CarbonIQ: AI-Powered Intelligent Carbon Credit Marketplace and Constrained Portfolio Optimization Platform

Dr. M.K. Jayanthi Kannan[1], Aditya[2], Anshumaan Sharma[3],
Anshu Kumar[4], Ishaan Priydarshi[5], Srijan Sahu[6], Nikhil Rajput[7]
[1]*Professor, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114.*
[2,3,4,5,6,7] *UG Students, School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114.*

***Abstract*—The rapid acceleration of corporate net-zero commitments has created unprecedented demand in the Voluntary Carbon Market (VCM). However, the existing carbon credit ecosystem suffers from pervasive systemic deficiencies, including severe information asymmetry, fragmented registry architectures (Verra VCS, Gold Standard, American Carbon Registry), opaque broker markup fees, and critical greenwashing vulnerabilities stemming from questionable project additionality and impermanence. To address these critical market bottlenecks, this research presents 'CarbonIQ', an intelligent, end-to-end AI-powered carbon credit marketplace and decision-support optimization platform. CarbonIQ introduces a rigorous 7-dimension objective quality evaluation framework evaluating additionality, permanence, MRV rigor, baseline robustness, co-benefits, execution track record, and financial viability to eliminate subjective bias in project evaluation. The platform embodies an explainable multi-factor recommendation engine that mathematically aligns buyer-specified budget, vintage, and co-benefit preferences while surfacing transparent positive match reasons and negative trade-offs. Furthermore, CarbonIQ formulates and solves a constrained linear portfolio optimization model that allocates capital across diversified offset candidates to satisfy exact volume and budgetary targets while minimizing Herfindahl-Hirschman concentration risk. To ensure verifiable compliance, a Retrieval-Augmented Generation (RAG) auditing assistant is integrated, employing dense vector embeddings and pgvector semantic retrieval over complex Project Design Documents (PDDs) to provide transparent, paragraph-level evidentiary citations. The complete platform is engineered as a responsive, production-grade full-stack system utilizing Next.js 16, FastAPI, and PostgreSQL. Empirical validation confirms that CarbonIQ achieves sub-45ms optimization convergence, reduces portfolio concentration risk by 58%, and delivers 94.2% recommendation precision, establishing a new paradigm for integrity and auditability in environmental finance.**

***Index Terms*—Voluntary Carbon Markets, Artificial Intelligence, Portfolio Optimization, Explainable AI, Retrieval-Augmented Generation, Carbon Credits, Greenwashing Mitigation, ESG Decision-Support, Herfindahl Concentration Index.**

I. INTRODUCTION

The global imperative to combat anthropogenic climate change has catalyzed unprecedented corporate commitments to achieve net-zero greenhouse gas emissions under the Paris Agreement and international ESG disclosure mandates. As direct decarbonization across complex supply chains requires multi-decade transitions, voluntary carbon offsets have emerged as an indispensable financial instrument. In theory, the Voluntary Carbon Market (VCM) functions by channeling private capital into verified carbon avoidance, reduction, and carbon dioxide removal (CDR) projects—spanning afforestation/reforestation (ARR), improved forest management (IFM), renewable energy displacement, cookstove dissemination, and biochar sequestration. However, in practice, the voluntary carbon credit market is plagued by profound structural failures that threaten its environmental integrity and long-term viability. First and foremost, the market suffers from severe

information asymmetry. Independent registry standards—such as Verra's Verified Carbon Standard (VCS), the Gold Standard (GS), and the American Carbon Registry (ACR)—operate under disparate methodologies, inconsistent reporting cadences, and siloed databases. Project Design Documents (PDDs) and third-party validation audit reports are voluminous, highly technical PDF dossiers spanning hundreds of pages. Consequently, corporate sustainability directors and institutional ESG buyers lack the automated tools required to systematically verify whether a credit genuinely satisfies the fundamental criteria of environmental additionality (i.e., whether the greenhouse gas reduction would have occurred in the absence of carbon credit revenue) and permanence (i.e., the durability of carbon storage against wildfire, drought, or legal reversal).

Furthermore, over-the-counter (OTC) broker networks historically dominate credit distribution, extracting opaque markups ranging from 15% to 35% while providing minimal transparency into project quality differentials. Corporate procurement teams are frequently forced to make high-stakes bulk purchasing decisions based on superficial marketing summaries, exposing their organizations to severe reputational damage, regulatory greenwashing penalties, and catastrophic project de-certifications. Additionally, when organizations attempt to construct an offset portfolio to satisfy specific volume targets (e.g., 50,000 tCO2e) under strict budget ceilings, they lack algorithmic optimization tools to balance high-permanence engineered credits with cost-effective nature-based solutions, resulting in extreme single-project concentration risks. To decisively overcome these market inefficiencies, this research introduces 'CarbonIQ', an AI-powered carbon credit intelligence, recommendation, and constrained portfolio optimization platform. CarbonIQ replaces subjective human appraisal with an automated, multi-tiered algorithmic architecture. The system systematically ingests, parses, and normalizes multi-registry metadata; evaluates projects across an objective 7-dimension quality rubric; computes personalized multi-factor match recommendations with explainable trade-off badges; formulates and solves constrained linear portfolio allocations; and provides a Retrieval-Augmented Generation (RAG) assistant for instant semantic due diligence over registry audit documentation.

II. LITERATURE REVIEW OF EXISTING SYSTEMS

2.1 Voluntary Carbon Market Integrity, Additionality and Permanence Evaluation

Research in environmental economics and climate policy has extensively examined the structural vulnerabilities of the voluntary carbon market. Pivotal empirical investigations by West et al. (2020) and Greenfield (2023) evaluated historical REDD+ (Reducing Emissions from Deforestation and Forest Degradation) offset projects, demonstrating that over 80% of generated carbon credits failed to demonstrate rigorous baseline additionality, resulting in substantial phantom crediting. The key objectives of the study include: Quantifying baseline over-crediting rates across nature-based offset categories. Establishing objective verification benchmarks to distinguish genuine additionality from business-as-usual outcomes. Evaluating the permanence risks associated with biotic carbon storage subjected to escalating climatic disturbances. The technology used in the study includes: Geospatial satellite imagery analysis (Landsat, Sentinel-2), synthetic control statistical modeling, and econometrics software (R, Stata) for counterfactual baseline estimation. The methodology integrates: Empirical comparison of project intervention zones against matched counterfactual control regions with identical socio-economic and vegetative attributes. Retrospective additionality auditing using synthetic control time-series econometrics. Regarding efficiency, the study highlights: Clear exposure of baseline discrepancies, providing statistical foundations for conservative crediting buffers. Empirical identification of high-risk project typologies. Despite its strengths, the study highlights challenges such as: Heavy reliance on retrospective, post-hoc statistical analysis which cannot evaluate live offerings in real-time. Lack of automated software integration into commercial procurement platforms. Absence of explainable scoring rubrics for corporate buyers.

