SEKYRA FOUNDATION
Ethics & Bioethics (in Central Europe), 2026, 16 (1–2), 77–90 DOI:10.2478/ebce-2026-0006

Rethinking corporate responsibility in the age of Artificial Intelligence

Martin Pazdera[1]

Abstract
The growing use of artificial intelligence (AI) in business decision-making is changing how people think about responsibility, accountability, and moral justification. As companies use AI systems for prediction, optimization, and automation, decisions are increasingly based on socio-technical systems in which data, models, organizational incentives, and human judgment work together to produce results. This article contends that traditional frameworks in business ethics and corporate governance are increasingly encountering conceptual challenges when applied to AI-mediated decisions, especially in contexts characterized by opacity, scale, and adaptive behavior that hinder traceability and contestability. The article constructs a comprehensive analysis of three ethical-responsibility domains: algorithmic decision-making (bias, opacity, and responsibility gaps), organizational governance (authority redistribution and oversight challenges), and stakeholder fairness (discrimination risks, trust, and legitimacy), drawing upon interdisciplinary research in business ethics, corporate governance, and AI ethics. The analysis takes place in a changing regulatory environment, including the EU Artificial Intelligence Act, the General Data Protection Regulation, and international standards and guidance from bodies such as the OECD and NIST. The article subsequently presents a governance-focused conceptual model for ethical AI in business, linking explainability, auditability, human oversight, and stakeholder participation as interdependent components. Instead of giving a set of rules to follow, the model is meant to be an analytical tool for corporate AI governance that goes beyond mere rule-following and supports long-term legitimacy and value creation.

Keywords: artificial intelligence; corporate responsibility; business ethics; corporate governance; accountability; stakeholder fairness; explainable AI

Introduction

It is a common mistake to think that artificial intelligence is still only on the fringes of business. These systems now actively control how opportunities and burdens are shared within organizations. AI now decides which candidates to interview, which customers receive credit, how dynamic pricing works, and which transactions are flagged as risky. This change has significant moral implications because these decisions are not just about speeding up internal processes. In a very real sense, they are deciding who gets what in life and whether the market remains fair.

The normative challenge isn't just that AI can be biased or hard to understand. AI changes the way responsibility is structured in business on a more basic level. Business ethics and corporate governance have always assumed that decisions can be traced back to specific individuals within formal authority structures. All stakeholder theory, CSR, and legitimacy approaches are based on the idea that organizations can explain and justify their choices to those affected (Freeman, 1984, pp. 24–26; Carroll, 1991; Suchman, 1995).

AI-driven systems are making these old ideas look less stable. Many of these tools work by recognizing statistical patterns in historical data, so their internal logic often relies on probabilities and, to be honest, is hard to understand (Burrell, 2016, pp. 1–12). Even when a company says it owns an outcome, it can be hard to determine how its original design choices led to the final real-world effect. In practice, we see a dangerous lack of accountability. Managers start to rely on the model's output, technical teams see every ethical issue as just a technical problem, and boards of directors often don't have the technical knowledge to really keep an eye on things. This is precisely how responsibility gaps emerge. It is not just a problem of who to blame when things go wrong; it is a fundamental crisis of institutional governance (Matthias, 2004; Kroll et al., 2017).

Using a bioethical framework makes it much clearer how serious these governance problems

[1] University of Prešov (Slovakia); email: martin.pazdera@unipo.sk; ORCID: 0009-0000-0606-4476

really are. Begin with the idea of autonomy. When algorithmic systems limit a person's options, like when they decide if someone can get a loan or get a job, without giving a clear reason or a way to appeal, it is fundamentally compromised (Beauchamp & Childress, 2019; Wachter, Mittelstadt & Floridi, 2017). We also need to consider the principle of non-maleficence. This principle is in danger whenever harms that can be predicted happen indirectly through statistical inference. In a technical sense, these harms might not be "nobody's fault" because they were not intended, but they can still be completely avoided if the right institutional protections are in place (Baracas & Selbst, 2016; Matthias, 2004; NIST, 2023).

Furthermore, the concept of beneficence is central here because the actual rewards of AI optimization are almost never distributed symmetrically. While firms usually capture the bulk of the efficiency gains, stakeholders are left to bear the risks, which can range from false positives and social exclusion to constant surveillance (Zuboff, 2019; O'Neil, 2016). Finally, the question of justice is implicated wherever these systems entrench structural inequality or generate disparate impacts via proxies and feedback loops (Rawls, 1971; Barocas & Selbst, 2016; Buolamwini & Gebru, 2018). All these points lead back to the article's central argument. Ensuring ethical AI in a business setting is not merely a technical design challenge. It is, at its heart, a governance problem that involves institutional responsibility in an environment defined by opacity, massive scale, and a deep asymmetry of power.

It is becoming increasingly clear that the people who make rules and set global standards see these problems as deeply rooted in the system. The EU Artificial Intelligence Act is a good example of this trend. It uses a risk-based approach to assign specific governance duties to high-risk systems. These duties include everything from strict risk management and record-keeping to the need for real human oversight (Regulation (EU) 2024/1689). The GDPR also sets important standards by requiring fairness and accountability and putting strict limits on how personal data can be used (Regulation (EU) 2016/679). The OECD AI Principles and the NIST AI Risk Management Framework are two international frameworks that are beginning to describe what trustworthy AI should look like when companies use it (OECD, 2019; NIST, 2023).

