sustainability

MDPI

Systematic Review

How Can Large Language Models Drive Environmental Sustainability? A Systematic Scoping Review

Xiaotong Su, Ting Liu, Patrick Pang *, Yiming Taclis Luo and Dennis Wong

Faculty of Applied Sciences, Macao Polytechnic University, Macao 999078, China; p2523603@mpu.edu.mo (X.S.); p2314982@mpu.edu.mo (T.L.); cwong@mpu.edu.mo (D.W.)
* Correspondence: mail@patrickpang.net

Abstract

Currently, Large Language Models (LLMs), exemplified by ChatGPT, are accelerating technological development across various domains, including the environmental domain, owing to their powerful text-generation and information-processing capabilities. With changes in global climate and environmental conditions, environmental sustainability has emerged as a major global challenge. Leveraging LLMs to advance environmental sustainability and mitigate current environmental problems is considered a valuable and effective approach. This study aims to systematically synthesize research progress and core challenges in current LLMs for promoting sustainability-related fields, and to comprehensively analyze the application contexts, impacts, and development potential of various LLMs within the environmental sector. Following the PRISMA-ScR guidelines, a comprehensive search was conducted across six databases: Web of Science (WOS), Scopus, ACM Digital Library, IEEE Xplore, ScienceDirect, and Google Scholar. A total of 20 articles were ultimately included for analysis. The findings indicate that LLMs play a positive role in maintaining environmental sustainability and promoting the low-carbon energy transition. The applications of LLMs span six core domains: the green transition, carbon emission management, air quality assessment, smart city operations, map analysis, and human cognition and behavioral observation. However, the training and operation of current LLMs consume considerable resources, which creates an inherent conflict with the goals of sustainable development. Future efforts must focus on developing a secure, equitable, and scalable LLM support system to advance environmental sustainability. This requires optimizing model energy efficiency and ensuring a balance between performance, reliability, and environmental impact. These endeavors are crucial for addressing environmental problems and guaranteeing the sustainable progression of LLMs across diverse environmental contexts.

Keywords: LLMs; large language models; environment; environmental sustainability; systematic scoping review

Check for updates

Academic Editors: Hossein Bonakdari, Ebrahim Ghaderpour and Silvio José Gumiere

Received: 16 March 2026
Revised: 23 April 2026
Accepted: 23 April 2026
Published: 27 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Large Language Models (LLMs) refer to deep learning models trained on vast amounts of textual data, enabling them to generate or comprehend natural language text [1]. By undergoing unsupervised training on large-scale datasets, LLMs learn the patterns and structures of natural language, thereby simulating human language cognition and generation processes to some extent [2]. In recent years, LLMs have experienced a process of rapid development, marked by advancements from Generative artificial intelligence's (GenAI)

GPT-3 with 175 billion parameters to the multimodal architecture of GPT-4, and the continuous emergence of open-source and closed-source models such as LLaMA and DeepSeek. These improvements signify major progress in artificial intelligence technology [3]. Leveraging their remarkable capabilities in complex reasoning, cross-modal integration, and unstructured data processing, current LLMs are no longer confined to being mere language processing tools but are gradually becoming highly versatile intelligent systems, Generative Pre-Trained Transformers (GPTs), providing a critical technological foundation for the digital transformation of various industries, including the environmental domain [4]. GenAI serves as the broadest category, referring to AI systems capable of creating new content. Underpinning GenAI are Foundation Models, which are large-scale models trained on vast datasets that can be adapted to a wide range of downstream tasks. LLMs represent a specific subset of foundation models optimized for processing and generating human language. Among these, GPTs refer to a specific architectural class of LLMs based on the Transformer framework. Concurrently, the global environment is facing severe challenges, ranging from extreme weather events triggered by climate change and the rapid loss of biodiversity to the severe exacerbation of resource depletion and environmental pollution. Environmental sustainability is a process of change in which the exploitation of resources, the direction of investments, the orientation of technological development and institutional change are all in harmony and enhance both current and future potential to meet human needs and aspirations, which is no longer a single-disciplinary issue but a systemic and challenging one involving global governance, economic transition, and social equity [5]. The international community has placed the Sustainable Development Goals (SDGs) and nation-level "Carbon Neutrality" strategies in a vital position [6]. In this pressing context, traditional environmental decision-making methods based on manual analysis and single models have increasingly shown their limitations. There is now an urgent need for intelligent systems capable of processing massive, high-dimensional, and heterogeneous data to accelerate the implementation of sustainability initiatives smoothly.

Due to their ability to efficiently integrate multi-source and heterogeneous data, LLMs are regarded as vital potential instruments for advancing environmental sustainability. Their value lies primarily in two dimensions: enhanced efficiency and expanded impact. At the application level, the value of LLMs has extended across multiple domains: they can automatically extract structured and validated KPIs and compliance information from massive amounts of scientific literature, policy documents, and corporate Environmental, Social, and Governance (ESG) reports, fundamentally improving the efficiency of environmental monitoring and reporting [7]. Concurrently, in urban operations, LLMs can process real-time traffic and energy sensor data to optimize power dispatch in low-carbon grids and route planning for intelligent transportation systems, thereby directly reducing carbon emissions [8]. In terms of climate risk communication, LLMs can translate complex climate model outputs and scientific findings into narratives that are easily understandable and highly targeted for the public, effectively fostering climate education, public engagement, and behavioral change, thereby bridging the knowledge gap between science and society [9].

The introduction of LLMs into research within the environmental sustainability domain holds profound significance for both environmental governance and advancing sustainability initiatives. LLMs are more than just efficiency tools; they represent an expansion of research methodologies and perspectives. Environmental sustainability issues are inherently highly complex, non-linear systems. The powerful reasoning capabilities of LLMs enable them to efficiently address and handle variable relationships that are difficult for traditional statistical models to manage, thereby allowing for more precise simulation and prediction of environmental intervention measures and policy outcomes [10]. Furthermore,

LLMs have lowered the application threshold for specialized data analysis and geospatial remote sensing technologies, making it easier for policymakers, business managers, and citizen scientists without professional backgrounds to participate in environmental monitoring and protection efforts through natural language interaction [11]. However, the environmental impact of LLMs has far surpassed that of traditional computing tasks, with their carbon footprint spanning the entire lifecycle from hardware manufacturing, model training, and inference services to the disposal of retired equipment. Relevant research indicates that training the GPT-3 model alone consumed approximately 190,000 kWh of electricity and generated 85,000 kg of carbon dioxide [12]. Of greater concern is that the energy consumption during the inference stage often exceeds the training itself, and as the model is called upon billions of times daily, the environmental load resulting from its continuous operation presents a rapid growth trend [13].

