Fairness and Accountability in AI-Driven Corporate Social Responsibility: Insights from Small and Medium-Sized Enterprises

Dorleta Urrutia-Onate[1,*,†], Julen Bollain[2,†], Enrique Onieva[3,†] and Asier Perallos[4,†]

[1]*University of Deusto, Bilbao, País Vasco, ES*
[2]*Business Faculty, Department of Economics and Finance, Mondragon University, Oñati, País Vasco, ES*
[3]*University of Deusto, Bilbao, País Vasco, ES*
[4]*University of Deusto, Bilbao, País Vasco, ES*

Abstract

The rapid advancement of Artificial Intelligence (AI) for Corporate Social Responsibility (CSR) intersects increasingly with fairness and accountability challenges, particularly under evolving European regulations such as the AI Act and the Corporate Sustainability Reporting Directive (CSRD). This article offers a qualitative exploration of AI-driven CSR initiatives within varied organizational settings, with special focus on small and medium-sized enterprises (SMEs). Through semi-structured interviews and document analysis, we identify how limited resources, data governance complexities, and lack of in-house AI expertise can constrain fairness and interpretability goals, underscoring the need for accessible, "white-box" solutions. Additionally, we examine how human oversight and explainable AI (XAI) frameworks foster stakeholder trust and ethical alignment, turning AI into a potential strategic differentiator in socially conscious markets. Our findings highlight that embedding fairness-oriented design, robust data governance, and co-regulatory support—particularly for resource-constrained firms—are critical for reconciling algorithmic innovation with societal expectations. In doing so, the study advances interdisciplinary dialogue on AI fairness, proposing tailored strategies that integrate technical, cultural, and policy dimensions to ensure AI solutions remain transparent, inclusive, and equitable.

Keywords

AI Fairness, CSR, Explainable AI (XAI), SMEs, Data Governance, European Regulatory Frameworks,

1. Introduction

In recent years, Artificial Intelligence (AI) has gained traction as a crucial driver for organizational innovation and competitiveness, particularly in the area of Corporate Social Responsibility (CSR). Scholars have highlighted that AI can help organizations streamline resource utilization, reduce carbon footprints, and improve social impact by processing data sets in real time and automating complex decision-making [1]. However, maximizing the transformative potential of AI for CSR requires navigating an increasingly complex environment—one marked by rising ethical, regulatory, and technological pressures. Issues such as AI bias, algorithmic transparency, and risk-based governance have emerged, demanding robust frameworks for responsible innovation [2].

Although advanced data analytics and machine learning promise significant gains—from monitoring resource consumption to guiding ethical decision-making—organizational readiness, interpretablity, and compliance hurdles remain. Such tensions become especially acute for small and medium-sized enterprises (SMEs), which frequently face tighter budgets and limited AI expertise [3]. Recent European initiatives— including the Artificial Intelligence Act (AIA), Non-Financial Reporting Directive (NFRD),

AIMMES 2025 Workshop on AI bias: Measurements, Mitigation, Explanation Strategies $|$ co-located with EU Fairness Cluster Conference 2025, Barcelona, Spain
*Corresponding author.
†These authors contributed equally.
dorleta.urrutia@opendeusto.es (D. Urrutia-Onate); jbollain@mondragon.edu (J. Bollain); enrique.onieva@deusto.es (E. Onieva); perallos@deusto.es (A. Perallos)
0000-0002-6862-6160 (D. Urrutia-Onate); 0000-0002-1298-5666 (J. Bollain); 0000-0001-9581-1823 (E. Onieva); 0000-0001-9661-3861 (A. Perallos)

and Corporate Sustainability Reporting Directive (CSRD)—further shape how AI can be deployed responsibly [4]. Although these measures aim to foster ethical and transparent systems, they impose additional conformity assessments, often challenging resource-constrained organizations [5].

Consequently, the present article examines how AI intersects with CSR under the realities of evolving governance and fairness demands, focusing particularly on SMEs. By analyzing the barriers and opportunities identified through qualitative inquiry, we seek to offer strategic and context-sensitive insights that integrate AI ethics, business objectives, and public policy objectives.

Following this Introduction, the Literature Review explores the ways AI fosters proactive CSR initiatives, addressing both data-driven benefits and the intricacies of European governance. The Methodology section outlines the qualitative approach, including semi-structured interviews, while the Results and Discussion highlight emergent themes—ranging from data governance to organizational readiness to XAI complexities. Finally, the Conclusions synthesize the overall contributions and discuss future directions for AI-driven CSR research.

2. Literature Review

Contemporary scholarship increasingly frames AI as a catalyst for CSR innovation, suggesting that advanced data analytics and machine learning can strengthen sustainable operations and address social challenges in more proactive ways [1]. From reducing resource consumption to guiding ethical decision-making, AI-powered systems provide a level of precision and scalability that manual approaches struggle to match [2].

