INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY
IJARIIT
ISSN: 2454-132X
Impact Factor: 6.078
(Volume 11, Issue 5 - V11I5-1251)
Available online at: https://www.ijariit.com

The Digital Catalyst for India's Green Transition: Applying AI-Driven Recommender Systems and Natural Language Processing to Enhance Sustainable Supply Chains in Tier 2 MSMEs

Purab Swarup
mypublishedpaper@gmail.com
Lotus Valley International School, Haryana

ABSTRACT

This paper proposes an AI-driven marketplace that leverages Natural Language Processing (NLP), Document Intelligence, and Learning-to-Rank (LTR) models to resolve market frictions of discovery, trust, and compliance. NLP structures fragmented product data, while document AI verifies sustainability claims aligned with frameworks like BRSR, EPR, and MSME ZED certification. LTR algorithms prioritise verified green suppliers, incentivising sustainable practices. The platform also ensures inclusivity through cross-lingual conversational agents, fairness audits, and interoperability via ONDC. By enabling verifiable "green matching," the system supports corporate Scope 3 emission reductions while advancing SDGs. The research demonstrates how AI marketplaces can serve as digital infrastructure for climate resilience and inclusive economic growth in India's green transition.

Keywords: *Artificial Intelligence, Natural Language Processing, Responsible AI.*

1. INTRODUCTION

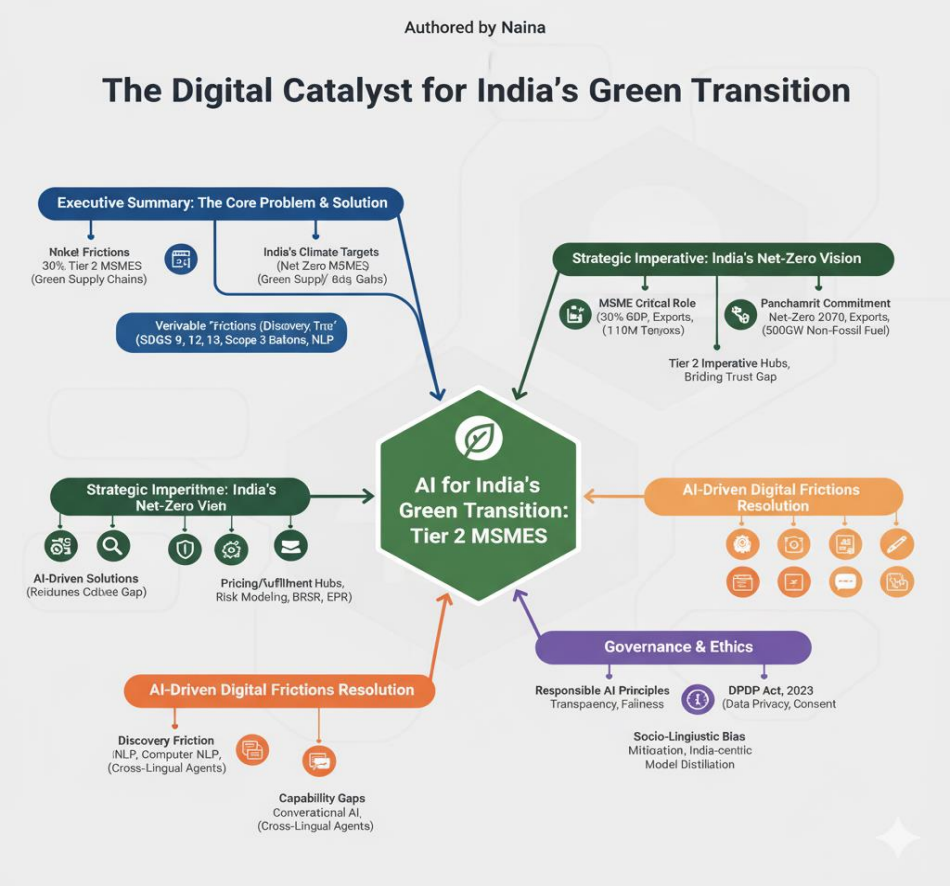
Achieving India's ambitious climate targets, specifically the Net Zero 2070 commitment, fundamentally relies on the rapid and scalable integration of Micro, Small, and Medium Enterprises (MSMEs) located in Tier 2 cities into verifiable green supply chains. This report argues that AI-driven recommender systems and Natural Language Processing (NLP) are essential digital infrastructures capable of resolving the pervasive market frictions that currently exclude Tier 2 suppliers, particularly concerning discovery, trust, and capability gaps. NLP is uniquely positioned to structure fragmented MSME digital content, transforming informal product descriptions into standardised, searchable assets. Document NLP and advanced risk models enable the automated verification of sustainability claims, leveraging nascent regulatory data streams like the Business Responsibility and Sustainability Reporting (BRSR) framework and Extended Producer Responsibility (EPR) obligations. Crucially, Learning-to-Rank (LTR) models actively steer market demand by optimising matches for sustainability attributes alongside traditional commercial factors like price and logistics. Operational success necessitates adherence to India's robust governance frameworks, including the Digital Personal Data Protection (DPDP) Act, 2023, and NITI Aayog's Responsible AI principles, with diligent focus on mitigating India-centric socio-linguistic biases. By facilitating verifiable "green matching," this AI marketplace operationalises key Sustainable Development Goals. It converts MSME resource efficiency into auditable corporate Scope 3 emission reductions, positioning the platform as a vital national tool for climate mitigation and inclusive growth.

