EuroMid Journal of Business and Tech-Innovation (EJBTI)
https://iejbti.com

Online ISSN: 3062-2131
Print ISSN: xxxx-xxxx

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

Kesmat AbdelAziz*
Arab Academy for Science, Technology & Maritime Transport, Cairo, Egypt.
E-mail: kesmatmyehia@gmail.com
*Corresponding Author

Received: June 2024; **Accepted**: December 2024

Abstract: The use of artificial intelligence systems by corporations in making important decisions, such as who to hire, who gets health care coverage, or who qualifies for credit, raises fundamental moral accountability issues that current business ethics frameworks do not adequately address. Who is morally responsible for a decision if a job application is rejected, a health insurance claim is denied, or a credit request is turned down? In this paper, we present a comparative case study of three firms — Amazon (automated hiring), UnitedHealth Group (AI-based healthcare benefit determinations), and Deutsche Bank (algorithmic credit scoring) — to explore how moral accountability is constructed, distributed, and contested in AI-mediated corporate decision environments. 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

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 – turn off the tool and tell recruiters not to rely on it – raised a question the company notably refused to answer: who was morally responsible for the harm done to the applicants the system had already judged.

Amazon's experience was neither isolated nor unique. In sectors and across the globe, companies have used AI systems to make or inform consequential decisions that affect people's access to employment, health care, credit, insurance, housing, and social services. Within each domain, the deployment of AI has produced observable patterns of harm – discriminatory outcomes, opaque refusals, systematic miscalibration against marginalized populations – that are remarkably resistant to attribution of moral responsibility. Engineers point to training data, data scientists point to business requirements, managers point to algorithm outputs, and executives point to regulatory compliance. The result is what we refer to as accountability diffusion, the systematic distribution of moral responsibility throughout technical systems, organizational hierarchies, and temporal sequences in ways that do not clearly make any individual or institution accountable for harmful outcomes.

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 paper has three original contributions to the business ethics literature. First, it develops the Accountability Diffusion Framework (ADF). The ADF identifies four mechanisms (opacity diffusion, temporal displacement, agential fragmentation, and normative laundering) through which the deployment of AI erodes moral accountability in corporate contexts. Second, it systematically employs this framework across three comparative cases to develop cross-case insights about the structural drivers of accountability diffusion that transcend any one firm or sector. Third, it proposes a set of accountability design principles – organizational, governance, and regulatory measures – that are specifically responsive to the accountability challenges 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 accountability, as it is defined in the philosophical literature, is the normative relationship that exists whereby an agent can be held responsible for the consequences of their actions or omissions – called to account, required to explain, and potentially blameworthy, sanctioned, or required to take remedial action (Scanlon, 1998; Strawson, 1962). According to standard accounts of moral accountability, three conditions must be met. First, the agent's action or omission must have made a difference to the outcome (causal contribution). Second, the agent must have known or should have known the relevant facts (epistemic access). Third, the agent must have been capable of acting otherwise (normative capacity) (Fischer and Ravizza, 1998; Wallace, 1994).

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

The use of AI systems in corporate decision-making introduces features that challenge the application of standard frameworks of moral accountability, as extensively documented in the emerging literature on algorithmic governance and AI ethics (Diakopoulos, 2016; Mittelstadt et al., 2016; Raji et al., 2020). There are three particularly important features. First, opacity: many high-performance AI systems—especially those built on deep-learning architectures—are uninterpretable, meaning that human observers cannot tell why a particular decision was reached. The 'black box' nature of such systems makes it difficult to establish the causal chain between algorithmic decision and harmful outcome, undermining the causal contribution condition of moral accountability.

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. This distributed development process produces multiple points at which responsibility can be deflected, creating what Nissenbaum (1996) famously termed the 'problem of many hands' -- a situation in which a harmful outcome is the product of many individual contributions, none of which is sufficient in itself to constitute full moral responsibility.

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 work applying virtue ethics to the AI context reinforces this position, arguing that organisational character — the cumulative disposition of a firm to act responsibly or irresponsibly in its AI deployments — is itself a proper object of moral evaluation (Farina et al., 2024). This view supports attributing moral responsibility to corporations for their AI uses, without having to hold any individual employee responsible for results beyond individual control.