For example, AI-powered Energy Management Systems (EMS) have demonstrated significant cost savings and a reduction in environmental footprint by optimizing energy consumption. A notable case is a manufacturing plant that implemented an AI-based EMS and achieved a 15% reduction in energy consumption within the first year of operation [6]. Blockchain integrated with AI enables end-to-end supply-chain traceability, ensuring labor and environmental standards are upheld and transparently reported [1]. Meanwhile, IoT sensors augmented with AI can deliver real-time insights on energy use or emissions, thereby expanding the organization's accountability to stakeholders [7]. Taken together, these applications illustrate how AI can overcome limitations of traditional CSR methods—often limited by fragmented data and delayed reporting—by promoting automated monitoring and continuous engagement with sustainability targets.

Beyond operational gains such as lower costs or reduced carbon footprints, AI integration in CSR provides intangible benefits—notably improved corporate reputation, heightened stakeholder trust, and stronger internal morale [8]. Transparent, data-driven metrics on progress toward social or environmental goals can reassure stakeholders that the company's stated commitments align with measurable action. In many cases, AI helps mitigate skepticism by offering real-time dashboards and automated audits that make social or environmental impacts clear and verifiable [9] . Employees, too, may experience greater motivation and job satisfaction when their organization demonstrates consistent ethical values and quantifiable sustainability outcomes [8]. Furthermore, advanced AI techniques such as Long Short Term Memory (LSTM) networks combined with Empirical Mode Decomposition (EMD) have demonstrated how complex sustainability indicators can be evaluated with greater precision, thus laying the groundwork for more accurate strategic decisions [10]. Altogether, these intangible benefits can contribute to both talent retention and competitive positioning—vital for businesses operating in socially conscious markets.

Despite the promise of AI-driven CSR, significant challenges persist. One major hurdle involves data governance, specifically balancing data-driven insights with privacy, security, and consent [9]. AI-based analyses often require substantial personal or sensitive information, raising ethical dilemmas around how these data are collected, stored, or shared. Another challenge revolves around organizational readiness, as many companies— especially SMEs—lack the technical workforce or financial resources to invest in robust AI infrastructures [5]. Resistance to change can compound these issues, with employees skeptical of algorithmic decision-making and managers uncertain about the return on AI investments [1].

Furthermore, explainable AI (XAI) metrics have gained importance as a way to ensure that algorithmic processes remain interpretable and trustworthy to both internal stakeholders and external regulators, yet implementing such metrics remains technically complex [10].

Building on the ethical and operational dimensions of AI-driven CSR, alignment with evolving regulatory frameworks emerges as another critical layer—especially in Europe, where legislators increasingly emphasize sustainability, ethics, and risk-based governance [4]. The AIA, building on data protection regulations like GDPR, seeks to harmonize AI oversight by classifying systems according to risk levels and mandating transparency and accountability—particularly challenging for SMEs with fewer technical and financial resources [11].

In parallel, AI-focused policies frequently intersect with CSR directives such as the NFRD and the CSRD. These demand clear, consistent disclosures on environmental and social impacts [12]. While these regulations reinforce responsible innovation and enhance stakeholder trust, they also bring added compliance hurdles—ranging from overlapping cybersecurity mandates to detailed conformity assessments—that risk hindering SME participation [5].

While these regulatory frameworks aim to foster responsible innovation, smaller firms often find compliance especially daunting due to their limited technical and legal resources [13]. Hence, the transformative potential of AI in CSR hinges not just on technological capability but also on regulatory alignment and organizational strategy— underscoring the need for context-sensitive models that integrate AI ethics, business objectives, and public policy aims [3].

Synthesizing these points, the literature converges around three core themes: (1) AI's capacity to optimize resource use and amplify stakeholder trust, (2) implementation hurdles involving data governance, expertise, and organizational culture, and (3) the intensification of regulatory and ethical demands, particularly acute for SMEs. These themes collectively frame the complex environment where AI meets CSR, establishing a need for empirical examination and strategic frameworks that can guide practitioners toward sustainable and responsible AI- based innovation.

3. Methodology

This study adopts a qualitative paradigm underpinned by an interpretive (or interpretivist) approach, augmented by phenomenological elements [14]. Qualitative inquiry seeks to understand reality based on the constructions created by the subjects involved [15]. Our focus on meanings, experiences, and social realities regarding AI adoption in CSR aligns well with this paradigm, which values contextualization, immersion in natural settings, and attention to insiders' perspectives [16]. Moreover, phenomenological elements guide our in-depth exploration of how participants live and experience AI in ways that illuminate "the essence" [17] of their engagement with CSR and technology.

3.1. Design and Data Collection Strategy

3.1.1. Semi-Structured Interviews

Semi-structured interviews [18] were selected, aimed at professionals in CSR and finance. Such interviews allow for a combination of predefined questions with the freedom to delve into emergent aspects and explore meaningful narratives for participants [16]. According to Seale (1999), a variety of perspectives helps strengthen credibility by generating "multiple angles of analysis" [19].

The interview guide includes questions oriented toward:

- AI Experience and General Perception: Inspired by Agee [18], includes questions such as "How would you describe the current level of AI adoption in your organization?" as an initial overview.
- View of CSR and the Role of AI: Drawing on interpretive concepts, it aims to capture subjectivity: "How do you see the balance between meeting sustainability regulations and seeking broader societal impact?"

