RESEARCH ARTICLE OPEN ACCESS

Ethical and Governance Challenges of AI in Sustainable Finance

Dr. Rachana Saxena*, Dr. Mohsina Hayat**, Hashim Khan***
*(Professor, School of Commerce, Jain (Deemed-to-be University), Bangalore
Email: dr.rachna.saxena@gmail.com orcid.org/0000-0003-3514-2757)
** (Assistant Professor, School of Commerce, Jain (Deemed-to-be University), Bangalore
Email: aissi.17@gmail.com orcid.org/0000-0003-1274-8713)
*** (Adjunct Professor, School of Commerce, Jain (Deemed-to-be University), Bangalore
Email: aissi.17@gmail.com, orcid.org/0009-0001-4015-7462)

--************************----------------------------------

Abstract:

Artificial intelligence (AI) is quickly reshaping the financial services industry, such as the up-and-coming and rapidly expanding field of sustainable finance. AI will help to multiply the positive impacts on the environment and the social sector by automating the assessment of the risks, improving ESG data analysis, and allowing a dynamic and sustainability-linked decision-making process. Nonetheless, applying AI to sustainable finance presents urgent ethical and governance risks: algorithmic discrimination and biases, obscurity and lack of explainability, quality and provenance of data, undermined accountability in the decision-making process, and increased greenwashing and perverse incentive risks. In this paper, the critical assessment of these issues is made, and technical, organizational, and policy-level solutions are suggested. Based on the recent regulatory initiatives and scholarly sources, we suggest that to close the disconnect in between AI capability and ethical governance, (a) strong data governance and model audits, (b) explicit legal and board-level accountability, (c) explainability and contestability strategies, specific to financial and ESG circumstances, and (d) policy coordination between financial regulation and the principles of AI governance are needed. The outcome is a list of recommendations in practice that institutions, regulators, and researchers wishing to implement AI should follow to implement it in a manner that actually contributes to sustainable finance and does not cause social harm disproportionately.

Keywords — **AI ethics, governance, responsible AI, sustainable finance policy.**

--************************----------------------------------

Introduction

The financial activity that involves the consideration of the environmental, social, and governance (ESG) is an area that has experienced a massive growth in the past decade. To process ESG disclosures, rate corporate sustainability, underwrite sustainability-linked loans, and build ESG-conscious portfolios, asset managers, banks, insurers, and corporate treasuries are growingly looking to automated instruments to do it. In line with this, new technologies in AI (machine learning, natural language processing, large language models) can be used to gain insight into noisy ESG data as well as predict climate risks and optimize capital allocation to green technologies.

Nevertheless, AI does not simply add to sustainable finance. The aspects that render AI appealing such as its capability to identify patterns in large, heterogenous data contribute to increased ethical and governance risks when used in sustainability decisions with impacts on markets and livelihoods. Bias within the training data can lead to discriminatory credit on lending or insuring; the unobservable models may hide the decision rationale to the investors, regulators and the beneficiaries; and misaligned incentives may make firms exaggerate the sustainability effects or suggest products that seem to be green but in practice will not have a significant impact on reducing environmental pollution. These issues can only be solved by having more powerful algorithms, but also by having well-formed governance institutions that bring accountability, transparency, and incentive alignment.

This paper critically discusses ethical issues, transparency issues, and bias-related issues that come to play in the implementation of AI on sustainability-related financial decisions. We integrate academic and policy literature, emphasize the examples, and provide specific recommendations comprehending technical controls, corporate governance, and the policies that people can do.

History and Recent Policy-Making

An increasing number of global bodies and controllers have harnessed the two-fold potential and the danger of AI in finance. Multilateral frameworks, including the OECD AI Principles have recently been revised to capture new technological and policy realities and focus on trustworthy AI that upholds human rights and democratic values. These values promote openness, responsibility and humanity in the application of AI.

Central banks and supervisory regulators have also started to put out guidance which explicitly focuses on the adoption of AI in the financial sector. Indicatively, the Bank for International Settlement released a report that provided a guiding principle on how central banks can detect and manage risks that are associated with AI, and it suggests auditability and risk-management frameworks that are unique to AI.

Both in their home jurisdictions and at international events, regulators have been categorical on whether AI-driven decisions are the responsibility of the corporations or not. The European Securities and Markets Authority (ESMA) made it clear that banks and investment companies are required to become responsible at board level in use of AI and cannot delegate legal responsibilities to technology providers-highlighting that management bodies have to learn and control the use of AI decisions.

