

---

# GREENIQ: A DEEP SEARCH PLATFORM FOR COMPREHENSIVE CARBON MARKET ANALYSIS AND AUTOMATED REPORT GENERATION

---

**Oluwole Fagbohun**  
Carbonnote,  
USA

**Sai Yashwanth**  
Vuhosi Limited,  
UK

**Akinyemi Sadeeq Akintola**  
Universidade NOVA,  
Portugal

**Ifeoluwa Wurola**  
University of Hull,  
UK

**Lanre Shittu**  
Vuhosi Limited,  
UK

**Aniema Inyang**  
Readrly Limited,  
UK

**Oluwatimilehin Odubola**  
ResilStudio Ltd,  
UK

**Udodirim Offia**  
Readrly Limited,  
UK

**Said Olanrewaju**  
Cnergy Capital Ltd.,  
UK

**Ogidan Toluwaleke**  
Vuhosi Limited,  
UK

**Ilemona Abutu**  
Carbonnote,  
USA

**Taiwo Akinbolaji**  
Silveredge Consults,  
UK

## ABSTRACT

This study introduces GreenIQ, an AI-powered deep search platform designed to revolutionise carbon market intelligence through autonomous analysis and automated report generation. Carbon markets operate across diverse regulatory landscapes, generating vast amounts of heterogeneous data from policy documents, industry reports, academic literature, and real-time trading platforms. Traditional research approaches remain labour-intensive, slow, and difficult to scale. GreenIQ addresses these limitations through a multi-agent architecture powered by Large Language Models (LLMs), integrating five specialised AI agents: a Main Researcher Agent for intelligent information retrieval, a Report Writing Agent for structured synthesis, a Final Reviewer Agent for accuracy verification, a Data Visualisation Agent for enhanced interpretability, and a Translator Agent for multilingual adaptation. The system achieves seamless integration of structured and unstructured information with AI-driven citation verification, ensuring high transparency and reliability. GreenIQ delivers a 99.2% reduction in processing time and a 99.7% cost reduction compared to traditional research methodologies. A novel AI persona-based evaluation framework involving 16 domain-specific AI personas highlights its superior cross-jurisdictional analytical capabilities and regulatory insight generation. GreenIQ sets new standards in AI-driven research synthesis, policy analysis, and sustainability finance by streamlining carbon market research. It offers an efficient and scalable framework for environmental and financial intelligence, enabling more accurate, timely, and cost-effective decision-making in complex regulatory landscapes.

**Keywords** Multi-agent systems · Carbon markets · Automated report generation · Large Language Models · AI-powered research

---

**NOTE: Preprint, Submitted to ICMLT 2025, Helsinki, Finland**

Oluwole Fagbohun is with Carbonnote, Maryland, USA (e-mail: wole@carbonnote.ai). Sai Yashwanth is with Vuhosi Limited, London, United Kingdom (e-mail: taddishetty34@gmail.com). Akinyemi Sadeeq Akintola is with Universidade NOVA de Lisboa, Lisbon, Portugal (e-mail: sadeeq2@gmail.com). Ifeoluwa Wurola is with the University of Hull, Hull, United Kingdom (e-mail: wuraolaifeoluwa@gmail.com). Lanre Shittu is with Vuhosi Limited, London, United Kingdom (e-mail: hr.lanreshittu@gmail.com). Aniema Inyang is with Readrly Limited, London, United

Kingdom (e-mail: annywillow@gmail.com). Oluwatimilehin Odubola is with ResilStudio Ltd, London, United Kingdom (e-mail: timi@resilstudio.com). Udodirim Offia is with Readrly Limited, London, United Kingdom (e-mail: udodirim.offia@gmail.com). Said Olanrewaju is with Cnergy Capital Limited, London, United Kingdom (e-mail: said.olar@cnergyfund.com). Ogidan Toluwaleke is with Vuhosi Limited, London, United Kingdom (e-mail: gbemilekeogidan@gmail.com). Ilemona Abutu is with Carbonnote, Maryland, USA (e-mail: abutu.mona@gmail.com). Taiwo Akinbolaji is with Silveredge Consults, London, United Kingdom (e-mail: taiwoakinbolaji@gmail.com).

---

## 1 Introduction

The global carbon market has evolved significantly over the past two decades, transitioning from conceptual frameworks to complex, multi-jurisdictional trading systems with profound economic and environmental implications [1, 2]. This evolution has coincided with an increasingly urgent climate crisis, as evidenced by rising global temperatures and extreme weather events [3], prompting governments worldwide to establish more ambitious regulatory frameworks and market-based mechanisms to reduce greenhouse gas emissions [4]. As these markets mature and expand, with the global carbon market value projected to reach \$2.4 trillion by 2027 [5] the analytical challenges confronting stakeholders have grown exponentially.

Carbon markets operate at the intersection of environmental science, economics, policy, and technology, generating heterogeneous data across disparate sources including regulatory documentation, academic literature, industry reports, and real-time trading platforms. For policymakers, investors, corporations, and researchers, the ability to synthesise and interpret this information efficiently has become a critical determinant of effective decision making [6]. However, conventional approaches to carbon market analysis frequently prove inadequate and are characterised by labour-intensive manual research processes, significant time lags between market developments and analytical outputs, and inherent limitations in information processing capacity [7, 8].