- Barriers and Expectations: Explores technological and ethical complexities, allowing deeper investigation of meanings and values. "Which ethical, technological, or cultural barriers have you encountered in AI adoption?"
- Explainability and Transparency: It explores aspects of "sincerity" and "credibility" [20], asking: "What do you think could help build more trust in AI systems within your organization?"

3.1.2. Ensuring Diversity in Sampling

To guarantee a meaningful range of perspectives on AI-driven CSR, we intentionally sought diversity along three dimensions: organizational size, sector, and level of digital maturity. First, we included both SMEs and larger firms, allowing us to observe how limited resources might affect AI investments and governance practices. Second, we approached multiple sectors (finance, manufacturing, and technology), recognizing that each domain could yield distinctive data-usage patterns and regulatory exposures. Third, we targeted participants exhibiting varying stages of digitalization—from early adopters experimenting with basic automation to more mature organizations employing advanced, explainable AI solutions. By actively combining these criteria when recruiting participants, we aimed to maximize heterogeneity and bolster the transferability of our findings, in line with interpretive principles [16] and qualitative best practices [21].

Participants were informed about the academic purpose, voluntary nature, and confidentiality of the study [21]. Access was facilitated via professional networks and institutional collaborations at Mondragon Unibertsitatea Enpresagintza, ensuring a purposeful sample.

3.1.3. Complementary Observations and Document Analysis

Although interviews serve as the main data collection method, non-participant observation (or indirect observation) was also considered, focusing on materials in corporate networks and forums, as well as organizational document analysis (CSR policies, internal manuals, etc.). This approach allows data triangulation [22] or, in postmodern terms, crystallization, fostering "expanded perspectives" [23] and ensuring a broad variety of evidence that complements participant testimonies.

3.2. Data Analysis

3.2.1. Transcription and Repeated Reading

All interviews were transcribed verbatim to capture communicative nuances [18]. Transcripts were then read repeatedly, allowing the researcher to become intimately familiar with participants' accounts. Each transcript was annotated with fieldnote comments, marking initial impressions and potentially significant quotations [16].

3.2.2. Open Coding and Categorization

In line with Strauss and Corbin's grounded theory tradition and interpretivist principles [24], we conducted open coding to locate emergent themes:

1. Initial Open Coding
   a) The first pass identified descriptive codes (e.g., "lack of digital skills," "pilot approach," "explainability," "data governance").
   b) A second pass refined these into more analytic codes, linking them to underlying concepts such as "fear of job displacement" or "fragmentation in AI adoption."
2. Focused Coding and Thematic Grouping
   a) Similar codes were clustered into broader categories or subthemes (e.g., "Knowledge Gaps and Resistance," "Importance of Data Quality and Governance").
   b) Iterative comparison of transcripts ensured that categories were not forced but rather reflected participants' own language and repeated patterns (Charmaz, 2014).

3. Establishing Thematic Relationships The final stage involved examining how categories related or informed each other (e.g., linking "Data Quality" with "Human Oversight and Explainability"). We revisited the transcripts throughout, thereby preserving interpretive fidelity and groundedness in participant testimonies.

Through this iterative and reflective process, six major categories emerged: (1) Fragmented and Incremental Adoption, (2) Knowledge Gaps and Resistance, (3) Potential Synergies with CSR, (4) Importance of Data Quality and Governance, (5) Value of Human Oversight and Explainability, and (6) Strategic Emphasis on Responsible AI.

3.3. Ensuring Methodological Quality

A concern for quality runs throughout the research process, combining several principles:

1. **Worthy Topic**: AI adoption in CSR is a timely and high-impact issue [20].
2. **Rich Rigor**: Combining semi-structured interviews, document analysis, and complementary observations enriches the variety of sources, ensuring "requisite variety" [16].
3. **Sincerity**: Continuous reflexivity regarding the researcher's position and potential biases. Careful documentation of each step and the challenges faced [19].
4. **Credibility**: Strengthened through thick description, triangulation, and the use of *member reflections* [20].
5. **Resonance**: The final presentation will be designed to be "meaningful" and "evocative". Transferability will be promoted with detailed narratives and concrete examples. [25]
6. **Significant Contribution**: The study seeks to offer new insights to the academic and professional community, contribute to designing CSR-related AI policies, and open new lines of research [18].
7. **Ethics**: In keeping with Ellis and university regulations, confidentiality and informed consent agreements are respected [26]. Situational and relational ethics are practiced, attending to relational impact with participants [27].
8. **Meaningful Coherence**: Each methodological step aligns with the interpretive and phenomenological perspective, ensuring that conclusions interconnect in a coherent way with the stated goals [20].

By incorporating diverse organizational sizes and digital maturity levels, the study's purposeful sampling sought to capture multi-faceted perspectives on AI-driven CSR—illuminating how constraints and opportunities materialize across different contexts. This design, combined with iterative analysis and reflexive rigor, supports the aim of generating context-sensitive findings that can inform both scholarship and practice.

4. Findings and Discussion

Following qualitative best practices [21], we present findings and interpretive discussion together to maintain a close linkage between empirical observations and theoretical or practical implications. This unified approach allows each theme to be immediately contextualized, clarifying both the data and its significance for AI-driven CSR—particularly in SMEs.

