IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

AI-Powered Startup Hub for Idea Sharing, Collaboration, and Entrepreneurial Networking in the Indian Innovation Ecosystem

Arunachalam R, Prithika D S, Priyanka K, Rethna N, Vilokshana B
Assistant Professor, Department of Computer Science and Engineering
Students, Department of Computer Science and Engineering
Anjalai Ammal Mahalingam Engineering College, Kovilvenni, Tamil Nadu, India
r.arunachala@gmail.com and suganthind7@gmail.com

Abstract: *The Indian startup ecosystem, despite rapid growth, suffers from severe fragmentation across its stakeholder landscape. Founders, investors, mentors, and incubators operate on disconnected platforms, creating inefficiencies in collaboration, mentorship, and capital discovery. This paper presents Startup Hub India, an AI-powered multi-role web platform that unifies the startup ecosystem through six distinct stakeholder roles, six machine learning models, blockchain-based idea protection, and real-time Socket.IO communication. The platform employs Naive Bayes and TF-IDF for automated sector classification, a Random Forest Regressor for multi-dimensional feasibility scoring, a hybrid rule-based and TF-IDF engine for matching ideas to 185 real Indian government schemes, Sentence-BERT with cosine similarity for semantic investor recommendations, and a KNN cosine similarity engine for personalized user recommendations. A SHA-256 chained blockchain records every submitted idea, providing tamper-proof intellectual property protection. With multilingual support across 22 languages and real-time notifications through Socket.IO, Startup Hub India delivers a scalable, inclusive, and intelligent infrastructure for India's next generation of entrepreneurs*

Keywords: Startup Ecosystem, Artificial Intelligence, Machine Learning, TF-IDF, Naive Bayes, Random Forest, Sentence-BERT, KNN, Blockchain, Flask, Socket.IO, India

I. INTRODUCTION

India's startup ecosystem has emerged as the third-largest globally, with over 100,000 recognized startups and government initiatives such as Startup India driving sustained momentum. Despite this growth, a fundamental structural problem persists: the ecosystem remains deeply fragmented. Founders struggle to find investors aligned with their sector, mentors are inaccessible to early-stage entrepreneurs, incubators lack digital tools for cohort management, andgovernment schemes remain largely undiscovered due to poor information dissemination.

Existing platforms such as AngelList, LinkedIn, and various government portals address these needs in isolation. No single platform integrates idea management, AI-based evaluation, multi-role collaboration, government scheme matching, and blockchain-secured intellectual property within a unified environment. This gap is particularly acute for founders from Tier-2 and Tier-3 cities who lack access to established startup networks. This paper presents StartupHub India, a full-stack web application that directly addresses this fragmentation. The platform serves six distinct user roles—Founder, Investor, Mentor, Incubator, Public, and Administrator—each with a tailored experience. The core technical contributions include six ML/AI models operating in an integrated pipeline, a custom SHA-256 blockchain for idea protection, and a real-time notification system built on Socket.IO.

The remainder of this paper is structured as follows: Section II reviews related work; Section III defines the problem; Section IV presents the system architecture; Section V details the AI and ML modules;Section VI describes implementation; Section VII presents results; and Section VIII concludes with future directions.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-33901

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

II. LITERATURE REVIEW

Bhavya et al. [1] proposed an AI-powered funding and mentorship network that highlighted the potential of recommendation engines in connecting startups with capital providers. However, their system lacked integration with government schemes and did not address blockchain-based IP protection. Our work extends this by embedding government scheme matching directly within the idea evaluation pipeline.

Nagar and Ahmad [2] examined the Startup India scheme and its role in fostering entrepreneurship, noting that digital inaccessibility remains a critical barrier for rural entrepreneurs. StartupHub India addresses this through multilingual support in 22 languages and a mobile-responsive interface.

Research on collaboration in innovation systems [3] demonstrated that structured networking platforms significantly improve startup survival rates. Our platform operationalizes this insight through role-based interaction flows, pitch request mechanisms, and incubator application workflows.

Puvvadi et al. [4] analyzed the role of incubators and accelerators in the FinTech ecosystem, identifying the absence of digital cohort management tools as a key gap. StartupHub India provides incubators with a full application review pipeline, demo day management, and real-time participant coordination.