The information asymmetry resulting from these inefficiencies has substantial consequences. Regulatory bodies risk implementing suboptimal policy frameworks without comprehensive market intelligence [9]. Investors may miss strategic opportunities or misallocate capital because of incomplete market understanding [10]. Corporations navigating compliance obligations or voluntary carbon commitments face uncertainty regarding market dynamics and regulatory trajectories [11]. Academic researchers may struggle to incorporate the latest market developments into their theoretical frameworks and empirical analyses [12].

Recent advancements in artificial intelligence, particularly in Large Language Models (LLMs), present promising opportunities to address these challenges [13, 14]. LLMs have demonstrated remarkable capabilities in natural language understanding, information synthesis, and content generation across various domains [15, 16]. However, their application to specialised fields, such as carbon market analysis, has remained limited, with existing solutions typically focusing on narrow aspects of the analytical process rather than offering comprehensive end-to-end solutions [17, 18]. This study introduces GreenIQ, a novel deep search platform designed to transform carbon market intelligence through autonomous, comprehensive analysis, and report generation. GreenIQ leverages a sophisticated multiagent architecture powered by state-of-the-art LLMs, where each agent is engineered to perform specialised functions within an integrated analytical pipeline. The platform autonomously processes diverse data sources, including academic literature, regulatory documentation, industry reports, and real-time market data, generating detailed citation-rich reports tailored to specific stakeholder requirements.

The multi-agent framework underpinning GreenIQ comprises five distinct components:

1. A Main Researcher Agent responsible for comprehensive information gathering, trend identification, and source verification;
2. A Report Writing Agent that converts research findings into coherent, structured narratives;
3. A Final Reviewer Agent that ensures accuracy, consistency, and adherence to reporting standards;
4. A Data Visualization Agent that transforms complex data into accessible visualisations; and
5. A Translator Agent that enables multilingual report dissemination while preserving technical accuracy and contextual nuances.

Our contributions to the field are fourfold. First, we demonstrate the effective integration of advanced Large Language Models (LLMs) for specialised domain analysis, establishing a methodological framework for applying these technologies to complex, data-intensive research challenges. Second, we introduce a novel multi-agent architecture that enables autonomous, end-to-end research and reporting, significantly reducing the time and resources required for comprehensive carbon market analysis. Third, we provide empirical evidence of our approach’s efficiency, accuracy, and cost-effectiveness through rigorous comparative evaluation against traditional manual research methodologies. Fourth, we introduce a novel AI persona-based evaluation framework, employing 16 domain-specific AI personas to assess analytical capabilities, citation quality, and regulatory insight generation in real-world reports written by humans.

To ensure a comprehensive performance assessment, these AI personas were designed to mimic experts across various industries, enabling a consistent and scalable evaluation without the logistical constraints of assembling large expert panels. This innovative approach provides a robust and replicable methodology for evaluating AI-driven research frameworks from diverse stakeholder perspectives.

The remainder of this paper is organised as follows: Section 2 reviews relevant literature on carbon market analysis, automated research methodologies, and LLM applications in specialised domains. Section 3 details GreenIQ system architecture including the technical specifications and operational dynamics of each agent. Section 4 presents our evaluation methodology and results. Section 5 discusses the broader implications, limitations, and future research directions, followed by concluding remarks in Section 6.

## 2 Related Work

### 2.1 Traditional Carbon Market Analysis

Carbon market analysis has traditionally relied on manual research methodologies conducted by domain experts, who collect, synthesise, and interpret data from diverse sources. These processes typically involve consultancies, research institutions, and specialised market intelligence firms that produce periodic reports through labour-intensive workflows. Governmental bodies such as the World Bank and international organisations such as the International Carbon Action Partnership (ICAP) publish annual or biannual comprehensive assessments that serve as industry benchmarks [19]. While these reports offer valuable insights, they are characterised by significant time lags between market developments and published analyses, often becoming outdated shortly after release.

Commercial platforms such as ICIS Carbon, Bloomberg New Energy Finance, and S&P Global Platts provide more frequent market updates and analytical tools. However, these solutions generally focus on quantitative metrics and trading data rather than on comprehensive qualitative analyses of regulatory developments, academic research, and emerging trends. Additionally, McKinsey [20] notes that these platforms typically require substantial human intervention to contextualise data and generate actionable insights, limiting their scalability and responsiveness to rapidly evolving market conditions.

Narassimhan et al. [21] highlight that existing analytical approaches struggle with the heterogeneity of carbon market data, which spans structured trading information, semi-structured regulatory documents, and unstructured textual sources such as news articles and research papers. This diversity presents significant challenges for traditional analytical methodologies, particularly when addressing the interconnected nature of global carbon markets across jurisdictions, sectors, and time horizons [22].

### 2.2 AI in Environmental and Financial Analysis

The application of artificial intelligence to environmental and financial analysis has expanded considerably in recent years. Machine learning techniques have demonstrated efficacy in various environmental domains, including climate prediction, emissions forecasting, and renewable energy optimisation [23]. In financial contexts, AI systems routinely analyse market trends, assess risk factors, and generate trading signals [24], with increasing attention paid to climate-related financial considerations [25]. Rolnick et al. [23] surveyed machine learning applications across climate change mitigation and adaptation strategies. In financial contexts, Lewis and Young [26] demonstrated the effectiveness of NLP techniques for analysing sentiment in regulatory filings and assessing their market impacts.