Even with new rules in place, many companies respond by focusing on documentation and formal processes rather than the ethical issues underlying their decisions or on how responsibility is shared throughout the company. The point of view in this work is one of carefulness with a critical edge. We are not saying that established ideas about business ethics or corporate governance have completely failed. The opposite is true: the argument is that AI creates specific and ongoing tensions that need to be resolved if the idea of corporate responsibility is to remain relevant in an era of automation and large-scale. The research questions here are meant to clarify concepts rather than seek a grand theoretical replacement. They are written like this.

The first question examines how AI-assisted decision-making makes it harder for traditional models of corporate responsibility based on human agency and actions that can be traced. After that, the focus shifts to how AI use changes who holds power in companies and the specific governance problems that result. A third area of concern is how AI systems might create or make unfairness worse for stakeholders, and how these effects might affect trust in organizations and social legitimacy.

The study also asks how businesses should view new tools for regulation and standard setting as a starting point for governance rather than a complete replacement for moral duty. Finally, I want to figure out which parts of governance are best at supporting ethical AI when the end goal is strong accountability that goes beyond mere compliance with the law.

The goal of this article is to be both integrative and focused on governance. It wants to connect the normative language of AI ethics, like fairness and openness, with the institutional logic of corporate governance, like oversight and control structures. It also links these to the

legitimacy issues that are so important to business ethics, such as stakeholder justification and the social license to operate. In this way, it shows how bioethical principles can help us better judge corporate AI as an institutional power that limits individual freedom, human dignity, and distributive justice.

Business ethics and corporate responsibility in an automated environment

For a long time, business ethics scholars have been actively fighting against the idea that shareholders should come first. Researchers have consistently underscored the moral significance of all stakeholders impacted by corporate decisions. For instance, stakeholder theory views the corporation as a complex network of relationships in which long-term stability and legitimacy depend on treating all interests fairly (Freeman, 1984, pp. 24–26; Donaldson & Preston, 1995; Freeman et al., 2010, pp. 3–10). Different frameworks for Corporate Social Responsibility also hold that a company's responsibilities extend beyond mere compliance with the law. These models contend that ethical and philanthropic initiatives are integral aspects of an organization's global responsibilities (Carroll, 1991; Garriga & Melé, 2004).

Legitimacy theory also suggests that an organization's very survival depends on how well it follows social norms and meets public expectations. A company must not only be profitable but also be perceived as fundamentally appropriate and trustworthy (Suchman, 1995, pp. 571–610). In these academic traditions, the idea of responsibility isn't just about blaming people for things that have already happened. It is also a process that is happening now and looking ahead. Now, organizations are expected to consider how their actions will affect others and take steps to protect themselves from risks so that their authority remains socially acceptable. This change is very important because what businesses do have a big impact on more than just market trends. They fundamentally delineate the genuine life opportunities of stakeholders, influencing aspects ranging from employment opportunities to access to essential services and overall vulnerability to risk.

The core of these traditions is the basic idea of justification. The main point is that a business should be able to explain why it did something and who is responsible for those actions. A significant moral assessment is predicated on the belief that actions transcend mere mechanical occurrences. They should instead be understood within a normative framework that considers both accountability and practical implications (Švaňa, 2023, pp. 72–82). In human-centered processes, justification is usually based on an individual's deliberation, personal choice, and specific job duties.

When AI is involved in making decisions, though, this act of justification becomes much more complicated. The reasons for a given result are often statistical rather than direct. Also, the process of making a choice is spread across a wide network of social and technical systems rather than being linked to just one person (Matthias, 2004; Kroll et al., 2017). We must also acknowledge that the rapidity and extensive scope of automated decisions can erode the social norms that typically uphold accountability. It is easy to lose careful thinking, situation-based decision-making, and the simple chance of a meaningful appeal (Diakopoulos, 2016, pp. 56–62). The risk is that justification becomes merely procedural. An organization might say its actions are based on a corporate policy or a model design, but it might not care at all whether the decision was fair to the person affected.

A common way for businesses to deal with these problems is to frame the whole thing as a technical issue. They work on improving a model, adding additional bias checks, or maybe even writing some documentation. Even though these steps may be useful on their own, they often miss the bigger ethical picture of the problem. The straightforward reality is that ethical responsibility cannot be confined to mere technical metrics or numerical values. Ethics entails intricate normative assessments regarding the essence of equitable treatment, the permissible trade-offs, and the requisite respect for affected individuals (Rawls, 1971, pp. 3–24; Sen, 2009,

pp. 7–9). So, the real moral problem is not just figuring out how to make better models. Instead, we need to figure out how to run these decision systems so that moral responsibility is clear and open to debate.

Corporate governance, accountability, and the board's role in AI oversight

When people study corporate governance, they usually examine the specific ways a company can ensure its actions align with its fiduciary duties and with society's expectations. At its core, agency theory emphasizes the need for monitoring and control to reduce the inherent conflicts of interest between owners and managers (Jensen & Meckling, 1976; Fama & Jensen, 1983). Governance now includes risk management and internal controls in addition to the board's responsibility. These are often formalized through international standards and professional guidance (OECD, 2015; COSO, 2017; Tricker, 2015).

Adding AI to this environment fundamentally changes these traditional structures, widening the gap between board-level responsibility and on-the-ground operations. Boards of directors are still legally responsible for overseeing risk, but the technical and organizational complexity of AI systems can make meaningful supervision almost impossible. In practice, boards often give management and technical teams the job of keeping an eye on things. This creates a governance bottleneck. In this case, responsibility is still formal on paper, but real control is only partial. This gap is important from an ethical perspective because it is often where harmful effects persist when governance fails. This happened not because no one cared about the result, but because the way the organization was set up made it hard to take real responsibility.