Another emerging trend of AI at a country level is sector-specific AI frameworks. A Reserve Bank of India committee suggested a thorough AI framework of the financial sector focusing on infrastructure, governance, and auditability, meaning that developing economies are also focusing on responsible AI in finance.

There are many different types of algorithmic discrimination and prejudice in the financial use, which have been reported in academic literature. The research classification of bias types and the steps of finding how the feature selection, historical discrimination in data and proxy variables generate unfair results. These theoretical and empirical roots are necessary in diagnosing the special ethical issues of sustainable finance.

The differences between sustainable finance and conventional finance are multiple and enhance the ethical issues surrounding AI:

Data Heterogeneity and Quality: ESG data are available in mixed formats: corporate disclosures, NGOs, satellite imagery, news, and alternative data. This heterogeneity increases the chances of low quality inputs, selective reporting and proxies which fail to reflect sustainability imperfectly. These errors can be ingrained and increased by the AI models trained on such data.

Measurement uncertainty: Sustainability is a multi-dimensional concept (carbon, biodiversity, labor practices). Results are usually late and difficult to quantify. Any AI that tries to use short-term indicators to predict long-term sustainability is subject to the chance of overfitting to non-significant changes.

Great Stakes and Externalities: The social and environmental externalities of sustainability-related choices (e.g. issuing of green bonds, lending contingent on reduction in emission) are extensive. The mistakes may result in capital misallocation, reputational damage, and systemic risks in case market participants will react in scale to erroneous AI-based cues.

Third-Party Data and Model Ecosystem: As often as not, financial firms are basing their models on third-party ESG ratings, NPL (non-performingloan) models, or trained language models. Reliance on non-transparent external models makes it less visible into biases, as well as limits the ability of firms to audit systemic risks.

Regulatory Complexity: Sustainable finance overlaps with securities, banking, corporate disclosure, and environmental law--that is, establishing legal multilayers. At the point of intersection between AI tools and these areas, legal responsibility is complicated to assign.

Collectively, these attributes render explainability, data provenance, model auditability, and governance mechanisms not a luxury addition but a core need of responsible deployment.

Core Ethical Challenges

There is a connection between algorithms and bias and discrimination, as demonstrated by Algorithmic Bias and Discrimination.

4.1 Problem.

Bias occurs when AI systems create systematically unfair results to specific groups. Finance In finance, historical inequalities may be reinforced by such biases (e.g. gender or race-based differences in access to credit). Sustainable finance is no exception, as bias in this field may also affect which companies or areas get access to green capital, which may disadvantage small companies or firms in the emerging markets whose ESG disclosures are less thick.

Mechanisms. Such standard methods of biasing are:

Historical bias: Training data are based on past discrimination (e.g., a redlining data set will reinforce those trends).

Proxy bias: This occurs in models that use variables that are related to the phenomena being measured (e.g., zip code is a proxy variable used in place of race, though not explicitly).

Measurement bias: ESG measures can deliberately include less informal-sector effects or not be available when used in small firms, distorting AI judgments.

Consequences. Sustainable finance is susceptible to algorithmic bias which can diminish the inclusion of financial systems towards those who might rightly deserve funding, misallocate capital, and worsen social injustices under the guise of technical objectivity.

4.2 Opacity and Explainability

Problem. Numerous AI models (deep learning, ensemble models) are black boxes, the reasoning of which is difficult to understand by humans. Financial regulatory standards and fiduciary obligations require decisions (e.g. refusal to give loans, pricing, etc.) to make sense to clients and supervisors. Sustainability-linked financial stakeholders further demand reasons as to why they would declare instruments green or make financing based on an emission target.

Trade-offs. The performance of models can often be traded off with their explainability. Nevertheless, in more serious sustainability situations this trade-off leans toward increased transparency: the stakeholders should be informed of how the ESG conclusions were made, and affected parties must have appeals.

Data Governance and Provenance: In a cloud-based platform, there exists a risk of losing control over the data (Bothou et al. 2019)

4.3 Data Governance and Provenance:

Data in a cloud-based platform can be lost (Bothou et al. 2019).The quality and provenance of the input data determine the integrity of the AI outputs. ESG data can be incomplete, inconsistent and even worst, manipulated. Certain companies can greenwash by selective disclosure, which AI systems can be trained on to heighten false positives resulting to mispriced sustainability products.