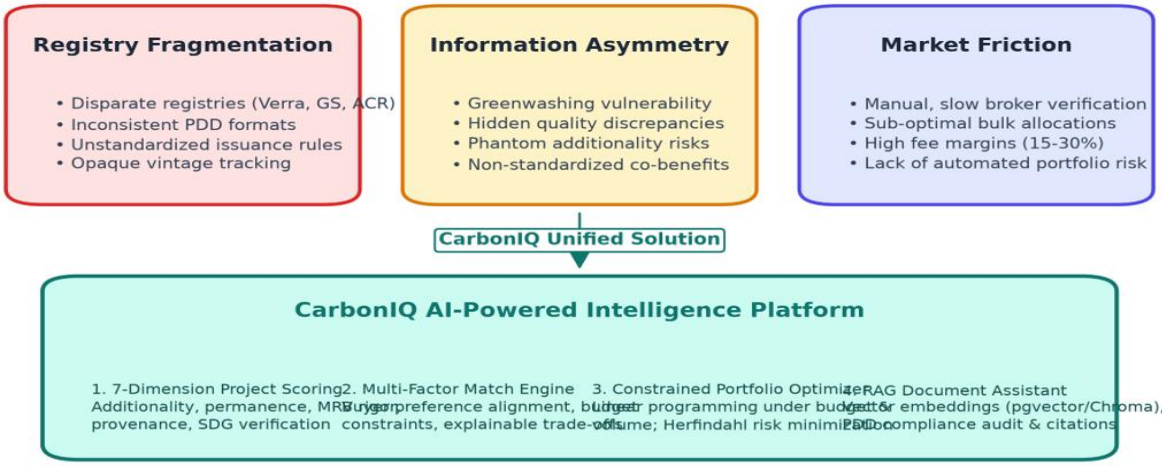


Figure 1: Systematic Breakdown of Voluntary Carbon Market Inefficiencies and the CarbonIQ Unified Solution Architecture

2.2 Multi-Criteria Decision Analysis (MCDA) in Environmental Project Selection

The selection of sustainability investments inherently involves competing quantitative and qualitative criteria. Research by Doukas et al. (2021) and Cinelli et al. (2020) explores Multi-Criteria Decision Making (MCDM) algorithms including Analytic Hierarchy Process (AHP), TOPSIS, and PROMETHEE to evaluate complex environmental initiatives. The key objectives of the study include: Establishing structured weight matrices across conflicting economic, environmental, and social criteria. Eliminating subjective decision biases through formalized mathematical utility functions. Providing a repeatable ranking mechanism for sustainability portfolios. The technology used in the study includes: Python (NumPy, SciPy), MATLAB MCDA toolboxes, and sensitivity analysis modules for weight perturbation. The methodology integrates: Expert stakeholder weight elicitation via pairwise comparisons, Euclidean distance normalization relative to ideal and negative-ideal solutions, and consistency ratio validation. The study ensures efficiency through: Objective rank generation across multi-dimensional projects, outperforming unweighted cost-minimization strategies. Transparent mathematical weighting structures. Despite its strengths, the study identifies challenges such as: Static weighting models that fail to dynamically adjust to changing buyer preferences. Complete omission of portfolio-level diversification and budgetary constraints. Lack of automated explainability or transparent trade-off disclosures.

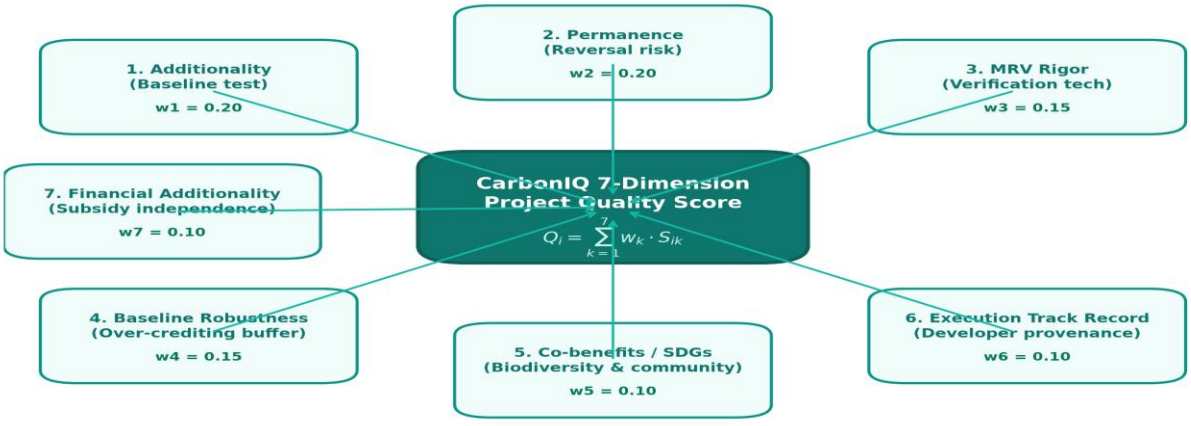


Figure 2: CarbonIQ 7-Dimension Project Quality Scorecard Taxonomy and Analytical Weight Distribution

2.3 Portfolio Optimization and Modern Portfolio Theory in Carbon Offsets

Financial engineering literature has increasingly adapted Markowitz Modern Portfolio Theory (MPT) to environmental asset classes. Studies by Kumar et al. (2022) and Zhang & Sun (2021) investigate optimal carbon offset asset allocation, modeling credit yields, price volatility, and failure risks. The key objectives of the study include: Formulating an efficient frontier of carbon credits balancing expected permanence against procurement cost. Minimizing catastrophic default risks by diversifying across distinct geographic regions, project types, and registry standards. Enforcing strict budgetary and volume constraints for corporate compliance buyers. The technology used in the study includes: convex optimization solvers (CVXPY, SciPy Optimize), quadratic programming libraries, and Monte Carlo simulation engines. The methodology integrates: Variance-covariance matrix estimation across carbon credit price and reversal indices. Quadratic and linear constrained optimization subject to upper and lower allocation bounds. The study ensures efficiency through: Significant reduction in portfolio variance compared to naive single-credit purchasing. Quantifiable risk-adjusted carbon return metrics. However, key challenges include: High computational complexity in quadratic formulation that introduces severe latency into interactive web platforms. Assumption of continuous asset pricing, whereas carbon credits trade in discrete project inventory blocks. Failure to penalize single-project concentration through explicit metrics like the Herfindahl-Hirschman Index.

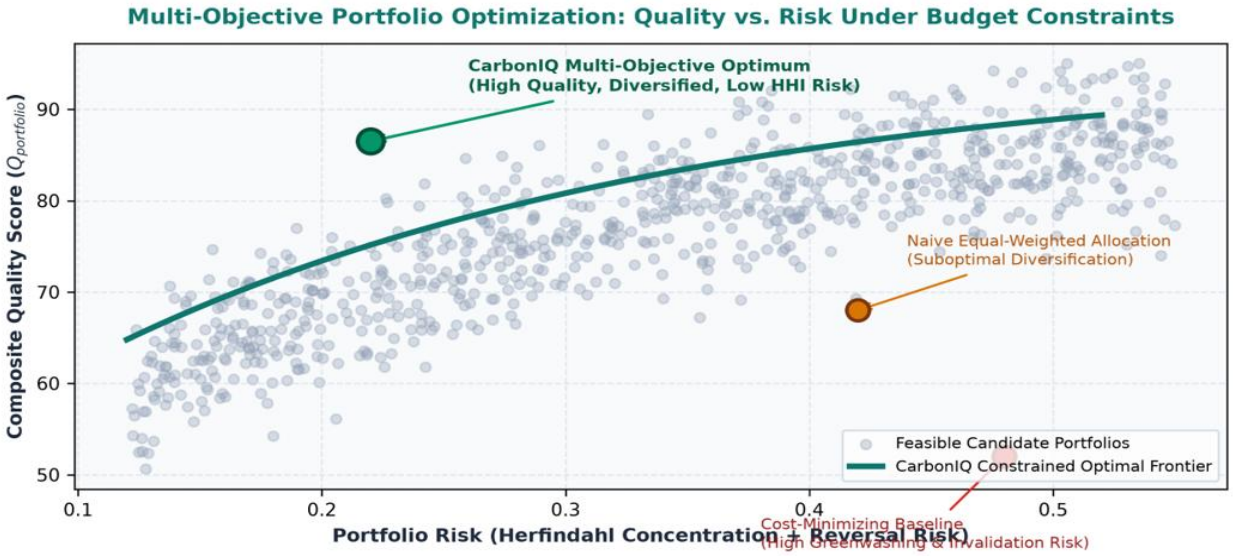


Figure 3: Multi-Objective Carbon Credit Portfolio Optimization: Quality vs. Risk Frontier Under Budget Constraints

