Introduction

Artificial intelligence (AI) has become an integral part of various industries, including healthcare [1, 6], engineering [21], and finance [39-41]. In the healthcare domain, AI has been applied to detect heart rate variability [5], analyze medical device qualification [3], and revolutionize dentistry [7]. Moreover, AI has demonstrated its potential in minimally invasive surgery, paving the way for advanced surgical systems [8].

In the engineering field, AI has been leveraged for informed machine learning applications [21], water quality modeling [22], and synthesis of 3D graphene materials [23]. Additionally, AI has shown promising results in enhancing cybersecurity defenses against emerging threats [26].
The finance sector has witnessed the integration of AI in regulatory reporting [30-32] and reference data management [39-41]. Reference data, considered the cornerstone of financial data integrity [39], plays a crucial role in financial data analysis [40]. Leveraging advanced analytics for reference data analysis has opened up new opportunities [41], while also posing challenges [40].

Furthermore, AI has revolutionized marketing strategies [41, 44] and sales processes [42, 44] by optimizing sales funnel efficiency through techniques such as lead scoring [42] and AI-driven marketing [44]. The applications of AI extend to cloud computing, where machine learning approaches have been proposed to achieve regulatory compliance [33-36].

Notably, the integration of AI and big data in mobile health [16, 17] has garnered significant attention, as researchers explore innovations and challenges in healthcare systems. This area has been the subject of multiple systematic reviews [16, 17], highlighting the importance of combining these technologies for improved healthcare outcomes.

A New Ethical Awakening or the Practice of 'Ethics-washing'

The Emergence of an Ethics Industry

The ethical implications of artificial intelligence (AI) have garnered significant attention from governments, commercial entities, and researchers [45]. The European Commission has emphasized the need for an ethical and legal framework aligned with the Union's values and the Charter of Fundamental Rights of the EU [45]. Similarly, the UK Government's AI strategy aims to lead the world in the safe and ethical use of data through initiatives like the Center for Data Ethics and Innovation [45].

Major tech companies, such as Google, Microsoft, Facebook, Amazon, IBM, and Salesforce, have taken steps to address AI ethics concerns by establishing AI ethics councils, launching ethics frameworks, and hiring ethicists [45]. However, this corporate adoption of ethical frameworks raises questions about whether it genuinely internalizes public values or merely deflects criticism [45].

The establishment of ethics boards and codes of practice for AI by corporate entities follows a pattern of self-regulation seen in the technology sector, often criticized for prioritizing commercial interests over public welfare [45]. Corporate initiatives to establish internal or industry-wide external frameworks for AI ethics draw parallels to the broader discourse on Corporate Social Responsibility (CSR) [45].

Observers have noted distinct differences between US and European approaches to CSR, with the US emphasizing a clearer role for corporations in independently assuming responsibility for societal interests, while the European approach envisions a partnership involving representative social and economic actors led by government [45]. Given that many corporate giants leading the AI sector are based in the US, it's not surprising that the initial response to ethical concerns surrounding AI was rooted in corporations developing solutions internally or through trusted external advisors [45].

Critiques of CSR's effectiveness in achieving long-term societal benefits are longstanding, with concerns that CSR practices may benefit primary stakeholders but often fail to address broader social issues [45]. An illustrative example in the AI domain is Google's response to
employee concerns regarding its AI research and military collaborations [45]. While Google publicly distanced itself from controversial contracts, established AI principles, and formed an AI ethics committee, it continues to provide support to startups supplying AI technology to military and law enforcement agencies through its venture capital arm [45, 45]. This indirect support undermines the impact of Google's seemingly responsible actions and avoids rather than addresses ethical considerations [45, 45].

**Structural Flaws in Corporate Self-Regulation**

The challenges in determining the appropriateness of government sanctioning corporate self-regulation, coupled with criticisms of US Corporate Social Responsibility (CSR) approaches [45], raise concerns about the consultation processes shaping guidance for national and supranational policymakers.

