather passionate, human-centric, and purpose-driven vision of RAI. Most emphasised short-term vision and views as compounding the problem. For instance, P11 denoted:

> *"You have a lot of use-and-drop technology and that's not very sustainable"*. (P11, Strategy Officer, Ethical bank, Germany)

P13 expressed a similar viewpoint in that it requires tedious efforts to identify the use cases and data to make sure that "the algorithms will last in time and not be replaced by, say, the next new version 3.5 or 4 or so on every 5 months". P3, P5, and P2 also emphasised the difficulty in effectively meeting sustainability goals as a key non-technical barrier. There was agreement that GenAI is very resource-intensive in terms of servers and memory. The inability to manage this not only increases costs but also has an environmental impact that is far from negligible. One particularly "wasteful" trend pointed out by P5 is the impact of a decentralised approach to AI deployment: every business unit developing and managing their own AI projects results in duplicity of resources for providing "pretty much the same services" with direct impacts on the cumulative carbon footprint.

4.3.7. Challenge #7: The Human Factor

Most (i.e., 13 out of 15) of the respondents highlighted people-related challenges ranging from mindset and culture to skills and attitudes. "People, people, people", "team", "staff", "everyone", "they", "we" were repeatedly used when highlighting the human effect as a non-technical barrier. One participant (P6) asserted that 80% of the success depends on people, while others agreed that convincing people to get on board with a shared, common

(AI) vision can indeed be a non-technical barrier. One powerful remark on this came from P5 in the context of sustainability, accountability, and ethics of AI:

> *"You can explain the importance of practising AI responsibly. Most people already know it but it is getting people to care to do the right thing"*. (P5, Data/Sustainability Officer, Insurance, France)

Similar sentiments were expressed by P12, P13, and P8. One way to engage people to care is to empathise with them, P12 suggested. According to P8, the problem stems from how guidelines are handed down, and not everyone views them from the same lens. While P7 said, demonstrating value that resonates with everyone about responsibility is "massively hard."

> *"It's easy to produce guidelines and standards for any organisations and then just tell everyone to follow it. In practice, how you wanna actually convince them that this is right for them and how you can show a return on investment. This is not an easy social skill"*. (P8, Security Officer, Retail Bank, UK)

Multiple participants pinpointed the skills and knowledge shortage in RAI as a severe human-orientated barrier to action.

> *"Talent is already rare (. . .) now the combination of AI and governance is (becoming) even more rare, niche, find"*. (P6, Chief Information Security Officer, Investment Bank, Switzerland)

Responses touching upon skills suggested that a major non-technical barrier unique to RAI compared to any other previous technologies is how every individual is expected to be accountable because otherwise, they risk jeopardising the whole organisation. Bringing everyone up to speed with digestible, easy-to-understand training programs and communication strategies so they all become responsible users at ground level appears to be a mighty task to move from theory to practice. People's skills, use of sharp judgement, emotional intelligence, and sense-checking are essential to identify bias in AI. In this regard, P12 explained this with an analogy of SatNavs:

> *"If it asked you to drive into a river because it's the fastest route between A and B, at that point you may as well turn it off because you don't trust it. So if AI tells you something, you still need skilled professional to validate that output"*. (P12, Cybersecurity Officer, General Bank, UK)

4.3.8. Challenge #8: Operationalising AI and Planning for What Happens Beyond Day Zero

Building, deploying, or adopting AI responsibly is only half the story. The barriers become murkier when practitioners have to think about what happens after Day Zero (implementation day). P4, P14, and P3 specifically consider not just how to introduce AI responsibly but also how to make it business-as-usual. One difficulty is creating processes and policies to allocate AI licenses appropriately and managing the different types of employees (joiners, movers, and leavers) and making sure it is funded within the right budget cycle and charged to the correct business unit. Additional hurdles had connections back to the fairness and inclusion principles, stressing the difficulty in rolling out AI "equitably" without making employees feel less privileged.

> *"How can I ensure no-one feels like a lower-valued employee because they can't have it?"*. (P3, Director/App Delivery Officer, Insurance, UK)

The laser focus on AI implementation means organisations are not considering barriers that they will stumble upon when pilots become successful and they have to make it business-as-usual. Addressing tough questions such as the impact on AI access policies,

redundant skills, job losses, or changing roles. This means when there are questions around why a particular user obtained access and another did not, there is a valid, consistent explanation for it so that it is not seen as creating a digital divide at a social level and in the workplace environment since AI has the potential to create social divides. Analysing responses indicates that almost all interviewed organisations have clear policies and guidelines and architectural principles in place but admitted their AI strategies may still not be fully risk-free.

> *"There are so many unknown unknowns, we can't plan for all scenarios*. (P2, Chief Technology Officer, Insurance, UK)

4.3.9. Challenge #9: Justifying and Securing Budgets

Multiple thoughts were shared on budgetary resources (or lack thereof) for implementing RAI. Respondents revealed that budget allocation for RAI practice is arbitrary, ranging from 2% to as high as 50%. Three participants refused to share information on budgets. Those with high spending expect it to go down in a couple of years when it stabilises. P5, P7 and P11, and P10 (representing entities from France, Switzerland, and Germany, respectively) indicated a higher budget compared to P4, P8, and P12 on the lower side.