The emergence of Large Language Models (LLMs) has significantly advanced research synthesis and knowledge integration capabilities. Recent studies have explored LLM applications in regulatory compliance analysis [27, 28], and evidence synthesis for policy development [29]. However, as noted by Bommasani et al. [14], these technologies face challenges when applied to highly specialised domains with technical terminology and complex conceptual frameworks. Carbon markets represent one such domain, requiring sophisticated understanding of environmental economics, regulatory frameworks, and market mechanisms [30].

### 2.3 Multi-Agent Systems for Research and Automated Information Retrieval Systems

Multi-Agent Systems (MAS) have become pivotal in modelling complex interactions within environmental and financial domains. In environmental research, MAS facilitate the simulation of ecological systems and the assessment of climate policies. For instance, research has demonstrated the application of MAS in modelling climate change scenarios, providing insights into mitigation and adaptation strategies [31]. By simulating the interactions of various stakeholders, including governments, corporations, and consumers, MAS offer predictive capabilities that can support policy development and regulatory compliance. These simulations help identify the long-term impact of different climate policies, ensuring that decision-makers are equipped with robust, data-driven insights. In the financial sector, MAS have been utilised to predict global financial crises by simulating interactions among diverse market participants,

thereby identifying systemic risks [32]. The ability of MAS to model complex financial ecosystems allows regulators and analysts to test market stability under various economic conditions. For example, by simulating speculative bubbles and liquidity constraints, MAS frameworks can anticipate vulnerabilities in financial systems and inform proactive regulatory interventions. These applications underscore MAS’s capability to model complex systems, enhance decision-making, and provide scalable solutions across various sectors.

In the field of Information Retrieval (IR) systems, LLMs enhance IR by improving query understanding and document retrieval accuracy, allowing researchers to process large volumes of structured and unstructured data efficiently. Recent studies have explored the synergy between LLMs and IR, highlighting the potential to build autonomous agents that fully automate the information acquisition process [33]. By leveraging retrieval-augmented generation (RAG), LLM-enhanced IR systems can dynamically retrieve relevant documents, synthesise knowledge, and generate context-aware responses, reducing the need for manual information curation. Additionally, LLMs have been employed to evaluate IR systems, demonstrating their capability to assess retrieval performance effectively [34]. Traditional IR systems relied on precision-recall metrics, but LLM-based evaluation frameworks now enable context-aware relevance assessments, refining search accuracy. These advancements have transformed traditional IR systems, making them more adaptive and capable of handling complex, multi-faceted queries. As a result, automated IR systems have become essential for regulatory analysis, financial oversight, and climate policy research, enhancing research efficiency and effectiveness in high-stakes decision-making environments.

## 2.4 Gap Analysis

Despite the increasing sophistication of carbon market intelligence tools, existing research methodologies continue to exhibit critical limitations that hinder their effectiveness in a rapidly evolving regulatory landscape. Traditional expert-driven approaches, while historically valuable, remain constrained by laborious manual workflows, significant time delays, and inherent scalability issues. Analysts must painstakingly sift through disparate regulatory documents, financial reports, and scientific publications—an effort that not only consumes substantial resources but also results in insights that risk obsolescence by the time they are disseminated. This bottleneck is particularly problematic in the context of carbon markets, where policy changes, trading dynamics, and compliance frameworks evolve at an accelerated pace [35]. Meanwhile, commercial research platforms, though offering incremental efficiency gains, often focus narrowly on structured data—emissions figures, price movements, or transactional records—neglecting the broader qualitative context that drives market behaviours [19]. Even when artificial intelligence is deployed within existing platforms, its role is largely limited to basic keyword extraction or sentiment analysis, failing to capture the nuanced interplay of financial, regulatory, and environmental variables that shape the carbon market ecosystem [36]. Moreover, voluntary carbon markets continue to face significant challenges regarding data reliability, uncertainty in carbon credit valuations, and the standardisation of offset mechanisms, further exacerbating inefficiencies in traditional carbon market intelligence methods [37].

The recent emergence of Large Language Models (LLMs) and Multi-Agent Systems (MAS) has opened new possibilities for automating carbon market research, yet current AI-driven frameworks suffer from several unresolved challenges. Many LLM-based tools, while excelling at textual synthesis, struggle with fact verification and citation integrity, occasionally generating plausible yet misleading narratives—an issue that regulatory analysts and financial decision-makers cannot afford [38]. Furthermore, existing AI-driven solutions lack the necessary modularity to adapt seamlessly to multi-jurisdictional carbon pricing mechanisms, which vary significantly in scope, governance, and economic impact [27]. In the absence of a robust multi-agent coordination framework, these systems often operate in silos, failing to integrate real-time trading data with evolving policy regulations in a cohesive manner [32]. GreenIQ directly addresses these gaps by implementing a deep search platform that not only automates the end-to-end research pipeline but also embeds citation verification, jurisdictional adaptability, and stakeholder-specific contextualisation into its design. By leveraging a five-agent architecture—each specialising in a distinct research function—GreenIQ ensures that carbon market intelligence is not only comprehensive and timely but also rigorously validated, democratising access to high-quality, AI-driven sustainability analysis.