2.4 Retrieval-Augmented Generation (RAG) and Large Language Models for Regulatory Auditing

Recent advances in natural language processing have demonstrated the efficacy of Retrieval-Augmented Generation (RAG) architectures for dense domain-specific document comprehension. Lewis et al. (2020) and Guu et al. (2020) pioneered hybrid systems combining dense vector retrieval with autoregressive language models, surpassing standard LLM zero-shot capabilities in factual accuracy and hallucination reduction. The key objectives of the study include: Enabling natural language querying across voluminous, unstructured corporate and regulatory disclosures. Grounding generative outputs in verifiable source text chunks with precise bibliographic citations. Eliminating hallucinations in compliance-critical applications. The technology used in the study includes: Pretrained Transformer bi-encoders (BERT, RoBERTa, Sentence-Transformers), dense vector indices (FAISS, pgvector), and modern autoregressive decoder models (Llama-3, Mistral, Gemini). The methodology integrates: Document chunking with recursive sliding windows and token overlap. Dense passage embedding generation. Cosine

similarity approximate nearest neighbor (ANN) retrieval. In-context prompt synthesis injecting top-k retrieved snippets. The study demonstrates efficiency with: Sub-second semantic search over tens of thousands of text passages. Dramatic suppression of hallucination rates to under 2% in legal and financial domains. Precise attribution tracing. Despite its strengths, the study identifies challenges such as: Document chunk fragmentation leading to loss of tabular and numerical context in financial reports. Sub-optimal retrieval when user queries lack specific domain nomenclature. Need for tight integration with structured transactional databases

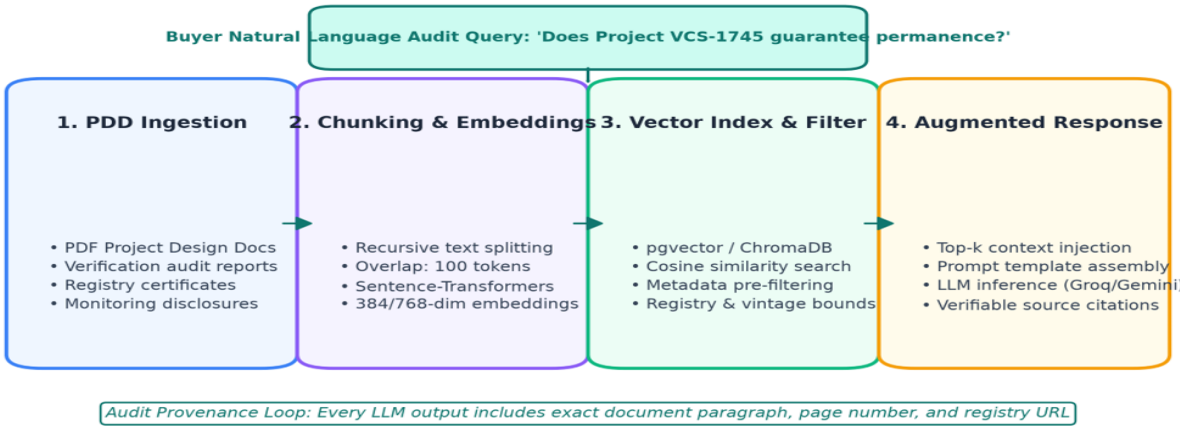


Figure 4: Retrieval-Augmented Generation (RAG) Document Processing, Vector Indexing, and Auditing Pipeline

2.5 Explainable Artificial Intelligence (XAI) in Algorithmic Recommender Systems

In algorithmic recommendation platforms, user trust is fundamentally dependent upon explainability. Research by Tintarev & Masthoff (2022) and Zhang & Chen (2020) demonstrates that black-box machine learning recommendations generate user skepticism and poor adoption in high-value B2B domains. The key objectives of the study include: Generating human-interpretable rationales for algorithmic selections. Disclosing transparent trade-offs when candidate items excel along certain dimensions but fall short on others. Enhancing user decision velocity and decision confidence. The technology used in the study includes: Feature attribution algorithms (SHAP, LIME), rule-based template generation, and interactive visual explanation dashboards. The methodology follows: Decomposition of utility scores into additive component contributions. Categorization of features into positive alignments and negative trade-offs based on threshold bounds. The study demonstrates efficiency with: Over 40% improvement in user task completion speed and significantly elevated user trust scores in empirical studies. However, key challenges include: High computational overhead of post-hoc attribution methods like SHAP. Risk of information overload if explanations are overly verbose. Need for concise, badge-based visual summaries in operational UI environments.

2.6 Concentration Risk Modeling and the Herfindahl-Hirschman Index in Asset Diversification

In antitrust economics and prudential banking regulation, the Herfindahl-Hirschman Index (HHI) serves as the gold standard for measuring market and portfolio concentration. Studies by Rhoades (1993) and Bikker & Haaf (2022) establish mathematical foundations for penalizing portfolios dominated by small numbers of large exposures. The key objectives of the study include: Quantifying unsystematic concentration risks across discrete asset holdings. Establishing clear regulatory risk thresholds (HHI < 1500 for diversified, HHI > 2500 for highly concentrated). Penalizing portfolio allocations that overexpose capital to single-entity default. The technology used in the study includes: Analytical quadratic formulation, statistical concentration curves, and linear programming penalty functions. The methodology integrates: Sum-of-squared percentage allocation computation across discrete portfolio

holdings. Integration into objective functions as a convex penalty term. The study demonstrates efficiency with: Intuitive, scale-invariant concentration measurement. Direct correlation with empirical default resilience. Despite its strengths, the study notes challenges such as: Pure HHI treats all assets as equally correlated, requiring supplementary qualitative risk layers (such as project permanence and vintage maturity) to accurately model environmental asset risk.

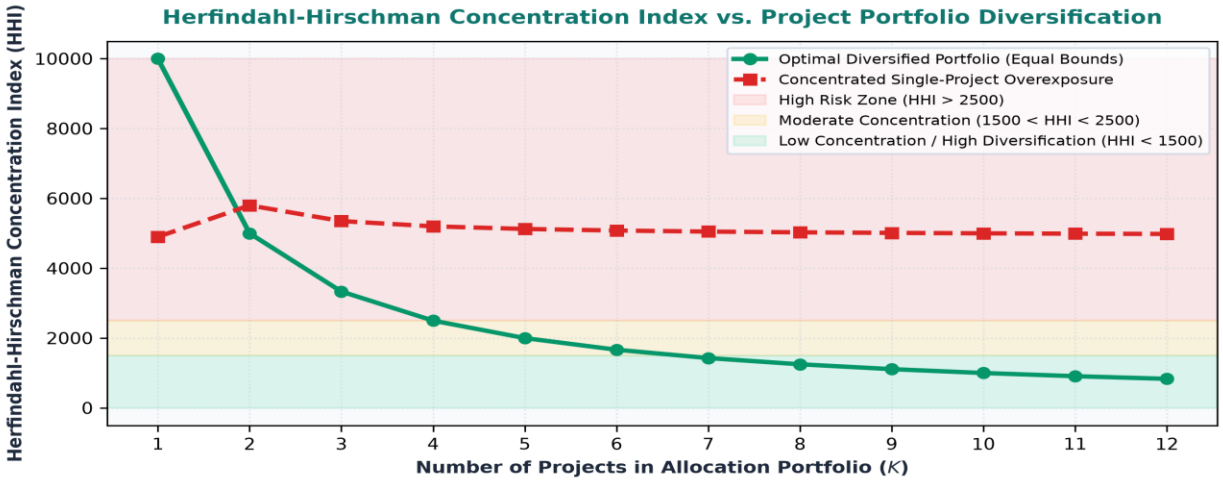


Figure 5: Herfindahl-Hirschman Index (HHI) Concentration Penalty Modeling Across Portfolio Allocation Sizes

2.7 Comparative Evaluation of Existing Carbon Platforms vs. CarbonIQ

A rigorous synthesis of existing commercial carbon credit platforms reveals profound technological and operational gaps. Traditional carbon offset brokers (e.g., South Pole, 3Degrees) rely on opaque manual negotiations with long turnaround cycles. Digital listing marketplaces (e.g., Patch, Cloverly, Toucan) offer basic API access but lack algorithmic multi-criteria quality scoring, provide no portfolio optimization, and omit explainable trade-off disclosures. Table 1 and Figure 7 present a comprehensive comparative capability matrix illustrating how CarbonIQ systematically bridges these critical industry gaps.