AI use requires heavy investment that smaller FIs find hard to undertake, risking their competitiveness. P13 highlighted it takes about GBP 50–70 k to take one use case fully scaled and ready after the PoC. In addition, there was little confidence in whether what each organisation is investing in is appropriate or not. Some participants were exercising caution in overinvesting into building sophisticated controls for GenAI at a time when there is so much hype around it, indicating that it may be a waste of resources if the adoption hits a ceiling sooner than expected. In fact, P3 asserted that stagnation or usage of GenAI tools for "the same use cases over and over again" is already visible nowadays. Likewise, P15 expects GenAI to work primarily on creative tasks and, in general, on anything where reliability and preciseness are not mandatory because of the difficulty in governing it. Similar viewpoints about the future were expressed by other executives in the context of operationalisation: one main observation was that there is no shining example within the financial services industry of how to embed AI and GenAI systems and completely transform their business. However, this is in stark contrast to examples of FIs like JPMorgan, Visa, or HSBC that have completely transformed how they overcome 'tough' business challenges such as money laundering and fraud management with AI.

This contrast of some financial services using AI more than others is down to attitudes and culture. Those interviewees that see AI as a hype (P2, P3, P5, P6, P7, P13, and P15) are cautious to experiment vs. those that see it as a transformational technology (P1, P4, P8, P9, P10, and P14).

P14 identified that some organisations are so worried about the things that can go wrong that they "don't do anything". His belief is that the technology progresses, and what AI does not process right today, it will process right in six months or a year. One piece of evidence of this is how HSBC or JPMorgan has been working and testing the use cases for years, identifying issues and co-developing solutions to address the issues before making the AI use case live.

> *"What's really important at the moment is to experiment and understand what is possible and be conscious of what is not."* (P14, AI Strategy Officer, Multi-Agent AI Services, UK)

But not everyone shares the optimism and wants assurances of tangible returns before allocating budgets.

> *"It is easy to fall for the hype. Analysing everything and understanding the benefits thoroughly before jumping into the trend is important."* (P6, Chief Information Security Officer, Insurance Bank, Switzerland)

P15 shared similar sentiments.

> *"... after the incredible hype that accompanied GenAI, I'm pretty sure its scope will shrink quite by much."* (P15, Head of AI & Data, Trading Platform, UK/Germany)

4.4. The Big Debate: Which Framework? ... and the Role of Corporate Digital Responsibility

One theme that stood out in the analysis was how different participants used different RAI frameworks. Three executives revealed that their organisation adopted the US NIST AI framework, which tried to make it actionable. The other twelve participants were following frameworks their FIs developed in-house by taking inspiration from GDPR, the EU AI Act, EU recommendations, the Chartered Institute for IT, and the OECD AI Principles (OECD [46] 2023b: 17), while the frameworks' dimensions covered aspects similar to the WEF's 7 principles (WEF, 2024b [24]). There was also no definitive approach on whether the AI approach is or should be centralised or decentralised. Some organisations that are large and with a global presence focused on a decentralised AI strategy, but equally, other multinational financial organisations opted for a centralised strategy to minimise challenges around sustainability. P3 said their organisation is developing their own framework as AI is evolving rapidly.

> *"We have developed our own AI framework and strategy rooted in our own risk framework, security framework and financial framework."* (P3, Director/App Delivery Officer, Insurance, UK)

4.5. CDR Intersection and Impact

Interestingly, two organisations that practice Corporate Digital Responsibility rooted their RAI practices within their CDR principles. Their representatives maintained that this allowed them to demystify some of the challenges identified by other participating executives, especially those barriers around dealing with unintended consequences. For example, P11 pointed out that the CDR framework "came in very handy" in identifying unintentional effects from AI, such as staff reduction, and helped them be proactively engaged in addressing newer threats. However, the challenge, as executives' responses denoted, is that an alignment between CDR and RAI can be a cumbersome, time-consuming, and thorough task (for every single principle through the lens of CDR), which can also slow down innovation potential and make certain stakeholders disgruntled. This further emphasises how delicate the balancing act such an alignment can be. Interestingly, those interviewees who were CDR practitioners displayed a stronger human-centric approach to RAI and deliberated more, using terms such as "idealist", "building consensus", "no margin for error" when it came to adopting AI responsibly (most likely because the very origins of CDR lie in corporate social responsibility and stakeholder theory, which prioritise human welfare, rights, and agency over strictly technical optimisation). CDR practitioners went beyond the quantitative aspects of the RAI framework, such as return on AI investments, to focus deeply on human-related consequences to shape their organisations' RAI actions. In their words, they considered the unintended consequences of AI upfront and employed CDR tenants to update HR and legal policies and related organisational processes to ensure value-first banking. One participant echoed how in a recent CDR group discussion on RAI, members felt that the discussion was "evolving from Responsible AI to actually more of a human rights conversation". The CDR practitioners also highlighted a growing, thriving, and collaborative community within their organisations that has RAI as an agenda item for discussion at every executive meeting. This helps in learning, sharing, and addressing

challenges together as opposed to every organisation trying to figure it out themselves, as is evident from other interviewees' statements.

For example, both P3 and P4 admitted they are still in the process of fully building out their RAI framework, although they are clear it is going to be their own, unique framework. For practical purposes, P4's organisation (which was not aware of CDR frameworks) is working through existing control policies:

> *"Currently, there are some big debates internally on whether we have a separate AI framework or do we work through existing control policies and strengthen those for AI risks to ensure that we have it all joined up. At this point, a decision hasn't been made. . ."*. (P4, Security Officer, Stock Exchange, UK)

Similarly, P7 said they developed their own RAI framework because they wanted a 'head start' in ensuring AI adoption is well-controlled in the organisation. However, putting the framework into practice was challenging, as every stakeholder works in their silos with their agenda items.