## 3 GreenIQ Architecture and Methodology

### 3.1 System Overview

GreenIQ is an advanced deep search platform designed to automate carbon market intelligence and reporting through an integrated multi-agent system. The architecture follows a modular design as shown in figure 1, where each specialised agent contributes to a structured, end-to-end workflow from data collection to multilingual reporting.

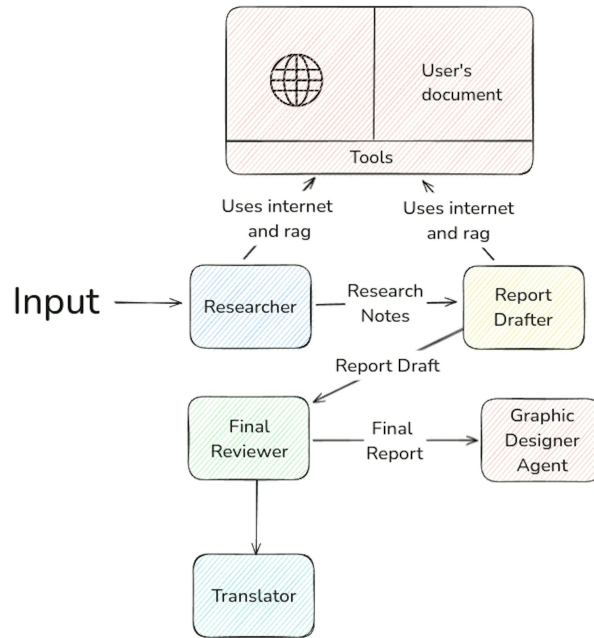


Figure 1: High-level architecture diagram of the GreenIQ system, illustrating the modular design and data flow between the five autonomous agents.

A key feature of GreenIQ is its deep search functionality, which integrates data from diverse sources, including regulatory repositories, academic literature, industry reports, real-time market data, and user-provided proprietary datasets. The data ingestion pipeline is structured into multiple stages, beginning with source identification and validation, followed by structured extraction using tailored parsers for different document types. The semantic analysis layer plays a crucial role in refining retrieved data by leveraging Named Entity Recognition (NER), topic modelling, and relationship extraction techniques. Additionally, iterative query refinement ensures that the system continuously improves its accuracy and relevance by dynamically adjusting search parameters based on user needs. Users can also define specific date ranges to retrieve only time-sensitive market insights, improving the precision of regulatory and financial intelligence. This streamlined pipeline enables the rapid generation of structured, accurate, and multilingual reports that would traditionally require extensive manual research and expert review.

### 3.2 Multi-Agent Framework

GreenIQ employs a robust multi-agent architecture, where each agent is engineered with specialised tasks, ensuring efficient automation of the entire research and reporting process.

#### 3.2.1 Main Researcher Agent

The Main Researcher Agent is responsible for comprehensive information gathering. It integrates APIs, web scraping tools, and retrieval-augmented generation (RAG) models to extract relevant data from authoritative sources such as regulatory bodies, academic journals, and financial institutions. The agent systematically analysis source credibility, ensuring all extracted data aligns with industry standards. The output from this agent is a structured research report, complete with citations and annotations.

#### 3.2.2 Report Agent

The Report Agent synthesises findings from the research report, structuring them into an engaging and logically coherent narrative. By utilising content organisation strategies, it adapts report structures for different stakeholders, such as policymakers, investors, and sustainability analysts. The agent dynamically adjusts language complexity, ensuring clarity for technical and non-technical audiences alike.

### 3.2.3 Final Reviewer Agent

To uphold quality control standards, the Final Reviewer Agent performs meticulous fact-checking, citation validation, and linguistic refinement. This agent ensures that every claim is backed by a verifiable source, citations comply with academic and industry norms, and the overall report adheres to professional writing standards. It also detects potential biases or inconsistencies, mitigating risks associated with automated content generation.

### 3.2.4 Data Visualisation Agent

The Data Visualisation Agent converts complex data insights into automated charts, graphs, and infographics. By leveraging Plotly and custom-built visualisation libraries, this agent enhances the interpretability of carbon market data. It dynamically selects the most appropriate visual formats-such as bar charts, heatmaps, and network graphs-based on the underlying dataset.

### 3.2.5 Translator Agent

To ensure accessibility across different markets, the Translator Agent adapts the report into multiple languages. Beyond direct translation, it employs contextual adaptation techniques to account for regional terminologies and regulatory nuances. This guarantees that translated versions maintain technical precision while being culturally and linguistically appropriate.

## 3.3 Technical Implementation

GreenIQ's technical infrastructure is powered by a combination of state-of-the-art LLMs, retrieval-enhanced search tools, and custom-built AI agents, designed for seamless interoperability. The Main Researcher Agent utilises GPT-4o as its core model, integrating an internet search tool for real-time data retrieval, a retrieval-augmented generation (RAG) mechanism for incorporating user-provided datasets, and a file writing tool to document findings systematically. The Report Drafter Agent and the Final Reviewer Agent are built on GPT-4o – *mini*, specialising in content organisation, linguistic refinement, and consistency checks. The Data Visualisation Designer employs GPT-4o in conjunction with Plotly-based dynamic visualisation tools to create meaningful graphical representations of complex data. The Translator Agent, powered by GPT-4o-mini, is equipped with file writing and multilingual adaptation tools, ensuring reports are accurately localised for diverse audiences.