Also, research in governance helps explain why strategies that focus solely on compliance are inherently weak. If AI governance is seen as just an extra thing separate from strategy, risk, and corporate culture, then ethical issues will always be on the outside looking in and will only respond to events. If AI governance is built into the systems that already give people power and make them responsible, on the other hand, it can really become part of how an organization sees its duty to its stakeholders and society as a whole.

Perspectives on AI ethics and bioethics

A lot of research has already been done on AI ethics on fairness and discrimination. We now have a much better idea of how different models can reflect, or even worsen, structural inequality by using biased data or proxy variables, even without anyone meaning to discriminate (Barocas & Selbst, 2016; Hardt, Price & Srebro, 2016; Buolamwini & Gebru, 2018). The lack of transparency remains a major concern, along with bias. Black box models can make things less clear and weaken the protection of due process. This is especially problematic when the people affected by these systems can't understand or question the logic behind a given outcome (Pasquale, 2015, pp. 3–18; Burrell, 2016; Wachter, Mittelstadt & Floridi, 2017). So, finding a designer is not enough to make algorithmic systems accountable. The question is whether the decisions that emerge from this process can be tracked, explained, and challenged in a real institutional setting (Diakopoulos, 2016; Kroll et al., 2017; Raji et al., 2020).

This leads us to the major risk, often called the "responsibility gap". When learning systems and their environments interact in ways that are hard to predict, our usual ways of assigning blame become less stable (Matthias, 2004, pp. 175–183). In a business setting, these gaps in responsibility often turn into gaps in governance. The main question is no longer just who is to blame. We should instead ask which types of institutions can ensure accountability even when things get complicated. This article is based on the fundamentals of AI ethics, but its main goal is to turn those vague ideas into a robust system of governance.

Bioethical perspectives are also very important here because they view institutional decision-making systems as more than just neutral tools. Instead, they think of them as moral agents that

shape how people are put at risk and how benefits are shared. At their core, these points of view examine whether people are treated as ends in themselves or merely as means to an end (French, 1979). In today's business world, AI systems are increasingly taking over areas that directly affect people's health and dignity. This includes everything from hiring and obtaining credit to setting insurance prices, detecting fraud, and managing at-risk customers. Because of this, the moral evaluation of these systems naturally overlaps with the most important issues in bioethics, such as protecting the weak, preventing harm, and seeking justice (Beauchamp & Childress, 2019; UNESCO, 2005).

One can easily connect a vocabulary focused on bioethical principles such as autonomy, non-maleficence, beneficence, and justice to the rules governing corporate AI (Beauchamp & Childress, 2019). The most important thing about autonomy is that people must be able to do something meaningful in response to decisions that affect them. In the context of AI-mediated business decisions, this means that there needs to be explanations that make sense, the real ability to challenge a decision, and a way to opt out or appeal (Wachter, Mittelstadt & Floridi, 2017; GDPR, 2016, art. 22; European Parliament & Council of the European Union, 2024). It is important to understand that technical transparency alone does not satisfy autonomy. It needs governance structures that really help people understand and give them real power over the system (Kroll et al., 2017; European Parliament & Council of the European Union, 2024).

In the same way, the principle of non-maleficence says that people should stay away from harms that can be predicted (Beauchamp & Childress, 2019). These harms are often indirect and based on chance when using AI systems. For instance, a fraud detection system that produces false positives could cause an account to be closed immediately, while a risk model that produces false negatives could leave customers open to offers that are clearly not in their best interest. There is also a chance that biased selection processes will exclude fully qualified candidates (Mittelstadt et al., 2016). Just because these harms happen through statistical inference doesn't mean they are morally accidental in any way. Instead, this reality necessitates that governments establish monitoring systems, intervention thresholds, and organized mechanisms to address problems (NIST, 2023; OECD, 2019). In this way, non-maleficence requires governance that includes testing a tool before it is put into use, monitoring it after it is put into use, and having a way to respond quickly when something goes wrong (NIST, 2023; European Parliament & Council of the European Union, 2024).

According to the principle of beneficence, organizations must demonstrate that using AI is a real benefit to stakeholders' welfare, rather than merely a means to make the company more efficient (Beauchamp & Childress, 2019; OECD, 2019). In many businesses, the term "benefit" is often used to mean little more than lower costs or faster processes. From a bioethical perspective, this narrow focus is insufficient if the optimization imposes burdens on others. A truly sensitive approach to beneficence must clearly explain how stakeholders benefit and the trade-offs involved. This means checking whether the system actively lowers discriminatory barriers, makes it easier to reach important services, or prevents real harm. For instance, a system could make it easier to detect fraud without unfairly excluding people (OECD, 2019; NIST, 2023). Beneficence also means that we need to find out whose definition of benefit is really shaping the system's design. Are shareholders, management, regulators, or stakeholders most affected (OECD, 2019)?

On the other hand, justice requires that risks and rewards be shared fairly and that there be strong protection against unfair practices (Beauchamp & Childress, 2019). It is well known that AI systems can reflect existing structural inequalities through biased data, the use of proxy variables, and dangerous feedback loops (Barocas & Selbst, 2016; Buolamwini & Gebru, 2018). Justice must also have a procedural side. People affected by these systems should never have to deal with an authority that seems arbitrary and lacks easy ways to fix problems (Kroll et al., 2017; GDPR, 2016, art. 22; European Parliament & Council of the European Union, 2024). In

bioethics, this kind of procedural justice is built into bodies such as ethics committees and oversight bodies that make decisions. In the context of corporate AI governance, similar tools include impact assessments, external audits, and stakeholder involvement to identify knowledge gaps within the institution (Kroll et al., 2017; NIST, 2023).