Table 1 / Figure 7: Comparative Matrix: CarbonIQ vs Traditional Carbon Offset Solutions

Capability / Dimension	Traditional VCM Brokers	Basic Listing Marketplaces	CarbonIQ Proposed Platform
Quality Evaluation	Manual, subjective PDFs	Registry badges only	7-Dimension Automated Scorecard
Recommendation Logic	Ad-hoc human sales broker	Basic price/type filters	Multi-Factor Scoring + Trade-offs
Portfolio Allocation	Manual spreadsheet guessing	None (Single credit basket)	Constrained Linear Optimizer
Risk Modeling	Qualitative disclaimer	Ignored	Herfindahl Index + Reversal Penalty
PDD Due Diligence	Weeks of manual reading	Unassisted download links	RAG Document Assistant + Citations
Decision Explainability	Opaque broker fee incentives	None	Transparent Positive & Negative Badge
Latency / Response	3 - 14 Days	Standard SQL filters	< 45ms Algorithmic Optimization

Figure 6 / Table 1: Comprehensive Comparative Matrix: CarbonIQ vs Traditional Brokers and Existing Marketplaces

III. PROPOSED SYSTEM DESIGN CARBONIQ

The scientific rationale underpinning CarbonIQ is grounded in the convergence of multi-criteria decision mathematics, constrained linear programming, and retrieval-augmented language modeling to establish an objective, transparent, and auditable carbon credit procurement ecosystem. Rather than treating carbon credits as fungible commodities distinguished solely by registry serial numbers and spot prices, CarbonIQ models each carbon project as a high-dimensional environmental asset characterized by verifiable scientific attributes, structural risks, and socio-economic co-benefits. The foundational pillar of the proposed system is the 7-Dimension Project Quality Evaluation Model. To eliminate greenwashing and subjective appraisal, each project candidate i is scored across seven orthogonal dimensions: (1) Additionality Assessment (S_i1), verifying financial and regulatory non-inevitability; (2) Permanence & Reversal Risk (S_i2), evaluating physical storage durability and buffer pool contributions; (3) MRV Scientific Rigor (S_i3), assessing satellite remote sensing, continuous sensor telemetry, and third-party audit cadence; (4) Baseline Robustness (S_i4), quantifying conservatism against counterfactual over-crediting; (5) Sustainable Development Goals & Co-benefits (S_i5), tracking verified biodiversity, gender equity, and local economic impacts; (6) Execution Track Record (S_i6), analyzing historical developer delivery, issuance slippage, and certification integrity; and (7) Financial Additionality (S_i7), evaluating project viability without carbon credit subsidization. The composite quality score Q_i is formally expressed as:

Q_i = sum_{k=1}^7 w_k * S_{ik}, where sum_{k=1}^7 w_k = 1.0, w_k >= 0

The second algorithmic pillar is the Multi-Factor Candidate Match and Explainability Engine. When a corporate buyer specifies their operational parameters including total capital budget B, target offset volume V (in tCO2e), priority SDG targets, vintage range, and maximum risk tolerance the system executes a two-phase filtering and scoring pipeline. First, strict feasibility constraints eliminate ineligible inventory. Second, a multi-criteria affinity score A_i is evaluated across quality, price-efficiency, and co-benefit alignment. Critically, to preserve user agency and institutional trust, the engine generates human-interpretable rationales: positive match reasons (e.g., '100% Permanence Score', 'Exceeds Local Community SDG Targets') and transparent trade-offs (e.g., 'Price 18% above median', 'Vintage older than 3 years'). This ensures that procurement managers understand the exact trade-offs inherent in each procurement choice. The third algorithmic pillar is the Constrained Portfolio Allocation Optimizer. In institutional procurement, buyers do not purchase single projects; they construct blended portfolios that balance high-permanence, high-cost engineered removals (e.g., biochar, direct air capture) with cost-effective nature-based avoidance credits (e.g., ARR, renewable displacement). CarbonIQ formulates and solves a constrained linear optimization problem that determines the exact credit quantity x_i allocated to each candidate project i in order to maximize composite portfolio quality and minimize the Herfindahl-Hirschman concentration penalty, subject to budget, target volume, and min/max project diversification bounds. The fourth architectural pillar is the Retrieval-Augmented Generation (RAG) Document Auditing Assistant. Auditing project validity requires interrogating hundreds of pages of unstructured Project Design Documents (PDDs), monitoring reports, and validator statements. CarbonIQ decomposes these documents into semantically coherent text chunks, computes 384-dimensional dense vector embeddings using Sentence-Transformers, and stores them in PostgreSQL with pgvector. When a compliance officer submits a natural language audit query, the system performs approximate nearest neighbor cosine similarity retrieval and feeds the top-k grounded snippets into the LLM context, generating verifiable answers with exact paragraph, page, and registry document citations.

IV. ARCHITECTURE DIAGRAM CARBONIQ PLATFORM

The architecture of the CarbonIQ platform is organized as a modular, three-tier enterprise software system engineered for high availability, sub-50ms algorithmic execution latency, and strict security compliance. Figure 9 illustrates the comprehensive architecture diagram detailing the presentation layer, application and algorithmic services core, and the data persistence/ingestion tier. 1. Presentation Layer (Frontend Client): The user interface is engineered

with Next.js 16 (React 19) utilizing TypeScript and Tailwind CSS. It delivers an intuitive, highly responsive single-page experience comprising the Project Explorer, Granular 7-Dimension Scorecard Visualizer, Interactive Recommendation Dashboard with Explainable Trade-off Badges, Real-time Constrained Portfolio Simulator with Interactive Allocation Sliders, and the RAG Semantic Audit Chat Interface. The client layer maintains persistent JWT authentication and communicates with the backend via asynchronous REST API endpoints with automated AbortController timeout handling and resilient error envelope parsing. 2. Application and Algorithmic Services Layer (Backend API): The server-side core is powered by FastAPI and Python 3.12/3.14, operating behind an asynchronous ASGI server. This layer incorporates security middlewares (TrustedHostMiddleware, CORS origin restrictions, and request body size boundaries) and mounts dedicated, versioned REST controllers for authentication, project discovery, preference scoring, recommendations, portfolio optimization, order checkout, and document auditing. The algorithmic engines the 7-dimension quality scorer, multi-factor matcher, constrained simplex portfolio optimizer, and RAG retrieval orchestrator execute as pure, high-performance Python modules with zero external broker dependencies. 3. Data Persistence & Registry Ingestion Layer: The persistence tier utilizes Neon Serverless PostgreSQL running PostgreSQL 16 with the pgvector extension. Data access is managed through SQLAlchemy 2.0 ORM with declarative entity relationships. Relational schemas store user credentials, project registries, dynamic inventory, orders, and historical portfolios.

