

Moral Accountability in AI-Driven Corporate Decision-Making: A Comparative Case Analysis of Amazon, United Health Group, and Deutsche Bank

Kesmat AbdelAziz*

Arab Academy for Science, Technology & Maritime Transport, Cairo, Egypt.

E-mail: kesmatmyehia@gmail.com

*Corresponding Author

Received: July 10, 2025; **Accepted:** December 27, 2025

Abstract: There are serious moral accountability issues involved in using artificial intelligence systems for critical corporate decisions, such as hiring, health insurance coverage, and credit approvals. Who is responsible morally when an application is not accepted, a claim for health insurance is refused, or a credit application is not approved? In this paper, we draw upon three corporate case studies, including Amazon with its automated hiring processes, UnitedHealth Group with AI and health insurance, and Deutsche Bank's algorithmic credit rating process in order to see how moral accountability is created, shared, and challenged in corporate fields that involve AI. Building on philosophical literature on collective moral responsibility, distributed agency, and algorithmic governance, we elaborate an Accountability Diffusion Framework (ADF) that describes four mechanisms through which the use of AI systematically undermines moral accountability in corporate contexts: opacity diffusion, temporal displacement, agential fragmentation, and normative laundering. In our analysis, all three companies have used these mechanisms, whether deliberately or not, to shield human decision-makers from responsibility for harmful AI-driven outcomes. We argue that restoring moral accountability in AI-driven organisations requires not only improved technical explainability, but fundamental changes to corporate governance, stakeholder engagement practices, and the regulatory architecture of AI deployment. The paper ends with a set of accountability design principles for ethically responsible AI governance.

Keywords: AI Ethics, Moral Accountability, Algorithmic Decision-Making, Corporate Governance, Responsible AI, Distributed Agency

Type: Review



© 2026 The Author(s). This article is an open-access article distributed under the terms of [the Creative Commons Attribution \(CC BY 4.0\) License](https://creativecommons.org/licenses/by/4.0/).

DOI: 10.51325/3yx5qe22

1. Introduction

On 25 October 2018, Reuters reported that Amazon had quietly shut down a machine learning recruitment tool it had developed over four years after discovering that the system systematically downgraded applications from women. The tool had been trained on historical hiring data, which reflected the male dominance across Amazon's technical workforce. In effect, the tool had captured and codified past discrimination in a predictive model and applied it at scale to

thousands of job applicants, who had no idea that they were being judged by an algorithm, or on what basis. Amazon's response, which involved discontinuing the tool and instructing recruiters not to rely on it, raised a question the company notably refused to answer: who bore moral responsibility for the harm caused to applicants the system had already assessed?

The case of Amazon was not a unique experience. Firms from different industries around the world are turning to AI systems in order to take important decisions impacting the access of citizens to jobs, health care, credit, insurance, housing, and social services. These systems often caused unintentional harm in the form of discriminatory decisions, refusal of service in the situations where such a decision was uncalled for and different forms of miscalibration affecting disadvantaged people. What is troubling about this process is that it is difficult to attach moral responsibility to a certain individual. Engineers refer to the training data, data specialists to business pressure, managers to the output of algorithms and top managers to compliance with regulations.

This paper examines accountability diffusion through a comparative case study of three firms that have been publicly scrutinized for AI-driven decision-making in high-stakes domains: Amazon's automated hiring system (employment), UnitedHealth Group's AI-driven healthcare coverage determination system (healthcare), and Deutsche Bank's algorithmic credit scoring infrastructure (financial services). The cases were selected to ensure a diversity of sectors including the technology, healthcare, and financial industries, but with the common feature of deploying AI in decisions with direct and significant consequences for the persons affected. There are also cases with a significant amount of public documentation, such as regulatory filings, investigative journalism, legal proceedings, and corporate disclosures, providing a strong evidence base for analysis.

The current paper presents three novel contributions to the field of business ethics. The first one is the establishment of the Accountability Diffusion Framework (ADF) developed by the authors. The ADF encompasses four mechanisms through which AI use leads to diminished moral accountability in business organizations: diffusion of opacity, temporal dislocation, fragmentation of agency, and normative laundering. The second contribution consists in systematic application of the ADF to three comparative case studies. The findings of such analysis reveal structural characteristics of accountability diffusion that cannot be associated with any particular organization or industry. The third contribution is the formulation of accountability design principles which include different organizational, governance, and regulatory policy options to solve the problems identified in the empirical analysis.