Specific risks include:

Greenwashing: Models based on veneered signs (e.g., press releases, self-reported targets) may overestimate sustainability performance.

Feedback loops: When AI models channel excessively capital to firms with particular digital footprints, they will become more likely to look sustainable due to their ability to invest in more disclosure or data practices, thus forming a reinforcing loop.

4.4 Accountability and Liability

Delegating decisions to AI makes it more complex who is responsible to the damages: model developers, data vendors, financial institutions or board members. Regulatory statements (e.g. ESMA on the focus on board responsibility) indicate that the avoidance of legal liabilities is impossible because of outsourcing algorithms.

Organizational implications. The firms require governance models that draw distinct responsibilities throughout procurement, model validation, compliance and top management, and the monitoring and remediation processes.

4.5 Alternative Data Privacy and Use.

In order to resolve data gaps, AI models tend to use alternative data (mobile phone records, satellite imagery, web scraping). Although such data can be very potent, their utilization poses privacy, consent, and cross-border data transfer dilemmas- especially when they are applied to draw sensitive attributes.

Regulatory cross-overs. The legislation on data protection (GDPR-type regimes) has an interface with AI governance: companies should use data in accordance with law, proportionately, and transparently and also grant the right to an explanation and deletion where reasonable.

The transparency and Explainability: Why They are important in Sustainable Finance.

Explainability and transparency are useful in a number of ways:

Investor protection and fiduciary duty: The investors and clients are entitled to see how sustainability claims are made and how AI is used to make investment decisions.

Market integrity: Transparent models decrease the possibility of systemic shocks that may occur due to the collective dependence on opaque signals.

Public trust and legitimacy: Since sustainability aims are typically of a political interest, an open AI promotes the social license of financial institutions to take on stewardship actors.

Explainability Methods Technical Methods Technical methods of explainability such as interpretable model classes (e.g., generalized additive models), post-hoc explainers (SHAP, LIME), and local explanations that defend single decisions. Explainability in sustainability is not entirely technical, however, the narratives should reflect the boundaries of model inferences, the lack of certainty in long-term environmental forecasts, and the assumptions made behind the proxy indicators.

Prejudice in Action: Cases and Proofs.

There are empirical studies and case reports of bias and algorithmic discrimination in finance. Researchers recognize various forms of bias and show their manifestation in the process of credit decision-making and pricing. To provide an example, peer-reviewed research on algorithmic discrimination in the credit space demonstrates that models that learn by observing past loan performance have the potential to reproduce discriminatory lending behaviour without being properly de-biased.

The history of the application of regulations also offers instructive lessons of situations where financial institutions have been fined due to discriminatory behaviour (though it might have existed before the advent of modern AI) to remind that algorithm systems based on contaminated data may also risk reviving past injustices in case the institutional checks are not in place.

The risk in sustainable finance particularly is more subtle: bias may lead to the allocation of disproportionate capital to firms and geographies with more positively digitized ESG disclosures (which are often large firms in advanced markets), discriminating against small businesses and emerging-market firms which, in actual fact, may be doing better sustainability wise but lack the ability to report.

Governance Problems and Regulatory Reactions.

7.1 Fragmentation and Overlapping of Regulations.

Sustainable finance cuts across a wide variety of regulation (securities law, banking prudential regulation, environmental regulation, consumer protection, data protection). The AI governance models are developing at a very fast rate, although they tend to remain isolated. Such fragmentation contributes to confusion regarding norms and adherence requirements in cases where AI systems are subject to multiple regulatory requirements (e.g., an AI credit model, which uses ESG scoring to price loans).

The international bodies (OECD, BIS) and country regulators are trying to align strategies. OECD AI Principles offer values-based initial location with focus on transparency and human control. oecd.ai The BIS has offered central bank-specific suggestions in dealing with AI risk. Bank for International Settlement There is some assistance, and still variances in areas of enforceability, area, and jurisdiction priorities are present.

7.2 Senior Management Responsibility and Board-level.

It has been noted that regulators have been putting more emphasis that board and senior management are still liable to use AI in financial services. The aspect of the statement by ESMA, that firms may not outsource legal obligations when applying third-party AI points to a more general tendency: governance is going up the ladder. Now boards need to have an idea of what AI can imply to client interests and market integrity and enforce supervisory structures (model risk committees, independent validation).