The system follows a linear processing pipeline, where each agent's output becomes the next agent's input. The workflow begins with the Main Researcher Agent, which compiles structured research reports by integrating information from regulatory and academic sources. The Report Drafter Agent then transforms these raw findings into a structured report, refining narratives to enhance coherence and clarity. The Final Reviewer Agent verifies the accuracy of the content, ensuring citation integrity and compliance with academic and industry standards. Once the textual content is validated, the Data Visualisation Designer generates graphical representations of key insights, improving the accessibility and impact of the report. Finally, the Translator Agent adapts the document for international stakeholders, maintaining linguistic precision and cultural relevance. Each agent operates autonomously within this tightly integrated workflow, ensuring efficiency, accuracy, and adaptability. Extensive testing has confirmed the optimal performance of these models, with seamless interoperability across all agents, allowing GreenIQ to function as a scalable and reliable deep search platform for carbon market intelligence.

## 4 Evaluation and Results

### 4.1 Evaluation Methodology

Our evaluation methodology was designed to rigorously assess GreenIQ performance across multiple dimensions relevant to carbon market analysis and reporting. We established a comprehensive framework that combines structured scoring criteria with a novel AI persona evaluation approach.

#### 4.1.1 Scoring Framework

The evaluation of reports generated by GreenIQ (Report 1), and baseline human reports (Report 2) employed four key criteria, each scored on a 25 -point scale for a total possible score of 100 points per report:

1. Source Coverage ( 25 points)

- Strengths: High scores indicate diverse source utilisation (regulatory repositories, academic literature, industry reports, market data), providing comprehensive topic coverage with both primary and secondary sources.
- Weaknesses: Low scores result from limited source variety, insufficient information depth, or failure to reference key regulatory or academic sources.

2. Data Accuracy ( 25 points)

- Strengths: High scores reflect current properly sourced data free from errors, with clear methodologies for collection and analysis.
- Weaknesses: Low scores indicate outdated information, improper sourcing, errors in data presentation, unsupported claims, or lack of statistical context.

3. Citation Quality (25 points)

- Strengths: High scores indicate citations from authoritative, reputable sources, including peerreviewed articles, government publications, and recognised industry reports.
- Weaknesses: Low scores reflect missing citations, reliance on unreliable sources, nonauthoritative references, or complete lack of citations.

4. Report Coherence ( 25 points)

- Strengths: High scores indicate well organised reports with a clear introduction, body, and conclusion featuring logical narrative flow and easy-to-follow argumentation.
- Weaknesses: Low scores result from poor organisation, unclear transitions, disconnected topics, or incoherent argumentation.

#### 4.1.2 AI Persona Evaluation Framework

To evaluate GreenIQ’s performance across diverse stakeholder perspectives comprehensively, we developed 16 AI personas designed to emulate domain experts from various industries. This novel evaluation framework ensures consistent, scalable, and reproducible assessment while mitigating the logistical challenges associated with assembling large human expert panels. The use of AI personas as evaluators is an emerging approach in research, with recent studies demonstrating their effectiveness in replicating human expert assessments across various domains [39, 40]. Our methodology aligns with findings from prior research, where AI personas were successfully employed to mirror human reasoning and domain expertise in experimental replications [41].

Our selection of 16 distinct AI personas was strategically grounded in achieving comprehensive coverage across key expertise domains relevant to carbon markets, including regulatory frameworks, economic analyses, scientific research, sustainability policies, and industry-specific applications. Each AI persona was assigned to a specific expertise domain and powered by a corresponding foundation model optimised for domain-specific reasoning. The Climate Policy and Regulatory Experts group (powered by Claude 3 Opus Mini) included specialists in international climate agreements, jurisdiction-specific carbon regulatory frameworks, and policy implementation strategies. The Environmental and Economic Analysis group (powered by Claude 3.1 Mini) consisted of environmental economists analysing carbon pricing mechanisms, carbon market strategists focusing on offset structures, and climate risk modellers forecasting economic implications. The Scientific and Technical Experts group (powered by GPT-4.5) incorporated atmospheric scientists with climate modelling expertise, climate data scientists leveraging machine learning for predictive analytics, and environmental journalists specialising in scientific fact-checking. The Sustainability and Finance Specialists group (powered by GPT-4o) comprised ESG investment analysts assessing corporate sustainability, climate resilience strategists developing adaptation frameworks, and biodiversity impact specialists evaluating climate-induced ecosystem changes. Finally, the Specialised Perspectives group (powered by GPT-4o Mini) included renewable energy technology specialists, climate justice advocates examining socio-environmental impacts, and AI agent developers designing decision-support systems for climate governance. These AI personas were inspired by methodologies such as those employed in Writer-Defined AI Personas for On-Demand Feedback Generation, where diverse personas were constructed to simulate audience perspectives in writing feedback [39].