Lastly, a bioethics lens emphasizes human dignity and the fact that people are weak (UNESCO, 2005). Corporate AI often works in places where people don't have much power to negotiate, like employees, low-income debtors, or already marginalized groups. Governance structures should be very careful about consent when the option to refuse is not available. There should be a clear focus on putting limits on surveillance, manipulative targeting, and personalization that feels like it takes advantage of people. From this point of view, ethical governance is more than just stopping bias. It is about putting an end to institutional practices that treat people as nothing more than tools for optimization.

Algorithmic decision-making: Gaps in bias, opacity, and responsibility

The main reason companies are interested in using algorithms to make decisions is that they promise to be very large and always the same. But that same scale can make unfairness even worse. When models are trained on historical datasets, they risk perpetuating past discrimination, especially when social inequalities are already evident in the data. This risk is not limited to overtly sensitive labels. Features that seem completely neutral can end up standing in for protected traits, with different effects on everyone (Barocas & Selbst, 2016; O'Neil, 2016, pp. 3–11). In a business setting, these effects can affect everything from hiring and promotions to credit granting, price setting, and even customer screening.

The problem is made worse by the fact that it is hard to see through. The idea of procedural fairness starts to fall apart if the people who are affected by a decision can't understand why it was made. A business may say it is making decisions based on pure logic, but from the stakeholder's point of view, a logical choice that isn't explained feels a lot like using power without reason. This means that the main ethical issue isn't just the chance of bias. The entire process becomes much harder to question or hold accountable (Pasquale, 2015, pp. 3–18; Burrell, 2016).

This is where the idea of a responsibility gap really starts to make sense in real life. A manager might say the model suggested shifting the blame, while a developer might say the system is just showing how the data really is. The organization, on the other hand, might say that it has followed all its own rules to the letter. We end up with a spread of responsibility across different roles, which makes it more likely that problems will keep happening without anyone doing anything to fix them. This kind of spread is the main reason any governance framework needs to prioritize traceability and thorough documentation. The emphasis must be on the system's accountability in the real-world post-deployment, rather than solely examining the algorithm during the design phase (Kroll et al., 2017; Mitchell et al., 2019; Raji et al., 2020).

It is possible to make a careful, but entirely reasonable, case here. As AI gets more popular, the need for governance systems that can turn raw statistical data into decisions that can be challenged and linked to institutional responsibility becomes even more urgent. We shouldn't assume that this means that every model must be a completely open book in a technical sense. But it does require that a company provide clear explanations, make it easy for people to appeal, and demonstrate ongoing oversight.

Governance and redistribution of authority: Oversight, accountability, and internal power

It's becoming increasingly clear that AI systems are doing more than just crunching numbers; they're changing the way power operates within organizations. Tasks that used to depend on a manager's personal judgment are now often based on the model's standardized outputs. This

change tends to take away people on the front lines' freedom and give more power to data science teams, outside vendors, or central analytics departments. What we see here is not just an increase in efficiency, but a huge shift in power. This frequently occurs without thorough contemplation regarding the rightful ownership of decision-making authority or the ultimate power to supersede the system when required (Kellogg, Valentine & Christin, 2020).

This change is very important from an ethical point of view because the idea of corporate responsibility depends on clear lines of responsibility. If authority changes but the system of accountability remain the same, an organization can't ensure its ethical promises are kept in practice. Even though boards and senior executives are still legally responsible for the results, the operational levers are often hidden within technical systems that are difficult to understand.

The resulting governance challenge is essentially twofold. First, supervision must be feasible in practice. This means that boards and top executives need to know enough about technology and have a good reporting system in place to really understand the risks, the limits, and the many ways a system could fail. Second, there needs to be a clear structure for accountability. There must be clear ownership of things like how well a model works, how good the data is, how often it is checked, and how to fix problems. This must also include the power to stop or change a system as soon as it starts to hurt people. Without these kinds of structures, a company could end up with power without responsibility, which is exactly what happens when institutional legitimacy starts to fade.

Fairness for stakeholders: Discrimination, exploitation, trust, and legitimacy

Just because an AI system doesn't show obvious signs of discrimination doesn't mean it's safe. These tools can still hurt stakeholders in big ways, though, through less obvious means like exploitation, manipulation, or an unfair distribution of risk. Constantly watching people gather data often leads to significant power imbalances and information asymmetry. This lets companies get a lot of value from people without giving them anything of equal value or getting anything like real consent (Zuboff, 2019). Automated profiling works in a similar way, breaking customers into groups that can target their specific weaknesses or slowly take away their freedom. This is especially bad when people can't figure out how the algorithmic targeting works or how to get out of it.

Differential error burdens can also lead to a very subtle form of unfairness. Even if a model does well on average, it can still make some groups pay more than others. Think about how false positives in fraud detection can lead to accounts being frozen, or how false negatives in credit risk can lead to complete exclusion. These results can last even if sensitive attributes are never directly included in the calculation. Additionally, systems intended to be personalized can use soft forms of coercion by carefully shaping the architecture of choice. They push stakeholders to make choices that are good for the company but bad for their own health or freedom, especially when there is a significant information gap (Thaler & Sunstein, 2008; Susser, Roessler & Nissenbaum, 2019; Zuboff, 2019).

These interventions may not technically deprive individuals of their formal freedom of choice, but they do raise serious moral questions about the integrity of human agency. They make us think about what conditions must be met for a decision to be truly independent in the first place (Špirková, 2023, pp. 55–71). From a broader moral perspective, the way technology subtly shapes the places where we make decisions could lead to a form of engineered behavioral steering that replaces careful, thoughtful judgment. This change calls into question our long-held beliefs about responsibility and how we judge morality (Brennan, 2024; Švaňa, 2023).