2. LITERATURE REVIEW

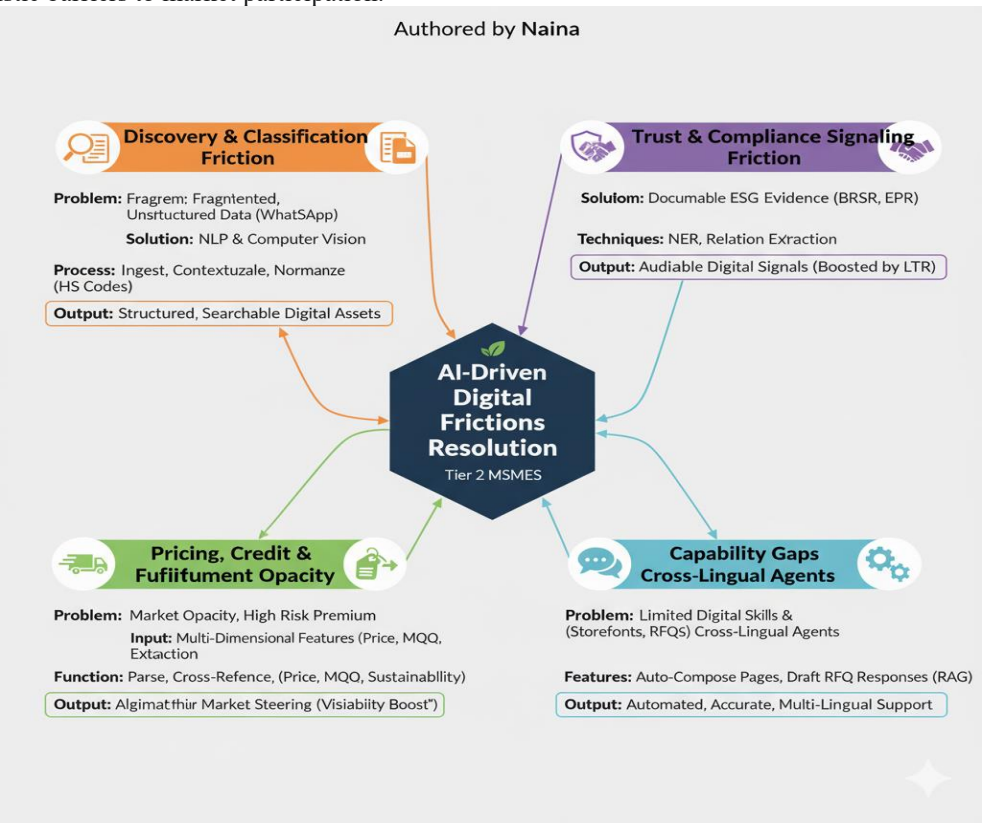
A primary impediment to integrating Tier 2 MSMEs is their product data's extreme fragmentation and lack of standardisation. Product names, specifications, and process descriptors often appear in mixed languages, inconsistent formats, and are communicated through informal channels, such as WhatsApp brochures. This unstructured environment creates a massive barrier to discovery for algorithm-driven procurement systems.