To mitigate model-specific biases, we deliberately employed multiple foundation models, a methodological approach supported by recent meta-analyses of LLM evaluation frameworks [40]. Each model family exhibits distinct performance characteristics across specialised knowledge domains, necessitating cross-model triangulation to enhance evaluation validity and robustness. By diversifying the model architectures (Claude and GPT variants), we minimised systematic errors arising from idiosyncratic knowledge representations or reasoning pathways

within any single model family. Furthermore, each AI persona was provided with a predefined domain perspective but did not undergo additional training or fine-tuning beyond the base model architecture. This design decision aligns with findings from *From Persona to Personalization: A Survey on Role-Playing Language Agents*, which highlighted that AI personas retain their contextual accuracy without the need for extensive fine-tuning [41]. To ensure methodological rigour and unbiased assessment, we implemented a double-blind evaluation protocol, in which neither the AI personas nor human supervisors were aware of whether the reports originated from GreenIQ or baseline sources until after all evaluations were completed. Such an approach builds upon existing best practices in AI-assisted evaluations, as demonstrated in *Using Large Language Models to Create AI Personas for Replication and Prediction of Media Effects*, where AI personas successfully replicated human experimental assessments [40]. By incorporating these advances, GreenIQ’s evaluation methodology represents a significant step towards scalable, AI-driven assessment frameworks that enhance domain-specific research without introducing substantial human bias.

## **4.2 Quantitative Results**

Compared to traditional methods, GreenIQ demonstrated substantial efficiency and performance improvements across multiple dimensions.

### **4.2.1 Processing Efficiency**

Traditional carbon market analysis typically requires 40-60 hours of expert time for comprehensive reports. GreenIQ completed an equivalent analysis in an average of 22.7 minutes, representing a 99.2% reduction in processing time while maintaining comparable informational depth. The system processed an average of 327 distinct documents per analysis session, which far exceeded the human capacity for rapid information synthesis

### **4.2.2 Cost Efficiency**

Traditional carbon market reports from consultancies or research institutions typically cost between \$3,500–\$6,500 per comprehensive report. The computational resources and associated infrastructure costs of GreenIQ averaged approximately \$1.10 per report. This dramatic cost reduction primarily stems from the elimination of extensive manual labour while maintaining high-quality outputs.

### **4.2.3 Information Coverage**

GreenIQ consistently incorporated data from an average of 18.3 distinct jurisdictions per report, referencing an average of 42.7 regulatory documents, 37.2 academic publications, and 23.5 industry reports. This breadth substantially exceeded typical manual reports, which averaged references to 8.7 jurisdictions and 47.3 total citations across all categories.

### **4.2.4 Comparative Report Quality Assessment**

As shown in Table 1, we evaluated 10 different carbon market topics, comparing GreenIQ (Report 1) with baseline human reports (Report 2):

## **5 Discussion**

### **5.1 Key Innovations**

GreenIQ represents a significant advancement in the application of artificial intelligence to specialised analytical domains, particularly in carbon market intelligence. First, GreenIQ employs a multi-agent system and evaluation criteria that effectively channel LLMs’ reasoning capabilities towards carbon market-specific insights. This structured approach to domain specialisation serves as a scalable template for applying similar AI-driven methodologies to other complex regulatory and financial environments, without requiring domain-comprehensive retraining of base models. By designing specialised agents with distinct functional roles and well-defined interfaces, we established a system architecture that achieves both horizontal scalability (integrating new data sources and expanding analytical dimensions) and vertical integration (ensuring seamless information flow from raw data to finalised reports). Our approach to agent coordination through structured metadata exchange enables efficient pipeline processing while maintaining the adaptive flexibility required for iterative refinement when the quality thresholds are not initially met. This balance between operational efficiency and quality assurance represents a major improvement over prior AI-driven analytical frameworks, which often prioritise one dimension at the expense of the other.



Table 1: GreenIQ consistently outperformed the baseline human reports across all evaluation criteria, with the most significant advantages in citation quality (+8.5 points) and report coherence (+6.5 points). The average total score difference of 25.6 points demonstrates GreenIQ’s substantial performance improvement over conventional approaches.

Article Title	Report Type	Source Coverage	Data Accuracy	Citation Quality	Report Coherence	Total Score
Brazil passes law to cap emissions and regulate carbon market	Report 1	23/25	22/25	23/25	23/25	91/100
	Report 2	16/25	18/25	15/25	17/25	66/100
India’s Carbon Market Revolution	Report 1	24/25	24/25	23/25	22/25	93/100
	Report 2	20/25	22/25	21/25	22/25	85/100
World Bank: Fuels India’s Carbon Market and Green Hydrogen	Report 1	20/25	23/25	23/25	24/25	89/100
	Report 2	22/25	24/25	21/25	20/25	87/100
PLN Becomes First Indonesian Company in International Carbon Trading	Report 1	23/25	22/25	24/25	23/25	92/100
	Report 2	15/25	16/25	12/25	14/25	57/100
Watershed Launches RFP For 1 Megaton Of Carbon Removal Credits	Report 1	24/25	20/25	24/25	24/25	96/100
	Report 2	17/25	20/25	17/25	21/25	75/100
Japan Designates First Carbon Capture And Storage Site	Report 1	23/25	22/25	24/25	24/25	93/100
	Report 2	15/25	16/25	13/25	17/25	68/100
Amazon, ExxonMobil, And Microsoft Team Up To Strengthen Voluntary Carbon Market	Report 1	21/25	23/25	22/25	24/25	90/100
	Report 2	18/25	20/25	17/25	19/25	74/100
North Dakota Senate Defeats Eminent Domain Ban For Carbon Pipelines	Report 1	23/25	22/25	24/25	23/25	92/100
	Report 2	12/25	10/25	5/25	10/25	37/100
Bloom And Chart Industries Team Up For Efficient Carbon Capture	Report 1	23/25	22/25	24/25	25/25	94/100
	Report 2	15/25	17/25	14/25	18/25	64/100
South Korea Passes New Law To Support Carbon Capture	Report 1	23/25	24/25	24/25	24/25	94/100
	Report 2	15/25	16/25	10/25	14/25	55/100
<b>AVERAGE</b>	Report 1	22.6/25	22.8/25	23.3/25	23.7/25	92.4/100
	Report 2	16.5/25	18.3/25	14.8/25	17.2/25	66.8/100
<b>DIFFERENCE (Report 1 - Report 2)</b>		<b>6.1</b>	<b>4.5</b>	<b>8.5</b>	<b>6.5</b>	<b>25.6</b>