7.3 Model Risk Management, Auditability and Validation.

Technical procedures necessary in AI governance are:

Pre-deployment validation: Fairness, robustness, and alignment of fiduciary duty model testing.

Continuous monitoring: Model drift monitoring, data modification monitoring and performance monitoring on groups being protected.

Independent audit trails: Recording of inputs, intermediate outputs and decision rationales so that the same can be reviewed ex-post.

Regulators demand auditability; companies need to invest in equipment, records, and staff to please overseers.

7.4 Third-Party Risk and Supply Chain Governance.

Financial companies tend to turn to data providers and AI sellers. This leads to contractual and supervisory issues: how can the models of vendors be fair and transparent, and how can it be done to perform due diligence on third-party datasets and architectures.

7.5 Cross-Border Data and Model Governance.

The flows of data across the borders are commonly needed by sustainable finance (e.g. satellite data, international corporate reporting). Cross-border compliance and data protection regimes make it difficult to design models. Companies should work on architectures that address the data privacy and AI accountability requirements.

Artificial Intelligence: Particular Traps in Sustainable Finance.

8.1 Greenwashing and Perverse incentives.

The AI scoring companies or suggesting investments that are green can unintentionally suggest greenwashing in circumstances where models use weak proxies (self-reported targets, press statements) instead of third party-validated results. Since AI can quickly increase the product offering, it can amplify those instruments that seem sustainable yet have little real-world effect, which invalidates the potential of sustainable finance.

8.2 The Short-term and the Long-term Sustainability.

Numerous AI systems maximize immediate signals (return, short-term risk) in the short term. Sustainability, in particular climate impact, is a long-term operation. Short-term optimized model types can either underprice long-term transition risks, or they do not provide resilience and adaptation, which is what sustainability-minded investors worry about.

8.3 Distribution Effects and Financial Inclusion.

Sustainable finance with AI may also be focused on the well-endowed firms and investors which decreases access to small businesses, the frontier markets and underserved communities. As an illustration, lenders based on AI to score ESGs may provide more favourable conditions to corporates in high-disclosure jurisdictions, and small businesses with no ESG reporting are pushed out of the market, generating capital flows that are not equal.

Mitigation: Policy, Organization and Technological.

The multi-layered interventions are necessary to address ethical and governance issues. One of these frameworks is summarized below, and it incorporates technical controls, corporate governance and public policy.

9.1 Technical Measures

Data provenance and quality controls: Provide metadata standards, provenance tracking and data lineage of all ESG and alternative datasets. Catalogues of data sources and verifiable credentials should be used to document the reliability of sources.

Bias detection and reduction: Perform systematic tests of model disparate impact on demographic and geographic groups. Apply pre-processing (rebalancing), in-processing (fairness constraints) and post-processing (output corrections) techniques based on the situation.

Explainability and uncertainty quantification: As much as possible, use inherently interpretable models to make high-stakes decisions. Add local explanations, counterfactuals, and uncertainty bands to convey confidence via supplement complex models.

Robustness and stress testing: Scenario test tail risks (e.g. climate extremes) and adversarial test to find out whether models generate brittle outputs in response to changes in distributions.

Human-in-the-loop design: Make sure that final decisions on high impact sustainability situations receive human scrutiny, there should be clear protocols on how to escalate decisions and record the reasoning behind the decision.

9.2 Organizational Measures

Governance at the board level AI: Introduce AI oversight boards or add AI oversight to risk boards. Demands management reporting of AI performance, harms and mitigation measures.

Three-lines-of-defense model of AI: Enforce explicit roles: front-line model owners, independent model validation teams, internal audit- reflects conventional model risk management to the AI idiosyncrasies.

Vendor due diligence and contractual clauses: Incorporate audit rights, explainability requirements and performance/security SLAs in vendor contracts. Make sure vendors of data offer provenance and consent data.

Transparency and stakeholder participation: Publish model documentation (what the model does, key inputs, limitations) in user friendly formats. Establish feedback and contestation means of stakeholders.

Upskilling of the workforce: Build in-house skills in AI ethics, ESG measurement, and model validation in order to diminish over-dependence on third-party vendors, as well as enhance in-house governance.