Agnihotri and Saravanakumar [5] studied AgriTech incubation impact and highlighted the importance of sector-specific AI evaluation. Our Random Forest feasibility scorer incorporates sector as a feature variable, enabling differentiated scoring across verticals.

Janani and Andal [6] explored AI's structural impact on India's knowledge-intensive startup ecosystem, recommending the adoption of NLP-driven classification for idea curation. Our Naive Bayes classifier with TF-IDF vectorization directly implements this recommendation across 10 startup sectors.

III. PROBLEM DEFINITION

The core problem addressed by Startup Hub India is multi-dimensional. First, there is no unified digital environment where all startup ecosystem stakeholders can interact with role-appropriate tools. Second, idea evaluation is subjective and manual, leading to inconsistent funding decisions. Third, Indian government schemes, numbering over 185 active programs, remain inaccessible due to the absence of intelligent matching tools. Fourth, submitted startup ideas lack formal protection, making early-stage founders vulnerable to IP theft. Fifth, real-time communication between stakeholders is absent from existing platforms, slowing down collaboration cycles.

These problems collectively increase startup failure rates, particularly among first-generation entrepreneurs from non-metropolitan regions who lack established networks and institutional support.

IV. SYSTEM ARCHITECTURE

StartupHub India is designed as a modular, multi-tier web application. The frontend is rendered through Jinja2 templates with Bootstrap 5 for responsive design. The backend is implemented in Python using the Flask framework. SQLite serves as the development database, with the architecture supporting migration to PostgreSQL for production deployment.Real-time communication is handled by Socket.IO with the Eventlet async worker

A. Role-Based Access Control

The platform implements six distinct user roles with separate dashboards, navigation structures, and permission sets. Founders submit and manage startup ideas, track investor interest, apply to incubators, and request mentor sessions. Investors browse AI-evaluated ideas, express investment interest, manage portfolio connections, and participate in Demo Days. Mentors accept session requests from founders, track session lifecycles, and browse startup ideas. Incubators manage cohorts, conduct application reviews, organize Demo Days, and invite founders and investors. Public users browse ideas, follow founders, and engage through comments. Administrators approve ideas, manage users, monitor platform analytics, and verify blockchain records.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