Facing the dual role of LLMs in environmental sustainability—as a potential solution and as a source of its own environmental burden, current research is characterized by fragmentation and rapid evolution, lacking a systematic synthesis and assessment of its overall benefits, technical bottlenecks, and limitations. Therefore, this study adopts the rigorous PRISMA-ScR methodology to conduct a systematic scoping review aimed at comprehensively examining the current application status of LLMs in the field of environmental sustainability research. It is important to note that while some individual obstacles identified in this review align with existing fragmented research, the primary contribution of this study does not lie solely in discovering entirely new barriers. Rather, the contribution is providing a systematic, objective, and reproducible taxonomy of these challenges using a scoping review framework and Natural Language Processing (NLP) techniques. This comprehensive mapping transitions the field from anecdotal, isolated qualitative observations to a structured landscape, clearly illustrating the relative weight, interconnectedness, and current research focus or technical bias, regarding LLMs in environmental sustainability. By adopting a systematic approach, we move beyond a simple descriptive summary to critically analyze the methodological rigor and citation impact of the current literature, and addressing the following research questions:

(1) What are the geographical and temporal trends of research utilizing LLMs for environmental sustainability?
(2) Which LLMs are most widely applied in the environmental sustainability domain?
(3) Which environmental sustainable development domains can LLMs be used in?
(4) What are the roles and impacts of LLMs in the field of environmental sustainability?
(5) What are the future potential and development trends for the process of environmental sustainability supported by current LLMs?

2. Methods

2.1. Search Strategy

This study screened relevant papers from six electronic databases: WOS, Scopus, ACM Digital Library, IEEE Xplore, ScienceDirect, and Google Scholar. We selected the above six databases as our primary databases to ensure that the collected information is comprehensive, as they include major publishers' repositories, including SpringerLink, Wiley Online Library, and Taylor & Francis. The search was conducted on 18 October 2025. Search terms included "LLMs," "ChatGPT," "Large Foundation Models," "AIGC," "GPT," "Sustainability," "Environmental Sustainability," and "Green Economy," among others (see Table 1). The PRISMA-ScR checklist is provided in Supplementary Materials.

Table 1. Selected Databases and Search Strategy.

Database	Search Formula
WOS	(TS=): ("LLM*" OR "Large Language Model*" OR "ChatGPT" OR "GPT*" OR "Foundation Model*" OR "AIGC") AND ("Sustainability" OR "Environmental Sustainability" OR "Green Economy" OR "Climate" OR "Environment*")
Scopus	(TITLE-ABS-KEY): ("LLMs" OR "Large Language Model*" OR "ChatGPT" OR "GPT*" OR "Foundation Model*" OR "AIGC") AND ("Sustainability" OR "Environmental Sustainability" OR "Green Economy" OR "Climate" OR "Environment*")
IEEE Xplore	(All Metadata): ("LLMs" OR "Large Language Model*" OR "ChatGPT" OR "GPT*") AND ("Sustainability" OR "Environmental Sustainability" OR "Green Economy" OR "Climate")
ACM Digital Library	(Full Text): ("LLMs" OR "Large Language Model*" OR "ChatGPT" OR "GPT*") AND ("Sustainability" OR "Environmental Sustainability" OR "Green Economy" OR "Climate")
ScienceDirect	(Title/Abstract/Keywords): ("LLMs" OR "Large Language Model*" OR "ChatGPT" OR "GPT*") AND ("Sustainability" OR "Environmental Sustainability" OR "Green Economy" OR "Climate")
Google Scholar	(LLMs OR Large Language Model OR ChatGPT OR GPT) AND (Sustainability OR Environmental Sustainability OR Green Economy OR Climate)

* is used to retrieve variations of a word root, including singular/plural forms and different parts of speech.

2.2. Data Selection and Extraction

All retrieved records were exported to the Zotero software (version 6.0), and duplicate entries were removed. Two independent authors (XS and TL) conducted the initial screening of article titles and abstracts based on pre-determined inclusion criteria. Any discrepancies between the two authors were resolved through consultation with a third author (PP). The inclusion criteria were as follows: (1) Research specifically focusing on environmental sustainability; (2) Research on LLM technology within the environmental sustainability domain; (3) Applied research utilizing LLM technology, defined operationally in this review as studies that (a) deploy, test, evaluate, or implement a specific LLM or LLM-based system within an environmental sustainability context; (b) report concrete input/output configurations, system architectures, or measurable performance outcomes; and (c) extend beyond theoretical argumentation or narrative description of LLM capabilities without empirical demonstration. Studies that rely solely on secondary data synthesis without applying an LLM as an active analytical or operational tool were assessed on a case-by-case basis by two independent reviewers. Discrepancies were resolved through consensus with a third reviewer, and the rationale for borderline decisions was documented in the data extraction table; (4) Research articles, full texts, and conference papers; and (5) Published in English. The design of the inclusion and exclusion criteria was primarily focused on the application of LLMs. Priority was given to empirical studies investigating the application of LLMs in environmental governance and sustainability implementation methods, thereby excluding research that primarily discussed the impacts and changes in environmental sustainability itself. Additionally, review articles (e.g., narrative or systematic reviews) were excluded as they synthesize existing literature rather than presenting original applications. The inclusion criteria covered qualitative, quantitative, and mixed-

methods studies to ensure comprehensive coverage of evidence regarding the practical implementation of LLMs. Interrater reliability revealed substantial to perfect agreement (κ = 0.703–0.815) between two independent reviewers for data selection, data extraction. These criteria are summarized in Table 2.

Table 2. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Research specifically targeting environmental sustainability	Research fields outside of environmental sustainability
Research on LLM technology within the environmental sustainability domain	Research on LLM technology outside of environmental sustainability
Applied research utilizing LLM technology	Studies focusing on attitudes, opinions, intentions, benefits, barriers, impacts, experiences, or usage demands regarding LLM technology
Research articles and conference papers (full text)	Review articles, theses/dissertations, non-academic publications, book chapters, etc.
Full text published in English	Full text published in other languages
Studies deploying LLMs as an active analytical or operational tool with documented input/output processes	Studies presenting solely conceptual, theoretical, or argumentative analyses of LLM potential without empirical demonstration or system implementation

2.3. Data Charting, Synthesis, and Reporting

A data extraction table was created based on the scoping review methodological guidelines provided by the PRISMA-ScR checklist [14]. The data extraction items included: author, year, country, research methodology, types of input and output, application domain, target population, significance, impact, potential, and future trends. All data were independently extracted by two authors. Any discrepancies encountered during the extraction process were resolved through consultation with a third reviewer. To synthesize the reliability and validity of the evidence, we conducted a rigorous assessment of the methodology and quality of each included study using the Mixed Methods Appraisal Tool (MMAT) [15]. The study findings were collected, summarized, and analyzed, with descriptive statistics employed to characterize the features of the sample articles. The descriptive findings are presented through figures and tables. The results were interpreted through a narrative synthesis to address the research questions posed in the review, and these interpretations were validated by all authors.