4.1. Overview of Participant Profiles

This section presents the main findings derived from the semi-structured interviews conducted with financial and strategic leaders responsible for adopting AI in their organizations. The presentation follows a thematic structure, detailing participants' profiles and the key themes that emerged from the data. Quotes or references to participant testimonies appear as E1–E7, in line with anonymization protocols.

1. **E1**: Financial manager in an industrial firm overseeing digital transformation and the implementation of AI-based systems to optimize financial and administrative processes.
2. **E2**: Financial lead in a higher-education institution, promoting sustainable values within strategic and economic management.
3. **E3**: Specialist in AI adoption with prior experience in sustainability consulting, advising companies from multiple sectors on RSE-related AI initiatives.
4. **E4**: Financial manager in an industrial cooperative, focusing on operational efficiency and modernization through automation.
5. **E5**: Financial head at a multinational metallurgy company, exploring AI for production optimization and emissions reduction.
6. **E6**: Central Services manager in a large cooperative group, spearheading data platforms for financial and non-financial reporting.
7. **E7**: Manager in a large social economic network service organization, with extensive experience in financial strategy and complex market initiatives.

Across these varied contexts—ranging from industrial settings to higher education—participants' perspectives on AI coalesce around multiple levels of adoption, organizational readiness, and challenges involving responsible innovation.

4.2. Emergent Themes

Building upon the variety of contexts outlined in the participant profiles, the interview data were subjected to an open-coding process that revealed six overarching themes. These themes, while applicable across organizational sizes, take on particular resonance in SMEs due to tighter budgets and limited technical staff.

In the subsections that follow, we detail each theme, present participants' insights, link them to relevant research, and discuss the implications for AI-driven CSR.

4.2.1. Fragmented and Incremental Adoption

A recurring thread in the interviews suggests that AI adoption is fragmented, often driven by individual champions rather than a cohesive organizational strategy. As E2 explains, "the use of AI arises from personal interest in different departments, not from top-down direction." Similarly, E4 mentions that "we have pilot projects in finance, but some units have not even started thinking about AI."

Several participants explained that SMEs often rely on opportunistic, small pilots due to budget constraints and minimal dedicated AI staff. E3 points out that finance frequently leads these initiatives given clear cost-saving opportunities, while HR or sustainability units tend to be less digitized and therefore slower to adopt AI tools. E2 specifically underscores that this department-by-department approach creates minimal synergy across initiatives, risking duplication of effort and inconsistent data practices. While larger organizations sometimes have the capacity to run multiple pilots in parallel, SMEs often implement only one or two small-scale projects, partially due to funding limitations (E2, E5). Consequently, these firms may not develop a formal AI roadmap, further accentuating fragmentation.

In addition, some interviews (E2, E3, E4) note that financial tasks—such as invoice processing or basic forecasting—are the first to benefit. Meanwhile, more strategic or CSR-focused AI uses remain aspirational.

This incremental approach resonates with research showing that emerging technologies commonly follow a phased path [1], beginning with limited demonstration or pilot deployments before gradually scaling. Yet smaller firms—with fewer managerial layers—face heightened difficulty coordinating resources across departments [28]. Although quick wins can bolster interest and justify initial AI spending, a lack of strategic coordination risks duplicating effort and undermining potential synergy between initiatives. Larger organizations may occasionally mitigate these issues by running parallel pilots under a broader framework, but SMEs confront tighter budgets that further diminish the scope for integrated AI solutions. Without a unifying vision, the resulting pockets of AI usage yield immediate but localized benefits, leaving cross-functional gains untapped.

4.2.2. Knowledge Gaps and Resistance

A second prominent theme involves inadequate digital training and emotional resistance to AI adoption. E2 emphasizes, "the biggest barrier is a simple lack of understanding of how AI can be used," while E3 and E5 point to emotional resistance—"fear" that AI might replace jobs. This anxiety appears closely tied to inadequate training in digital or analytical skills, which E2 describes as "the biggest barrier" to understanding AI's real capabilities. Limited budgets compound the issue—particularly in SMEs—where the hiring of data scientists or AI engineers is often out of reach, leaving organizations dependent on ad hoc or third-party solutions (E2, E5, E7). Additionally, certain interviewees (E2, E5) pointed out that negative coverage or sensationalist media stories further confuse employees about AI's "black box" nature, eroding trust and discouraging exploration. Consequently, in many workplaces, staff adopt a wait-and- see attitude rather than proactively experimenting with new systems.

When budgets are tight, SMEs often cannot afford extended training programs or specialized tech hires—magnifying the impacts of knowledge gaps and fueling a "wait-and-see" attitude. E2 specifically commented that "ignorance is the biggest barrier" in resource-limited environments.

Such resistance and skill deficits reflect previous research indicating that cultural and emotional factors play a critical role in AI adoption [29]. Smaller firms have particular difficulty securing technical expertise or organizing extensive upskilling, human capital constraints are especially acute in lean organizations [5]. Moreover, negative media coverage often dramatizes algorithmic failures, heightening staff anxieties.