The rest of this paper is organized as follows. Section 2 reviews relevant literature on moral accountability, distributed agency, and algorithmic governance. The research methodology and the rationale for the case selection are presented in Section 3. Section 4 extends the Accountability Diffusion Framework. The three case analyses are discussed in Sections 5, 6, and 7. Section 8 explains the cross-case analysis and accountability design principles, while Section 9 concludes.

2. Theoretical Background

2.1. Moral Accountability and Its Conditions

Moral responsibility is a normative relation, explained in philosophical literature as the state where an agent is held liable for the outcomes of his/her actions and omissions. This could include being brought into account, being required to explain, being blamed, being punished, and being required to take corrective measures (Scanlon, 1998; Strawson, 1962). Regular views generally admit three criteria of moral responsibility. First, the act or omission is in some way involved in the events (causal contribution). Second, the agent knew or should have known the relevant facts (epistemic access). Third, the agent was able to act differently (normative capacity).

Moral responsibility in corporate contexts is inherently shared: the conduct of corporations is the result of complex organizational processes involving many individuals, all making partial contributions to the results that no single individual fully controls or anticipates (French, 1984; Velasquez, 1983). In the philosophical literature, there are sophisticated accounts of attributing responsibility to organizations as moral agents without succumbing to the collective action problem of responsibility diffusion, where everyone is slightly responsible, and no one is fully responsible (Bratman, 1992; List and Pettit, 2011). But these accounts were designed to deal with human collectives in the first place, and they struggle with the new features of AI-mediated decision-making, i.e., the need to somehow accommodate the agency of non-human systems in the accountability analysis.

2.2. Algorithmic Governance and the Accountability Gap

Incorporating AI systems into business decision-making opens the door to aspects that question the traditional notions of moral accountability, as widely discussed in the latest studies in algorithmic governance and AI ethics (Diakopoulos, 2016; Mittelstadt et al., 2016; Raji et al., 2020). One of these aspects is opacity; many AI systems with high predictive accuracy, mostly those designed with the help of deep learning technologies, cannot be easily understood by humans. Hence, it is impossible to clarify how the system arrived at a certain decision. In this context, the black-box nature of the system may obscure the reasons behind the algorithm-based decision and a harmful result and complicate the estimation of what kind of cause and effect effect is necessary to understand the events.

Second, scale and speed. AI systems can make millions of decisions a day, far beyond any human oversight capacity. An automated hiring system can make hundreds of decisions in a year, the equivalent of ten years' worth of decision-making by a human HR department. At this scale, the moral landscape changes significantly. Harms that may be trivial from the perspective of a single decision become enormous in the aggregate, and the temporal compression of AI-driven decision-making collapses the feedback loop between deployment and harm in ways that challenge traditional models of governance (Binns, 2018; Zarsky, 2016).

Third, distributed development: AI systems are rarely the work of a single designer or single organization. Engineers build them, others collect the data they are trained on, others validate them, others integrate them into organizational processes, and they are deployed in environments shaped by regulatory

frameworks created by governments and industry bodies. The distributed nature of the development process brings about many possibilities to evade responsibility, thus resulting in what Nissenbaum (1996) referred to as the 'problem of many hands'. In these circumstances, a harmful outcome emerges from the collective actions of many people without any of them being individually responsible for it.

2.3. Corporate Moral Agency and AI

The question of whether corporations can be true moral agents and, if so, whether their AI systems can be viewed as extensions of that agency has been debated vigorously in business ethics (Donaldson, 1982; French, 1984; Velasquez, 1983). For the purpose of this paper, we adopt a modified version of the corporate moral agency position: corporations are moral agents insofar as they can be held accountable for systemic patterns of conduct that reflect organizational choices, even where those choices are distributed across many individuals and partially implemented through technical systems. Recent studies utilizing virtue ethics in relation to AI validate this viewpoint based on the notion that organizational character can be defined as a company's overall tendency to act responsibly or irresponsibly in employing AI. This viewpoint reinforces the idea that companies can be held morally accountable for their usage of AI even if employees are not held accountable for situations that are beyond their control. (Farina et al., 2024)

Recent academic work has pointed to the importance of algorithmic accountability as a specific governance challenge that warrants specific institutional responses (Diakopoulos, 2016; Kearns and Roth, 2019; Raji et al., 2020). The notion of 'responsible AI' has gained traction in both academic and practitioner literatures, resulting in a proliferating set of principles, frameworks, and guidelines that broadly agree on the importance of fairness, transparency, explainability, and human oversight but vary widely in how these values should be operationalized in practice. A common critique of the responsible AI literature is that it tends to focus on technical design and individual firm practice, while missing the structural and governance aspects of accountability – the question not just of how firms should build ethical AI systems but what institutional mechanisms are necessary to ensure that they are held accountable when they do not.