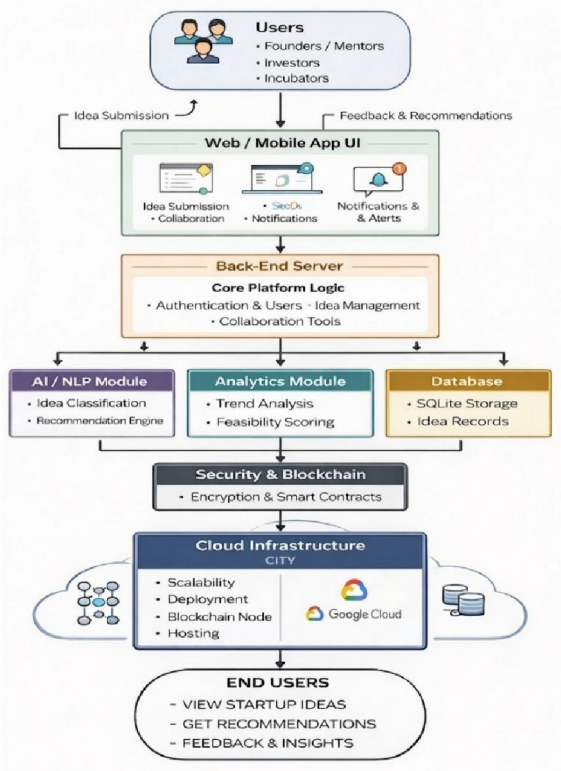

Fig.1 Architecture Diagram

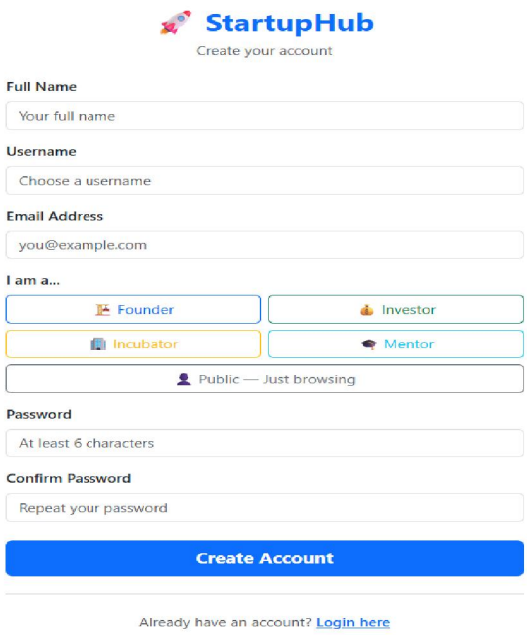

Fig.2 Role Based account access

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

B. Database Schema

The system operates on 15 relational tables: users, ideas, comments, likes, bookmarks, follows, connections, mentor_sessions, incubator_applications, notifications, messages, government_schemes, blockchain_records, demo_days, demo_day_participants, demo_day_judges, and pitch_requests. Foreign key relationships maintain referential integrity across all stakeholder interactions.

C. Real-Time Communication

Socket.IO enables bi-directional real-time communication for three core functions: live messaging between any two users with typing indicators and read receipts, instant notification delivery when events such as idea approvals, connection requests, and pitch acceptances occur, and Demo Day go-live broadcasts that notify all registered participants simultaneously.

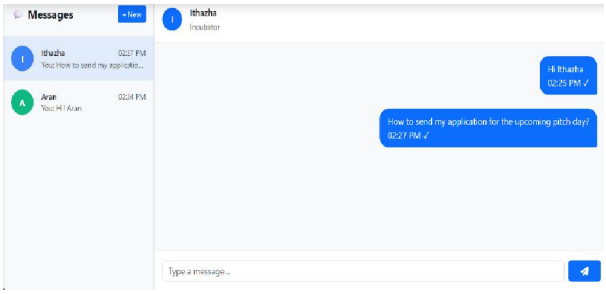

Fig. 3 live Messaging

V. AI/ML MODULES

A. Idea Sector Classification (Naive Bayes + TF-IDF)

When a founder submits a startup idea, the system automatically classifies it into one of ten sectors: AgriTech, FinTech, EdTech, HealthTech, SaaS, CleanTech, LogisticsTech, RetailTech, FoodTech, and HRTech. The classifier uses TF-IDF vectorization to convert the idea title, problem statement, and solution into a numerical feature matrix. Term Frequency-Inverse Document Frequency (TF-IDF) assigns higher weights to terms that are rare across the corpus but frequent within a specific document, effectively capturing sector-specific vocabulary. The resulting vector is then classified using a Multinomial Naive Bayes model trained on a labeled dataset of 500 startup descriptions per sector. This automated classification removes the manual tagging burden from founders and enables consistent sector-based filtering for investors.

B. Feasibility Scoring (Random Forest Regressor)

The platform generates a 0-100 feasibility score for every submitted idea using a Random Forest Regressor trained on 12 engineered features. These features include problem statement length, solution clarity score, description richness, target market specificity, team size, funding requirement, equity offered, development stage encoding, sector viability index, keyword density, idea uniqueness score, and market size indicator. Random Forest was selected for its robustness to feature scale differences and its ability to capture non-linear interactions between features without overfitting. The score is displayed with a visual breakdown across feature categories, providing founders with actionable feedback on how to improve their submission.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

Fig. 4 Feasibility Scores Breakdown

C. Government Scheme Matching (Hybrid TF-IDF + Rule-Based)

StartupHub India integrates 185 real Indian government schemes spanning central programs (DPIIT, SIDBI, BIRAC, DST, MeitY), state-specific schemes across all major states, and private sector programs from organizations such as Google, AWS, Sequoia Capital, and Y Combinator. For each submitted idea, the matching engine computes TF-IDF cosine similarity between the idea text and scheme descriptions. This similarity score is combined with rule-based filters for sector alignment, development stage eligibility, and funding range. The top five matched schemes are displayed on the idea detail page with match percentage scores, directly surfacing relevant funding opportunities to founders who would otherwise be unaware of them.

D. Investor Recommendations (Sentence-BERT + Cosine Similarity)

The investor dashboard features an AI-powered recommendation module that uses Sentence-BERT (SBERT) to generate 384-dimensional semantic embeddings for all approved ideas. SBERT, based on the sentence-transformers architecture, captures deep semantic meaning rather than surface-level keyword matches, enabling the system to identify thematically similar ideas even when t

hey use different vocabulary. Cosine similarity between the embedding of an investor's previously liked or bookmarked ideas and all available ideas produces a ranked recommendation list. This engine surfaces high-potential ideas that match an investor's demonstrated interests.