Without proactive training or clear communication about AI's potential benefits, employees may remain skeptical and managers uncertain about the real-world returns of these technologies. In consequence, participants (E2, E3) emphasized that consistent capacity- building, guided by "knowledge-sharing consortia" or government incentives, can be essential to bridging these skill gaps—especially for SMEs seeking to avoid falling behind in their digital transformation journeys. By addressing employee fears and clarifying AI's value-add, organizations may foster a more receptive, innovative culture prepared to integrate AI responsibly.

4.2.3. Potential Synergies With CSR

Across interviews, participants envision robust synergies between AI and sustainability initiatives—often referred to as "CSR." E2 and E3 consider AI "a tool that can significantly advance sustainability monitoring," particularly in supply chains or environmental metrics. Meanwhile, E5 notes that AI can "open doors to measure carbon footprints far more precisely." illustrating a shift away from sporadic reporting toward more real-time analytics. In practice, respondents referenced a range of possible uses, including tracking CO_2 emissions, energy consumption, or water usage, enhanced supply-chain traceability to ensure compliance and monitor labor conditions, and the potential for AI-based dashboards that highlight social metrics like diversity, equity, or community engagement.

Despite these benefits, SMEs frequently remain at exploratory stages. One participant (E3) noted that smaller firms "could truly benefit from real-time data on resource usage but lack accessible platforms." This shortfall underscores the need for lower-cost, user-friendly AI solutions that help small businesses integrate sustainability data without excessive complexity or financial burden. Still, interviews also reveal lower adoption in CSR functions compared to finance. E3 suggests "CSR lacks a tradition of data analytics," and E2 sees "few integrated solutions linking sustainability to AI."

Such perspectives align with prior scholarship demonstrating how AI can transform everything from supply-chain traceability to resource consumption tracking [10]. Although these possibilities are exciting, participants acknowledged varying levels of adoption maturity: E3 reported that many smaller companies "could truly benefit" from data-driven sustainability dashboards but lack accessible platforms. E2 similarly noted that "CSR lacks a tradition of data analytics," limiting the uptake of more advanced AI solutions.

Hence, while participants generally see AI as integral to achieving organizational sustainability targets, the cost and technical complexity of advanced systems often slow adoption—especially in SMEs [3]. Some see low-cost, user-friendly AI solutions as a potential "game-changer" for bridging that gap, allowing smaller firms to harness real-time insights on carbon footprint, labor conditions, or community engagement. However, as E2 concluded, "Without simpler tools, the concept remains aspirational for most."

4.2.4. Importance of Data Quality and Governance

A prominent theme for participants (E4, E6, E7) is the essential role of data quality and robust governance. E6 highlights building "a centralized platform for standardized data input," while E4 warns about the consequences of mixing old, unstructured data with new AI processes.

Organizations like E4 suggested that SMEs face heightened difficulties adopting robust data governance frameworks, owing to time and staffing constraints. Without clear standards, AI can underperform or produce flawed outputs, undermining attempts at capturing and showcasing CSR metrics. E4, additionally, warned that mixing old, unstructured data with newly automated processes "leads to flawed outputs," jeopardizing the reliable tracking and reporting of CSR indicators; she further highlighted the need for "recurrent audits" to continuously validate data accuracy.

Interviewees consistently noted that such procedures not only enhance AI's performance but also strengthen accountability—particularly crucial for CSR metrics, where transparency and comparability of data are pivotal for both internal stakeholders and external oversight.

Such comments echo the well-known adage of "garbage in, garbage out" [10], highlighting that AI's value hinges on well-curated and interoperable data. However, participants (E4, E6) note that many SMEs face time and staffing constraints, leaving them with fragmented spreadsheets or outdated infrastructures. Consistent data standards and best practices can be especially elusive in decentralized setups [30].

If organizations—particularly smaller ones—cannot unify data dictionaries or establish ongoing validation mechanisms (e.g., "recurrent audits" per E4), AI may generate unreliable analyses or fail to integrate with broader sustainability objectives. Thus, shared guidelines or cross-firm collaborations could lighten the governance burden. In so doing, smaller companies might adopt AI more confidently, using transparent metrics to substantiate their CSR claims, rather than grappling with half-measures or siloed data repositories.

4.2.5. Value of Human Oversight and Explainability

Nearly all interviews include references to the need for interpretability (E2, E7) and the necessity for humans to verify AI outputs (E4, E5). In E7's words, "Employees more readily trust automated suggestions if they can verify or understand them."

As E2 remarked, "We must understand how IA works if we're to accept it." reflecting a reluctance to rely on opaque models. Several participants also described a hybrid approach—referred to by some as "supervised IA"—in which algorithms propose solutions, but humans finalize decisions. From their perspective, this oversight goes beyond mere functional checks; it ties directly to ethics and accountability, ensuring that biases or errors are caught early and aligning the organization's AI practices with broader moral responsibilities.

Several interviewees (E2, E7) pointed out that smaller firms frequently lack the specialized staff needed to implement advanced interpretability frameworks, making lighter, more transparent AI solutions preferable. This trade-off emphasizes the role of low-complexity yet explainable tools, particularly where decisions carry ethical weight or pose higher regulatory risks.