The solution relies on advanced Natural Language Processing (NLP) and Computer Vision (CV) techniques. The AI system first ingests raw data, including product titles, descriptions, and accompanying images from supplier catalogues. NLP models perform contextual analysis to decode ambiguous terms, for instance, distinguishing whether "adapter" refers to electrical hardware or software.

Large corporate buyers and export anchors require verifiable evidence to satisfy their ESG due diligence requirements, including certifications (e.g., recycled content), test reports, Extended Producer Responsibility (EPR) compliance letters, or disclosures linked to the SEBI Business Responsibility and Sustainability Reporting (BRSR) regime. The BRSR regime, though directly mandated for large listed firms, has subsequently pushed ESG requirements across all tiers of the value chain, increasing the market value of verifiable data for MSMEs. Furthermore, obligations like EPR, particularly for plastic waste management, demand detailed documentation.



Market opacity regarding input prices, lead times, and logistical reliability limits competitive participation and often raises the risk premium associated with sourcing from small suppliers. To overcome this, the marketplace employs Learning-to-Rank (LTR) models, a form of machine-learned ranking (MLR) that uses supervised machine learning to construct complex ranking models. For B2B engagement, AI tools, often leveraging Retrieval-Augmented Generation (RAG), can parse complex RFQ documents and auto-draft accurate, context-specific responses by pulling facts from the MSME's verified internal content library. This architecture ensures that the efficiency gains of automated response do not compromise accuracy. RAG, by grounding its output in the MSME's trusted, verified documents, becomes a required security feature for reliable B2B transactions. Critically, to support inclusion across the diverse Indian market, these AI solutions must be cross-lingual, capable of fluently conversing in and seamlessly switching between English, Hindi, and regional languages, including code-mixed conversations, thereby lowering the technical and linguistic barriers to market participation.



3. METHODOLOGY

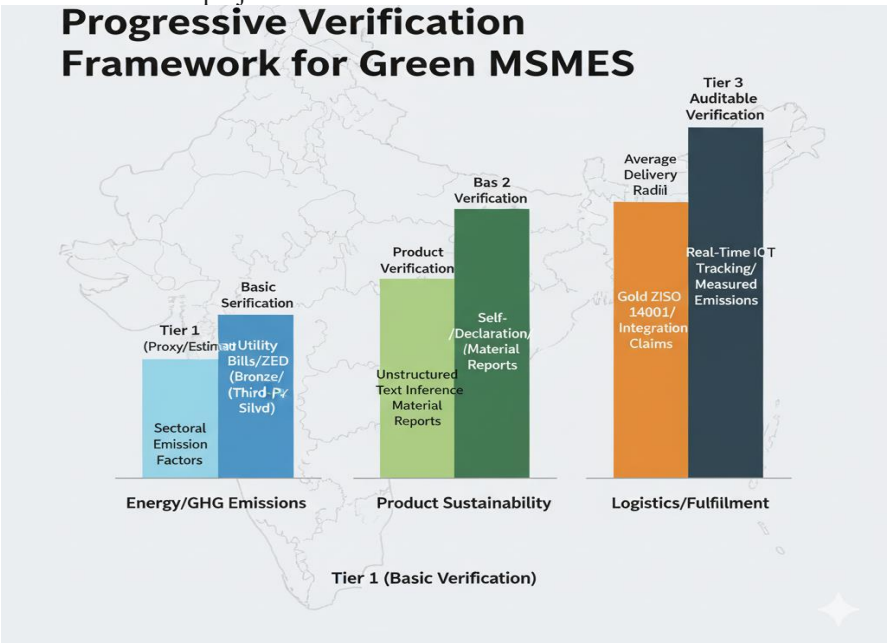
The goal of equitable market access for Tier 2 MSMEs is significantly enhanced by leveraging India's public digital infrastructure. The Open Network for Digital Commerce (ONDC) is designed to dismantle monopolistic tendencies and lower entry barriers for small businesses. ONDC standardises various operations, including cataloguing, inventory, and order management. A major challenge for Tier 2/3 MSMEs is the frequent lack of formalised, auditable data trails necessary for ESG due diligence, such as metered energy use or effluent process logs. In many cases, especially where centralised SME data is unavailable, using proxy data and estimation methodologies is the only practical starting point.