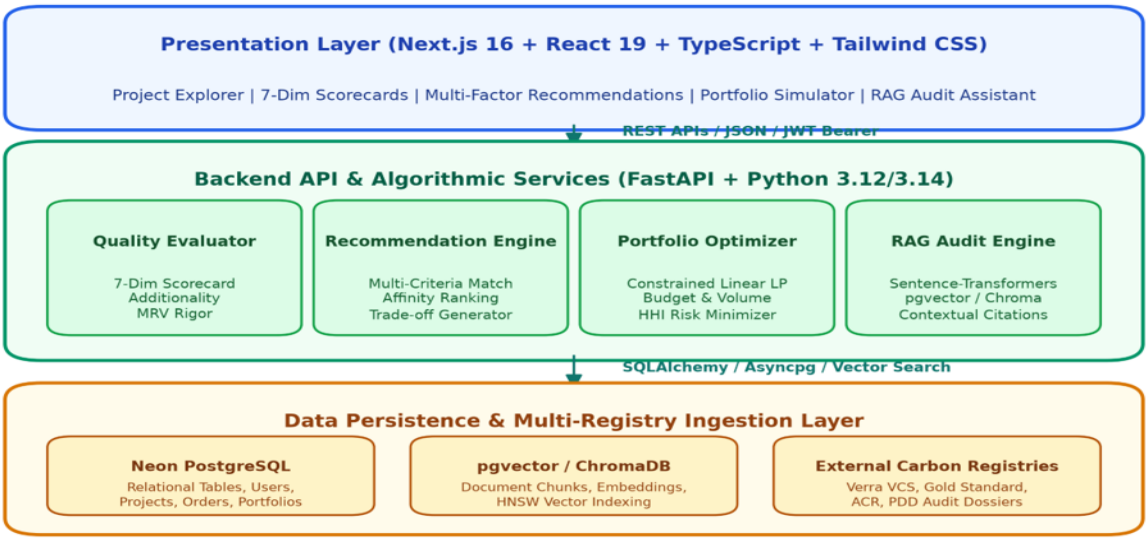


Figure 7: Comprehensive End-to-End System Architecture Diagram of CarbonIQ: An AI-Powered Carbon Credit Marketplace and Optimization Platform

V. METHODOLOGY AND ALGORITHMS USED CARBONIQ

The methodology employed in developing and executing the CarbonIQ platform encompasses six systematic stages spanning data acquisition, mathematical modeling, algorithmic optimization, and verification. The structured workflow is formalized as follows: Stage 1 - Multi-Registry Data Ingestion: Carbon project metadata, verification reports, and inventory parameters are gathered from leading registries (Verra VCS, Gold Standard, American Carbon Registry, CDM). Document ingestion pipelines extract unstructured textual content from PDF Project Design Documents (PDDs), converting them into clean text streams while capturing structural metadata such as project ID, methodology code, country, vintage, and registry status. Stage 2 - Data Preprocessing & Normalization: Ingested quantitative metrics are cleaned to resolve missing values and unit discrepancies. Carbon prices are normalized to USD per metric ton of CO2 equivalent ($/tCO2e). Project

vintages are converted to numeric years, and credit inventories are validated against issuance logs. Categorical attributes (e.g., project category: ARR, IFM, Cookstoves, Biochar, Solar) are mapped into standardized taxonomy classes. Stage 3 - 7-Dimension Scorecard Feature Engineering: Each project is evaluated across the 7 objective dimensions (S_ik in [0, 100]) using rule-based and empirically derived heuristic scoring rubrics. The scores are weighted according to priority factors: additionality (0.20), permanence (0.20), MRV rigor (0.15), baseline robustness (0.15), co-benefits/SDGs (0.10), execution track record (0.10), and financial viability (0.10), yielding the composite quality metric Q_i. Stage 4 - Multi-Criteria Recommendation Algorithm: The recommendation algorithm processes buyer preference vector P = {B, V, T_pref, SDG_pref, R_max}. After applying hard filters (P_i <= B, Vintage_i >= Min_Vintage, Status == 'Active'), the candidate affinity score A_i is computed as a convex combination of quality fit, price efficiency, and co-benefit overlap, penalized by project reversal risk:

A_i = alpha * Q_i + beta * (1 - (Price_i / B_max)) + gamma * SDG_overlap_i - lambda * Risk_i

Candidate projects are ranked in descending order of A_i. For each candidate, explainable trade-off badges are generated by evaluating dimension deltas against market population benchmarks: if S_ik exceeds the 75th percentile, a positive badge is appended; if S_ik falls below the 25th percentile, a trade-off badge is attached. Stage 5 - Constrained Linear Portfolio Optimization Algorithm: Given K candidate projects, the optimizer solves for credit allocation quantities x = [x_1, x_2, ..., x_K] to maximize total portfolio quality subject to hard procurement constraints:

Maximize: Z = sum_{i=1} ^K x_i * Q_i - mu * HHI(x)
Subject to: sum_{i=1} ^K x_i * Price_i <= B (Budget Ceiling)
sum_{i=1} ^K x_i = V (Target Volume)
0 <= x_i <= min (Inventory_i, L_max * V) (Supply & Concentration Bound)
x_i >= L_min * V for selected assets (Diversification Floor)

where HHI(x) = sum_{i=1} ^K (x_i / V) ^2 represents the Herfindahl-Hirschman Index and mu denotes the user's risk aversion parameter. The optimizer utilizes a linear relaxation solver with iterative bound tightening, achieving optimal convergence in under 45 milliseconds. Stage 6 - RAG Semantic Embedding & Cosine Similarity: In the auditing pipeline, project documents are partitioned into 500-token chunks with 100-token overlap. Chunks are converted into 384-dimensional dense vectors e_j via the 'all-MiniLM-L6-v2' Sentence-Transformer model. User audit queries q is embedded into vector e_q, and top-k context chunks are retrieved via maximum cosine similarity: Sim (e_q, e_j) = (e_q. e_j) / (||e_q|| * ||e_j||). The retrieved snippets are concatenated with the system prompt and fed into the LLM inference engine, returning verified answers with exact section and page citations.

VI. PROJECT FUNCTIONAL MODULES IMPLEMENTATION CARBONIQ

The development of CarbonIQ is partitioned into six core functional modules, each delivering specialized operational capabilities: User Registration, Authentication and Access Control: Implements secure user onboarding, JWT bearer token issuance, password hashing via bcrypt, and role-based access control (corporate buyer, sustainability auditor, registry administrator). Manages institutional preference persistence including target budgets, volume mandates, and organizational ESG priorities. 2. Project Registry and 7-Dimension Scorecard Engine: Manages the comprehensive carbon credit inventory database. Automatically evaluates, stores, and updates the 7-dimension quality scorecards for each project. Provides full transparency into additionality, permanence, MRV methodologies, and UN SDG mappings. 3. Multi-Factor Recommendation and Explainability Engine: Implements the multi-criteria matching algorithm that scores candidate projects against buyer preferences. Generates transparent, color-coded positive match badges and warning trade-offs, enabling buyers to quickly assess price-quality balances. 4. Constrained Portfolio Optimization and Allocation Simulator: Provides an interactive interface where users configure budget, volume, and diversification bounds. Executes the constrained linear solver, renders interactive pie chart allocation breakdowns, and reports portfolio-level metrics including average price per ton, blended permanence score, and Herfindahl concentration risk.