To address potential methodological inconsistencies arising from the heterogeneous nature of included studies, we introduced an additional classification dimension to distinguish between empirical studies, those directly applying an LLM system to real-world environmental data with verifiable inputs and outputs, and conceptual-applied studies, defined here as studies that engage LLMs as a central subject of applied analysis (e.g., evaluating an LLM's carbon footprint, assessing its output quality on domain-specific tasks, or analyzing its behavior through structured data collection), even where the primary data are partially secondary in origin. This distinction was documented in the data extraction table and informed our narrative synthesis. Studies classified as conceptual-applied were retained in the review because they address concrete operational dimensions of LLM deployment; however, their findings are interpreted with explicit acknowledgment of the inferential limitations associated with secondary data reliance. This classification procedure

was independently applied by two reviewers, with inter-rater agreement documented (κ = 0.703–0.815).

2.4. Quality Assessment of Included Studies

The methodological quality of the 20 included studies was evaluated using the MMAT. As shown in Appendix A, the overall quality of the literature is high. Among the included studies, 11 studies (55%) achieved a perfect score of 100%, while the remaining 9 studies (45%) scored 85.71%.

Specifically, within the category of Randomized Controlled Trials, common minor limitations were related to the blinding of outcome assessors (Criterion 2.4). For Qualitative Studies, most met all criteria except for occasional ambiguity regarding the influence of the researcher on the study (Criterion 1.5). For Non-Randomized Studies, all 8 included publications demonstrated robust methodology, each achieving a 100% score except for one study [16]. This high overall scoring suggests that the evidence synthesized in this review is derived from methodologically sound research.

3. Results

The systematic search initially retrieved a total of 6264 documents, as illustrated in Figure 1. After removing duplicate entries using the Zotero software, 3540 articles remained. Two reviewers independently screened the article titles and abstracts, resulting in the exclusion of 3451 articles deemed not directly relevant to the research topic, along with 3 non-English articles. Subsequently, the remaining 158 articles were comprehensively assessed by the two reviewers. Of these, 69 articles were excluded due to the inaccessibility of the full text. The final remaining 89 articles were then reviewed. Among these, 69 articles focused on opinions, expectations, and attitudes rather than specific technological applications. Since these studies were not the focus of the current research, they were excluded. Consequently, a final total of 20 documents were included in the scope of this systematic review (see Table 3).

To ensure the transparency, reproducibility, and rigor of our findings, this review was conducted in strict accordance with the PRISMA-ScR statement. We intentionally chose the PRISMA-ScR framework over scoping review methodologies (Arksey and O'Malley) because our research objective is to synthesize specific outcomes and evaluate the academic impact of included studies rather than merely mapping the conceptual boundaries of the field. The PRISMA-ScR guidelines provide a structured four-phase flow (Identification, Screening, Eligibility, and Inclusion), which allows for a more controlled and bias-reduced selection process, ensuring that only the most relevant and high-quality studies are synthesized.

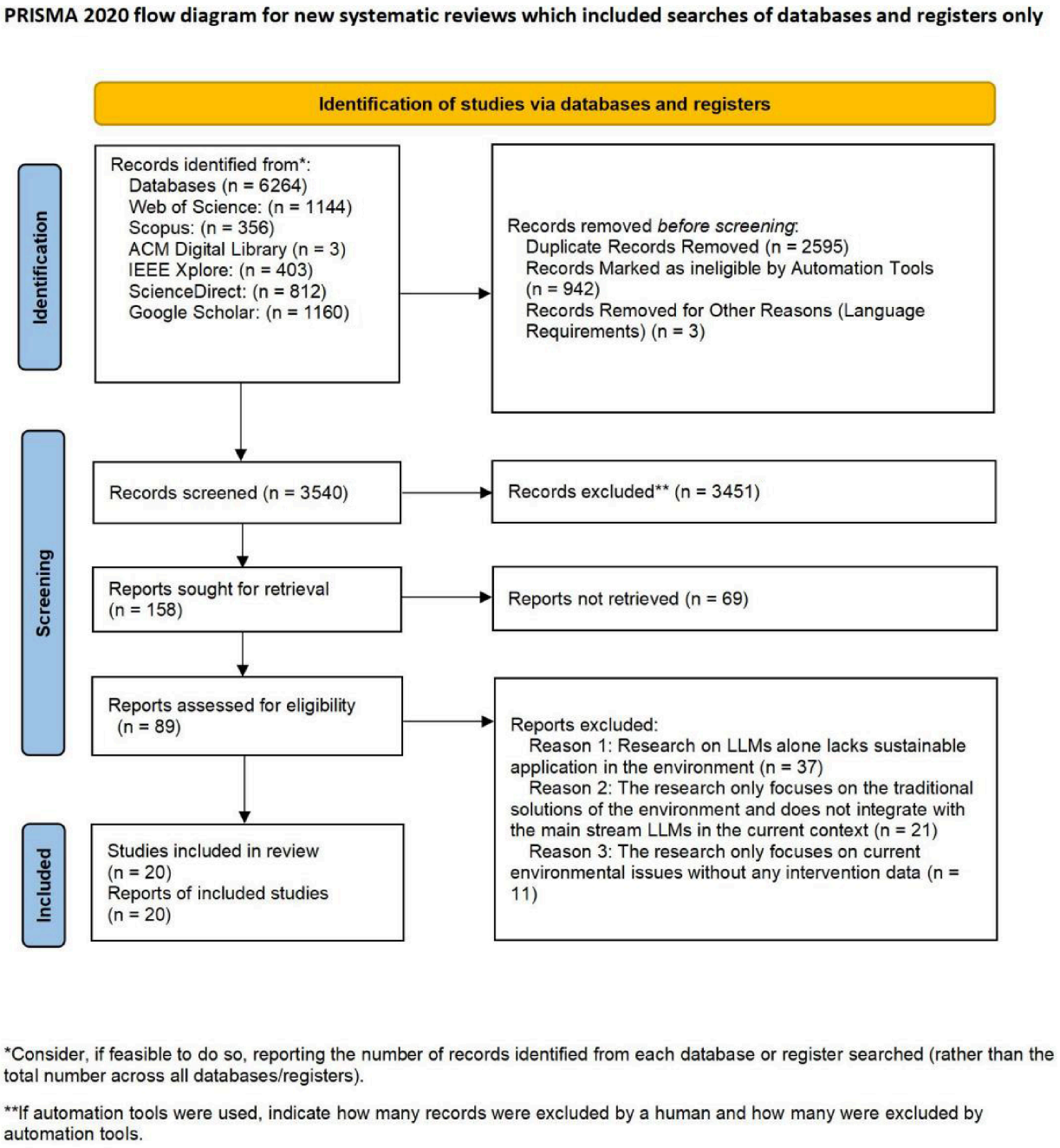

Figure 1. PRISMA flowchart.