One can't ignore how significant these problems are in undermining the legitimacy of institutions. People now judge organizations not only by how well they follow the law, but also by how fair and trustworthy their actions seem to be. Legitimacy is a weak thing, and it tends to fade quickly when people see algorithmic choices as random or impossible to question

(Suchman, 1995). Also, any public debate about the harm of an algorithm can quickly lead to more damage to its reputation and more attention from regulators. So, setting up ethical AI governance is not a moral luxury. It is a basic requirement for keeping what is sometimes called a "social license to operate". From a governance perspective, legitimacy also depends on whether stakeholders have reasonable expectations about how decisions are made and whether a company always makes things right when things go wrong.

In this case, stakeholder theory makes a lot of sense because it views corporate responsibility as relational and future oriented. We need to ask not only whether a decision is legal, but also whether it treats stakeholders fairly and provides them with a clear reason for the organization's decision (Freeman, 1984, pp. 24–26; Freeman et al., 2010, pp. 3–10). In the world of AI, this means making systems that can turn technical governance into actions that benefit stakeholders. This means that there is real transparency, real ways to appeal, and real ways to participate that go beyond just making a statement.

Because algorithms make decisions in ways that aren't clear, it's hard for a board or management to keep an eye on things. Leaders have a hard time understanding why certain results occur or how a model's behavior might change over time due to drift (Burrell, 2016; Pasquale, 2015; NIST, 2023; Sculley et al., 2015). Weak governance, on the other hand, makes it much more likely that biased systems will stay in place because accountability is spread too thin across different roles and no one feels like they own the actions that need to be taken (Kroll et al., 2017; Raji et al., 2020; Matthias, 2004). A lot of the time, the unfairness that stakeholders feel isn't caused by just one bug in the code. Instead, it comes from the combined effects of several interacting weaknesses: systems that are hard to understand and built on weak structures of accountability, used on a huge scale, and seen by people as something random and final.

This interaction effect is very important for how governance is set up. If an organization sees fairness as just a number in a technical audit, transparency as just another piece of paper, and accountability as just a box to check for compliance, then the whole system is still ethically weak (Selbst et al., 2019; Raji et al., 2020). Instead, a governance-oriented approach aims for a sense of order. People who are affected by explanations must be able to understand them; auditing must be linked to a real authority that can step in; and oversight must include the power to raise issues or even shut down a system. Lastly, any input from stakeholders must lead to real change in the process (Kroll et al., 2017; NIST, 2023; European Parliament & Council of the European Union, 2024). In short, each part needs to be closely connected to the others to work well.

The rules and policies that are in place

The current regulatory environment is beginning to treat AI governance as a basic organizational duty rather than just another optional ethical initiative. The EU Artificial Intelligence Act is a good example of this change. It establishes a risk-based framework that requires high-risk systems to follow specific rules. These duties include strict risk management, thorough documentation, constant human oversight, and monitoring after the tool is available for sale (Regulation (EU) 2024/1689). The GDPR, like this, sets important and basic rules for how data can be processed legally, who is responsible for it, and how decisions based on data must be fair (Regulation (EU) 2016/679; Voigt & von dem Bussche, 2017). The OECD Recommendation on AI (OECD, 2019) outlines values-based principles for developing AI that people can trust. The NIST AI Risk Management Framework (NIST, 2023) provides a structured approach for mapping, measuring, and managing risks throughout a system's lifecycle.

For those in charge of businesses, these tools should be seen to set minimum standards while still allowing significant ethical freedom. Compliance can help an organization avoid certain legal risks, but it doesn't automatically fix the deeper issues of fairness, accountability, and

social legitimacy. In this way, regulation should not be seen as a complete replacement for moral duty, but rather as a starting point for how to run things.

We need to understand that regulation only sets minimum standards and can never fully cover the ethical duties of businesspeople. Legal categories are, by their very nature, very broad. On the other hand, ethical responsibility is often situation-specific and looks to the future. Organizations often must make decisions that may not break any laws but still raise serious issues about personal freedom, human dignity, and fair resource distribution. For example, worker monitoring systems are a clear example. Even if the law requires data protection and notice, a system of widespread behavioral monitoring can still violate people's dignity by making work feel like constant surveillance. Personalization systems may be legally acceptable, but they are ethically dubious when they exploit the weaknesses of impulsive borrowers or compulsive consumers through unclear profiling. These cases show exactly why just following the rules is not enough to be ethical.

Also, keep in mind that regulation is often reactive by nature. On the other hand, governance should be proactive. Organizations can anticipate potential harm by conducting structured impact assessments, testing different scenarios, and actively involving stakeholders. In practice, this means companies need to commit to integrating ethical questions directly into their business strategy and risk management, rather than leaving them to a compliance process that happens after the fact.

Suggested governance-focused framework for ethical AI in commerce

The main goal of this article is to demonstrate how ethical AI can be integrated into corporate governance as a key part of an overarching conceptual model. This framework is based on four complementary pillars: explainability, auditability, human oversight, and stakeholder involvement.

In this model, explainability is understood in a practical sense, directly related to governance needs. The main point is that an organization needs to be able to provide good reasons for its choices. These explanations should be carefully tailored to what stakeholders need to know and to the level of risk associated with the application (Doshi-Velez & Kim, 2017; Rudin, 2019). Auditability, on the other hand, means that a company can check how well its models are working, how good its data is, and how well its decisions are working overtime. This should also include the option of independent evaluations when an outside view is needed (Raji et al., 2020; Mitchell et al., 2019).