3. Research Methodology

3.1. Case Study Design

In this study we adopt a comparative case study methodology based on the framework of Yin (2018) and adapted for the normative aims of business ethics research (Beveridge, 2012; Hartley, 2004). The methodology is appropriate since we are ultimately interested in how moral accountability emerges and diminishes in particular organisational and technological contexts. Answering that question requires close scrutiny of the context and cannot be pursued through quantitative methods alone. The comparative approach gives us the opportunity to identify common elements and patterns across cases while allowing for attention to the specificities of each situation.

Purposeful theoretical sampling was utilized to select the three cases of Amazon, UnitedHealth Group, and Deutsche Bank in order to maximize diversity

in terms of two aspects: the industry sector (technology, healthcare, and financial services) and the type of decision involving AI (hiring, benefits determination, and credit scoring). All selected cases involve the AI systems making or having a considerable impact on the decisions which have serious implications for the people involved. In addition, all cases have major public reporting available for in-depth investigation of evidence. The present cases do not represent the case of AI deployment in their relevant industries. Instead, they act as theoretically-generating cases demonstrating the accountability mechanisms established in ADF.

3.2. Data Sources and Analysis

The evidence base for each case draws on multiple documentary sources: regulatory filings and enforcement actions (SEC disclosures, CFPB enforcement orders, EU regulatory decisions); investigative journalism (Reuters, ProPublica, The New York Times, Der Spiegel); legal proceedings (class action complaints, court judgments); corporate disclosures (annual reports, sustainability reports, AI ethics statements); and academic research on the specific systems in question. The triangulation of sources allows for cross-validation of factual claims and reduces dependence on a single perspective. Before inclusion in the analysis, all sources were critically appraised for bias, reliability, and weight of evidence.

The cases were analyzed using a structured thematic framework based on the Accountability Diffusion Framework developed in Section 4. After completing our investigation of the different situations, we found proof of four different accountability diffusion mechanisms, these were: opacity diffusion, temporal displacement, agential fragmentation, and normative laundering. We also looked into how businesses responded in cases where a dispute occurred regarding their accountability as a result of public scrutiny. Our comparative analysis showed how similar or different the cases were and how they contributed to improving the principles of accountability design, which are described in Section 8.

4. The Accountability Diffusion Framework

We propose the Accountability Diffusion Framework (ADF) as a conceptual framework for analysing the erosion of moral accountability in corporate decision-making environments where AI is deployed. The framework identifies four mechanisms of diffusion of accountability that can operate independently or in combination to insulate human decision-makers from moral responsibility for harmful AI-driven outcomes.

4.1. Opacity Diffusion

Opacity diffusion is the use of technical, commercial, or organizational opacity to obscure the causal chain between AI-driven decisions and harmful outcomes, thereby undermining the epistemic access condition of moral accountability. Technical opacity arises from real limits to the interpretability of complex machine learning systems, which can prevent even their creators from fully explaining any given decision. Commercial opacity is produced through the invocation of intellectual property protections and trade secrecy to obfuscate algorithmic systems from external scrutiny. Organizational opacity arises because

relevant knowledge is distributed across several teams, functions, and hierarchical levels, so that no single person has the full picture.

Opacity diffusion is morally relevant because it exploits the epistemic access condition of moral accountability: if an AI system made a decision, and neither the firm nor the individual affected can explain why the system made this decision, it is very hard to demonstrate that a moral agent knew or should have known that the decision would lead to a harmful outcome. This does not imply that moral responsibility vanishes entirely. Companies that intentionally create a lack of clarity to protect themselves from being held accountable have even greater moral blameworthiness. Nevertheless, a lack of clarity creates significant challenges to accountability, which companies may take advantage of to reduce their likelihood of facing legal responsibility.