Table 3. Overview of Study Characteristics.

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Shoaib et al. 2023 Singapore [17]	Qualitative	Text/Conference papers	Intelligent Transportation Systems (ITS), Vehicle Technology, Smart Cities	LLM, ITS Domain	LLMs not only facilitate the development of autonomous vehicles but also contribute to smart cities by alleviating congestion and optimizing traffic routes through ITS. They address the challenge of fundamental models and frontier AI not efficiently solving problems	Researchers need to proactively use regulatory frameworks to mitigate potential harms and the unique challenges associated with integrating frontier AI and fundamental models into ITS	Future research in this area should delve into the temporary response capabilities of these AI technologies when confronting new challenges. This involves refining their real-time decision-making in traffic management, tailoring responses to the specific needs of smart cities, and encouraging interdisciplinary collaboration to fully unlock the potential of LLMs in creating sustainable, smart, and human-centered transportation ecosystems
Song et al. 2025 Hong Kong, China [18]	Quantitative (Atmospheric Simulation Systems, Retrieval-Augmented Generation (RAG) Model, AirGPT Framework)	Literature Corpus, Related Professional Data/Journal Articles	Air Quality Assessment	LLM	Demonstrates exceptional ability in providing accurate regulatory information, executing basic data analysis, and generating location-specific management advice. AirGPT outperforms others in professionalism and accuracy across most query categories. Although GPT-40's quality score is comparable, its broader knowledge base occasionally leads to more detailed answers	The system aids researchers by providing accurate regulatory information, executing basic data analysis, and generating location-specific management advice. It avoids hallucinations by strictly adhering to validated scientific sources, which is especially critical when addressing air quality management policies that affect public welfare	Although AirGPT exhibits strong analytical capabilities, it is a decision support tool, not a substitute for expert judgment. Future research advises users to exercise professional discretion and verify key recommendations, particularly those affecting public health and environmental policy. Future work needs to incorporate safeguards in training data to prevent potential biases and continuously monitor system output to ensure alignment between LLMs and established scientific consensus in environmental science and air quality management

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Cappendijk et al. 2025 Netherlands [19]	Quantitative (Peak and Floating Point Operations (Flops))	Energy-Efficient Code/Conference papers	Atmospheric Greenhouse Gas Emissions	Code Llama-70b, Code Llama-70b-Indict, Code Llama-70b-Python, DeepSeek-Coder-33b-BASE, and DeepSeek-Coder-33b-Indict	LLMs are largely capable of generating code with lower energy consumption than human-written code	For-Loop optimization often leads to lower energy consumption compared to baseline solutions. Replacing native Python 3.8 code with more efficient library functions allows some LLM-generated solutions to outperform baseline solutions	AI-generated code is not always superior to human-generated code, especially since the AI-generated code setups were simple (single prompt, no hyperparameter tuning). In future coding processes, the same result might sometimes lead to lower or higher energy consumption
Lin et al. 2025 Taiwan, China [16]	Quantitative (Python mapping libraries, JSON module, Folium Heatmap plugin)	Geographic Information Maps, Taiwan CDC Open Data Platform/Journal Articles	Dengue Fever Epidemic and Earthquake Intensity Maps	ChatGPT 4, Copilot1.8, Claude, Chatbot UI 2.0, Code Lats, Code Llama, Gemini, and BigCode	Enhances the autonomy of geographical monitoring for dengue fever and earthquakes, saving researchers substantial time and labor costs. Autonomously generated maps provide a clear understanding of the distribution and clustering of disasters in different regions. Improves the timeliness, consistency, scalability, and standardization of data processing while automating repetitive tasks and assisting with data analysis	Highlights the broad applicability of LLMs in text generation and data visualization, and offers valuable reference for research in automated disaster monitoring, prevention, and mitigation	Challenges remain in ensuring the accuracy of geographically related code generated by LLMs and effectively interpreting results. Future research still requires targeted training of LLMs in the geospatial domain

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Lee et al. 2024 USA [20]	Mixed (Sampling Survey, Silicon Wafer Sample Data Collection, Survey Measurement, MAF)	Questionnaire/Journal Articles	Human Attitudes and Behavior, Global Warming Opinions	GPT-4	LLMs can effectively reproduce presidential voting behavior but not global warming opinions, apart from issue-related covariates. When demographic and covariates are considered, GPT-4 demonstrates higher accuracy in predicting beliefs and attitudes toward global warming	Provides valuable insights into the algorithmic fidelity and bias of LLMs when simulating public opinions on global warming. Offers targeted practical guidance on conditional prompting and model selection to maximize fidelity in social science applications, while emphasizing the importance of validating LLMs, especially for minority groups	Future researchers need to pay close attention to whether massive training data favors certain attitudes towards climate change, or whether post-training adjustments guided by human feedback lack representativeness. A nuanced approach is needed for human attitudes and behavior to utilize the low-cost management capabilities while addressing their limitations through proactive algorithmic auditing and bias mitigation
Cecconi et al. 2024 Italy [21]	Qualitative (Predictive Modeling, NLP	Legislative Documents/Book Chapter	Policies related to Climate Change Mitigation Efforts	GPT-4	Through NLP, LLMs can help interpret complex climate model results, translating technical model outputs into understandable narratives that describe potential future states of the climate system under different policy pathways. Enhancing public engagement and education: LLMs play a crucial role in transforming complex climate science into content that is understandable and engaging for the public	The integration of LLMs into climate change research and policymaking represents a promising convergence of technology and environmental science. By leveraging the power of LLMs, researchers, policymakers, and advocates can enhance citizens' understanding of climate dynamics, improve the formulation of effective policies, and promote greater public participation in climate action	As the urgency to address global warming intensifies, current research techniques are temporarily unable to efficiently support environmental research. Integrating complex AI technologies like LLMs into environmental research is a direction researchers need to prioritize