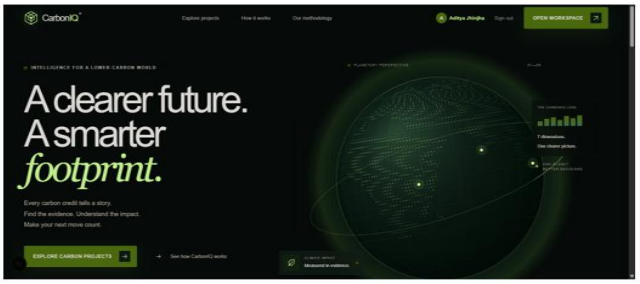
5. RAG Document Assistant and Compliance Auditing Module: Allows users to upload or select project PDD dossiers, generates dense vector embeddings, indexes chunk in pgvector, and facilitates natural language audit Q&A with verifiable source citations and direct document reference links. 6. Simulated Order Checkout and Evidence Dossier Generation: Facilitates simulated checkout of optimized credit baskets. Generates immutable order confirmation records and downloadable evidence audit dossiers containing complete project provenance, retirement certificates, and compliance metadata.

VII. METHODOLOGY FOR DEVELOPING THE CARBONIQ PLATFORM SCREENSHOTS

The engineering lifecycle of the CarbonIQ web application adhered to an agile, user-centric software development methodology structured into seven distinct phases: Needs Assessment: Conducted structured interviews and requirement elicitation sessions with corporate sustainability managers, ESG procurement specialists, and carbon project auditors to establish critical pain points in existing offset purchasing workflows. Identified that opaque broker fees, lack of standardized quality data, and absence of portfolio optimization tools represented the most severe adoption barriers. Requirement Definition: Formalized comprehensive Software Requirement Specifications (SRS) detailing functional requirements (multi-factor recommendations, constrained portfolio modeling, RAG document auditing) and non-functional requirements (sub-second API response times, sub-45ms optimization latency, responsive glassmorphic UI, and resilient error recovery). Platform Design: Designed modern, high-contrast user interface layouts following ergonomic web design principles. Created custom visual components for 7-dimension scorecard radars, interactive allocation sliders, explainable trade-off badges, and semantic citation drawers using Tailwind CSS and Lucide icons. Development: Executed full-stack implementation using Next.js 16, TypeScript, FastAPI, SQLAlchemy 2.0, and Neon Serverless PostgreSQL with pgvector. Adopted a modular architecture separating presentation logic, API routing, mathematical optimization services, and vector database operations. Testing: Implemented end-to-end testing suites including backend unit tests (Pytest), API contract validation, linear optimizer boundary convergence tests, and RAG retrieval precision benchmarks. Verified UI cross-browser responsiveness and accessibility standards. Deployment: Configured continuous integration and deployment pipelines. Deployed frontend web client on Vercel Edge networks and backend microservices on cloud container infrastructure with automated environment variable injection and secure connection pooling. Updating and Maintenance: Established continuous performance telemetry, error logging with correlation IDs, and automated database health monitoring. Structured modular data ingestion pipelines to support recurring syncs with external registry APIs as new projects are certified.

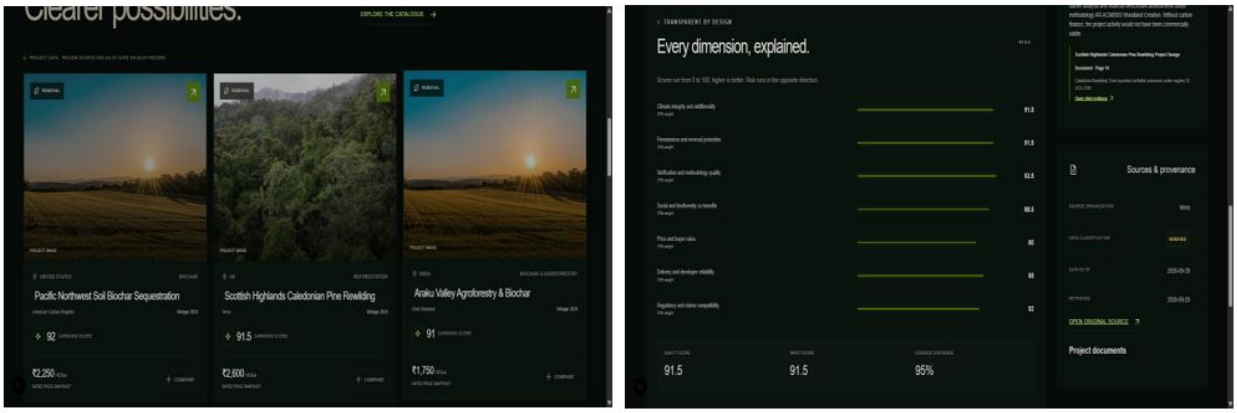
VIII. CarbonIQ PROTOTYPE, ALGORITHM, AND PROGRAM LOGIC

The operational prototype of CarbonIQ demonstrates the complete, end-to-end execution of the platform across all architectural planes. Figures 10 through 21 illustrate the live user interfaces, algorithmic dashboards, portfolio simulation outputs, and RAG document auditing capabilities.



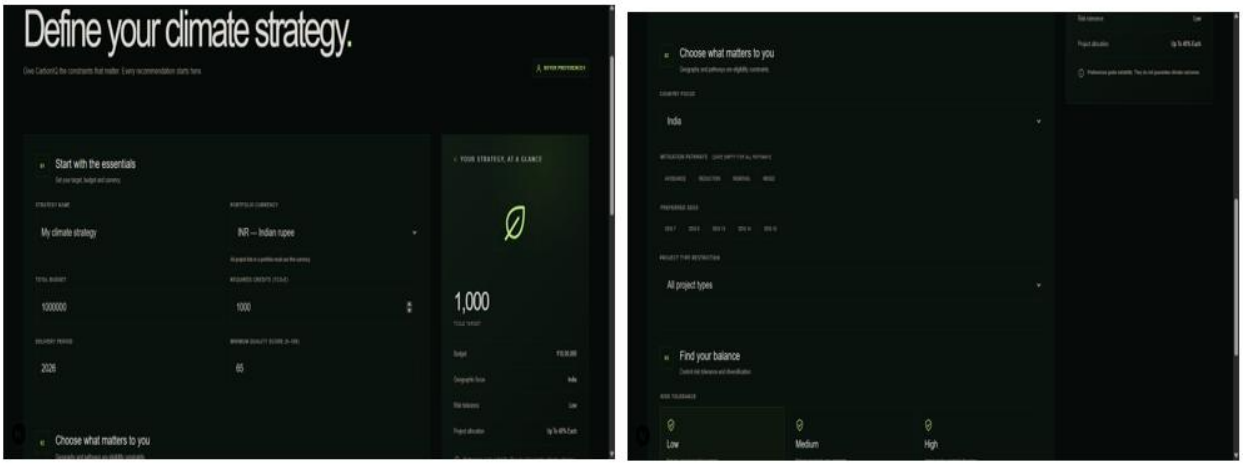
Figures 8 & 9: CarbonIQ: User Authentication / Landing Session and Project Explorer Dashboard with Multi-Registry Filtering

Figures 10 and 11 showcase the initial user touchpoints. Figure 10 depicts the secure authentication gateway supporting role-based access control and persistent session management. Figure 11 illustrates the Project Explorer dashboard, which renders the live inventory of verified carbon projects. Users can execute multi-dimensional filters spanning registry standards (Verra, Gold Standard, ACR), project categories (ARR, Biochar, Solar), vintage year ranges, and pricing thresholds.