4.2. Temporal Displacement

Temporal displacement is the exploitation of the temporal gap between the time at which consequential design decisions are made about an AI system and the time at which the harmful consequences of those decisions become visible. For long-lived AI systems, this gap may be years or even decades, and it creates conditions in which the people who made the original design decisions are no longer in their roles, the organizational context in which those decisions were made no longer exists, and the documentary record of the decision-making process is incomplete or inaccessible. Temporal displacement thus exploits the causal contribution condition of moral accountability: by the time we identify the harm, the causal chain leading back to the original design choices has become practically untraceable.

4.3. Agential Fragmentation

Agential fragmentation is a term that refers to the sharing of responsibility for the design, creation, use, and administration of AI systems between many different organizational actors including engineers, data scientists, product managers, compliance officers, business users, and high-level managers. As a result, it is not possible to find just one person who has complete control and can be held responsible for harmful outcomes. Some degree of fragmentation is unavoidable in complex AI development, since it is a collaborative process. Nonetheless, it is also possible to include fragmentation in the organizational structure, function of the actors, and the system of governance, which will create accountability gaps in every phase of development.

4.4. Normative Laundering

Normative laundering is likely to be the most dangerous among the four methods outlined above. It can be defined as the process of adopting various formal ethical frameworks, responsible AI principles, bias audits, and compliance processes in order to give the impression of having addressed the accountability issues when in fact nothing has been done to alter the system of accountability in AI governance. An organisation might have AI ethics principles published, arrange the third-party audit of its AI systems, and take part in the process of responsible AI with various stakeholders while still being able to use its algorithmic systems that cause major harm to people.

The four mechanisms are summarized in Table 1, also showing the condition of moral accountability to which each mechanism primarily responds, and the governance response most directly required of each.

Table 1: The Accountability Diffusion Framework

Mechanism	Accountability Condition Targeted	Primary Effect
Opacity Diffusion	Epistemic access	Prevents causal tracing of harmful decisions
Temporal Displacement	Causal contribution	Disconnects the decision from consequence across time
Agential Fragmentation	Normative capacity	No individual bears a sufficient causal share
Normative Laundering	All three conditions	Legitimizes continued harm through ethical performance

5. Case One: Amazon, Automated Hiring and the Disappearing Discriminator

5.1. Context and System Description

Amazon first began testing the use of AI technology in its recruitment process in 2014, when machine learning specialists from Edinburgh created a tool that would allow them to screen applications for technical positions. This tool was founded on ten years of recruitment history of CVs and hire decisions, which allowed the team to draw a picture of the qualities linked to success in job performance at Amazon. By 2015, this tool was reportedly rating candidates on a scale from one to five stars and that being able to select promising candidates from thousands of different applications at once. Internal audits conducted during the same period showed that the system gave biased results, such as the fact that CVs with terms such as “women’s” (e.g. “women’s chess club”) were marked down, as well as all those CVs received from graduates of colleges that also had only female students. Ultimately, this project was shut down by 2017, but the results of the research became known only in 2018.

Amazon’s case serves as a valuable lesson by revealing more about the different accountability systems implemented when using AI. It is known that gender discrimination through automated recruitment processes can be found in many companies and systems (Albaroudi et al., 2024; Sheard, 2025). Amazon had discovered the discrimination internally and tried to modify the program before it was discontinued altogether. In the case presented here, there is no evidence of any public announcements made about it, informing individuals involved in the process, taking actions to find solutions to possible harm, or establishing any mechanisms that ensure those who created and applied the programs are held responsible for discrimination caused by it.

5.2. Accountability Diffusion Analysis

Amazon’s case clearly displays all four mechanisms of the ADF. Opacity spread on both a technical and commercial level: the system’s decision logic was not interpretable for the recruiters who used it, and Amazon’s IP protections effectively shielded its architecture from outside scrutiny until the Reuters investigation. The practical consequence was that applicants whose CVs were

downgraded by the system had no grounds to challenge the decision and no access to any explanation as to why their applications had been rejected.

The temporal dislocation is reflected in the hiatus between the system's development (2014–2015), the internal discovery of bias (c. 2015–2017), and the public disclosure (2018). By the time the Reuters story was published, the system had been decommissioned, the engineers who built it had moved to different roles, and the organizational context in which the decision to deploy a biased training dataset was made was no longer clearly constructible. The time sequence made it difficult to pinpoint and hold responsible the people who had made the decisions that led to the discriminatory harm.