By clarifying how AI arrives at conclusions, organizations meet regulatory demands (e.g., AI Act) while building internal acceptance—a theme that also resonates in CSR contexts (E2, E3).

From a theoretical standpoint, such calls for XAI reflect broader debates on black-box decision-making and the responsibilities of algorithmic accountability [20]. Yet for SMEs with fewer specialized staff, implementing advanced interpretability frameworks can be challenging. E2 indicated that they "lack in-house expertise to interpret complex outputs," suggesting more lightweight solutions or "white-box" algorithms might better suit smaller-scale usage.

Additionally, participants tie human oversight to a form of moral responsibility. E5 remarked that "we can't let AI run everything," especially in socially significant decisions that might shape public or employee trust. This balance between automation and oversight might foster compliance with emerging regulations (e.g., the AI Act) and mitigate potential biases or mistakes. However, it also raises questions of resource allocation—meaning that smaller entities must choose carefully which processes to automate and which to keep firmly under human control.

4.2.6. Strategic Emphasis on Responsible AI

Finally, several interviewees (E3, E5, E7) articulate a vision of Responsible AI as a strategic differentiator, linking corporate values to advanced technology. E7 describes it as a "powerful brand statement: we are adopting AI ethically to remain a trusted partner.". E1 and E5 both emphasized that ethically aligned AI helps "foster trust" among clients and external stakeholders by reducing potential reputational risks and ensuring AI usage remains fair, transparent, and socially beneficial. In E7's words, "AI with strong ethical guardrails" becomes a unique selling proposition for customers.

E3 specifically noted that while large organizations might invest heavily in compliance and brand-building around Responsible AI, SMEs have fewer resources for formal initiatives of this kind. Nevertheless, E3 added, if integrated thoughtfully, such an ethically grounded approach could become a unique selling proposition for smaller players operating in specialized or niche markets—delivering trust alongside innovation.

Such perspectives echo scholarship tying AI ethics to brand reputation [8], suggesting that robust data protection, fairness measures, and transparent reporting can evolve from compliance tasks into market advantages. However, E3 cautioned that achieving these higher standards often requires significant investments in compliance and brand-building—feasible for large multinationals but onerous for SMEs. Fulfilling local data-protection laws, equality plans, and other EU-level frameworks can prove disproportionately taxing for smaller firms [11].

Thus, while responsible AI can serve as a unique selling proposition—particularly in specialized or niche markets—participants (E3, E5, E7) concluded that SMEs must often collaborate or form alliances to navigate the complexity of new regulations and the cost of best-practice frameworks. Otherwise, they risk being sidelined by compliance burdens or overshadowed by bigger players with more resources.

5. Conclusions

5.1. Synthesis of Key Findings

Taken together, the interviews depict a multilayered landscape of AI adoption in CSR, reflecting both optimism and practical hurdles. Many organizations—particularly SMEs—engage with AI through fragmented, pilot-driven projects, often spearheaded by some departments while others remain slower to digitize. This situation yields minimal synergy between initiatives, as well as widespread fears tied to job displacement and inadequate digital training. Despite these concerns, participants also envision strong synergies between AI and sustainability efforts, including more precise monitoring of carbon footprints, labor conditions, and supply-chain fairness.

Although organizations of all sizes share aspirations to enhance data-driven decision-making and build stakeholder trust, SMEs specifically encounter heightened obstacles related to resource constraints, limited in-house skills, and rapidly evolving regulatory mandates. A recurring obstacle centers on data quality and robust governance. Interviewees emphasized that mixing unstructured legacy data with new AI processes impedes reliable CSR metrics, reinforcing the adage that "garbage in, garbage out" can undermine advanced analytics. Likewise, human oversight and explainability (XAI) consistently emerged as vital for building trust in automated decisions. Yet smaller firms, constrained by limited budgets and specialized staff, find advanced interpretability frameworks challenging to implement. Finally, a strategic emphasis on Responsible AI emerged, with participants describing it as both a market differentiator and an ethical imperative. Larger entities can often afford formal brand-building around AI ethics, while SMEs must frequently rely on simpler, cost-effective solutions or alliances to navigate compliance burdens.

5.2. Practical and Policy Implications

The interviews confirm that fragmentation and pilot-driven adoption remain the norm, underscoring the need for well-coordinated strategies that move beyond siloed projects. Explaining and governing AI effectively demands both accessible technology (to match SMEs' budgets and staffing realities) and cultural readiness—ensuring that employees comprehend AI's potential benefits rather than focusing solely on job-security fears. Another core insight is that robust data governance with standardized dictionaries and recurring audits is crucial for producing valid CSR metrics. Given SMEs' leaner structures, policy interventions—such as cooperative alliances, government-backed AI sandboxes, or training subsidies—can significantly ease the transition toward responsible AI usage.

Furthermore, participants repeatedly suggested that adopting Responsible AI fosters intangible advantages like improved brand reputation and deeper stakeholder engagement. In specialized or niche markets, ethically aligned AI can serve as a unique selling proposition, enabling smaller firms to stand out by emphasizing trust and transparency. However, the flipside is that compliance and accountability frameworks (e.g., AI Act, data-protection laws) can disproportionately tax SMEs unless tailored support is provided—validating broader arguments that new regulations may unintentionally penalize resource-constrained organizations if left unaddressed.