To address this without imposing prohibitive data burdens, AI systems must adopt a progressive, tiered verification model.

Data Gap Type	Tier 1 (Proxy/Estimate)	Tier 2 (Basic Verification)	Tier 3 (Auditable Verification)
Energy/GHG Emissions	Sectoral emission factors (e.g., GLEC values, India-specific BEE data), process type estimates.	Submission of validated utility bills, MSME ZED Certification (Bronze/Silver).	Smart meter integration, third-party energy audit report, verifiable carbon accounting.
Product Sustainability	Unstructured text inference (e.g., "recycled fabric" extracted by NLP).	Upload of self-declaration or material test reports; EPR mandate compliance evidence.	Verified certifications (e.g., Gold ZED, ISO 14001, audited recycled content claims).
Logistics/Fulfilment	Average delivery radii, generalised service reliability scores.	Integration with basic ONDC logistics providers for trackable fulfilment data.	Real-time IoT tracking, optimised route data shared with the buyer, and measured emissions data for the last-mile.

For Tier 1 verification, AI systems utilise inferred attributes, such as typical emissions intensity by process or material. To mitigate the risk of inaccurate environmental signalling, these estimation algorithms must embed highly granular, India-specific factors, such as India-specific Global Logistics Emissions Council (GLEC) values for freight and subnational variability in power grid emission factors. This commitment to granular proxy data minimises the risk of unfairly penalising efficient MSMEs. The platform must clearly separate these estimates from verified values.

The progression toward Tier 3 verification (auditable data) often requires significant capital expenditure from the MSME (e.g., purchasing smart meters, installing dedicated ESG software). The AI platform must provide a clear financial justification for this investment by using LTR data to show the projected economic benefit.



To boost adoption, the AI platform must be tightly integrated with existing national initiatives. The government's MSME Sustainable (ZED) Certification Scheme, which promotes 'Zero Defect Zero Effect' practices, offers tiered certification (Bronze, Silver, Gold) along with substantial subsidies (up to 80% for Micro units). The AI platform must recognise and prioritise ZED-certified units within its recommender logic.

Similarly, the Bureau of Energy Efficiency (BEE) implements long-standing programs focused on energy efficiency (EE) and Renewable Energy (RE) adoption in MSME clusters, which leads to significant cost savings and reduced greenhouse gas emissions. India's policy guidance emphasises "AI for All," centred on core principles including safety, accountability, transparency, fairness, and inclusivity. Operationalising these principles requires strict implementation within the AI marketplace:

i. **Transparency and Accountability:** The platform must provide meaningful explanations for ranking and matching decisions and document data provenance. Furthermore, systems must enable human override mechanisms in high-impact decisions, such as supplier delisting.
ii. **Fairness Assessment:** It is necessary to systematically assess bias across different languages, regions, and firm sizes to ensure equitable access.
iii. **Governance Frameworks:** To facilitate responsible deployment, especially during rapid Tier 2 rollout, the platform should adopt lightweight governance scaffolding, such as detailed model cards, process checklists, and registers for tracking incidents, aligning with guidance issued by NITI Aayog.