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 paper, we employ a comparative case study approach, based on the canonical framework for case study research by Yin (2018) and adapted to the normative purposes of business ethics scholarship (Beveridge, 2012; Hartley, 2004). The case study methodology is appropriate for this research because our main concern is to understand the processes by which moral accountability is constructed and lost in particular organizational and technological circumstances – a question that demands the kind of detailed, contextually aware analysis that quantitative methods are not capable of providing. "Comparative design allows for cross-case pattern recognition while maintaining sensitivity to the particularities of each case".

The three cases – Amazon, UnitedHealth Group, and Deutsche Bank – were selected through purposive theoretical sampling to maximize variation on two dimensions: industry sector (technology, healthcare, financial services) and the nature of the AI-driven decision at issue (hiring, benefits determination, credit scoring). All three involve AI systems making or materially influencing decisions with significant consequences for the people affected, and all three have been the subject of significant public documentation that can be subjected to rigorous evidence-based analysis. The cases are not intended to be representative of AI deployment practices across their respective industries, but are selected as theoretically generative examples that shed light on the accountability mechanisms identified in the 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. We examined each case for the presence of the four accountability diffusion mechanisms – opacity diffusion, temporal displacement, agential fragmentation, and normative laundering – and for corporate responses to accountability struggles that emerged through public scrutiny. Cross-case analysis then revealed patterns of similarity and difference that inform the accountability design principles developed 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 is not to say that moral accountability is abolished – firms that intentionally foster opacity to insulate themselves from accountability have a greater, not lesser, degree of moral culpability – but it does generate major practical barriers to accountability that can be exploited by firms seeking to reduce their liability exposure.

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 the dispersion of the design, development, deployment, and governance of AI systems across multiple organizational actors – engineers, data scientists, product managers, compliance officers, business users, senior executives – such that no individual action is enough to constitute full moral agency over the harmful outcome. Part of the agential fragmentation is an inevitable consequence of the collaborative nature of complex AI development. But it can also be intentionally designed – through organizational structures, role definitions and governance arrangements that ensure that accountability gaps are present at all levels – to prevent the attribution of clear moral responsibility to any identifiable individual.

4.4. Normative Laundering

The most insidious of the four mechanisms may be normative laundering. It refers to the adoption of formal ethical frameworks, responsible AI principles, bias audits, and regulatory compliance processes to produce the appearance of normative accountability without altering the underlying accountability structures of AI governance. A company that issues detailed AI ethics principles, commissions third-party audits of its algorithmic systems, and engages in multi-stakeholder responsible AI initiatives can simultaneously deploy AI systems that cause significant harm, without establishing clear lines of moral accountability for that harm. The ethical apparatus legitimizes continued deployment—by showing

good faith engagement with accountability concerns—while the substance of accountability remains elusive.

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 — Summary

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	Legitimates continued harm through ethical performance

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

5.1. Context and System Description

Amazon started experimenting with AI-assisted recruitment around 2014, when machine learning teams in Edinburgh developed a tool to automate screening of job applications for technical roles. The system had been trained on about ten years of historical CV data and hiring decisions, from which it was supposed to learn the attributes that predicted successful performance at Amazon. As of 2015, the system was reportedly rating candidates on a five-star scale and was being used to surface promising candidates from large pools of applications. And the system's biased outputs – systematically dinging CVs that contained words like 'women's' (as in 'women's chess club') and penalizing graduates of all-female colleges – were identified internally around 2015, and the project was ultimately abandoned by 2017, but the discovery wasn't publicly reported until October 2018.

Amazon's case is instructive not so much because of the discriminatory outputs the system produced — gender bias in automated hiring has been widely documented across multiple firms and systems (Albaroudi et al., 2024; Sheard, 2025) — but because of what the corporate response to the discovery reveals about the accountability structures surrounding AI deployment. When the bias was internally discovered, Amazon responded by attempting technical fixes and ultimately decommissioning the tool. There was no public announcement, no notice to the affected applicants, no systematic effort to identify and remedy harms to women rejected by the system, and no formal accountability mechanism by which those responsible for the system's design and deployment were required to justify its discriminatory outputs. This pattern of organisational evasion mirrors broader findings on how social and structural processes — not only technical design choices — generate and sustain algorithm-facilitated discrimination in hiring contexts (Sheard, 2025).