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Bhaskar et al. 2024 India [22]	Quantitative	Digital Articles, Theses, Reputable Journals, Government Websites, Statistical Websites, etc./Book Chapter	Environmental Impact, Carbon Footprint, Water Footprint, Greenhouse Gas Emissions	ChatGPT	Emphasizes the importance of sustainability in AI development and reveals the negative environmental impacts of ChatGPT. Highlights the excessive energy consumption associated with training and running ChatGPT models. The carbon footprint generated by the training process raises concerns about AI development's contribution to climate change and environmental degradation	Longitudinal studies using LLMs over extended periods can help researchers monitor changes in the environmental impact of AI models. Provides researchers with insightful information on the precise environmental impact of AI models across various industries or applications. Future focus can be placed on actual settings, considering factors like data centers, power supply, and energy-saving technologies	The study relies solely on secondary data sources, meaning the actual findings are limited by the availability and quality of existing studies and reports. Future research should focus on providing more comprehensive and accurate empirical data on the environmental impact of AI models like ChatGPT
Wang et al. 2025 China [23]	Quantitative (Points of Interest (POI) calculation, Descriptive Statistics, Mean Absolute Percentage Error (MAPE), Coefficient of Determination, and A20 index)	POI, Road Network Data, Housing Prices, and Population Data/Journal Articles	Built Environment; Public Transit Ridership	DeepSeek-R1	Offers valuable insights for transportation planners and policymakers, assisting government and transportation departments in urban traffic layout planning and providing guidance for environmental sustainability governance policies	Research findings can lead to land use intensification, an increased modal share for public transport, reduced traffic congestion, thereby lowering associated carbon emissions from traffic, and ultimately enhancing the overall sustainability of the transportation system	The mechanism by which the built environment influences various modes of transportation remains a vital topic requiring further research. With the continuous improvement of LLM reasoning capabilities in future research, LLMs can be utilized by researchers to consider a wider range of environmental sustainability impact factors and explore the joint influence between environmental sustainability and other factors

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Martín-Domingo et al. 2025 Ireland [7]	Mixed (Sustainability Reporting, Systematic Review)	Environmental Key Performance Indicators (KPIs)/Journal Articles	Environmental Sustainability Indicators and Regulatory Compliance	GPT-4.0, o3mini, and DeepSeek R1	Expands the scope of previous research to include a broader range of sustainability indicators, different transport modes or domain, and the use of multiple languages and further fine-tuning for Small Island Developing States (SIDS). Precisely evaluates the accuracy and reliability of LLMs in extracting emission-related KPIs from European airline sustainability reports	The model analysis covers multiple data sources, extraction strategies, and model architectures, providing a comprehensive overview of factors affecting automated KPI extraction performance. Using commercial LLMs may lead to data being publicly disclosed outside the responsible organization's jurisdiction, potentially violating GDPR or other regulations when processed by cloud-hosted LLMs, and risks exposing sensitive information to external systems	Future research needs to further explore the integration of LLMs with structured data extraction tools. Cost–benefit and business scalability assessments will help promote the automation of ESG data extraction and support constantly changing regulatory requirements. Facilitating collaboration across the interdisciplinary boundaries of AI, sustainability, and compliance will lay the groundwork for future sustainable digital transformation and standards
Chew et al. 2024 Malaysia [24]	Quantitative (Machine Learning (ML), Data Analysis)	Keetch-Byram Drought Index (KBDI), Soil Moisture, Temperature, Wind Speed, Land Surface Temperature (LST), Palmer Drought Severity Index (PDSI), Normalized Difference Vegetation Index (NDVI), Land Cover, and Rainfall, etc./Journal Articles	Forest Fire Investigation	Google Earth Engine Integrated Framework, ChatGPT	Provides valuable insights into the fire scenarios in Peninsular Malaysia. Preliminary analysis of the annual average of forest fires concludes that the main factors	Lowers the threshold for data scientists, allowing users to apply their analytical skills directly to datasets extracted by the Global Environmental Research Centre, thereby reducing the need for in-depth remote sensing knowledge	No manual coding was performed during the analysis; the analysis and Python scripts for generating the results were created through simple prompts in the ChatGPT interface. As technology continues to evolve, researchers leading future studies should consider adopting this technique to improve their methods and analysis

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Stein 2025 USA [25]	Qualitative	Digital Data of the Internet, Social Media, and AI/Journal Articles	Water, Energy, Carbon, Waste, and Land Use	GenAI	GenAI can bring more net benefits in our collective pursuit of a more sustainable environment, while stimulating innovation and sustainability potential, offering more targeted innovation and upgrades compared to previous research	GenAI, like all major industries, places pressure on the world's limited resources while bringing convenience. Data centers training and running GenAI models generate non-negligible negative impacts on water, energy, carbon, waste, and land use	Most AI users are distant from these impacts. Future involvement from government or industry is needed to minimize the socially borne costs of data centers. The GenAI industry is better suited to record its environmental impact and integrate sustainability into corporate ethical commitments. Future research in related fields needs to strive to minimize the negative environmental impact of GenAI
Bibri et al. 2025 Switzerland [26]	Quantitative	Heterogeneous Real-time Data from Urban Systems/Journal Articles	Cognitive Enhancement, Resource Efficiency, Network Traffic, Cybersecurity and Anomaly Detection, Resource Sustainability, Resource Efficiency	GenAI, AI + Internet of Things (AIoT) Systems	The study integrates current GenAI and AIoT, further emphasizing domain-specific advancements and their synergy. It promotes the development of sustainable smart cities by fostering a smarter, more energy-efficient, adaptive, secure, robust, and autonomous AIoT ecosystem through the strategic application of Generative intelligence	Provides practical guidance for policymakers, urban planners, system designers, and technology developers, helping researchers utilize GAIoT to enhance the resilience, sustainability, and operational capabilities of smart cities	Future research should consider simulation-based testing, prototyping, and real-world implementation to validate the practical value of the framework. This includes specifying data flows, control algorithms, feedback mechanisms, and system interactions to convert the high-level conceptual framework into a functional architecture

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Huang et al. 2025 Switzerland [27]	Quantitative	Lausanne's Blue City Project Data/Journal Articles	Innovative Urban Development Solutions	GenAI, Foundation Models (FMs), and Urban Digital Twin (UDT) Framework	Enhances decision-making processes, supports evidence-based planning and design, promotes integrated development strategies, and enables the development of more efficient, resilient, and sustainable urban environments. It advances the theory and practice of AI-driven, environmentally sustainable urban development through the implementation of GenAI and FMs within the UDT framework	Provides complex decision tools and valuable insights for urban planners, designers, policymakers, and researchers, helping them address the complexities of modern cities and accelerate the transition towards a sustainable urban future	N/A
Cheng et al. 2024 China [28]	Quantitative (Time Series Analysis, Regression Analysis)	Real-time Meteorological Data, Electricity Market Data, and Equipment Operating Status Data/Conference papers	Low-Carbon Power Scenarios, Carbon Market Dynamics, Climate Risk Assessment, and Urban Planning Strategies	LLM	Low-carbon energy management can drive innovation and sustainability in the low-carbon energy transition, and is capable of optimizing grid operations, integrating renewable energy, and predicting demand patterns	LLMs can facilitate better decision-making for researchers, optimize resource allocation, and accelerate the development of innovative low-carbon technologies. It also maximizes the impact of low-carbon energy management in the low-carbon energy transition	Future research may face challenges in data quality and availability, cybersecurity, legal management training costs, interdisciplinary collaboration, and evaluation benchmarks. Further exploration of land use management applications in various low-carbon energy transition scenarios and its combination with other advanced technologies will help unlock its full potential in driving sustainable development