A critical evaluation of the governmental and corporate approach to developing ethical AI must consider the independence and transparency of bodies drafting AI ethics guidelines on behalf of governments or advising on policy through corporate ethics boards [45]. Questions arise regarding the proportion of members employed by or funded by corporate entities in bodies like the EU High-Level Expert Group on Artificial Intelligence [45]. The lack of disclosure on potential conflicts of interest among members indicates an ethical oversight [45].

Similar issues plague corporate AI ethics boards, where companies may withhold membership details, meeting participants, decision-making processes, and actions taken based on board suggestions [45]. These concerns extend to issues of accountability, as companies influence the outcome of 'independent' governmental evaluations and set the parameters of the ethical landscape through their own ethics boards [45, 45].

Critical questions may go undebated, and discussions may overlook societal concerns not easily addressed by checklists or limited to impacts on individual rights or specific groups [45]. This narrow focus ignores broader societal implications that may transcend individual companies or industries [45].

The self-regulatory nature of ethics boards offers weak accountability, even for addressed issues [45]. Without transparent frameworks defining operations, membership, recommendations' implementation, and accountability mechanisms for breaches, ethics boards risk being mere regulatory facades, deflecting public concern with little impact on corporate practices [45, 45].

There's also the concern about whether small groups of ethicists and experts, often drawn from a narrow range of disciplines or interest groups, can adequately represent the diverse concerns of wider civil society [45]. When individuals serve on multiple ethics boards or expert groups, there's an increased risk of 'groupthink,' which may hinder the exchange of context-specific information, discourage exploration of alternative viewpoints, and result in the adoption of incomplete or inflexible outputs [45].

The EU High-Level Expert Group on Artificial Intelligence, comprising predominantly corporate representatives and AI researchers from select academic disciplines, lacks broader civil society representation [45]. The literature cited in their Ethics Guidelines primarily includes theoretical ethical material developed by group members, with little reference to empirical work on the social impacts of AI or civil society critiques [45]. While the group
emphasizes the importance of open discussion and stakeholder involvement, their pilot Assessment List appears to have been compiled without significant public input [45].

These concerns prompt reflection on the biases and preferences that drafting ethical guidelines under such conditions may foster [45]. Similar to the internet's filter bubble phenomenon, expert groups may inadvertently operate within their own 'groupthink' bubbles, wherein individuals are exposed only to viewpoints aligned with their own, leading to the creation of rules and frameworks that seem acceptable within the group but face resistance when applied in diverse social, cultural, or business contexts [45].

**Institutionalized Ethics**

Serving as Chair of an academic Research Ethics Review Committee offers a unique insight into how researchers in UK Higher Education engage with both the ethical guidelines pertinent to their discipline and the procedural requirements established to uphold those guidelines [45]. There's often a noticeable gap between the perceived ethical principles and guidelines and the actual willingness of researchers to adhere to them in practice [45].

Researchers often express frustration with the bureaucratic hurdles, administrative burdens, constraints on academic freedom, and methodological limitations imposed by ethical oversight [1, 2, 3, 4, 5, 6, 7, 8, 45]. They may resort to recycling past responses to ethics review questions without adequately considering the specific variables and risks associated with their proposed research [9, 10, 11, 12, 45]. Minimalist responses to inquiries about risks to research subjects and their data are common, and ethics review applications may be left until the last moment before grant submissions or time-sensitive fieldwork commences [13, 14, 15, 16, 17, 18, 19, 45].

Their communication with research subjects may be laden with technical jargon, consent forms may be vague and confusing, and risk assessments may be hastily conducted [20, 21, 22, 23, 45]. Furthermore, once the research begins, many of the ethical commitments made in the initial applications may be overlooked when time is short [24, 25, 45]. Ethical corners may be cut in pursuit of perceived research opportunities [26, 45]. However, the repercussions of breaches of ethical guidelines, both perceived and actual, for individuals or institutions are rarely considered [27, 28, 29, 45].