> *"Developers say 'we have a vision for AI and we want all these data sources', but privacy officer declines. So, AI trials within our RAI framework was of little use as it was based on dummy data giving very different results."* (P7, DevOps Engineer, Pensions Fund, The Netherlands)

Having documented non-technical barriers and CDR's mediating effects, we next contextualise these findings within scholarly debates on AI governance and further reflect on this study's implications, limitations, and avenues for future research.

5. Discussion and Concluding Remarks

The findings to answer the main research purpose of this paper highlight nine non-technical barriers: (i) complexities in determining accountability; (ii) trickiness in preparing and managing unintended consequences quickly; (iii) difficulty in ensuring fairness and inclusion, particularly because defining fairness remains vague; (iv) limitations of existing business approaches, processes, and models to cope with AI; (v) difficulties in balancing trade-offs; (vi) managing AI's carbon impact; (vii) people's attitudes, skills, and cultures; (viii) managing unforeseen challenges when AI is operational; and (ix) securing investments for RAI. Approaches to address certain non-technical challenges, such as how many resources to dedicate to AI governance or whether the approach should be centralised or decentralised, were largely non-homogenous. Interestingly, addressing people-orientated barriers was identified as a high-impact solution in implementing RAI. Those executives from entities with CDR practices in place prioritised the need to resolve people-orientated barriers and sustainability considerations, acknowledging those as important elements of RAI success. Such prioritisation aligns with observations from Elliot et al. [5] that "CDR creates cultural change to avoid invoking "tick-box" compliance" (p. 185) and materialises Lewis's statement that "business will never be any more ethical than the people who are in the business," (Lewis [47]: 377).

The interview findings suggest that Responsible AI is non-negotiable but remains a work in progress in the financial services sector, with GenAI concerns superseding AI concerns. Moreover, responses reveal that non-technical barriers around GenAI are identified as trickier to address, given the inherent nature of the technology. Executives expressed a common approach for financial institutions to overcome such challenges by being fast-followers and not early-adopters, with terms such as "not being at the bleeding edge" and "wait, watch, and get on" (to mitigate risks) used during the interviews. Retaining human oversight was stressed as an imperative, and participants insisted on making AI a human augmenter and not a decision-maker to retain control; similar sentiments are raised

by Mikalef et al. [2] in their discussion of human oversight, highlighting the concept of "conjoined agency to balance the strengths of humans and machines in symbiotic relationships" (p. 263). In this regard, taking a staged approach can help in making Responsible AI digestible and actionable, according to certain interviewees.

It was also revealed that the technical and non-technical barriers are intertwined, requiring practitioners to understand and address both to effectively practice RAI. For example, fairness or transparency principles depend not only on technical features, such as explainability or algorithm monitoring, but also on non-technical aspects, such as AI skills and literacy and the organisation's policies on data input sources. Practitioners find technical challenges more tangible to address and adopt a broad-stroke AI governance approach. This is making AI challenges overwhelming with no clear starting point or step-by-step approach. Among the interviewees, those with technical jobs or security responsibilities tend to view RAI implementation through a strong technical lens, whereas those with CDR practices take a creative, idealistic, and thorough approach, indicating that for some organisations, making a start and learning along the way is more important, while for others, getting it right is a prerequisite to starting. A similar observation has recently been made by Wang et al. [20] in their study, highlighting how "organisational macro-motivators" such as "company culture and individuals" are "overlooked" aspects that motivate some companies to (de)prioritise RAI (Wang et al. [20]: 3). In a similar vein, as Lobschat et al. [27] emphasise that digital responsibility must be embedded into a company's digital strategy and culture, the paper corroborates this by showing that where CDR mindsets were institutionally reinforced, there was a stronger alignment between RAI intentions and governance practices. In addition, this study's findings align with and build on the growing body of the literature exploring the integration of ethical, social, and environmental considerations into AI governance. Specifically, our study empirically grounds the conceptual works of Elliot et al. [5] and Tóth and Blut [48], who propose that CDR offers a moral compass for AI deployment by extending traditional legal and risk frameworks and where "(...) accountability and the division of responsibilities should be made as clear as possible" so that financial services materially "demonstrate their commitment to responsible AI use via CDR" (Tóth and Blut [48]: 6–7). Moreover, the findings lend support to the view of Olatoye et al. [6], who argue that CDR and RAI can form a "symbiotic framework" primarily in sectors where high public accountability pressures and stakeholder expectations occur. Lastly, the need for interdisciplinarity defining the governance models in the digital economy [49] is reflected in the qualitative findings, demonstrating how cross-functional AI boards, anchored in CDR principles, may serve as viable governance mechanisms that bridge technological innovation with ethical deliberation. Taken together, these insights suggest that CDR is not merely supportive of RAI but can be foundational to its meaningful implementation, especially in sectors like finance, where trust, risk, and fairness are critical to legitimacy [1,50]. Rather than treating RAI and CDR as parallel or competing frameworks, this study proposes integrated models in which CDR functions as a practical mediator that translates ethical aspirations into organisational processes, cultural practices, and measurable outcomes [5,50]. In this respect, insights from these interviews lend support to previous warnings stressing that without "(...) a viable value proposition integrated into a company's overall goals and model, responsible practices are unlikely to be properly incentivised or prioritised against other competing demands" [51], which indicates that practitioners, even if they use all their subject-matter expertise to practice RAI, its effectiveness may be limited, given inhibitory factors and gaps setting back broader business model designs.