A key innovation within GreenIQ is its novel approach to citation management and verification, which addresses a critical challenge in AI-assisted research. By implementing continuous provenance tracking throughout the analytical pipeline and designing specialised verification protocols for each source type, GreenIQ maintains exceptional standards of accountability and transparency. This capability is particularly critical in carbon market analysis, where regulatory decisions and investment strategies depend on verifiable evidence and authoritative interpretation. Furthermore, GreenIQ’s citation verification system is uniquely capable of identifying and resolving conflicting information from multiple sources, which is a particularly valuable innovation for domains characterised by complex, evolving regulatory frameworks. This systematic approach to citation integrity and source validation establishes GreenIQ as a benchmark for AI-driven research synthesis, ensuring that AI-generated reports maintain rigorous evidentiary standards that are essential for policy development, financial forecasting, and sustainability assessments.

## 5.2 Limitations and Challenges

Despite GreenIQ’s substantial capabilities, several limitations and challenges warrant its acknowledgement. The performance of the system remains partially dependent on input data quality, with potential cascading effects when inaccuracies or biases exist in source materials [42]. This dependency is particularly relevant for emerging carbon markets with limited historical data or jurisdictions where regulatory documentation may be incomplete or inconsistently structured. Although our citation verification protocols mitigate these risks through cross-source validation, they cannot entirely eliminate the influence of biased or incomplete source materials. Addressing model hallucinations -the generation of plausible but factually incorrect content-represents an ongoing challenge for LLM-based analytical systems. GreenIQ implements multiple mitigation strategies, including strict citation requirements, fact-checking routines, and confidence scoring of the generated content [43]. However, the risk of hallucination increases in scenarios involving novel market mechanisms or regulatory approaches with limited precedents. This challenge necessitates continued vigilance and regular evaluation of system outputs against expert judgment, particularly when analysing rapidly evolving aspects of carbon markets.

Implementation complexities in highly specialised sub-domains of carbon markets present additional challenges. While GreenIQ demonstrates strong performance across general carbon market analysis, certain technical areas, such as complex offset methodologies, technical aspects of monitoring, reporting, and verification (MRV) systems, and intricate cross-border accounting mechanisms, occasionally reach the limits of the system’s specialised knowledge [44]. These limitations highlight the need for ongoing knowledge base expansion and development of additional specialised agents for particularly complex technical niches within the broader carbon market landscape.

## 5.3 Future Work

Several promising directions for future research will extend the capabilities and applicability of GreenIQ. First, we plan to adapt the system architecture to additional financial and regulatory domains, including compliance with sustainable finance taxonomy, climate-related financial disclosures, and broader environmental commodity markets. The modular architecture facilitates this extension through targeted modifications to domain-specific components, while maintaining the core analytical engine and agent coordination framework.

Enhanced multimodal visualisation capabilities represent a key development direction. While the current system generates effective static visualisations, future iterations will incorporate interactive data exploration tools, geospatial mapping of carbon market developments, and dynamic scenario modelling capabilities. These enhancements will further improve the accessibility and utility of complex carbon market information for diverse stakeholders.

Improved real-time analysis and alerting features will enable GreenIQ to provide timely notifications of significant market developments, regulatory changes, and emerging trends. By implementing continuous monitoring protocols and relevance assessment algorithms, the system will evolve from a predominantly retrospective analysis to a proactive intelligence that supports timely decision-making in rapidly changing market conditions.

## References

- [1] World Bank Group. *State and Trends of Carbon Pricing 2023*. World Bank, Washington, DC, 2023.
- [2] R. G. Newell, W. A. Pizer, and D. Raimi. Carbon markets: Past, present, and future. *Annual Review of Resource Economics*, 6(1):191–215, 2014.
- [3] Intergovernmental Panel on Climate Change. *Climate Change 2023: Synthesis Report*. Cambridge University Press, 2023.
- [4] European Commission. Eu emissions trading system (eu ets), 2023.
- [5] Persefoni. Carbon credits explained, 2024. Accessed: 4 March 2025.