AI recommenders and NLP modules inevitably handle sensitive business data and personal information (names, contacts, invoices). India's Digital Personal Data Protection (DPDP) Act, 2023, requires adherence to principles of lawful purpose, informed consent, purpose limitation, and data minimisation.

Marketplaces acting as Data Fiduciaries must implement robust compliance measures, including consent dashboards for users to manage opt-in and withdrawal, and strict data retention controls. A key tension exists between the need for robust verification (which requires data) and the DPDP's data minimisation requirement. All notices and compliance documents must be provided in local languages to align with DPDP requirements.

A significant challenge in deploying large language models (LLMs) across India is the presence of systemic biases learned from training data, which often perpetuate harmful stereotypes related to caste, religion, and regional identity. These biases directly threaten the platform's mandate for inclusivity. If NLP systems exhibit bias against regional dialects or specific socio-cultural identifiers, the resulting misclassification or poor service quality will disproportionately affect Tier 2 MSMEs, creating digital exclusion. Therefore, bias assessment is not just an ethical consideration but a functional requirement for inclusive growth.

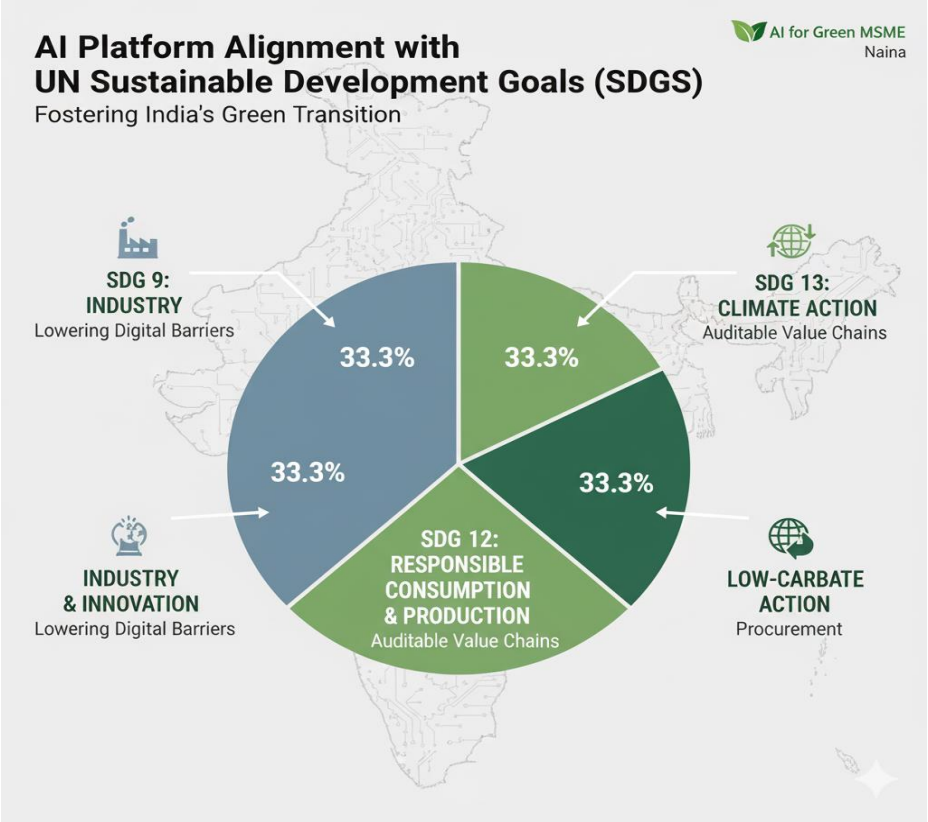
Choosing energy-efficient model architectures, such as distilled models (smaller, more efficient versions) and Retrieval-Augmented Generation (RAG) models, substantially reduces energy consumption during inference while maintaining accuracy. Moreover, responsible operational choices, such as implementing edge inference, sharing models across multiple tenants to avoid redundant training cycles, and employing carbon-aware scheduling (aligning compute demand with peak renewable energy availability), materially reduce the overall environmental footprint of the marketplace.