5.3. Future Research

Looking ahead, further empirical exploration of how data governance frameworks, organizational readiness, and regulatory alignments converge will be critical to harnessing the transformative potential of AI in CSR contexts. For now, these insights demonstrate the nuanced interplay of challenges and opportunities facing AI-based CSR initiatives, suggesting that well-coordinated strategies—encompassing technical, cultural, and ethical dimensions—are essential for harnessing the true transformative potential of AI.

In particular, quantitative studies could probe correlations between AI investments and specific CSR outcomes—such as emissions reduction, supply-chain fairness, or stakeholder engagement—offering a more causal basis for understanding AI's tangible impact. Likewise, longitudinal designs may track how organizations progress from pilot-driven experimentation to advanced, integrated deployments, illuminating the evolving interplay of technical, cultural, and ethical factors. Finally, cross-sector and cross-cultural investigations—extending beyond finance or industrial settings and beyond European regulatory environments—would refine our grasp of how local policies, institutional norms, and resource constraints shape AI's responsible adoption. Through such research, both smaller and larger enterprises can better identify the cohesive, multidimensional strategies needed to maximize AI's potential for enhancing sustainability, transparency, and trust.

6. Acknowledgments

This research was funded by the "Ayudas para la Responsabilidad Social Empresarial Euskadi 2024" program, supported by the Department of Economy, Labour and Employment of the Basque Government.

Declaration on Generative AI

During the preparation of this work, the author(s) used X-GPT-4 and Gramby in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] A. Atanasov, G. Chipriyanova, R. Krasteva-Hristova, Integration of Digital Technologies in Corporate Social Responsibility (CSR) Activities: A Systematic Literature Review and Bibliometric Analysis, Journal of Risk and Financial Management 16 (2023) 373. URL: https://www.mdpi.com/1911-8074/16/8/373. doi:10.3390/jrfm16080373.

[2] M. Wynn, P. Jones, Corporate Responsibility in the Digital Era, Information (Basel) 14 (2023) 324. doi:10.3390/info14060324, place: Basel Publisher: Basel: MDPI AG.

[3] N. Perucica, K. Andjelkovic, Is the future of AI sustainable? A case study of the European Union, Transforming Government: People, Process and Policy 16 (2022) 347–358. URL: https://www.emerald.com/insight/content/doi/10.1108/TG-06-2021-0106/full/html. doi:10.1108/TG-06-2021-0106.

[4] V. Carlsson, M. Rönnblom, From politics to ethics: Transformations in EU policies on digital technology, Technology in Society 71 (2022) 102145. URL: https://linkinghub.elsevier.com/retrieve/pii/S0160791X2200286X. doi:10.1016/j.techsoc.2022.102145.

[5] P. Bisconti, D. Orsitto, F. Fedorczyk, F. Brau, M. Capasso, L. De Marinis, H. Eken, F. Merenda, M. Forti, M. Pacini, C. Schettini, Maximizing team synergy in AI-related interdisciplinary groups: an interdisciplinary-by-design iterative methodology, AI & SOCIETY 38 (2023) 1443–1452. URL: https://link.springer.com/10.1007/s00146-022-01518-8. doi:10.1007/s00146-022-01518-8.

[6] Osinachi Deborah Segun-Falade, Olajide Soji Osundare, Wagobera Edgar Kedi, Patrick Azuka Okeleke, Tochukwu Ignatius Ijomah, Oluwatosin Yetunde Abdul-Azeez, Assessing the transformative impact of cloud computing on software deployment and management, Computer Science & IT Research Journal 5 (2024) 2062–2082. URL: https://fepbl.com/index.php/csitrj/article/view/1492. doi:10.51594/csitrj.v5i8.1492.

[7] N. Finnerty, R. Sterling, D. Coakley, S. Contreras, R. Coffey, M. M. Keane, Development of a Global Energy Management System for non-energy intensive multi-site industrial organisations: A methodology, Energy 136 (2017) 16–31. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360544216314839. doi:10.1016/j.energy.2016.10.049.

[8] M. Cazacu, S. Dumitriu, I. Georgescu, D. Berceanu, D. Simion, A. A. Vărzaru, C. G. Bocean, A Perceptual Approach to the Impact of CSR on Organizational Financial Performance, Behavioral Sciences 13 (2023) 359. URL: https://www.mdpi.com/2076-328X/13/5/359. doi:10.3390/bs13050359.

[9] S. Quach, P. Thaichon, K. D. Martin, S. Weaven, R. W. Palmatier, Digital technologies: tensions in privacy and data, Journal of the Academy of Marketing Science 50 (2022) 1299–1323. URL: https://link.springer.com/10.1007/s11747-022-00845-y. doi:10.1007/s11747-022-00845-y.

[10] S.-L. Lin, H.-W. Huang, Improving Deep Learning for Forecasting Accuracy in Financial Data, Discrete Dynamics in Nature and Society 2020 (2020) 1–12. URL: https://www.hindawi.com/journals/ddns/2020/5803407/. doi:10.1155/2020/5803407.