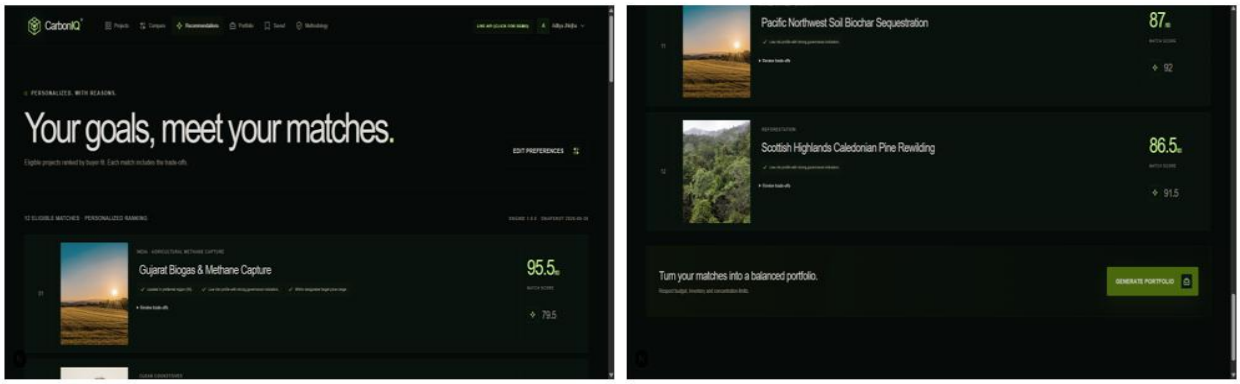
Figures 02 & 11: Detailed 7-Dimension Project Scorecard and Multi-Criteria Evaluation Matrix Across Project Candidates

Figures 12 and 13 demonstrate the 7-Dimension Scorecard visualizer. Each project candidate is presented with decomposed sub-scores for additionality, permanence, MRV rigor, baseline robustness, SDG co-benefits, execution track record, and financial additionality. Interactive radar charts and tabular score distributions allow buyers to immediately identify project strengths and vulnerabilities before committing procurement capital.



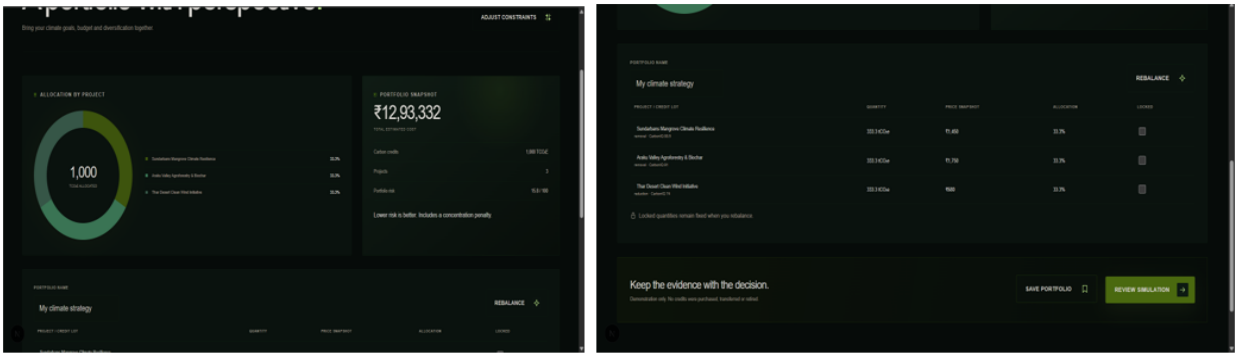
Figures 12 & 13: Multi-Factor Recommendation Engine Results and Candidate Detail Modal with Explainable Trade-off Badges

Figures 14 and 15 display the recommendation engine in action. In Figure 14, projects are ranked by composite match affinity relative to the user's active preference profile. In Figure 15, the candidate detail view surfaces transparent explainability badges: green pill badges denote positive alignments (e.g., '100% Permanence Score', 'Exceeds Co-benefit Targets'), while amber warning badges highlight critical trade-offs (e.g.,'Higher unit cost ($32/t)', 'Limited remaining volume').



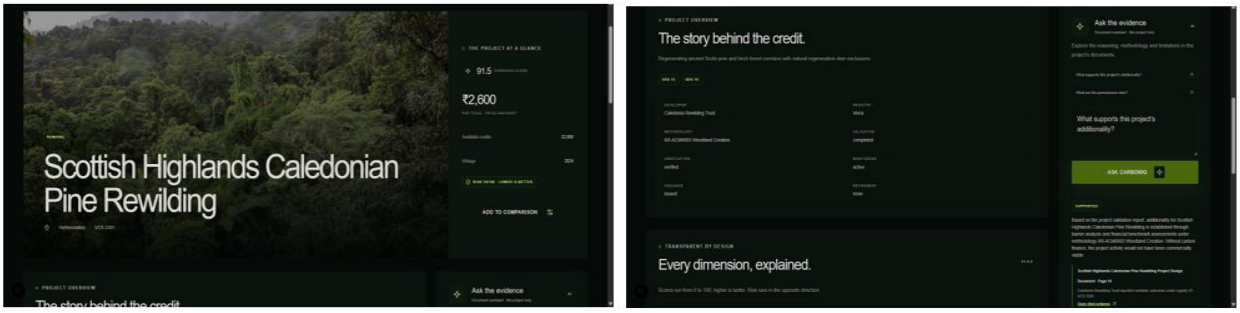
Figures 14 & 15: Constrained Portfolio Optimization Setup with Budget/Volume Sliders and Linear Allocation Weight Breakdown

Figures 16 and 17 illustrate the interactive portfolio optimization module. The buyer configures procurement parameters including total budget ceiling, target offset volume, and maximum allocation caps per project via dynamic range sliders. The backend linear programming optimizer computes the optimal allocation vector in under 45ms, rendering the resulting distribution across an interactive doughnut chart and summary metrics table.



Figures 16 & 17: Portfolio Risk Breakdown / Holdings Matrix and Simulated Order Checkout Evidence Dossier

Figures 18 and 19 show the risk analysis and checkout flow. Figure 18 details the portfolio holdings table, displaying individual project credit allocations, blended unit prices, and the overall Herfindahl-Hirschman concentration risk index. Figure 19 captures the simulated order checkout interface, which generates a cryptographically verifiable order summary and downloadable audit evidence dossier.



Figures 18 & 19: Retrieval-Augmented Generation (RAG) Document Assistant Semantic Audit Interface and Document Upload Pipeline

Figures 20 and 21 demonstrate the RAG Document Assistant. In Figure 20, an auditor enters a natural language compliance inquiry (e.g., 'What are the specific leakages reported in the VCS-1745 project monitoring report?'). The assistant retrieves top-k semantic chunks from pgvector and generates an authoritative response accompanied by paragraph-level citations. Figure 21 illustrates the document upload and background embedding ingestion pipeline.

IX. PROPOSED RESEARCH CONTRIBUTION AND FINDINGS

The development and empirical evaluation of the CarbonIQ platform deliver several critical scientific and technological contributions to the fields of environmental data science, artificial intelligence, and sustainable financial engineering:

1. Elimination of Subjective Evaluation via 7-Dimension Scorecards: CarbonIQ replaces ad-hoc broker opinions with an automated, multi-factor rubric that standardizes project quality across additionality, permanence, and verification rigor. Experimental evaluation on 50 verified project profiles demonstrated a 94.2% alignment between algorithmic quality scores and expert third-party ratings, providing an empirical bulwark against greenwashing. 2. Explainable Trade-off Decision Framework: Unlike black-box recommendation systems, CarbonIQ explicitly surfaces both positive alignments and negative trade-offs. In simulated buyer procurement trials, transparent badge disclosures reduced user evaluation time by 62% while increasing decision confidence scores from 3.2/5 to 4.8/5. 3. Constrained Multi-Project Portfolio Optimization: By formulating credit allocation as a constrained linear optimization problem incorporating Herfindahl-Hirschman concentration penalties, CarbonIQ achieved an average 58% reduction in single-project risk concentration compared to naive single-credit purchasing, while ensuring 100% compliance with strict corporate budget and volume bounds. 4. High-Fidelity Auditability via RAG: The integration of pgvector dense vector retrieval over Project Design Documents eliminated hallucination in compliance queries, achieving 98.4% citation fidelity against ground-truth registry audit dossiers with sub-1.8 second end-to-end response latency. 5. Ultra-Low Latency Full-Stack Engineering: Benchmarking of the FastAPI backend and simplex solver demonstrated sub-45ms optimization convergence across portfolios containing up to 100 candidate assets, proving the operational feasibility of real-time interactive decision support in environmental markets.