To demonstrate how CDR principles address the barriers identified in the interviews, in Table 2, we attempt to link each non-technical challenge to a corresponding CDR do-

main and illustrate examples of concrete mechanisms either explicitly described by participants or synthesised from closely related statements that emerged during thematic analysis and abstracted to reflect the underlying logic of the participants' responses (while aligning with recognised CDR concepts). Such mapping reinforces the role of CDR as a practical mediator of responsible AI adoption, with the CDR principles shown in Table 2 drawn from established frameworks found in the literature (e.g., Lobschat et al. [27]; Mihale-Wilson et al. [50]) and adapted inductively to reflect how practitioners can frame their own relevant digital governance practices.

Table 2. Mapping non-technical RAI challenges to CDR principles and organisational practices.

RAI Challenge-Barrier	Relevant CDR Principle	Operational Example
1. Accountability Ambiguity	Data Stewardship and Shared Responsibility	Creation of AI Governance Boards with clearly defined RACI matrices; each model assigned an accountable "owner".
2. Unintended Consequences	Ethical Risk Assessment	Mandatory pre- and post-deployment Ethical Impact Assessments (EIAs); simulation of GenAI misuse scenarios.
3. Fairness and Inclusion	Stakeholder Engagement and Digital Inclusion	Establishment of cross-departmental AI Stakeholder Councils to review model outputs for fairness and accessibility.
4. Legacy Processes	Adaptive Governance	Implementation of quarterly reviewed "CDR Playbooks" to update AI policies in line with fast-changing technologies.
5. Stakeholder Trade-Offs	Value-Conflict Resolution	Use of Value Impact Scorecards to compare competing priorities (e.g., privacy vs. efficiency) across departments.
6. Sustainability Concerns	Environmental Responsibility	Deployment of "AI Carbon Budgets"; routing model training to green data centres and optimising compute usage.
7. Human Factors and Skills Gaps	Culture and Training	Launch of an RAI–CDR Training Academy with mandatory modules tied to AI tool access and performance evaluations.
8. Operationalisation Beyond Day Zero	Life cycle Management	Formal adoption of "Day-Zero Plus" protocols with scheduled regular reviews for deployed models.
9. Budget Constraints	Transparent Ethics Investment	Integration of CDR-specific budget lines in AI project proposals, subject to ethical and environmental review.

This approach has several implications. From an academic standpoint, it invites scholars to deepen their focus into how varying organisational interpretations of CDR influence the scope, depth, and effectiveness of RAI initiatives across contexts [6,25]. It also demonstrates how high-level ethical guidelines translate (or fail to embed) into day-to-day governance challenges. By foregrounding the role of CDR as a governance lens, the rich, interview-based data illuminate that future research should develop more integrated, interdisciplinary models that weave together ethics, organisational behaviour, and information systems theory. Additionally, it empirically extends Fjeld et al.'s [1] "thorny gap" between RAI principles and practice by introducing a sector-specific taxonomy of critical barriers. Likewise, it addresses Merhi's [3] call for research on organisational dimensions of RAI and provides the first empirical evidence supporting Elliot et al.'s [5] proposition that CDR

"demystifies governance complexity", revealing that human-operational barriers dominate practitioner challenges (against an RAI literature that overemphasises technical solutions and persistently privileges concerns such as bias mitigation and explainability). In this respect, by documenting how GenAI supersedes traditional AI concerns due to its opacity and multimodal risks, this study responds to the OECD's [43] call for industry-specific risk assessments, positioning GenAI as a distinct research priority in finance.

In terms of practical implications in finance, it offers a nuanced roadmap for embedding RAI and suggests that AI governance strategies should be shaped within wider digital responsibility agendas, not simply tacked on as standalone ethical checklists, aligning with calls for value(s)-led digital transformation [27,52]. For instance, CDR's efficacy in resolving human-centric barriers (e.g., ethics training and consensus-building) offers a replicable model for better embedding RAI in the enterprise. Moreover, the suggested barrier taxonomy equips practitioners with a diagnostic tool to prioritise interventions and underscores the necessity of holistic governance architectures that balance innovation speed with robust risk controls to meet ESG commitments (such as carbon-footprint metrics for AI workloads to address critical underlying gaps in the current EU AI Act). The demonstrated value of CDR frameworks suggests that firms could achieve better alignment between ethical aspirations and operational realities by formally codifying espoused values around data stewardship, environmental impact, and employee empowerment and operationalising RAI through integrated policies (e.g., HR-legal alignment), urging its adoption beyond those pace-setting business entities. In practice, this means establishing cross-functional AI governance bodies, embedding and regularly conducting input/output audits, designing and applying clear responsibility matrices, and allocating budgetary and training resources for the post-implementation ("day-after") maintenance of AI systems. Organisations that adopt such a socio-technical approach are more likely better positioned to realise (Gen)AI's productivity and customer-service benefits while safeguarding trust, accountability, and legitimacy with governance protocols, such as cross-functional "risk rapid-response" teams, as guardrails to address emergent threats like algorithmic creativity misuse.