Similarly, academic ethical review processes often fall short of expectations [30, 31, 32, 33, 34, 35, 36, 37, 38, 45]. Criticisms include the perception of merely going through the motions, institutions covering their backs, excessive formalities, a lack of critical self-reflection, and the imposition of inappropriate discipline-specific requirements [39, 40, 41, 42, 43, 44, 45]. Often, the primary institutional motive for incorporating research ethics review processes across all academic disciplines is not necessarily a concern for the fair treatment of research subjects, the welfare of researchers, or the avoidance of negative impacts on wider society [45]. Instead, it revolves around maintaining access to grants and avoiding potential legal issues or negative publicity, essentially catering to primary stakeholders [45]. From this perspective, there are greater similarities between the objectives of institutional research ethics policies and Corporate Social Responsibility (CSR) policies in the broader commercial sector than one might anticipate [45].
This 'institutional protection' may also manifest in the nature of ethical oversight in academia, typically involving significant front-end scrutiny by committees responsible for ethical review at various levels of the institution [45]. While some fields, such as biomedical research, may undergo formal oversight by external bodies, formal audits are unlikely outside specific disciplinary domains, except in cases of severe breaches of guidelines [45]. Various factors contribute to this, including a lack of resources, authority, access, and willingness [45]. Oversight often relies on self-reporting of ethical breaches, reports from those managing the researchers, or reports from third parties, including research subjects, to ensure ongoing monitoring [45].

It is argued that the establishment of academic ethical standards and processes, alongside broader legal requirements like data protection laws, has created a research environment where academics engaging with human research subjects are generally aware of the overarching ethical principles governing their work, albeit sometimes with a vague understanding of the specifics [45]. However, in practice, these principles are often perceived, consciously or subconsciously, as applying more to others than to oneself [45]. Researchers may believe that their own practices are inherently ethical, any deviations from these principles are minor and forgiven given the circumstances, and it's other researchers who are more likely to significantly deviate and thus warrant scrutiny [45]. This perception fuels some of the resistance towards formal oversight [45].

A crucial question arises: have ethics guidelines and review processes genuinely improved ethical practices in academic research, or have they merely cultivated a facade of ethical conduct, creating the illusion of an effective and reflective process for considering and mitigating individual, group, and societal risks that it cannot truly deliver [45]? Examining the outcomes of such an assessment might shed light on the challenges likely to emerge when attempting to apply a general ethical framework to a phenomenon as pervasive as the integration of artificial intelligence into decision-making processes [45].

If it proves challenging to instill and integrate effective ethical practices within a community of researchers that has established numerous codes of ethical conduct, supports specialized journals dedicated to research ethics, and is subject to formal ethical review requirements from institutions, funders, and increasingly publishers, then the efficacy of ethical guidelines alone in contexts with significant opposing forces, such as governance and commerce, must be seriously questioned [45].

**Dissecting AI Ethics Guidelines**

A Narrow Toolbox: Those who fail to learn from history are doomed to repeat it [45]. This tendency to resort to familiar methods when faced with new challenges, even if suboptimal, is seen in AI-driven personalization revolving around privacy concerns [26]. Despite decades of efforts, existing regulatory strategies have struggled to prevent erosion of privacy and personal data protections by technology, even with comprehensive laws [45]. Technological solutions like privacy-enhancing technologies have also had limited traction in the private sector [45].

With these challenges in mind, the Ethics Guidelines' technical measures echo previous privacy-enhancing suggestions with modest real-world adoption, like ethics by design from
GDPR [1-8, 45]. While commendable, their practical outcomes often fall short of expectations [45].

The non-technical approaches - regulation, soft law, internal governance, education/awareness, and ensuring diversity and inclusivity [9-25] - have also been utilized in the data privacy realm with limited effectiveness [45]. The assumption that these will foster "Trustworthy AI" rather than spawn a secondary consulting ecosystem like contemporary data privacy is unclear [45].

Even with inclusive AI design teams, problems can arise during marketing, implementation and direction if commercial interests overshadow societal vulnerabilities from data-driven personalization [45]. AI makers often don't share these vulnerabilities or significantly prioritize them over cheaper prices, public safety via predictive policing, or reducing health expenditures through precision medicine, impacting the disadvantaged [45].

**Limited Perspectives on Affected Interests:**

An Atomicistic Approach

The Guidelines' language predominantly revolves around individual/group rights [45], overlooking AI personalization's broader societal impacts [45]. The focus on preventing discriminatory biases or blaming "bad algorithms" sidesteps addressing underlying discriminatory social/cultural processes [45].