[11] G. Lazcoz, P. De Hert, Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities, Computer Law & Security Review 50 (2023) 105833. URL: https://linkinghub.elsevier.com/retrieve/pii/S0267364923000432. doi:10.1016/j.clsr.2023.105833.

[12] F. Caputo, S. Pizzi, L. Ligorio, R. Leopizzi, Enhancing environmental information transparency through corporate social responsibility reporting regulation, Business Strategy and the Environment 30 (2021) 3470–3484. URL: https://onlinelibrary.wiley.com/doi/10.1002/bse.2814. doi:10.1002/bse.2814.

[13] F. Bostoen, L. M. Santos, A. V. D. Veer, Not “big is bad” but “closed is bad”? Exploring dynamic competition in generative AI, Market and Competition Law Review (2024) v. 8 n. 1 (2024). URL: https://journals.ucp.pt/index.php/mclawreview/article/view/16137. doi:10.34632/MCLAWREVIEW.2024.16137, publisher: Market and Competition Law Review.
[14] J. W. Creswell, W. E. Hanson, V. L. Clark Plano, A. Morales, Qualitative Research Designs: Selection and Implementation, The Counseling Psychologist 35 (2007) 236–264. URL: https://journals.sagepub.com/doi/10.1177/0011000006287390. doi:10.1177/0011000006287390.
[15] E. G. Guba, Y. S. Lincoln, Fourth generation evaluation, 13. print ed., Sage Publ, Newbury Park, Calif., 2003.
[16] U. Flick, Qualitative Methoden in der Evaluationsforschung, Zeitschrift für qualitative Forschung (????) 9–18.
[17] M. Van Manen, Beyond assumptions: Shifting the limits of action research, Theory Into Practice 29 (1990) 152–157. URL: http://www.tandfonline.com/doi/abs/10.1080/00405849009543448. doi:10.1080/00405849009543448.
[18] J. Agee, Developing qualitative research questions: a reflective process, International Journal of Qualitative Studies in Education 22 (2009) 431–447. URL: http://www.tandfonline.com/doi/abs/10.1080/09518390902736512. doi:10.1080/09518390902736512.
[19] C. Seale, Quality Issues in Qualitative Inquiry, Qualitative social work, 1(1), 97-110. (????). doi:https://doi.org/10.1177/147332500200100107.
[20] O. Tracey, R. Irish, Explainability for All: Care Ethics for Implementing Artificial Intelligence, in: 2023 IEEE International Symposium on Technology and Society (ISTAS), IEEE, Swansea, United Kingdom, 2023, pp. 1–8. URL: https://ieeexplore.ieee.org/document/10306199/. doi:10.1109/ISTAS57930.2023.10306199.
[21] M. B. Miles, A. M. Huberman, Qualitative data analysis: an expanded sourcebook, 2. ed., [nachdr.] ed., Sage, Thousand Oaks, Calif., 200.
[22] N. K. Denzin (Ed.), Strategies of qualitative inquiry, number 2 in Handbook of qualitative research, 3. ed ed., Sage Publications, Los Angeles, 2008.
[23] L. L. Ellingson, Engaging crystallization in qualitative research: an introduction, Sage Publications, Thousand Oaks, 2009.
[24] A. L. Strauss, J. Corbin, E. Zimmerman, Bases de la investigación cualitativa: técnicas y procedimientos para desarrollar la teoría fundamentada, Universidad de Antioquía, Medellín (Colombia), 2002. OCLC: 954957839.
[25] A. P. Bochner, Criteria Against Ourselves, Qualitative inquiry, 6(2), 266-272 (????). doi:https://doi.org/10.1177/107780040000600209.
[26] C. A. Ellis, R. L. Miller, V. D. Calhoun, A Systematic Approach for Explaining Time and Frequency Features Extracted by Convolutional Neural Networks From Raw Electroencephalography Data, Front Neuroinform 16 (2022) 872035. doi:10.3389/fninf.2022.872035, place: Switzerland Publisher: Switzerland: Frontiers Research Foundation.
[27] C. G. Christians, Ethical theory in communications research, Journalism Studies 6 (2005) 3–14. URL: http://www.tandfonline.com/doi/abs/10.1080/1461670052000328168. doi:10.1080/1461670052000328168.
[28] S. Fassiaux, Preserving Consumer Autonomy through European Union Regulation of Artificial Intelligence: A Long-Term Approach, European Journal of Risk Regulation 14 (2023) 710–730. URL: https://www.cambridge.org/core/product/identifier/S1867299X23000582/type/journal_article. doi:10.1017/err.2023.58.
[29] S. Du, C. Xie, Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities, Journal of Business Research 129 (2021) 961–974. URL: https://linkinghub.elsevier.com/retrieve/pii/S0148296320305312. doi:10.1016/j.jbusres.2020.08.024.
[30] M. Veale, F. Zuiderveen Borgesius, Demystifying the Draft EU Artificial Intelligence Act, 2021. URL: https://osf.io/38p5f. doi:10.31235/osf.io/38p5f.