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Karlsson 2025 Sweden [29]	Mixed	Sustainability Reports, Corporate Websites, Social Media, and News Articles/Book Chapter	Green Transition	GenAI	Microsoft Copilot and ChatGPT achieved similar results in identifying environmental prospects, with only minor differences observed between the AI models	GenAI adds value to exploration tasks, but the usefulness of its models depends on the implementation strategy. The models' conclusions are accurate enough for the task but should be continuously monitored. The suggested tools are not meant to replace human judgment but to assist the sales process and drive green transition	Researchers should adopt alternative methods, emphasizing strong initial economic partnerships before focusing on sustainability intentions
Krzyżewska 2025 Poland [30]	Quantitative (Cloud Identification and Classification)	Official World Meteorological Organization (WMO) Cloud Atlas/Journal Articles	WMO Cloud Classification; AI Map Interpretation	ChatGPT o3-mini, o1, 4.0, 4.0; Gemini Advanced 1.5 and 2.0; Copilot; Pplexity; DataAnalyst; Consensus; ScholarGPT; SciSpace; Claude; and DeepSeek	Current systems in the meteorological field offer tremendous support in areas such as cloud classification, map interpretation, and literature review support, but their performance remains inconsistent and varies across models and tasks	Standardized protocols must be established to evaluate their performance over time. Repeating tests across model updates and platforms will help determine if these tools can achieve the consistency and reliability required for broader adoption in meteorology and climate science	Future research should focus on improving the ability of AI models to interpret structured geo-scientific data, such as time series, gridded weather data, and integrated model outputs. Specialized evaluation benchmarks are also needed to reflect the complexities of specific domains, such as the WMO classification system or geospatial metadata interpretation

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Rivero et al. 2025 Spain [31]	Qualitative (Theoretical and Comparative Analysis)	Publicly Available Technical Literature, Academic Literature, and Official Policy Documents related to LLM Development and Deployment/Journal Articles	Environmental Sustainability	ChatGPT and DeepSeek	The study highlights certain security risks associated with the DeepSeek distilled model. It indicates that sustainability is no longer a marginal issue but is increasingly viewed as a crucial factor in the geopolitical agenda	While it is too early to definitively conclude that LLMs are the decisive axis of technological competition, current findings suggest that China is progressively adjusting its strategic focus toward more responsible innovation in the field of environmental sustainability	Parts of the study rely on developer information and white papers, which may not fully reflect the technical specifications or energy consumption data of ChatGPT and DeepSeek. Inherent opacity limits full comparability between models. The academic community must remain focused on empirically validated new data in the future. Although the article positions sustainability as a potential strategic advantage axis, this hypothesis has not been empirically tested through deployment or real-time performance measurement
Hou et.al 2025 China [32]	Quantitative (Questionnaire Survey, Back-translation Method)	Questionnaire from 260 High-tech Manufacturing Enterprises in China/Journal Articles	Decarbonization Capability, Environmental Performance	GenAI	Given the dual impact of GenAI on the environment, achieving sustainability through LLMs requires careful management of the technology's footprint, posing a key challenge for engineering managers	GenAI contributes to the technology-driven management literature in environmental sustainability and provides valuable insights to help companies take steps toward achieving carbon neutrality	Future research should collect historical data on sustainability initiatives and incorporate other research methods to further improve the accuracy of conclusions. It emphasizes the boundary conditions of environmental digitization but does not consider some contextual factors, including pressure from customer engagement, policy support, and industry pollution levels. Future research may include more contextual factors to systematically explore the boundary conditions for GenAI in unlocking environmental sustainability

Table 3. *Cont.*

Author/Year/Country	Methodology	Input and Output Content Type/Publication	Application	Model	Significance	Impact	Potential & Future Trends
Cheng et al. 2025 Taiwan, China [33]	Quantitative	Carbon emission formulas for various modes of transportation in 2021/Conference papers	Public Transport Carbon Emissions, Carbon Emissions, Carbon Reduction	ChatGPT	The study demonstrates the complexity and challenges of carbon emission calculation, emphasizing the importance of comparing and validating formulas from different sources. The potential of the ChatGPT model as an NLP technology is showcased in the context of carbon emission calculation	Potential to provide more accurate carbon emission estimation methods and recommendations, helping governments and relevant organizations formulate effective emission reduction policies and measures	Future calculations of carbon emissions for different transportation modes should consider factors like vehicle type and driving conditions in greater detail to improve calculation accuracy. A meticulous assessment of accuracy, reliability, interpretability, and practicality is required when selecting carbon emission calculation formulas
Zhang et al. 2025 China [34]	Quantitative (Integrated Energy System Model)	Problems of Supply–Demand Imbalance in Energy Systems, Operational Optimization Problems in Grid-Interactive High-Efficiency Commercial Buildings/Journal Articles	Efficient Energy Conversion and Utilization	Deep Reinforcement Learning (DRL), Integrated Energy System (IES), LLMs	Provides a new approach for improving the optimization and decision-making results of intelligent evolutionary systems in environmental sustainability, enhancing efficiency in environmental sustainability governance	The combined mechanism is specifically designed to support dynamic transactions and enhances decision performance	Consumer satisfaction, which was not explicitly reflected in the reward function in the research findings, leads to an incomplete analysis. Future research needs to design specific consumer satisfaction evaluation segments to ensure coverage of various consumer groups

3.1. Characteristics of Studies

This study included 20 final research documents for analysis, examining the publication trends and regional distribution characteristics of GenAI research in environmental sustainability (Figures 2 and 3). Temporally, these studies exhibit a significant exponential growth trend: the average publication volume increased from just 1 in 2023 to 5 in 2024, and then surged to 14 in 2025, with the 2025 output accounting for 70% of the total publications. This data signifies that the application of GenAI in the environmental sustainability domain has entered a rapid development phase. The low publication rate in the initial period (2023) can be attributed to the limited maturity and relatively low adoption rate of early GenAI technologies. The marked increase in 2025 is the result of a confluence of three major factors: LLM technological breakthroughs, the continuous aggravation of environmental challenges, and the gradual implementation of strategic national policies emphasizing urban environmental improvement.