Agential fragmentation can be seen in the way accountability has been split across the various groups involved in the development and deployment of the system, such as the engineering, HR tech and business management groups. The engineers responsible for creating the model did not make the decision to implement it, the managers who approved the implementation did not design how the model works, and the recruiters who made use of its results are not responsible for its functioning. This division of labour makes it hard to trace the link between people's actions and the discrimination outcomes produced by those actions, thus complicating moral accountability of the actors.

Normative laundering is less visibly documented here than in the other cases but is implicit in Amazon's broader commitment to diversity and inclusion. This commitment provided a legitimating narrative frame within which the deployment of a discriminatory hiring system was rendered invisible until uncovered by investigative journalism. The presence of formal diversity commitments did not prevent the deployment of a discriminatory system but may have contributed to the organizational complacency that allowed the system to operate for a number of years without the kind of substantive ethical scrutiny that might have detected and remedied its harms earlier.

6. Case Two: United Health Group — Algorithmic Care Denial and the Illusion of Clinical Judgment

6.1. Context and System Description

UnitedHealth Group, the largest U.S. health insurer by revenue, relied on an AI tool known as the nH Predict model, developed by NaviHealth, which it acquired in 2020, to determine coverage for post-acute care for Medicare Advantage patients. The system was created to anticipate the expected patient recovery trajectories for patients needing post-acute care (rehabilitation, skilled nursing, home health services) and to guide coverage decisions on the duration and intensity of care that would be reimbursed. An investigative report published by STAT News in 2023 and subsequent reporting by The Guardian and Reuters said the system had an error rate of about 90% in cases in which appeals were filed over its coverage denials and alleged it was used to systematically cut off care for elderly patients before their physicians deemed them medically ready for discharge.

A November 2023 class action lawsuit alleged that UnitedHealth was using the nH Predict model as a “predetermined” tool to deny coverage, overruling the clinical judgments of treating physicians in ways that caused direct patient harm and, in some cases, sped the deaths of elderly patients who were discharged

prematurely from post-acute care. According to UnitedHealth, the system was not utilized to simply refuse coverage and the decisions were always taken by licensed clinical evaluators who took into account the findings of the program along with other data. Such characterisation which presents a picture whereby AI only plays a complementary role in making decisions showcases the dynamics of the accountability dilution which the ADF seeks to establish.

6.2. Accountability Diffusion Analysis

As for UnitedHealth, the diffusion of opacity functions primarily at the level of clinical-algorithmic integration. The company claimed that the nH Predict model was an informational input to human clinical reviewers, not the operative decision-maker. This effectively immunized the system's decision logic from accountability by putting nominal accountability in the clinical reviewer and actual decision-making power in the algorithm. Patients who were denied coverage had no right to know that an AI system had predicted their trajectory to recovery or that this prediction had influenced the determination of their reviewer. The signature of the clinical reviewer on the denial letter humanized an algorithmically driven decision, meeting the formal requirements of human oversight, but undermining its substance.

The issue of temporal displacement is especially serious. After being discharged earlier than necessary, harmful effects may manifest in a patient's health only days or weeks following a denial of coverage. These outcomes may also be attributable to the patient's condition before the denial of coverage. This complexity makes it hard to determine the connection between the denial of coverage and what happens afterwards concerning patient outcomes, and this makes even more difficult to evaluate the epistemic access element relevant for moral responsibility.

Agential fragmentation is structurally embedded in the architecture of UnitedHealth's coverage determination process. The chain of actors involved in any given coverage denial include the data scientists who built and calibrated the nH Predict model, the NaviHealth employees who integrated the model into clinical workflows, the clinical reviewers who nominally made coverage decisions with the model as an input, the UnitedHealth compliance officers who approved the process, and the senior executives who made the strategic decision to deploy algorithmic tools in coverage determinations. When asked to explain a harmful outcome, each of these actors can point to another in the chain, creating the classic problem of many hands in a particularly acute form.

The matter in question may be an example of a problematic type of normative laundering. UnitedHealth's commitment to ensuring improved health care efficiency, patient-centric approach, and quality of clinical decisions made it possible for the company to present its algorithm usage in coverage determinations as fitting its mission of bettering health results. The formal process of clinical reviews made the decision-making process more legitimate, even though there were signs proving that the use of the algorithm might have led to negative consequences. The claims regarding compliance with the rules provided more legitimacy to the actions undertaken thus reducing the probability of scrutiny until independent journalists made those based on allegations known to the public.