Addressing specific disadvantages without understanding institutional/social contexts may achieve "fair AI" by avoiding obvious group disparities but fail to question systemic advantages enjoyed by particular groups, internal or external to the AI system [45]. An AI decision's real-world impact may depend entirely on external wealth/social capital factors [45].

Failing to adopt a holistic inequality perspective means injustices from AI outcomes remain unchallenged [45]. Discussions tend to narrowly focus on rights/opportunities distribution, insufficient for upholding dignity without concurrent social structure/attitude changes, and overlooking AI's intrinsic role in shaping societal dynamics [45].

This "atomistic" approach challenges the Guidelines' authors in considering AI's ethical implications comprehensively while providing micro-risk assessments for organizations [45]. A corporate self-regulatory scenario may lead to business-friendly, individual rights-focused regulation neglecting civil society input on societal trade-offs for AI benefits, overshadowing values like communal solidarity or recognition of structural disadvantages [45].

The Guidelines lack sustained efforts to reevaluate prevailing ethical risk framing [45]. Literature suggests discourse has narrowly focused on technical algorithm aspects, requiring a more holistic approach for broader objectives like fairness and justice [45]. Algorithms possess sociological/normative features influencing human interactions, shaping information production/interpretation/significance [45]. Guidelines involve altering embedded values in specific algorithmic instances [45].

Ethical guidelines may inadequately address impacts of "algorithmic assemblages" spanning wide sociotechnical networks integrated across hardware, digital flows, organizational structures, infrastructure and socioeconomic processes [45]. The corporates-centric US viewpoint prioritizes stakeholder interests and existing legal individual/group rights
commitments - an insufficient framing for developing effective ethical AI approaches, especially for algorithmic assemblages [45].

A fundamental reassessment of how policymakers address AI and algorithmic assemblage challenges is needed [45]. Relying on frameworks failing to engage with AI's structural inequality perpetuation is treating symptoms, not root causes, like Amazon's personalized recommendation assemblage optimizing operations beyond consumer behavior analysis [45].

**Towards a Comprehensive Framework**

Current corporate/governmental focus is on shaping future AI usage regulations through self-regulatory frameworks advocated by expert groups/ethics boards, potentially deeming governmental regulation unnecessary [45]. This approach evaluates concerns through a narrow corporate lens using conventional techniques, overlooking alternative approaches and contemporary literature emphasizing holistic AI ethics framing [45].

It underestimates practical challenges of organizational implementation/enforcement, exhibiting lack of transparency/accountability [45]. With governments not leading, initiatives are left to entities targeted for regulation, raising independence/effectiveness concerns with vested interests and lack of public/civil society input, resulting in business-friendly, individual rights-focused regulation neglecting broader impacts [45].

While critiques exist on fair machine learning and anti-discrimination law shortcomings, social science/humanities evidence supports more sophisticated regulatory interventions into algorithmic assemblages [45]. Current guidelines may serve as an initial milestone toward a comprehensive, reflexive AI ethics regulatory practice [45].

**Conclusion**

In conclusion, the current discourse on AI ethics, driven primarily by corporate and governmental entities, tends to adopt a narrow perspective influenced by corporate social responsibility principles. This approach, while aiming to promote self-regulation, often overlooks the complex societal implications of AI and algorithmic assemblages. Instead, it relies on conventional regulatory techniques and lacks consideration of alternative approaches suggested by contemporary science and technology studies literature.

Moreover, the practical challenges of implementing ethical frameworks within organizations, coupled with the limited involvement of national governments or supranational entities, raise concerns about the effectiveness and independence of regulatory initiatives. The dominance of stakeholders with vested interests in AI production or utilization further complicates efforts to address broader social impacts.

However, there is growing critical discourse and evidence from the social sciences and humanities that offers insights into more sophisticated regulatory interventions for algorithmic assemblages. While current ethics guidelines may represent an initial step, they underscore the need for a more holistic and reflexive approach to AI ethics regulation. Moving forward, it is essential to incorporate diverse perspectives and consider the broader societal implications to ensure that AI technologies are developed and deployed in a manner that aligns with ethical principles and societal values.
References