It's not just a vague slogan about having a person in the loop when we talk about human oversight. This is instead framed as a serious promise to keep decision-making power accountable. Such a commitment must encompass unequivocal escalation procedures and the explicit authority to override or even suspend a system when deemed necessary (Kroll et al., 2017). Moreover, stakeholder involvement ensures that governance transcends a limited group of internal participants. An organization can strengthen its social legitimacy and make sure its actions are open to challenge by including the views of people who are affected during the design, evaluation, and review stages (Jobin, Ienca & Vayena, 2019; Suchman, 1995).

This model operates on a conceptual level through two fundamental layers. Explainability and auditability constitute the epistemic foundations of governance, enabling an organization to comprehensively understand a system's functioning and the rationale behind specific outcomes. In the meantime, human oversight and stakeholder participation provide the institutional and normative bases. These measures ensure that people can still hold decision-makers accountable and that the public can still investigate them. This model ensures that ethical AI is never seen as just compliance paperwork by linking technical practices directly to governance duties.

From a governance perspective, implementation is primarily about adding these specific

parts to the company's existing structures. As part of their oversight of enterprise risk and strategy, boards of directors should view AI systems as a key part of their work. This means that reports should clearly explain the purpose of a model, its limitations, the effects on stakeholders, and any results from ongoing monitoring (OECD, 2015; COSO, 2017). Management is also responsible for ensuring that different roles have clear responsibilities and that there are reliable processes for handling appeals, responding to incidents, and improving. The basic moral argument here is straightforward. If there is no sense of institutional ownership and no commitment to continuous monitoring, any ethical promises will remain dreams rather than become reality.

When we look at how businesses really work, the most important thing to remember is that ethical AI governance can't be left to just technical teams or departments focused on following the rules. The company must see AI as a main part of governance instead. This means that AI is a system that actively affects stakeholders' results and requires a structured way to hold people accountable. Board members don't have to become data scientists, but they do need to understand governance well. This means being able to ask the right questions, having the power to demand reports that are genuinely useful, and having the will to ensure oversight is real and not just a show.

There are trade-offs that can't be avoided in this field. In some cases, pushing for more explainability may hurt predictive performance, just like strict auditing processes can slow down a launch. However, it would be erroneous to perceive these tensions as mere external obstacles impeding progress. Instead, they are an important part of the moral and strategic thinking that goes into responsible innovation. A business that only cares about speed and optimization runs the risk of creating hidden liabilities. These often manifest later as stakeholder trust loss, sudden regulatory actions, or long-term damage to the company's reputation.

Ethical implications for society

The way companies run AI has big effects on fairness, personal freedom, and even the trust we have in our institutions. Increasingly, AI-powered systems are acting as gatekeepers for jobs, credit, and basic services. When these big decisions are hard to understand or cannot be challenged, the person starts to feel like they are under the weight of arbitrary authority. For this reason, ethical governance is not only a business issue but also a democratic one. It protects the principles of due process, allows people to challenge decisions, and prevents structural forms of discrimination (Barocas & Selbst, 2016; Pasquale, 2015, pp. 3–18).

Moral minimalism is a big problem in this area. This happens when companies think that following the law is enough to be ethical. But being truly responsible often means dealing with harm that may be perfectly legal but are still unfair. We need to look at practices that are good for the company but clearly unfair to the stakeholders. This article's governance model is meant to keep this ethical weight visible and active so that it doesn't just turn into a pile of bureaucratic paperwork.

This work is mostly a conceptual analysis,[2] so it doesn't give an empirical evaluation of the model itself. So, there is a clear need for more research into how different organizations actually use explainability, auditability, oversight, and participation across different areas. We also need to know how these habits affect measurable goals, such as reducing the risk of discrimination and increasing the organization's legitimacy. Also, looking at how legal minimums interact with

[2] This article is a conceptual and theoretical analysis. It synthesizes interdisciplinary literature from business ethics, corporate governance, and AI ethics, and interprets regulatory and policy documents as normative reference points. The method is appropriate to the research problem because AI governance challenges are partly conceptual: they concern how responsibility should be attributed and institutionalized under conditions of automation, opacity, and scale. The goal is not empirical generalization but conceptual clarification and model development to guide future normative and empirical research.

more ambitious ethical choices across different regulatory environments would help clarify things.

This study shows that our usual ways of thinking about responsibility, which usually focus on the individual agent, need to be expanded to include institution-centered approaches. We can't just look at what one person does or says to figure out who is responsible anymore, because AI now mediates corporate actions. We need to look at the organizational structures that guide decision-making, allocate power, and set the rules for recognizing and addressing harms as the cause of this.

This change in perspective doesn't mean people are no longer responsible. Instead, it means that individual action is now part of a bigger system of governance. So, any ethical evaluation must examine how companies plan and manage their decision-making power throughout an AI system's life. This last point strengthens the main point of this work. The ethical soundness of AI does not depend solely on a model's features. It depends on whether the governance structures around it can maintain accountability, allow for contestability, and provide a rationale that really respects the stakeholder.

Conclusion

Making decisions with the help of AI does much more than just introduce new technical risks. It fundamentally questions how corporate responsibility is set up and explained today. The rules we have for business ethics and corporate governance are still very useful, but they are clearly under increasing pressure. This is mostly because systems that make decisions are now both opaque and adaptable, operating on a huge scale. This article has shown that problems such as gaps in responsibility, power redistribution, and concerns about fairness among stakeholders are best understood as governance challenges. Because of this, they need more than just technical fixes; they need institutional responses. This work changes the way we think about ethical AI by giving us a conceptual model that connects explainability, auditability, human oversight, and stakeholder participation. Compliance is no longer just a side job; it's now a key part of corporate responsibility and long-term institutional legitimacy. It does this by giving organizations and researchers a keyway to assess how to use AI in ways that are fair, accountable, and create long-term value.