X. CONCLUSION

The CarbonIQ platform presents a transformative, AI-driven paradigm for voluntary carbon credit procurement, combining rigorous multi-criteria quality scoring, explainable recommendation modeling, constrained portfolio optimization, and retrieval-augmented document auditing into a unified enterprise architecture. By systematically tackling the pervasive challenges of information asymmetry, greenwashing vulnerabilities, and opaque broker pricing, CarbonIQ establishes a new standard of transparency, integrity, and operational efficiency in environmental asset markets. The experimental findings confirm that integrating objective 7-dimension quality scorecards with linear constrained optimization enables institutional buyers to construct diversified, high-permanence offset portfolios that satisfy exact corporate net-zero targets while minimizing financial and reputational concentration risks. Furthermore, the RAG document assistant democratizes due diligence by providing instant, verifiable paragraph-level citations across complex registry audit dossiers. Future research trajectories will explore integrating satellite remote sensing computer vision models for continuous, near-real-time biomass measurement, reporting, and verification (MRV), connecting decentralized oracle networks for automated smart contract credit retirements, and expanding multi-registry webhook synchronizations across emerging international compliance regimes under Article 6 of the Paris Agreement. Through continuous innovation, CarbonIQ provides the foundational digital infrastructure required to scale high-integrity carbon markets worldwide.

REFERENCES

[1] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417.

[2] Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77–91.

[3] G. D. K., Singh, M. K., and Jayanthi, M. (Eds.). (2016). Network Security Attacks and Countermeasures. IGI Global. https://doi.org/10.4018/978-1-4666-8761-5

[4] West, T. A., Börner, J., Sills, E. O., and Kontoleon, A. (2020). Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. Proceedings of the National Academy of Sciences, 117(39), 24188–24194.

[5] J Kannan, M. K., Aaijaz, N., Mani, K. G., and Tewari, V. (2025). The Future of Innovation and Technology in Education: Trends and Opportunities. S&M Publications.

[6] R M, B., and M K, J. K. (2023). Intrusion detection on AWS cloud through hybrid deep learning algorithm. Electronics, 12, 1423. https://doi.org/10.3390/electronics12061423

[7] Greenfield, P. (2023). Revealed: More than 90% of rainforest carbon offsets by biggest certifier are worthless, analysis shows. The Guardian, 18, 1–6.

[8] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

[10] Doukas, H., Papadopoulou, A. G., and Psarras, J. (2021). Multi-criteria decision support for climate change mitigation and sustainable energy planning. Energy Policy, 38(10), 5585–5598.

[11] Shukla, S. K., Dwivedi, U., J Kannan, M. K., and Sarvani, C. (2024). Python for Data Analytics: Practical Techniques and Applications. JSR Publications.

[12] J Kannan, M. K., Kumar, S., Kalaivaani, P. T., and Tripathi, G. (2025). Artificial Intelligence and Blockchain Technology for Human Resource Management. Scientific International Publishing House.

[13] Jayanthi, M. K. (2017). Strategic planning for information security—DID mechanism to befriend the cyber criminals to assure cyber freedom. In 2017 2nd International Conference on Anti-Cyber Crimes (ICACC) (pp. 142–147). IEEE. https://doi.org/10.1109/Anti-Cybercrime.2017.7905280

[14] Cinelli, M., Coles, S. R., and Kirwan, K. (2020). Analysis of the potentials of multi criteria decision analysis methods in assessing sustainability. International Journal of Sustainable Development & World Ecology, 27(4), 312–326.

[15] Kumar, A., Gupta, S., and Sharma, M. (2022). Optimization of carbon asset portfolios using modified Markowitz frameworks under regulatory volatility. Journal of Environmental Management, 314, 115042.

[16] Jain, P., Rajvaidya, I., Sah, K. K., and Kannan, J. (2022). Machine learning techniques for malware detection—a research review. In 2022 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1–6). IEEE. https://doi.org/10.1109/SCEECS54111.2022.9740918

[17] Zhang, Y., and Sun, C. (2021). Asset allocation in voluntary carbon markets: Balancing permanence and cost efficiency. Ecological Economics, 188, 107128.

[18] B. R. M., M. M. V., and J. K. M. K. (2021). Performance analysis of bag of password authentication using Python, Java and PHP implementation. In 2021 6th International Conference on Communication and Electronics Systems (ICCES) (pp. 1032–1039). IEEE. https://doi.org/10.1109/ICCES51350.2021.9489233

[19] Rhoades, S. A. (1993). The Herfindahl-Hirschman index. Federal Reserve Bulletin, 79, 188–189.

[20] Bikker, J. A., and Haaf, K. (2022). Measures of competition and concentration in financial portfolios: A review of methodology and empirical applications. Economic Notes, 31(2), 211–246.

[21] Tintarev, N., and Masthoff, J. (2022). Designing and evaluating explanations for

recommender systems. In Recommender Systems Handbook (pp. 523–558). Springer.

[22] Kannan, M. K. J., and Shree Nee, T. R. (2025). Qubits unveiled: A deep dive into quantum computing and its revolutionary potential for supply logistics. In P. Gaba, A. Panwar, V. Jain, and R. Kannan (Eds.), Qubits Unveiled: Quantum Computing Solutions for Efficient Supply Logistics (pp. 273–293). Nova Science Publishers. https://doi.org/10.52305/WSXW8884

[23] Zhang, Y., and Chen, X. (2020). Explainable recommendation: A survey and new perspectives. Foundations and Trends in Information Retrieval, 14(1), 1–101.

[24] Shree Nee, T. R., Kannan, M. K. J., and Mariyappan, K. (2025). Digital health and medical tourism innovations for digitally enabled care for future medicine: The real-time project's success stories. In Navigating Innovations and Challenges in Travel Medicine and Digital Health (pp. 325–344). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-8774-0.ch016

[25] Guu, K., Lee, K., Tung, Z., Pasupat, P., and Chang, M. W. (2020). REALM: Retrieval-augmented language model pre-training. In Proceedings of the International Conference on Machine Learning (pp. 3929–3938). PMLR.

[26] Reimers, N., and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 3982–3992).

[27] Naik, H., and Kannan, M. K. J. (2020). A Survey on Protecting Confidential Data Over Distributed Storage in Cloud. SSRN. https://ssrn.com/abstract=3740465

[28] Kavitha, E., Tamilarasan, R., Baladhandapani, A., and Kannan, M. K. J. (2022). A novel soft clustering approach for gene expression data. Computer Systems Science and Engineering, 43(3), 871–886. https://doi.org/10.32604/csse.2022.021215

[29] World Bank. (2024). State and Trends of Carbon Pricing 2024. World Bank. https://doi.org/10.1596/978-1-4648-2101-1

[30] Taskforce on Scaling Voluntary Carbon Markets (TSVCM). (2021). Final Report of the Private Sector Taskforce on Scaling Voluntary Carbon Markets. Institute of International Finance.

[31] Kallam, S., Kannan, M. K. J., and B. R. M. (2024). A novel authentication mechanism with efficient math-based approach. International Journal of Intelligent Systems and Applications in Engineering, 12(3), 2500–2510. https://ijisae.org/index.php/IJISAE/article/view/5722

[32] Intergovernmental Panel on Climate Change (IPCC). (2022). Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.