7. Case Three: Deutsche Bank — Algorithmic Credit Scoring and the Laundering of Systemic Bias

7.1. Context and System Description

Deutsche Bank's algorithmic credit scoring infrastructure is a textbook example of what we call second-order accountability diffusion: the use of technically sophisticated AI systems to make or substantially influence credit decisions in ways that replicate and amplify pre-existing patterns of structural disadvantage, while creating the appearance of objective, merit-based assessment. In 2019, Deutsche Bank was under intense regulatory and media scrutiny in Germany and the United States following reports that its credit-scoring algorithms were producing systematically lower creditworthiness assessments for applicants from certain postal codes—predominantly those with high proportions of immigrant and working-class residents—in ways that correlated strongly with race and ethnicity even after controlling for conventional creditworthiness indicators.

The bank defended its scoring method, saying postal code is a valid proxy for credit risk, and that the residential location statistically correlates with past default rates. This defense, which has been widely used by financial institutions that employ geographic variables in credit scoring, is one example of what Pasquale (2015) calls the 'black box' legitimization strategy: the invocation of actuarial objectivity to insulate discriminatory outcomes from accountability by embedding these outcomes in the apparently neutral language of risk management. Deutsche Bank then conducted an internal review of its scoring models and implemented changes it said would lead to fairer outcomes, but it did not make the methodology or results of that review public.

7.2. Accountability Diffusion Analysis

What is different about Deutsche Bank is that opacity diffusion is used in a sophisticated manner to facilitate normative laundering. The bank's dependence on complex algorithmic models that use hundreds of variables, combined in non-linear ways, such that it is almost impossible to trace the causal contribution of any single factor to a particular credit decision, created a kind of technical opacity that was at once an apt description of real complexity and a resource for the avoidance of accountability. Credit applicants who were denied credit received a standard rejection letter that referenced unspecified credit scoring criteria; regulators who inquired about discriminatory patterns were told the model architectures were proprietary and could not be fully disclosed for fear of losing competitive advantage.

Temporal displacements in credit scoring assume a particular structural form. The training data that credit-scoring models are built on is a reflection of historical lending patterns, which themselves are the product of past discriminatory practices – redlining, discriminatory lending, and exclusionary neighborhood policies that shaped residential patterns for decades. An algorithmic model that learns that residents in certain postal codes have historically had higher rates of default is learning a statistical pattern that is itself the product of historical injustice. It then forwards this historical injustice into future credit decisions, creating a temporal loop where past discrimination is laundered through present-day algorithmic 'objectivity'. This temporal projection

of historical injustice effectively absolves it of its moral responsibility: no individual designer is responsible for the historical patterns encoded in the training data; no individual decision-maker is responsible for the algorithmic outputs those patterns generate.

In the case of Deutsche Bank, agential fragmentation is structurally similar to the case of Amazon, but with the additional layer of outsourced model development. Many of the algorithmic components of Deutsche Bank's credit scoring infrastructure were developed by third-party vendors, further distributing the causal chain and the accountability landscape. The relationship with these vendors created a governance structure in which the bank could disclaim responsibility for algorithmic decisions, because the relevant design choices were made by external parties, while those external parties could disclaim responsibility, because they were building tools to the bank's specifications, and the bank decided to deploy them.

For Deutsche Bank, normative laundering works through the replacement of substantive accountability with regulatory compliance. The bank's credit scoring practices were in line with the relevant German and European financial regulations, which at the time of the controversy did not explicitly ban the use of geographic variables in credit assessment and did not mandate the kind of algorithmic impact assessment that might have uncovered the discriminatory patterns before they became public. Thus, reliance on an inadequate regulatory framework served as a justification for the implementation of a system producing outcomes that were inconsistent with the bank's stated commitments to fair and non-discriminatory access to financial services.

8. Cross-Case Analysis and Accountability Design Principles

8.1. Cross-Case Patterns

Several consistent patterns emerge across the three cases that illuminate the structural drivers of accountability diffusion in AI-deploying organizations. First, all three firms deployed AI systems in decision domains that naturally produced circumstances conducive to accountability diffusion because of their scale, speed, and technical complexity, and where no appropriate internal or regulatory governance frameworks were in place to manage these circumstances, which were instead exploited. This suggests that accountability diffusion is not primarily a product of bad actors, but of governance architecture systematically inadequate to the accountability demands of high-stakes AI deployment.