Beyond individual business entities, findings from research endeavours, such as ours, have broader social resonance as they stimulate policymakers to incorporate CDR dimensions and non-technical factors into emerging AI regulatory frameworks [24,53,54]. As financial institutions deploy AI/GenAI at scale, failures in fairness, trust, transparency, or sustainability can erode public trust, exacerbate inequalities, and contribute to environmental externalities. The risk of a "digital divide" (whereby certain employee cohorts may gain AI fluency and decision-making power) mirrors larger societal concerns about digital exclusion and/or algorithmic bias, underscoring a need for bottom-up participatory RAI design by engaging civil society in key interventions to prevent marginalisation. By advocating for a human-centric, consensus-driven CDR orientation, this study amplifies the social imperative that AI systems serve collective well-being, not just organisational optimisation. Policymakers can draw on such insights to enact regulations and disseminate best-practice guidelines that mandate organisational readiness metrics that ensure AI's benefits accrue across society and communities in a transparent (e.g., explainable credit denials), equitable (e.g., mitigating GenAI's language biases), and sustainable (e.g., enforcing carbon standards for AI infrastructure) manner, and thus, ensure social acceptability and support.

Indeed, by its nature, qualitative research is open-ended and subject to bias, as "...researchers need to remember that patterns discovered in such data may come from informants as well as from investigators' recording biases" (Ryan and Bernard [38]: 100). Although care was taken, Ryan and Bernard call out that the researcher acts as a subconscious theme filter, potentially limiting findings. The amount of data collected has been

overwhelming, distracting, and sometimes difficult to analyse systematically. Efforts to avoid "describing the data rather than analysing it" (Bell et al. [33]: 530) also helped navigate through data that are interesting but not relevant to the research objective. Interviews depend on participants' own expertise, experiences, and priorities, as well as what strikes them as significant or important, which is decided by the researcher. The small sample size is practical for this study, but the scope/scale of the findings can be limited. However, this study aims to bring diverse viewpoints from participants with deeper practical knowledge and those with direct stake/engagement in the topic. While the value of small samples in making "empirical generalisations" is questioned (Bell et al. [33]: 376), the value of important inferences derived from the qualitative data cannot be discounted. Defending qualitative interviews and case studies in generalisation, Flyvbjerg [55] highlights the importance of being "problem driven and not methodology driven" (p. 242) and sees methodology as the best available means to "best help answer the research questions at-hand".

Thus, while this study offers novel insights into emerging AI domains, it is not without limitations. First, the sample was restricted to 15 professionals from European financial institutions, selected through purposive-convenience sampling. Although participants occupied diverse roles (e.g., AI leads, risk managers, and ethics officers), their perspectives may reflect organisational or cultural biases specific to the financial sector in Western regulatory settings. This limits the transferability of findings to other regions or industries with different governance dynamics or levels of digital maturity. Second, reliance on a single qualitative data source, i.e., semi-structured interviews, means that we were unable to triangulate themes with other forms of evidence, such as organisational documentation, AI policy artefacts, or direct observation. While we employed reflexive thematic analysis and maintained an audit trail of coding decisions, we acknowledge the absence of formal cross-validation, which may affect the confirmability of interpretations. Third, gender and positional diversity within the sample was uneven; the majority of interviewees were male and held senior positions. This may have led to an underrepresentation of frontline or non-dominant voices, particularly with regard to fairness, inclusion, or human-centric governance concerns. We encourage future studies to intentionally oversample underrepresented groups to broaden the ethical and experiential scope of relevant insights.

Future research should build on these findings in several ways. First, scholars could confirm and expand on the non-technical barriers identified and extend the scope of inquiry beyond the financial sector to examine how RAI frameworks are applied (or struggle to gain traction) in other domains such as healthcare, manufacturing, or the public sector. Comparative, cross-sectoral studies would help assess the generalisability and contextual specificity of CDR-informed RAI governance models, as it is essential to "include a greater diversity of actors in AI governance activities" (Attard-Frost and Lyon [56]: 17). Second, empirical studies could further develop and test the operational components of CDR by using quantitative or mixed-method approaches. For instance, surveys, action research, or longitudinal case studies could evaluate how specific CDR mechanisms (e.g., AI carbon footprint management, stakeholder management) impact ethical outcomes, trust, or innovation performance across firms. In this respect, researchers could also examine how evolving legal frameworks (e.g., the EU AI Act) influence the practical adoption of CDR principles and whether they promote or hinder human-centric governance in AI ecosystems. Third, there is a need for deeper theoretical exploration of the interplay between RAI and CDR, particularly regarding how organisational culture, leadership, and regulation mediate their integration.

By acknowledging the aforementioned limitations and outlining targeted paths forward, studies such as ours lay the groundwork for a more rigorous and empirically grounded understanding of complexities in AI deployment, aligning technological ambi-

tion with responsibility principles and ensuring not only compliance but also legitimacy and trust in the digital age.

Author Contributions: Conceptualisation, A.V.; methodology, A.S. and A.V.; investigation, A.V.; writing—original draft preparation, A.S.; writing—review and editing, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the guidelines of the Declaration of Helsinki and approved by the Research Ethics Committee of the University of Reading (ref. code: SREC-HBS-20240928-ARVE1689; date of approval: 30 September 2024).

Informed Consent Statement: Informed consent was obtained from all executives participating in this study.

Data Availability Statement: The raw data supporting the findings of this article can be made available by the authors upon reasonable request.

Acknowledgments: We are grateful to the three anonymous reviewers for their constructive comments and recommendations and to the 15 participants, without whom this study would not have been possible.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. Interview Guide/Protocol for Semi-Structured Interviews

Thank you for participating in this research. As explained in the Information Sheet, this study explores one of the pressing industry topics today—AI. Much is said about its value and pitfalls and the importance of Responsible AI. This research particularly focuses on the financial services sector and aims to identify the non-technical barriers in taking Responsible AI from theory to practice.