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 distribution of accountability over the various teams – data engineering, HR technology, business leadership – engaged in the development and deployment of the system. The engineers who built the model weren't responsible for the decision to deploy it, the managers who approved deployment weren't responsible for the technical architecture, and the recruiters who used the outputs weren't responsible for what the system was doing under the hood. This diffusion meant that no one individual's contribution was sufficiently large to amount to clear moral agency over the discriminatory outcomes.

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: UnitedHealth 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. UnitedHealth said the system was not used to automatically deny coverage and said determinations were always made by qualified clinical reviewers who used the model as one input among many. This defense – that the AI system did not make the relevant decisions but merely informed them – provides a particularly clear example of the accountability diffusion dynamics that the ADF is designed to identify.

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 temporal displacement here is particularly disturbing. The harms of premature discharge from post-acute care — declining health outcomes, hospital readmissions, patient deaths — may not be apparent until days or weeks after the coverage denial that set them in motion and may be blamed on the patient's pre-existing condition rather than on the inadequacy of the care they received. This causal complexity presents serious obstacles to establishing the link between an algorithmic coverage denial and a patient outcome, further undermining the epistemic access condition of moral accountability.

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.

Perhaps this case represents the worst form of normative laundering. UnitedHealth's public commitments to value-based care, patient-centered outcomes, and quality clinical decision-making provided a legitimating framework within which the deployment of a coverage-denial algorithm was framed as consistent with – indeed, as serving – the company's stated mission to improve health outcomes. Formal clinical review processes existed that provided

procedural legitimacy for algorithmic determinations that the underlying data appeared to be generating systematically harmful outcomes. The company's compliance with Medicare Advantage regulatory standards — which do not mention the use of AI in coverage determinations — gave the practice regulatory cover that shielded it from scrutiny until independent journalism uncovered the extent of the problem.

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' legitimation 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.

Second, all three cases demonstrate the importance of normative laundering to the process of accountability diffusion. In each case, formal ethical commitments—to diversity and inclusion, to patient-centered care, to fair financial access—offered a legitimating narrative in which clearly harmful algorithmic practices were rendered invisible or acceptable until exposed by

external investigation. This finding is at odds with the responsible AI literature's depiction of ethical principle adoption as a positive sign of progress toward accountability: in the absence of meaningful enforcement mechanisms, principles are not only insufficient but actively harmful, providing a veil for continued harm while stalling the institutional changes that true accountability would require.

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 question in our title — who is responsible when the algorithm decides? — does not have an easy answer. Our analysis of Amazon, UnitedHealth Group, and Deutsche Bank shows that the deployment of AI in high-stakes corporate decision-making creates systematic conditions for the diffusion of moral accountability across technical systems, organizational hierarchies, and temporal sequences in ways that leave affected individuals with harms and no identifiable accountable party. The Accountability Diffusion Framework developed in this paper, which conceptualizes four key mechanisms of diffusion – opacity diffusion, temporal displacement, agential fragmentation, and normative laundering – provides a conceptual language for analyzing and responding to these conditions that transcends any one firm, sector, or regulatory jurisdiction.

The practical implication of our analysis is that, to restore moral accountability in AI-driven organizations, more is required than technical solutions to the explainability problem. It calls for fundamental changes to organizational governance — the way accountability is structured, allocated, and enforced within AI-deploying firms. It demands regulatory innovation — the development of frameworks that are specifically tailored to face the accountability challenges of high-stakes AI deployment rather than merely extending existing consumer protection and anti-discrimination law. And it requires normative reorientation — a shift from the responsible AI discourse's focus on the adoption of ethical principles towards a genuine accountability culture in which the costs of harmful AI-driven decisions are borne by the ones who made the decisions to deploy, not only by those who suffered their consequences.

This reorientation has profound implications for business ethics. The literature on corporate moral agency has long argued that corporations can be real moral agents, responsible for the systemic consequences of their actions. Our analysis suggests that the deployment of AI systems is a particularly acute test of

this claim: firms that use the technical complexity and organizational diffusion of AI governance to evade accountability for harmful algorithmic decisions are not merely failing a technical compliance requirement – they are abdicating the moral agency that gives corporate claims to legitimacy their foundation. The business ethics appropriate to the age of AI must insist, and help firms understand how to achieve, the indissoluble connection between the exercise of corporate power through algorithmic systems, and the moral accountability that the exercise of power demands.

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). R*esponsibility 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 Value*s, 41(1), 118-132.