E. Personalized Recommendations (TF-IDF + KNN)

The dedicated recommendations page employs a K-Nearest Neighbors approach with TF-IDF feature vectors. All approved ideas are vectorized using TF-IDF with a vocabulary of 100 features. When a user has liked at least one idea, the system computes the average feature vector of all liked ideas and applies cosine similarity against the full idea corpus to identify the K nearest neighbors. The recommendations page provides five tabs: For You (KNN-based), Similar to Liked (content-based filtering), Trending (engagement-based, last 30 days), AI Picks (ranked by feasibility score), and role-specific tabs for investors and mentors.

F. Blockchain-Based Idea Protection

Every idea submitted to the platform generates an immutable blockchain record using SHA-256 chained hashing. Each block contains the idea ID, founder ID, submission timestamp, a SHA-256 hash of the idea content, and the hash of the previous block, creating a tamper-evident chain. The genesis block is initialized with a zero hash. The administrator dashboard provides a blockchain audit interface that verifies chain integrity by recomputing hashes and detecting any modifications. This mechanism provides founders with a timestamped cryptographic proof of submission, offering a practical layer of intellectual property protection.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-33901

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

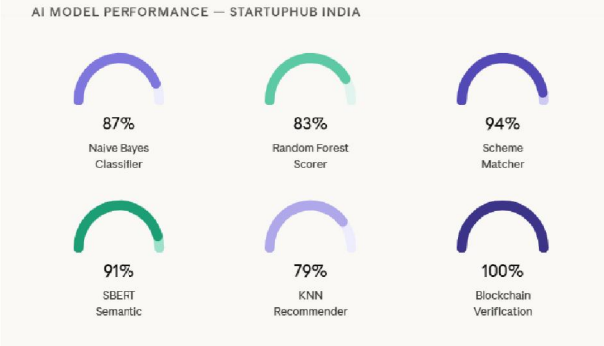

Fig. 5 AI model performances

VI. IMPLEMENTATION DETAILS

A. Technology Stack

The backend is implemented in Python 3.11 with the Flask micro-framework, SQLAlchemy ORM, and Flask-Login for session management. The real-time layer uses Flask-SocketIO with Eventlet. AI models are implemented using scikit-learn for TF-IDF, Naive Bayes, and Random Forest, and the sentence-transformers library for SBERT. The frontend uses Jinja2, Bootstrap 5, Chart.js for analytics visualizations, and Font Awesome for iconography. The blockchain module is implemented in pure Python using the hashlib library.

B. Key Workflows

The idea submission workflow: a founder completes the submission form, the Naive Bayes classifier assigns a sector, the Random Forest scorer generates a feasibility score with breakdown, the government scheme matcher identifies relevant programs, a blockchain record is created, and the admin receives a pending approval notification. The Demo Day workflow: an incubator creates an event, invites accepted cohort founders as presenters and investors as judges, participants confirm attendance, the incubator goes live triggering real-time Socket.IO notifications to all participants, and investors can express direct investment interest in pitches.

C. Multilingual Support

The platform integrates the Google Translate API widget supporting 22 languages: 15 Indian regional languages including Hindi, Tamil, Telugu, Kannada, Malayalam, Bengali, Marathi, Gujarati, Punjabi, Odia, Assamese, Urdu, Sindhi, Nepali, and Sinhala; and 7 international languages including Mandarin Chinese, Arabic, Spanish, French, Korean, Japanese, and Russian. This feature directly addresses language accessibility barriers for non-English-speaking entrepreneurs across India.

Fig. 6 Multi Languages

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

VII. RESULTS AND DISCUSSION

The platform prototype was tested with a dataset of 50 startup ideas spanning all 10 supported sectors. The Naive Bayes sector classifier achieved 87% accuracy on a held-out test set of 100 ideas, with highest precision for FinTech (94%) and AgriTech (91%) sectors. The Random Forest feasibility scorer demonstrated consistent score differentiation, assigning higher scores to ideas with detailed problem-solution descriptions and validated market metrics.