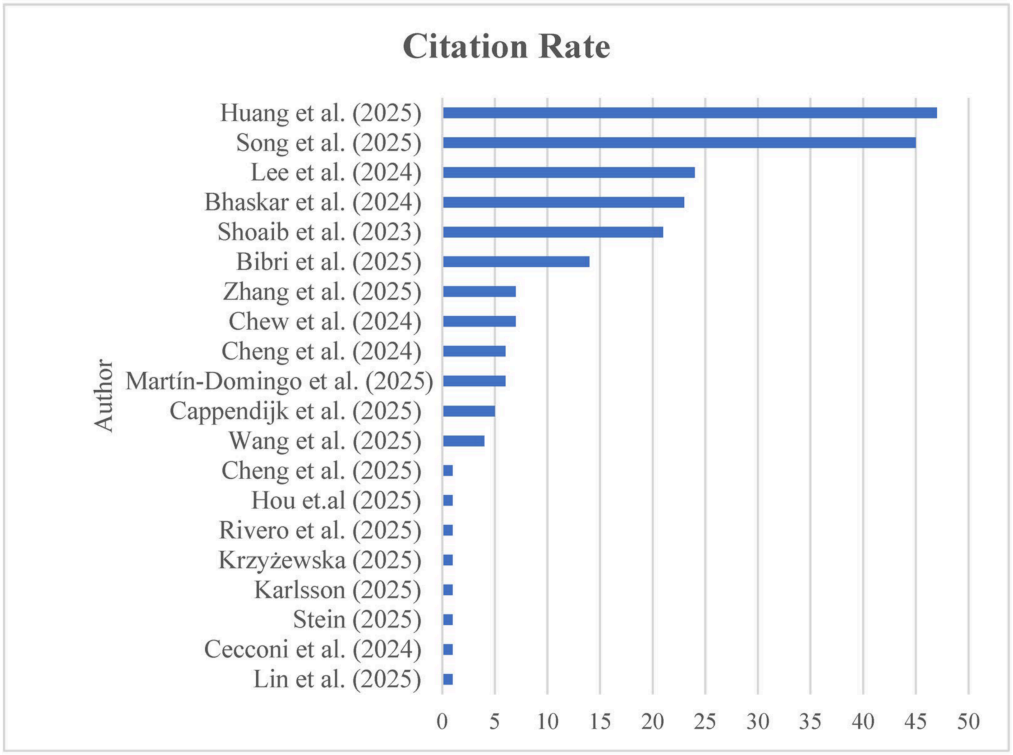

Figure 2. Citation Rate [7,16–34].

Specifically, the recent LLM technological breakthroughs include the shift from purely text-based processing to multimodal architectures (e.g., GPT-4 and beyond) and the significant enhancement of contextual reasoning and specialized fine-tuning techniques (such as RAG). These advancements enable LLMs to efficiently integrate and interpret the diverse data sources required for environmental monitoring and governance, including text, real-time sensor streams, and geospatial imagery, making them more practical, reliable, and cost-effective for professional environmental tasks. Simultaneously, this surge in research is driven by the intensifying global environmental crisis, characterized by the escalating frequency and severity of extreme weather events and the implementation of increasingly stringent ESG disclosure regulations. These urgent, high-stakes challenges create a robust demand for capabilities such as real-time monitoring, complex scenario modeling, and automated compliance reporting, which the latest generation of LLMs are uniquely positioned to address. The alignment of technological feasibility, the urgent necessity for

environmental sustainability, and supportive governmental policy drivers collectively fuels the acceleration of research in this domain.

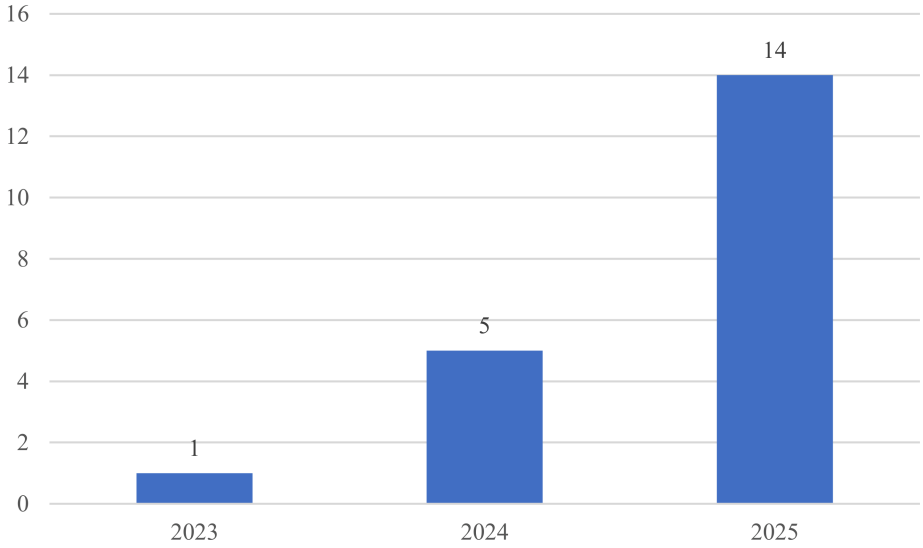

Figure 3. Annual number of publications.

It should be noted that the bibliometric data presented in this section, including the publication types distribution (Figure 4), are intended to descriptively map the current literature landscape rather than to establish statistical correlations or evaluate the substantive efficacy of LLMs in environmental contexts.

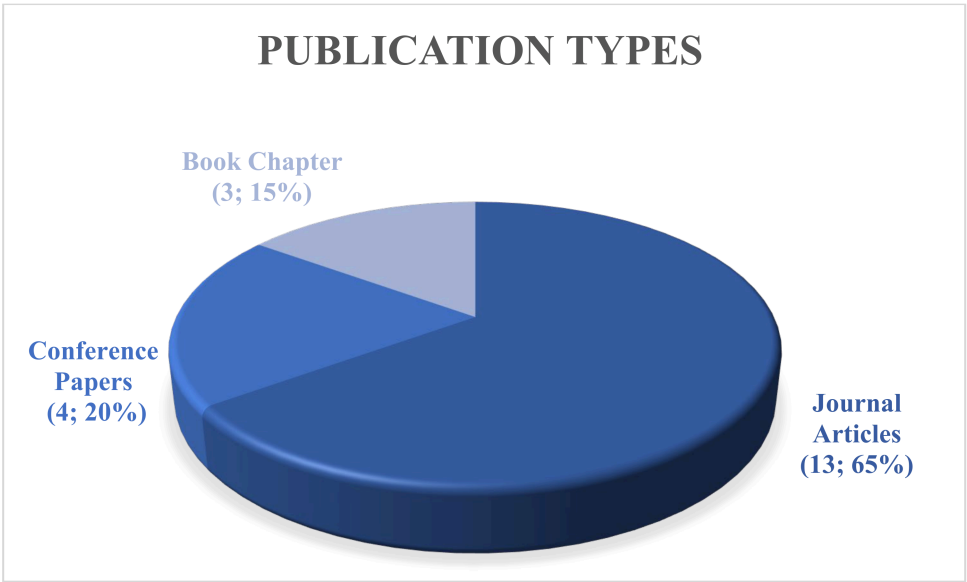

Figure 4. Publication types.