I am honoured that you agreed to be interviewed, as your views will help us improve our research insights.

Would it be OK to record or transcribe this interview, please?

Opening questions:

Q. Could you please tell me about yourself and about your organisation?

Q. How is your organisation using or planning to use AI or GenAI?

Q. What are your and your team's roles in your organisation's AI strategies?

Topic-orientated questions:

Q. Could you share what your organisation's vision is about using AI and GenAI?

Q. What are the use cases, and how strategic, according to you, are AI and GenAI for your organisation?

Q. Do you and others in your organisation have concerns about using AI or GenAI Responsible AI?

Responsible AI is a formal approach to guide the implementation of AI methods at scale. Responsible AI guidelines not only focus on technical capability but also include considerations around fairness, accountability, and privacy. It has become more commonly accepted to guide AI implementation.

Q. Are there any Responsible AI frameworks or guidelines that your organisation is using?

Follow-on questions on why a framework was selected:

Q. What is the goal of your Responsible AI journey?

Q. Who are the stakeholders that determine the AI governance or Responsible AI approach in your organisation?

Further probing questions to understand practical limitations of the frameworks:

Q. How successful have you been so far in implementing Responsible AI principles?

Pauses and echo probing:

Practical Barriers:

Q. What, according to you, are the practical challenges in taking the Responsible AI guidelines to practice?

Follow-up question on why:

Q. Which precisely are the non-technical barriers?

Follow-up questions to probe and have a deeper understanding of the non-technical barriers and challenges in practicing Responsible AI:

Q. Are they the same for AI and GenAI, or are they considerably different?

Echo probing on non-technical barriers for GenAI and identifying new or emerging challenges:

Q. Why have you cited these, and could you give 1–2 examples to bring these to light, please?

Understand the examples to illustrate practical challenges.

Q. According to you, how are these non-technical barriers restricting your organisation from meeting your AI/GenAI goals?

Q. Are the non-technical barriers widely understood in the organisation?

Pauses and repetition to establish whether there is a shared understanding of the barriers

Measures to address the barriers, action so far, and results:

Q. What is the benefit of addressing the non-technical barriers that you have highlighted?

Q. How will that help in taking Responsible AI from theory to practice?

Q. Among the highlighted non-technical barriers, which are harder to solve?

Q. Why?

Pauses and echo probing:

Follow-up question on whether addressing any barrier will have a multiplier effect:

Q. According to you, resolving which type of barriers (technical, legal, or non-technical) is a top priority for your organisation, and why?

Q. Who is/should take accountability?

Q. Could you give some examples of how you and other teams are working towards overcoming the non-technical barriers? E.g., are you hiring skilled people with governance experience, and are you increasing budgets for AI training and awareness, etc.?

Probing questions to understand the starting point or strategies to address the non-technical challenges:

Q. Is your organisation tweaking the Responsible AI principles to specifically include and address the non-technical barriers?

Corporate Digital Responsibility and the Link between CDR and RAI:

Q. Does your organisation practice or have Corporate Digital Responsibility (CDR) strategies or a charter?

If not,

Q. Are you aware of CDR? Will you be exploring it to address your non-technical barriers?

If yes,

Q. Has your organisation considered using CDR elements to address the RAI barriers that you have highlighted?

If yes, probing questions to understand the hows and whys and to help compare responses:

Q. In your opinion, does CDR have a strong role/potential in making Responsible AI more actionable?

Follow up on how and why, with an example:

Open-ended Q&A

Q. Is there anything else you would like to highlight that we may not have covered but you think will be relevant for my research topic?

Pause:

Q. Are you anticipating any future changes to RAI that will have a big impact?

Q. What is your personal view/vision/objective/observation on non-technical challenges of RAI?

Debriefing:

Thank you very much for sharing your views and experiences in our interview. I have no further questions. If there is nothing else from your side too, then I would like to take this opportunity to thank you once again for your inputs.