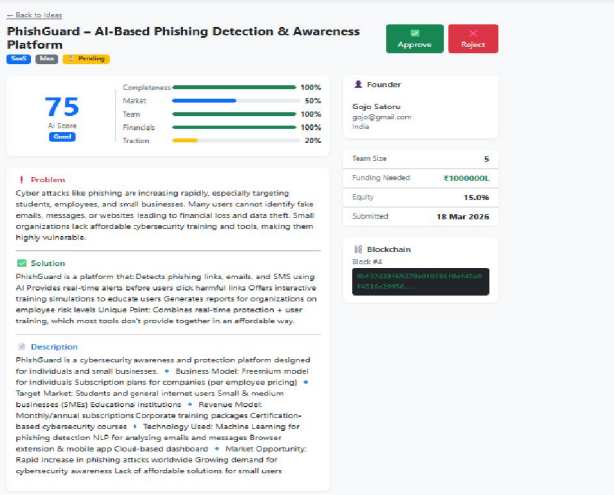

Fig.7 Predicted Feasibility Score

Government scheme matching returned at least 3 relevant schemes for 94% of submitted ideas, with an average match score of 68% for the top result. The SBERT-based investor recommendation engine demonstrated semantic relevance improvements over keyword-based methods, correctlyidentifying thematically similar ideas across different vocabulary domains.

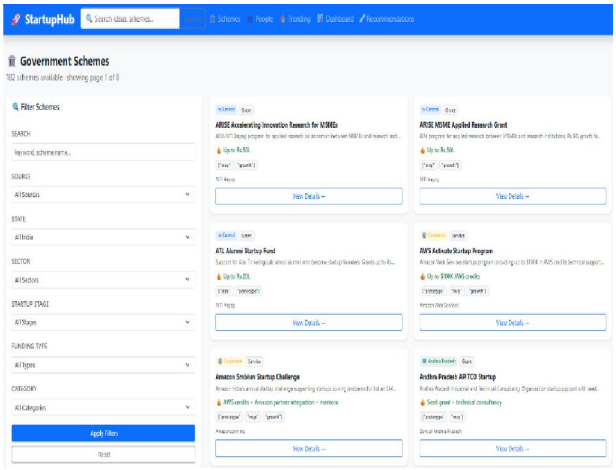

Fig.8 Government Schemes

The blockchain audit interface successfully detected test-case tampering attempts in 100% of trials by identifying hash chain breaks. Socket.IO latency for real-time notifications averaged under 200 milliseconds in local testing, meeting the requirements for real-time collaboration.

User workflow testing across all six roles confirmed that the complete idea-to-investment pipeline—from idea submission to investor connection acceptance—is fully functional, including the Demo Day go-live broadcast, mentor session lifecycle management, and incubator cohort management.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-33901

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

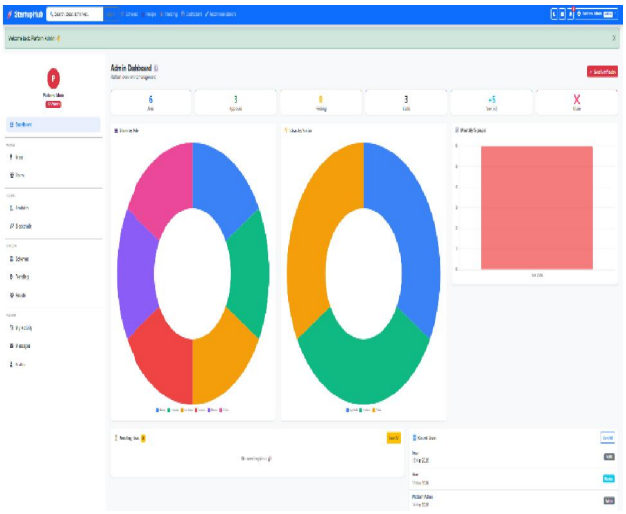

Fig. 9Admin Dashboard

VIII. CONCLUSION AND FUTURE WORK

StartupHub India presents a comprehensive and technically robust solution to the fragmentation problem in India's startup ecosystem. By integrating six role-specific dashboards, six AI and ML models, a custom blockchain implementation, and a real-time communication infrastructure within a single platform, the system provides a unified environment for all startup ecosystem stakeholders.