This article also adds to the discussion by bringing together ideas that connect business ethics and corporate governance with AI ethics in the specific context of AI use in business. It makes clear how using AI to help make decisions can create gaps in responsibility and shift who is in charge, in ways that test traditional ways of holding people accountable. The integrative model proposed here combines explainability, auditability, human oversight, and stakeholder participation into a governance framework in which these elements support one another. This model moves the conversation forward by framing ethical AI as an issue of organizational responsibility and legitimacy instead of just a technical problem or a box to check for compliance.

The model connects technical practices directly to governance duties and stresses that ethical commitments must be put into practice. This needs clear ownership, constant monitoring, and good ways to escalate and fix problems. From a bioethical perspective, these same elements of governance support several important principles. They support autonomy by enabling people to understand and fight back. They help non-maleficence by keeping an eye on things and stepping in when they see problems, which lowers the risk of harm. They promote beneficence by making it clear who will benefit and what trade-offs will be made. Lastly, they promote fairness by making any unfair effects clear and actionable for institutions. In businesses where AI is increasingly responsible for deciding who gets what resources and opportunities, responsibility must remain clear, debatable, and enforceable through the creation of real institutions.

References

BAROCAS, S. & SELBST, A. D. (2016): Big data's disparate impact. In: *California Law Review*, 104(3), pp. 671–732.

BEAUCHAMP, T. L. & CHILDRESS, J. F. (2019): *Principles of biomedical ethics* (8th ed.). New York: Oxford University Press.

BUOLAMWINI, J. & GEBRU, T. (2018): Gender shades: Intersectional accuracy disparities in commercial gender classification. In: *Proceedings of Machine Learning Research*, 81, pp. 1–15.

BRENNAN, D. (2024): Revisiting Czech philosophical critiques of science in the age of generative AI and big data. In: *Ethics & Bioethics* (*in Central Europe*), 14(3–4), pp. 235–238.

BURRELL, J. (2016): How the machine "thinks": Understanding opacity in machine learning algorithms. In: *Big Data & Society*, 3(1), pp. 1–12. [online] [Retrieved January 14, 2026] Available at: https://doi.org/10.1177/2053951715622512

CARROLL, A. B. (1991): The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. In: *Business Horizons*, 34(4), pp. 39–48. [online] [Retrieved December 22, 2025] Available at: https://doi.org/10.1016/0007-6813(91)90005-G

COMMITTEE OF SPONSORING ORGANIZATIONS OF THE TREADWAY COMMISSION (COSO) (2017): *Enterprise risk management: Integrating with strategy and performance.* New York: AICPA.

DIAKOPOULOS, N. (2016): Accountability in algorithmic decision making. In: *Communications of the ACM*, 59(2), pp. 56–62. [online] [Retrieved December 26, 2025] Available at: https://doi.org/10.1145/2844110

DONALDSON, T. & PRESTON, L. E. (1995): The stakeholder theory of the corporation: Concepts, evidence, and implications. In: *Academy of Management Review*, 20(1), pp. 65–91. [online] [Retrieved December 26, 2025] Available at: https://doi.org/10.5465/amr.1995.9503271992

DOSHI-VELEZ, F. & KIM, B. (2017): Toward a rigorous science of interpretable machine learning. In: arXiv. [online] [Retrieved January 15, 2026] Available at: https://arxiv.org/abs/1702.08608

EUROPEAN PARLIAMENT & COUNCIL OF THE EUROPEAN UNION (2016): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation). In: *Official Journal of the European Union*. [online] [Retrieved January 2, 2026] Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj

EUROPEAN PARLIAMENT & COUNCIL OF THE EUROPEAN UNION (2016): Regulation (EU) 2016/679 (GDPR), Article 22 (Automated individual decision-making, including profiling). [online] [Retrieved January 16, 2026] Available at: https://gdpr-info.eu/art-22-gdpr/

EUROPEAN PARLIAMENT & COUNCIL OF THE EUROPEAN UNION (2024): Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). In: *Official Journal of the European Union*. [online] [Retrieved December 27, 2025] Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

FAMA, E. F. & JENSEN, M. C. (1983): Separation of ownership and control. In: *Journal of Law and Economics*, 26(2), pp. 301–325. [online] [Retrieved December 30, 2025] Available at: https://doi.org/10.1086/467037

FREEMAN, R. E. (1984): *Strategic management: A stakeholder approach.* Boston: Pitman.

FREEMAN, R. E., HARRISON, J. S. et al. (2010): *Stakeholder theory: The state of the art.* Cambridge: Cambridge University Press.