The second point is that all three cases illustrate the significance of normative laundering in the distribution of accountability. In the three cases, formal commitments to ethics concerning diversity and inclusion, patient-centered caring, or fair access to financial services provide a legitimization narrative that hides and even normalizes harmful algorithmic practices until external investigations unveil them. This observation punishes the tendency in the responsible AI literature to view the adoption of ethical principles as evidence of progress towards accountability. In the absence of meaningful enforcement mechanisms such principles could be more than ineffective; in fact, they may be capable of both hiding ongoing damage and postponing the needed institutional changes necessary for real accountability.

Third, all three cases show a shared pattern of response to accountability diffusion: strategic acknowledgement of the problem with little substantive accountability. Amazon killed a tool that it had already decided to kill, without telling affected applicants or acknowledging the harm done. UnitedHealth defended its processes while secretly changing its scoring model. Deutsche Bank launched an internal review, the methodology and findings of which were kept secret. In no case was there a systematic attempt to identify and address harms caused to affected individuals. In no case was there a requirement for the individuals responsible for relevant design and deployment decisions to be publicly accountable for those decisions.

8.2. Accountability Design Principles

Based on the cross-case analysis and the theoretical framework developed above, we propose five accountability design principles for ethically responsible AI governance in corporate contexts.

Principle 1: Accountability Assessment Needed Impact-Forward: Before any AI system is deployed in a high-stakes decision domain, firms should be required to conduct a structured accountability assessment that maps the causal chain from design choices to potential harmful outcomes, identifies the individuals and organizational units responsible at each stage, and establishes ex ante governance mechanisms for monitoring, detecting, and responding to harmful outcomes. This assessment should be made available to relevant regulators and to potentially affected stakeholders in an appropriately anonymized form.

Principle 2: Interpretability, a non-negotiable governance requirement: The employment of non-interpretable AI systems in decisions that have material consequences for individuals should be viewed as a prima facie governance failure rather than a mere technical limitation. Where real interpretability constraints exist, for example, in complex deep learning architectures, companies should be required to use proxy explanatory techniques to provide affected individuals with a meaningful account of the factors influencing their assessment and bear the accountability costs of deploying systems whose decision logic cannot be fully explained.

Principle 3: Longitudinal Accountability Registers: Firms should keep detailed records of the design choices, characteristics of the training data, validation results, and deployment decisions for each AI system used to make high-stakes decisions. These records should be retained for a period long enough to enable accountability for temporally displaced harms, at a minimum, the period during which affected individuals retain legal rights to challenge the relevant decisions, and should be made available to regulators and affected individuals upon request.

Principle 4: Hire AI Accountability Officers with Real Authority: This problem of agential fragmentation calls for organizational responses that integrate, rather than disperse, accountability for AI governance. Companies should designate senior AI accountability officers with clear responsibility and real authority for the ethical governance of AI deployments across the organization. These roles should be separate from, but in dialogue with, technical AI development and compliance

functions and should have direct reporting lines to the board level (Cihon et al., 2021). “They should be held formally accountable, and their performance publicly reported.”

Principle 5: Mechanisms for stakeholder redress that are meaningful: Third, firms that employ AI in high-stakes decisions should put in place mechanisms that are accessible and transparent for affected persons to challenge algorithmic decisions, get real explanations, and have access to appropriate remedies when harmful outcomes are identified. These mechanisms should be designed with the interests of affected individuals, not the operational convenience of the deploying firm, as the primary design criterion, and be subject to independent audit. Where class-level harms are found, redress mechanisms should function both at the collective and individual levels.

9. Conclusion

The subject of the title, ‘Who is responsible when the algorithm decides?’, is not easy to answer. We have carried out an analysis of Amazon, UnitedHealth Group, and Deutsche Bank highlighting how AI implementation in important corporate decisions creates structural conditions for accountability to be distributed among technical systems, organizations, and time. This creates a situation where victims will get harmed but will be unable to find anyone to turn for accountability claims. In this article, we have outlined the Accountability Diffusion Framework, which consists of four mechanisms, namely: the diffusion of opacity, temporal dispersion, fragmentation of agency, and laundering of norms. These mechanisms allow for analyzing the accountability problem in any sector no matter whether a firm does business in one country or the other.