9.3 Policy and Regulatory Measures.

Sectoral standards of ESG data: Advance standardized ESG reporting frameworks that would allow comparability and would not use noisy proxies.

Regulatory direction on AI in finance: Financial regulators ought to provide explicit foresees on AI auditability, elucidation, and responsibility- based on cross-jurisdictional guidelines such as OECD principles and central bank guidelines.

Disclosure of mandatory models on sustainability claims: To products that are marketed as being sustainability related, require disclosure of model logic and material assumptions employed to support sustainability claims (balancing IP issues).

Remedies and redress mechanisms: Provide remedies and contestation mechanisms to ensure that affected parties get access to redress mechanisms and remedies in instances where AI-driven sustainability decisions are harmful.

International cooperation: Favor international agreement on standards to avoid regulatory arbitrage and to deal with international data/model governance.

A Blueprint of Governance Practical Institutional.

A brief practical roadmap that can be implemented by financial institutions in order to regulate AI in sustainable finance is as follows:
AI & ESG Inventory: Have a current list of all AI models and ESG data sources applied in products with sustainability-linked products, including metadata around the vendor and training-data.
Risk Categorization: Division of AI applications according to the risk (low, medium, high) based on the potential harm and reversibility. Use supplemented controls on risky sustainability decisions (e.g. loan denials, green certification).
Model Cards and Datasheets: On each of the models, release internal model cards with purpose, data sources, performance measures, fairness tests, constraints and owners. To enhance external transparency, create shortened versions of the information that point to important information seen by the user.
Pre-Deployment Ethical Review: Form an ethics review board including risk, ESG, legal, technical and stakeholder representatives to assess the possible harms and the mitigation measures.
Continuous Monitoring and Drift Detection: Put in place automatic monitoring of model regression, the appearance of bias, and modification of data pipeline. Automatic audits on the crossing of key thresholds.
Human Oversight Protocols: Stipulate the limits of decision making at which machine intervention is necessary and provide the appropriate tools and information to the reviewers to counter-decide machine results.
Incident Response & Redress: Establish mechanisms to research, rectify, and report incidents (e.g. algorithmic harm, greenwashing claims), (including communication instructions to regulators and victims).
Vendor Governance: Have vendors provide documentation, allow auditing and meet the obligations of fairness and explainability in the contracts.
Stakeholder Engagement & Transparency: Involve civil society, investor groups, and impacted communities in the design and review, and report on the use of AI in sustainable finance products periodically through transparency.
Regulatory Liaison & Compliance Roadmap: Chart relevant AI, data protection, and sustainability policies and a sustainability plan. Leading: Be proactive to experiment with new forms of governance.

Scope of Future Research

Although this has been achieved, there are still a number of gaps in research and practice like lack of sustainability outcomes as a causal inference. Numerous models are based on correlation. Studies should focus on causal frameworks to help to assess the impact of financing on sustainability measures in the long run.
Measures of fairness in ESG settings (equalized odds, demographic parity) should be extended to multi-dimensional ESG performance and settings where labels of the protected groups are not accessible, or are legally prohibited. Green outcome auditable standards: Establish third party verifiable standards to curtail self-reporting. Interdisciplinary analysis like integrated climate science, social impact assessment, and economics should include better systemic effects.

Conclusion and Recommendations

The power of AI to enable sustainable finance can be enhanced in three ways: through better risk analysis, greater transparency, and targeted capital flows to climate transition and social-impact projects. Nevertheless, unless monitored by stringent authorities, AI is likely to intensify bias, blur responsibility and facilitate greenwashing. The heterogeneous data, the ambiguity of measurement, and high externalities are the peculiarities of sustainable finance that implies that ethical protection should be the focus of AI implementation.

REFERENCES

[1] OECD. *AI Principles* (overview and updates). oecd.ai

[2] Bank for International Settlements. *Governance of AI adoption in central banks* (BIS report, Jan 2025). Bank for International Settlements

[3] Reuters. EU watchdog says banks must take full responsibility when using AI (ESMA guidance, May 2024). Reuters

[4] Reuters. India central bank committee recommends AI framework for finance sector (RBI committee, Aug 2025). Reuters

[5] Garcia ACB, et al. *Algorithmic discrimination in the credit domain* (Springer, 2024). SpringerLink

[6] Wang X., et al. Algorithmic discrimination: examining its types and mechanisms (PMC, 2024).