Framework	**Applicable Area**	**Key Design Obligation for Marketplace**	**Relevant Stakeholders**
DPDP Act, 2023	Data Privacy & Consent	Consent dashboards, purpose limitation, data minimisation, local language notices, and grievance redressal mechanisms.	MSMEs (Data Principals), Platform (Data Fiduciary), Buyers.
SEBI BRSR Regime	Sustainability Disclosure	Facilitating documentation and verification streams for value chain partners (especially transaction value for listed buyers).	Large Buyers (Listed Entities), MSMEs (Value Chain Partners).
NITI Aayog RAI	Algorithmic Ethics	Bias assessment across regions/languages, provision of meaningful explanations for ranking/matching decisions, and human override capability.	Platform Developers, Data Scientists, Tier 2 MSME users.
MSME ZED Scheme	Sustainable Practices	Integration of ZED Certification status into the recommender system logic, offering tools to facilitate assessment for Bronze/Silver tiers.	MSMEs, Government/Ministry of MSME.

4. RESULT AND ANALYSIS

The deployment of an AI-driven marketplace is positioned to be a highly effective catalyst for India's climate transition by resolving foundational market failures. The platform systematically operationalises sustainable trade by steering commercial demand toward verifiably greener suppliers, making sustainable choices the path of least resistance for Tier 1 buyers.

The technology successfully resolves the trilemma facing Tier 2 MSME participation: granting **visibility** through NLP-driven standardisation; establishing **trust** via Document NLP and BRSR compliance validation; and ensuring deep **inclusion** through cross-lingual AI support and a focus on Tier 2 geography. By enabling the substitution of energy-intensive inputs with low-carbon alternatives, optimising logistics, and improving resource efficiency (SDG 12), the aggregated effect of the AI marketplace contributes directly to resilient supply chains and India's long-term climate action pathway (SDG 13). The platform advances SDG 9 (Industry, Innovation, and Infrastructure) by integrating MSMEs into advanced digital networks, promoting resilient industrialisation. It directly operationalises SDG 12 (Responsible Consumption and Production) by steering demand towards resource-efficient and circular practices, ensuring that green claims are auditable end-to-end.



5. CONCLUSION

This study recommends four key strategies. First, policymakers should mandate interoperable and fair data standards through ONDC to ensure sustainable data portability and reduce fragmentation. Second, governments and financial institutions must incentivise progressive verification by subsidising affordable ESG tools that help MSMEs advance from proxy data to auditable ZED certification. Third, a national AI fairness audit mechanism should be established to certify the ethical performance of trade-related AI systems and mitigate socio-linguistic and regional biases. Finally, public procurement policies should prioritise providers committed to Green AI principles, emphasising energy-efficient models and carbon-aware practices to lower the digital sector's environmental footprint.

REFERENCES

[1] Role of MSMEs in Enabling Growth in Tier 2 and Tier 3 Cities in India, updated on February 18, 2025, https://kinaracapital.com/msmes-growth-in-tier-2-and-tier-3-cities/
[2] Digital platform to support MSME climate action launches today in India, accessed August 31, 2023, https://smeclimatehub.org/digital-platform-to-support-msme-climate-action-launches-today-in-india/
[3] Green MSMEs in India: Small Businesses Powering Big Climate Action, accessed June 16, 2025, https://netzeroindia.org/green-msmes-in-india/
[4] India's path to net zero: a work in progress | British Safety Council India, accessed April 8, 2025, https://www.britsafe.in/safety-management-news/2025/india-s-path-to-net-zero-a-work-in-progress