FRENCH, P. A. (1979): The Corporation as a Moral Person. In: *American Philosophical Quarterly*, 16(3), pp. 207–215. [online] [Retrieved January 15, 2026] Available at: https://www.sci.brooklyn.cuny.edu/~schopra/Persons/French.pdf

GARRIGA, E. & MELÉ, D. (2004): Corporate social responsibility theories: Mapping the territory. In: *Journal of Business Ethics*, 53(1–2), pp. 51–71. [online] [Retrieved January 15, 2026] Available at: https://doi.org/10.1023/B:BUSI.0000039399.90587.34
HARDT, M., PRICE, E. & SREBRO, N. (2016): Equality of opportunity in supervised learning. In: D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon & R. Garnett (eds.): *Advances in Neural Information Processing Systems*, vol. 29. Red Hook, NY: Curran Associates, pp. 3315–3323.
JENSEN, M. C. & MECKLING, W. H. (1976): Theory of the firm: Managerial behavior, agency costs and ownership structure. In: *Journal of Financial Economics*, 3(4), pp. 305–360. [online] [Retrieved January 16, 2026] Available at: https://doi.org/10.1016/0304-405X(76)90026-X
JOBIN, A., IENCA, M. & VAYENA, E. (2019): The global landscape of AI ethics guidelines. In: *Nature Machine Intelligence*, 1(9), pp. 389–399. [online] [Retrieved January 14, 2026] Available at: https://doi.org/10.1038/s42256-019-0088-2
KELLOGG, K. C., VALENTINE, M. A. & CHRISTIN, A. (2020): Algorithms at work: The new contested terrain of control. In: *Academy of Management Annals*, 14(1), pp. 366–410. [online] [Retrieved December 27, 2025] Available at: https://doi.org/10.5465/annals.2018.0174
KROLL, J. A., HUEY, J. et al. (2017): Accountable algorithms. In: *University of Pennsylvania Law Review*, 165(3), pp. 633–705.
MATTHIAS, A. (2004): The responsibility gap: Ascribing responsibility for the actions of learning automata. In: *Ethics and Information Technology*, 6(3), pp. 175–183. [online] [Retrieved January 14, 2026] Available at: https://doi.org/10.1007/s10676-004-3422-1
MITCHELL, M., WU, S. et al. (2019): Model cards for model reporting. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*'19)*, pp. 220–229. New York: ACM. [online] [Retrieved January 14, 2026] Available at: https://doi.org/10.1145/3287560.3287596
MITTELSTADT, B. D., ALLEN, L. et al. (2016): The Ethics of Algorithms: Mapping the Debate. In: *Big Data & Society*, 3(2), pp. 1–21. [online] [Retrieved January 15, 2026] Available at: https://journals.sagepub.com/doi/10.1177/2053951716679679
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST) (2023): *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* NIST AI 100-1. Gaithersburg, MD: U.S. Department of Commerce. [online] [Retrieved January 2, 2026] Available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
OECD (2015): *G20/OECD Principles of Corporate Governance.* Paris: OECD Publishing.
OECD (2019): Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). OECD Legal Instruments. [online] [Retrieved January 30, 2026] Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
O'NEIL, C. (2016): *Weapons of math destruction: How big data increases inequality and threatens democracy.* New York: Crown.
PASQUALE, F. (2015): *The black box society: The secret algorithms that control money and information.* Cambridge, MA: Harvard University Press.
RAJI, I. D., SMART, A. et al. (2020): Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*'20)*, pp. 33–44. New York: ACM. [online] [Retrieved January 16, 2026] Available at: https://doi.org/10.1145/3351095.3372873
RAWLS, J. (1971): *A theory of justice.* Cambridge, MA: Harvard University Press.
RUDIN, C. (2019): Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. In: *Nature Machine Intelligence*, 1, pp. 206–215. [online] [Retrieved January 15, 2026] Available at: https://doi.org/10.1038/s42256-019-0048-x
SEN, A. (2009): *The idea of justice.* Cambridge, MA: Harvard University Press.

SELBST, A. D., BOYD, D., FRIEDLER, S. A., VENKATASUBRAMANIAN, S. & VERTESI, J. (2019): Fairness and Abstraction in Sociotechnical Systems. In: *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (FAT*)*, pp. 59–68.
SUCHMAN, M. C. (1995): Managing legitimacy: Strategic and institutional approaches. In: *Academy of Management Review*, 20(3), pp. 571–610. [online] [Retrieved January 20, 2026] Available at: https://doi.org/10.5465/amr.1995.9508080331
SUNSTEIN, C. R. (2016): *The ethics of influence: Government in the age of behavioral science*. Cambridge: Cambridge University Press.
SUSSER, D., ROESSLER, B. & NISSENBAUM, H. (2019): Online manipulation: Hidden influences in a digital world. In: *Georgetown Law Technology Review*, 4(1), pp. 1–45.
ŠPIRKOVÁ, S. (2023): Free will, moral responsibility and automatisms. In: *Ethics & Bioethics (in Central Europe)*, 13(1–2), pp. 55–71.
ŠVAŇA, L. (2023): War and peace as consequences of human nature? In: *Ethics & Bioethics (in Central Europe)*, 13(1–2), pp. 72–82.
THALER, R. H. & SUNSTEIN, C. R. (2008): *Nudge: Improving decisions about health, wealth, and happiness.* New Haven: Yale University Press.
TRICKER, R. I. (2015): *Corporate governance: Principles, policies, and practices* (3rd ed.). Oxford: Oxford University Press.
UNESCO (2005): *Universal Declaration on Bioethics and Human Rights.* Paris: United Nations Educational, Scientific and Cultural Organization. [online] [Retrieved January 20, 2026] Available at: https://unesdoc.unesco.org/ark:/48223/pf0000146180
VOIGT, P. & VON DEM BUSSCHE, A. (2017): *The EU general data protection regulation (GDPR): A practical guide.* Cham: Springer. [online] [Retrieved January 20, 2026] Available at: https://doi.org/10.1007/978-3-319-57959-7
WACHTER, S., MITTELSTADT, B. & FLORIDI, L. (2017): Why a right to explanation of automated decision-making does not exist in the GDPR. In: *International Data Privacy Law*, 7(2), pp. 76–99. [online] [Retrieved January 16, 2026] Available at: https://doi.org/10.1093/idpl/ipx005
ZUBOFF, S. (2019): *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* New York: PublicAffairs.