Our analysis carries important practical implications for the efforts to restore moral accountability in organisations employing AI. It goes beyond the basic need for the technical solutions to the problem of explainability. It requires a complete overhaul in the architecture and distribution of accountability inside companies that use AI. It also necessitates a major regulatory innovation in terms of introducing the instruments that will be capable of addressing the accountability issues arising from the use of consequential AI, rather than just applying existing rules of consumer protection and antidiscrimination saving their form, so they did not change radically.

This change means a lot for ethics in business. Studies show that businesses are viewed as moral actors that can be held accountable for negative effects of their decisions. We believe that using artificial intelligence will be an interesting test of this assertion. Companies that take advantage of the technological complexity and distributed nature of their operations cannot simply avoid following the rules, but also lose their moral right to operate. Therefore, ethics of business today should underline the tight bond between the use of the power of companies through algorithms and the ethical responsibility of companies.

References

- Albaroudi, E., Mansouri, T. and Alameer, A. (2024). A comprehensive review of AI techniques for addressing algorithmic bias in job hiring. *AI*, 5(1), 383–404. <https://doi.org/10.3390/ai5010019>
- Beveridge, R. (2012). Consultants, depoliticisation and arena-shifting in the policy process: Privatising water policy in Berlin. *Policy Sciences*, 45(1), 47–68. <https://doi.org/10.1007/s11077-011-9144-4>
- Binns, R. (2018). *Fairness in machine learning: Lessons from political philosophy*. Proceedings of the 2018 Conference on Fairness, Accountability and Transparency, PMLR 81, 149–159.
- Bratman, M. (1992). Shared cooperative activity. *Philosophical Review*, 101(2), 327–341. <https://doi.org/10.2307/2185537>
- Cihon, P., Schuett, J. and Baum, S.D. (2021). Corporate governance of artificial intelligence in the public interest. *Information*, 12(7), 275. <https://doi.org/10.3390/info12070275>
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62. <https://doi.org/10.1145/2844110>
- Donaldson, T. (1982). *Corporations and Morality*. Prentice-Hall, Englewood Cliffs.
- Farina, M., Zhdanov, P., Karimov, A. and Lavazza, A. (2024). AI and society: a virtue ethics approach. *AI & Society*, 39(3), 1127–1140. <https://doi.org/10.1007/s00146-022-01530-y>
- Fischer, J.M. and Ravizza, M. (1998). Responsibility and Control: A Theory of Moral Responsibility. *Cambridge University Press*, Cambridge. <https://doi.org/10.1017/CBO9780511814594>
- French, P.A. (1984). *Collective and Corporate Responsibility*. Columbia University Press, New York. <https://doi.org/10.7312/fren90672>
- Hartley, J. (2004). *Case study research*. In C. Cassell and G. Symon (Eds.), *Essential Guide to Qualitative Methods in Organizational Research* (pp. 323–333). Sage, London. <https://doi.org/10.4135/9781446280119.n26>
- Kearns, M. and Roth, A. (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press, Oxford.
- List, C. and Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press, Oxford.
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21. <https://doi.org/10.1177/2053951716679679>
- Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics*, 2(1), 25–42. <https://doi.org/10.1007/BF02639315>
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, Cambridge. <https://doi.org/10.4159/harvard.9780674736061>
- Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D. and Barnes, P. (2020). *Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing*. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33–44. <https://doi.org/10.1145/3351095.3372873>

- Reuters (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters News Agency.
- Scanlon, T.M. (1998). *What We Owe to Each Other*. Harvard University Press, Cambridge.
- Sheard, N. (2025). Algorithm-facilitated discrimination: A socio-legal study of the use by employers of artificial intelligence hiring systems. *Journal of Law and Society*, 52, 269–291. <https://doi.org/10.1111/jols.12535>
- Strawson, P.F. (1962). *Freedom and resentment*. *Proceedings of the British Academy*, 48, 1-25.
- Velasquez, M.G. (1983). Why corporations are not morally responsible for anything they do. *Business & Professional Ethics Journal*, 2(3), 1-18. <https://doi.org/10.5840/bpej19832349>
- Wallace, R.J. (1994). *Responsibility and the Moral Sentiments*. Harvard University Press, Cambridge.
- Yin, R.K. (2018). *Case Study Research and Applications: Design and Methods* (6th ed.). Sage, Thousand Oaks.
- Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, 41(1), 118-132.