References

1. Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.C.; Srikumar, M. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research Publication No. 2020-1. 2020. Available online: https://ssrn.com/abstract=3518482 (accessed on 7 July 2024).
2. Mikalef, P.; Conboy, K.; Lundström, J.E.; Popovič, A. Thinking responsibly about responsible AI and 'the dark side' of AI. *Eur. J. Inf. Syst.* **2022**, *31*, 257–268. [CrossRef]
3. Merhi, M.I. An Assessment of the Barriers Impacting Responsible Artificial Intelligence. *Inf. Syst. Front.* **2022**, *25*, 1147–1160. [CrossRef]
4. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
5. Elliott, K.; Price, R.; Shaw, P.; Spiliotopoulos, T.; Ng, M.; Coopamootoo, K.; van Moorsel, A. Towards an Equitable Digital Society: Artificial Intelligence (AI) and Corporate Digital Responsibility (CDR). *Society* **2021**, *58*, 179–188. [CrossRef]
6. Olatoye, F.O.; Awunga, K.F.; Mhlongo, N.Z.; Ibeh, C.V.; Elufioye, O.A.; Ndubuisi, N.L. AI and ethics in business: A comprehensive review of responsible AI practices and corporate responsibility. *Int. J. Sci. Res. Arch.* **2024**, *12*, 1433–1443. [CrossRef]
7. Riemer, S.; Strauß, M.; Rabener, E.; Bickford, J.K.; Hilbers, P.; Kalra, N.; Kapoor, A.; King, J.; Palumbo, S.; Pardasani, N.; et al. A Generative AI Roadmap for Financial Institutions. 2023. Available online: https://www.bcg.com/publications/2023/a-genai-roadmap-for-fis (accessed on 15 July 2024).
8. HSBC. Harnessing the Power of AI to Fight Financial Crime. 2024. Available online: https://www.hsbc.com/news-and-views/views/hsbc-views/harnessing-the-power-of-ai-to-fight-financial-crime (accessed on 7 September 2024).
9. Bhatnagar, S.; Mahant, R. Unleashing the Power of AI in Financial Services: Opportunities, Challenges, and Implications. *Int. J. Adv. Res. Sci. Commun. Technol.* **2024**, *4*, 439–448. [CrossRef]
10. Ali, H.; Aysan, A.F. What will ChatGPT revolutionize in the financial industry? *Mod. Financ.* **2023**, *1*, 116–129. [CrossRef]
11. Chui, M.; Roberts, R.; Yee, L. Generative AI Is Here: How Tools Like ChatGPT Could Change Your Business. McKinsey & Company 2023. Available online: https://www.mckinsey.com/capabilities/quantumblack/our-insights/generative-ai-is-here-how-tools-like-chatgpt-could-change-your-business (accessed on 15 July 2024).
12. Botunac, I.; Parlov, N.; Bosna, J. Opportunities of Gen AI in the Banking Industry with regards to the AI Act, GDPR, Data Act and DORA. In Proceedings of the 2024 13th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 11–14 June 2024.
13. McKinsey. Setting Up Generative AI Pilots Is Easy; Scaling Them to Capture Material Value Is Hard. A Recipe for Success Is Emerging. 2023. Available online: www.mckinsey.com/industries/financial-services/our-insights/capturing-the-full-value-of-generative-ai-in-banking (accessed on 6 September 2024).
14. McCarthy, J. *What Is Artificial Intelligence*; Formal Reasoning Group: Stanford, CA, USA; Stanford University: Stanford, CA, USA, 2007.
15. Robertson, A. Google Apologizes for 'Missing the Mark' After Gemini Generated Racially Diverse Nazis. *The Verge*. 2024. Available online: https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical (accessed on 21 February 2024).
16. Kenthapadi, K.; Lakkaraju, H.; Rajani, N. Generative AI meets Responsible AI: Practical Challenges and Opportunities. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), Long Beach, CA, USA, 6–10 August 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 5805–5806.
17. Schiff, D.; Rakova, B.; Ayesh, A.; Fanti, A.; Lennon, M. Principles to Practices for Responsible AI: Closing the Gap. *arXiv* **2020**. [CrossRef]