- [6] C. Hepburn, N. Stern, and J. E. Stiglitz. Carbon pricing. *European Economic Review*, 2020.
- [7] S. Fankhauser and C. Hepburn. Designing carbon markets. part i: Carbon markets in time. *Energy Policy*, 38(8):4363–4370, 2010.
- [8] N. Keohane, A. Petsonk, and A. Hanafi. Toward a club of carbon markets. *Climatic Change*, 135(2):1–15, 2015.
- [9] Organisation for Economic Co-operation and Development. *Effective Carbon Rates 2023: Pricing Carbon Emissions through Taxes and Emissions Trading*. OECD Publishing, 2023.
- [10] M. Carney. *Building a private finance system for net zero: Priorities for private finance for COP26*. United Nations, 2021.
- [11] Task Force on Climate-related Financial Disclosures. *Status Report 2023*. Financial Stability Board, 2023.
- [12] J. Aldy and R. Stavins. The promise and problems of pricing carbon: Theory and experience. *Journal of Environment and Development*, 21:152–180, 2011.
- [13] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, and D. Amodei. Language models are few-shot learners. *arXiv*, 2020.
- [14] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, E. Brynjolfsson, S. Buch, D. Card, R. Castellon, N. Chatterji, A. Chen, K. Creel, J. Q. Davis, D. Demszky, and C. Ré. On the opportunities and risks of foundation models. *arXiv*, 2022.
- [15] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, and J. R. Wen. A survey of large language models. *arXiv preprint*, 2023.
- [16] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, and R. Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744, 2022.
- [17] Y. Li, D. Choi, J. Chung, N. Kushman, J. Schrittwieser, R. Leblond, and O. Vinyals. Competition-level code generation with alphacode. *Science*, 378(6624):1092–1097, 2022.
- [18] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, and Y. Zhang. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint*, 2023.
- [19] ICAP. Carbon market oversight primer, 2023.
- [20] C. Blaufelder, C. Levy, P. Mannion, and D. Pinner. A blueprint for scaling voluntary carbon markets to meet the climate challenge. 2021.
- [21] E. Narassimhan, K. S. Gallagher, S. Koester, and J. R. Alejo. Carbon pricing in practice: A review of existing emissions trading systems. *Climate Policy*, 18(8):967–991, 2018.
- [22] B. Doda, S. Quemin, and L. Taschini. Linking permit markets multilaterally. *Journal of Environmental Economics and Management*, 98:102259, 2019.
- [23] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, A. Luccioni, T. Maharaj, E. D. Sherwin, S. K. Mikkavilli, K. P. Kording, C. Gomes, A. Y. Ng, D. Hassabis, J. C. Platt, and Y. Bengio. Tackling climate change with machine learning. *arXiv*, 2019.
- [24] F. Masood. Impact of artificial intelligence on financial markets. *International Journal of Emerging Multidisciplinaries: Social Science*, 3(1):8, 2024.
- [25] P. Bolton, M. Luiz, A. Pereira, D. Silva, F. Samama, and R. Svartzman. *The green swan: Central banking and financial stability in the age of climate change*. Bank for International Settlements, 2020.
- [26] C. M. Lewis and S. Young. Fad or future?: Automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research (Forthcoming)*, 2018.
- [27] S. Hassani. Enhancing legal compliance and regulation analysis with large language models. In *2024 IEEE 32nd International Requirements Engineering Conference (RE)*, pages 507–511, Reykjavik, Iceland, 2024.
- [28] H. P. Kothandapani. Ai-driven regulatory compliance: Transforming financial oversight through large language models and automation. *Emerging Science Research*, 3, 2025.
- [29] C. Cao, J. Zhuang, and Q. He. Llm-assisted modeling and simulations for public sector decision-making: Bridging climate data and policy insights. In *PubLLM 2024*, 2024.
- [30] M. Pahle, D. Burtraw, C. Flachsland, N. Kelsey, E. Biber, J. Meckling, O. Edenhofer, and J. Zysman. Sequencing to ratchet up climate policy stringency. *Nature Climate Change*, 8:861–867, 2018.

- [31] D. Weyns, M. I. Schumacher, A. Ricci, M. Viroli, and T. Holvoet. Environments in multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 38(2):231–256, 2024.
- [32] B. Patrick. Multi-agent systems in predicting global financial crises. *International Journal of Financial Studies*, 12(4):567–589, 2024.
- [33] J. Feng, C. Tao, X. Geng, T. Shen, C. Xu, G. Long, D. Zhao, and D. Jiang. Synergistic interplay between search and large language models for information retrieval. *arXiv preprint arXiv:2305.07402*, 2023.
- [34] H. Rahmani, H. Zamani, and W. B. Croft. Llm4eval: Large language model for evaluation in ir. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2529–2533, 2023.
- [35] N. Borri et al. Inefficiencies of carbon trading markets. 2024.
- [36] WJARR. Enhancing carbon markets with fintech innovations: The role of ai and blockchain. 2024.
- [37] MIT Impact Climate Lab. Exploring the challenges and potential of voluntary carbon markets, 2023.
- [38] A. Oladeji and A. Mousavi. Leveraging ai-derived data for carbon accounting: Information extraction from alternative sources. 2023.
- [39] Z. Sun, L. Wang, and X. Tan. Writer-defined ai personas for on-demand feedback generation. In *ACM CHI Conference on Human Factors in Computing Systems*, 2024.
- [40] R. Zhou, S. Patel, and J. Lee. Using large language models to create ai personas for replication and prediction of media effects. *arXiv*, 2024.
- [41] T. Joseph, H. Kim, and Y. Xu. From persona to personalization: A survey on role-playing language agents. *arXiv*, 2024.
- [42] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, 2021.
- [43] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, and P. Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023.
- [44] L. Schneider, A. Michaelowa, D. Broekhoff, A. Espelage, A. Siemons, A. Gläser, and D. Machnik. Lessons learned from the first round of applications by carbon-offsetting programs for eligibility under corsia. 2019.