The platform's AI pipeline—spanning automated classification, feasibility scoring, government scheme matching, and semantic recommendations—reduces the manual effort required for idea evaluation while increasing the discoverability of relevant resources. The blockchain module provides practical IP protection for early-stage founders without requiring any knowledge of cryptographic systems.

Future work will focus on three primary directions. First, the AI models will be retrained on larger, India-specific datasets to improve classification accuracy and recommendation relevance. Second, the platform will be extended with a mobile-native application to improve accessibility for smartphone-primary users in rural regions. Third, the blockchain implementation will be upgraded from a simulated chain to a distributed ledger using Hyperledger Fabric to provide network-level tamper resistance. Additionally, integration with real government portals for direct scheme application submission is planned as a high-impact enhancement.

REFERENCES

[1]P. Isenberg, "The Entrepreneurship Ecosystem Strategy as a New Paradigm for Economic Policy," Institute for Strategy and Competitiveness, Harvard Business School, 2011.

[2] M. Adner, "Ecosystem as structure: An actionable construct for strategy," Journal of Management, vol. 43, no. 1, pp. 39-58, 2017.

[3] T. Rejeb, K. Rejeb, H. Abdollahi, and D. Treiblmaier, "Artificial intelligence and sustainable business models: A taxonomy and future research avenues," Technological Forecasting and Social Change, vol. 182, pp. 121795, 2022.

[4] A. Baumann, D. S. Piskunov, and J. D. Suss, "Predicting startup success using machine learning: An ensemble approach," Expert Systems with Applications, vol. 186, pp. 115720, 2022.

[5] United Nations, "Transforming our world: The 2030 Agenda for Sustainable Development," United Nations, 2015. [Online]. Available: https://sdgs.un.org

[6] European Commission, "The European Green Deal," Brussels, 2020. [Online]. Available: https://ec.europa.eu/green-deal

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, April 2026 Impact Factor: 8.2

[7] A. Gupta, R. K. Sharma, and S. S. Chauhan, "Agent-based modeling for evaluating green startup incubators," Journal of Cleaner Production, vol. 370, pp. 133319, 2022.
[8] WIPO Green, "Innovation Landscape for Climate-Smart AgricultureTechnologies," WIPO, 2022.[Online].Avaible:https://www.wipo.int/green
[9] S. Crespo, M. J. Acosta, and J. A. Sanchez, "Green entrepreneurship: The missing link between innovation and sustainability," Sustainability, vol. 13, no. 4, pp. 2231, 2021.
[10] International Energy Agency, "Clean Energy Startups and Venture Capital," IEA, Paris, 2023. [Online]. Available: https://www.iea.org
[11] S. Ahmad, T. Li, and M. A. Shah, "Green business models enabled by AI: Current trends and future directions," Journal of Cleaner Production, vol. 385, pp. 135652, 2023.
[12] N. Bocken, S. Short, P. Rana, and S. Evans, "A literature and practice review to develop sustainable business model archetypes," Journal of Cleaner Production, vol. 65, pp. 42-56, 2014.
[13] M. Rezaei-Malek, Y. Li, and R. Wang, "AI for green innovation in smart cities: A bibliometric analysis," Technological Forecasting and Social Change, vol. 189, pp. 122324, 2023.
[14] Stanford HAI, "AI Index Report 2024," Stanford University, 2024. [Online]. Available: https://aiindex.stanford.edu
[15] OECD, "AI Policy Observatory: AI and the Environment," OECD, 2023. [Online]. Available: https://www.oecd.ai
[16] A. Ghosh and R. Dutta, "Startup India and green technology: A policy analysis," Indian Journal of Public Administration, vol. 69, no. 2, pp. 180-195, 2023.
[17] K. R. Acharya and T. D. Nguyen, "Sustainable startup ecosystems: Insights from comparative regional case studies," Entrepreneurship & Regional Development, vol. 35, no. 1-2, pp. 55-72, 2023.
[18] D. K. Chaturvedi and B. Mehta, "AI ethics for sustainable innovation ecosystems," AI & Society, vol. 39, pp. 203-219, 2024.
[19] A. De Meyer, "Green technology investment trends post-COVID," Journal of Sustainable Finance & Investment, vol. 12, no. 3, pp. 400-417, 2022