18. Rakova, B.; Yang, J.; Cramer, H.; Chowdhury, R. Where Responsible AI meets Reality. *Proc. ACM Hum.-Comput. Interact.* **2021**, *5*, 1–23. [CrossRef]
19. Dhake, S.P.; Lassi, L.; Hippalgaonkar, A.; Gaidhani, R.A.; Jyothi, N.M. Impacts and Implications of Generative AI and Large Language Models: Redefining Banking Sector. *J. Inform. Educ. Res.* **2024**, *4*, 248–257. [CrossRef]
20. Wang, A.; Datta, T.; Dickerson, J.P. Strategies for increasing corporate responsible AI prioritization. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, Santa Clara, CA, USA, 21–23 October 2024; Volume 7, pp. 1514–1526.
21. Powell, J.; Kleiner, A. *The AI Dilemma: 7 Principles for Responsible Technology*, 1st ed.; e-book; Berrett-Koehler Publishers: Oakland, CA, USA, 2023.
22. Tornatzky, L.G.; Fleischer, M.; Chakrabarti, A.K. *The Processes of Technological Innovation*; Lexington Books: Lanham, MD, USA, 1990.
23. WEF. The Presidio Recommendations on Responsible Generative AI. 2023. Available online: www.weforum.org/publications/the-presidio-recommendations-on-responsible-generative-ai/ (accessed on 14 March 2025).
24. WEF. AI for Impact: The PRISM Framework for Responsible AI in Social Innovation. 2024. Available online: https://www.weforum.org/publications/ai-for-impact-the-prism-framework-for-responsible-ai-in-social-innovation/ (accessed on 26 April 2025).
25. de Sio, F.S.; Mecacci, G. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philos. Technol.* **2021**, *34*, 1057–1084. [CrossRef]
26. Bietti, E. From Ethics Washing to Ethics Bashing: A view on tech ethics from within moral philosophy. In Proceedings of the FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; pp. 210–219.
27. Lobschat, L.; Mueller, B.; Eggers, F.; Brandimarte, L.; Diefenbach, S.; Kroschke, M.; Wirtz, J. Corporate digital responsibility. *J. Bus. Res.* **2021**, *122*, 875–888. [CrossRef]
28. Mueller, B. Corporate Digital Responsibility. *Bus. Inf. Syst. Eng.* **2022**, *64*, 689–700. [CrossRef]
29. Office of the CDR Initiative. Taking on Responsibility Together—The CDR Initiative. Office of the CDR Initiative, Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV). Available online: https://cdr-initiative.de/en/initiative (accessed on 21 October 2024).
30. Edelman. 2019 Edelman Trust Barometer. 2019. Available online: https://www.edelman.com/sites/g/files/aatuss191/files/2019-02/2019_Edelman_Trust_Barometer_Global_Report.pdf (accessed on 16 September 2024).
31. Cheng, C.; Zhang, M. Conceptualizing Corporate Digital Responsibility: A Digital Technology Development Perspective. *Sustainability* **2023**, *15*, 2319. [CrossRef]
32. Holstein, K.; Vaughan, J.W.; Ill, H.D.; Dudik, M.; Wallach, H. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? In Proceedings of the CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; Association of Computing Machinery: Cambridge, MA, USA, 2019; pp. 1–16.
33. Bell, E.; Bryman, A.; Harley, B. *Business Research Methods*, 6th ed.; Oxford University Press: Oxford, UK, 2022.
34. Azungah, T. Qualitative research: Deductive and inductive approaches to data analysis. *Qual. Res. J.* **2018**, *18*, 383–400. [CrossRef]
35. Braun, V.; Clarke, V. *Thematic Analysis: A Practical Guide*, 1st ed.; SAGE Publications Ltd.: London, UK, 2022; ISBN 9781473953246.
36. Ryan, G.W.; Bernard, H.R. Techniques to Identify Themes. *Field Methods* **2003**, *15*, 85–109. [CrossRef]
37. European Commission. *Women in Digital Scoreboard 2024*; Publications Office of the European Union: Luxembourg, 2024.
38. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
39. Sekaran, U.; Bougie, R. *Research Methods for Business: A Skill-Building Approach*, 7th ed.; John Wiley & Sons: Chichester UK, 2016.
40. Hennink, M.M.; Kaiser, B.N.; Marconi, V.C. Code saturation versus meaning saturation: How many interviews are enough? *Qual. Health Res.* **2017**, *27*, 591–608. [CrossRef]
41. Bank of England. How Will Increasing Business Use of Artificial Intelligence (AI) Affect UK Labour Demand? 2024. Available online: https://www.bankofengland.co.uk/bank-overground/2024/how-will-increasing-business-use-of-ai-affect-uk-labour-demand?sf203014121=1 (accessed on 22 November 2024).
42. BBC News. Bacon Ice Cream and Nugget Overload Sees Misfiring McDonald's AI Withdrawn. *BBC News.* 2024. Available online: https://www.bbc.com/news/articles/c722gne7qngo (accessed on 29 November 2024).
43. OECD. *Generative Artificial Intelligence in Finance*; OECD Artificial Intelligence Paper 9; OECD Publishing: Paris, France, 2016.
44. Capraro, V.; Lentsch, A.; Acemoglu, D.; Akgun, S.; Akhmedova, A.; Bilancini, E.; Bonnefon, J.-F.; Brañas-Garza, P.; Butera, L.; Douglas, K.M.; et al. The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS Nexus* **2024**, *3*, 191. [CrossRef] [PubMed]
45. Amnesty International. Dutch Childcare Benefit Scandal an Urgent Wake-Up Call to Ban Racist Algorithms. 2021. Available online: https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/ (accessed on 29 November 2024).
46. OECD. *Advancing Accountability in AI: Governing and Managing Risks Throughout the Lifecycle for Trustworthy AI*; OECD Digital Economy Paper 349; OECD Publishing: Paris, France, 2023.
47. Lewis, F.V. Defining 'Business Ethics': Like Nailing Jello to a Wall. *J. Bus. Ethics* **1985**, *1*, 377–383. [CrossRef]

48. Tóth, Z.; Blut, M. Ethical compass: The need for Corporate Digital Responsibility in the use of Artificial Intelligence in financial services. *Organ. Dyn.* **2024**, *53*, 101041. [CrossRef]
49. Mihale-Wilson, C.; Hinz, O.; van der Aalst, W.; Weinhardt, C. Corporate digital responsibility: Relevance and opportunities for business and information systems engineering. *Bus. Inf. Syst. Eng.* **2022**, *64*, 127–132. [CrossRef]
50. European Commission. *Proposal for a Regulation Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) (COM(2021)206)*; European Commission: Brussels, Belgium, 2021.
51. Akbarighatar, P. Operationalizing responsible AI principles through responsible AI capabilities. *AI Ethics* **2025**, *5*, 1787–1801. [CrossRef]
52. Mihale-Wilson, C.; Zibuschka, J.; Carl, C.; Hinz, O. Corporate digital responsibility—Extended conceptualization and empirical assessment. In Proceedings of the 29th European Conference on Information Systems (ECIS 2021), Marrakech, Morocco, 14–16 June 2021; Available online: https://aisel.aisnet.org/ecis2021_rp/80 (accessed on 8 June 2025).
53. WEF. From Climate to Coding, AI's Impact Is Ramping Up. These 7 Principles Ensure It Remains Human-Centric. 2024. Available online: www.weforum.org/stories/2024/01/7-principles-integrate-artificial-intelligence-impact/ (accessed on 4 December 2024).
54. Binns, R.; Van Kleek, M.; Veale, M.; Lyngs, U.; Zhao, J.; Shadbolt, N. 'It's Reducing a Human Being to a Percentage'—Perceptions of Justice in Algorithmic Decisions. In Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018.
55. Flyvbjerg, B. Five Misunderstandings About Case-Study Research. *Qual. Inq.* **2006**, *12*, 219–245. [CrossRef]
56. Attard-Frost, B.; Lyons, K. AI governance systems: A multi-scale analysis framework, empirical findings, and future directions. *AI Ethics* **2024**, *5*, 2557–2604. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.