From a geographical perspective (Figure 5), China and the United States, as the world's two largest economies and technological leaders, collectively contributed nearly half of the publications, establishing their dominant position in the field of environmental sustainability. China's emergence as a crucial research front for LLM-enabled sustainability solutions is underpinned by its rapidly developing digital technology industry, robust national policies like the "Dual Carbon" strategy, and increasing governmental emphasis on environmental issues and the advancement of sustainability initiatives. The United States, leveraging its deep foundation in basic AI research, its mature green technology

innovation ecosystem, and long-term experience in addressing climate change and promoting sustainable development, has taken the lead in exploring how LLMs can serve environmental, social, and governance goals. Although smaller in size, Switzerland also published two articles, demonstrating strong competitiveness in specific frontier areas due to its reputation in sustainable finance, environmental protection technology, and an excellent research system.

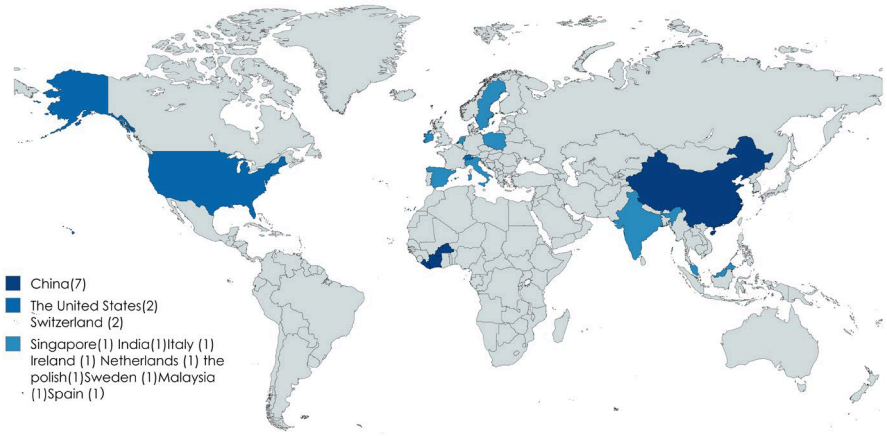

Figure 5. Countries of publication.

The remaining countries, including Singapore, India, Italy, Ireland, the Netherlands, Poland, Sweden, Malaysia, and Spain, each contributed one article. While the overall research volume from these nations is not comparable to China, the US, or Switzerland, they each present unique regional research characteristics. Certain developed European countries, such as Sweden, the Netherlands, Italy, Spain, and Ireland, often link their research closely with stringent EU environmental regulations and the Green Transition strategy, focusing on the cutting-edge exploration of LLMs in niche areas like the circular economy, sustainable supply chain management, climate modeling, and ethical alignment. Singapore, as a smart city and technology hub, naturally focuses its research on utilizing LLMs to optimize urban energy efficiency and resource management. Conversely, emerging economies like India and Malaysia, although their research is nascent, reflect the gradual diffusion of LLM applications in the environmental and sustainability domains towards the Global South. These countries are more concerned with leveraging LLMs as a cost-effective tool to solve localized sustainability challenges, such as addressing environmental pollution, enhancing natural disaster warning capabilities, or mitigating deficits in environmental regulatory capacity. Research from Eastern European nations like Poland is likely closely tied to its own energy structure transformation and the overall policy direction of the EU, hence its focus on the intersection of LLMs and environmental sustainability.

The global distribution pattern of LLM and sustainability research is collectively shaped by a multitude of factors, including technological research and development strength, the urgency of national sustainable development strategies, economic and research resources, and specific societal needs. The recent surge in research output correlates with actual breakthroughs in LLM technology and the growing global concern over the climate crisis.

3.2. Research Methodology

Regarding the research methodologies employed, this study primarily adopted quantitative methods, supplemented by qualitative and mixed-methods designs (see Table 4). According to the literature statistics, quantitative research methods were widely utilized, appearing a total of 13 times. Specific quantitative techniques included atmospheric simula-

tion systems, RAG models, the AirGPT framework, carbon emission formulas, IES models, machine learning data analysis, time series analysis, and regression analysis, among others. These methods were mainly used for model construction, environmental impact assessment, energy consumption forecasting, or simulating the effects of policy interventions, which reflects the environmental sustainability domain's emphasis on data-driven validation and model performance verification. Quantitative research serves as the primary evaluation framework for GenAI applications in environmental sustainability, leveraging structured data (e.g., descriptive statistics, relevant indicators) to directly verify the efficacy and underlying mechanisms of the tools. Specific statistical techniques included data analysis, POI calculation, descriptive statistics, MAPE, Coefficient of Determination, and the A20 index. Notably, assessment and validation methods (including MAPE, comparative discussions, and sampling surveys) appeared in three studies, highlighting a focus on standardized quantification of effects.

Table 4. Research Methodologies.

Methods	Type Frequency (n)	Specific Methodologies
Quantitative	13	Atmospheric simulation systems, RAG models, AirGPT framework, LLMs, Peak and Flops, Python mapping libraries, JSON modules, Folium Heat map plugins, data analysis, POI calculation, descriptive statistics, MAPE, Coefficient of Determination, A20 index, machine learning, flowchart, conceptual framework, large flow models, time series analysis, regression analysis, cloud identification and classification, carbon emission formulas, IES models
Qualitative	4	Predictive modeling, NLP, comparative discussion, theoretical and comparative analysis
Mixed	3	Sampling surveys, silicon wafer sampling, data collection, survey measurements, sustainable investigation reports, systematic review, LLMs, and manual control groups

In contrast, only 4 studies adopted qualitative research methods, mainly covering predictive modeling, NLP, and theoretical and comparative analysis. These studies typically focus on theoretical construction, comparison of technical pathways, or the in-depth exploration of the potential impacts of LLMs in sustainable governance. Mixed methods, utilized in 3 studies, combine the strengths of both quantitative and qualitative approaches. Typical implementations include combining sampling surveys with systematic reviews, or analyzing text data via LLMs supplemented by manual verification. This approach demonstrates good adaptability in assessing sustainable behavior, analyzing corporate environmental reports, or integrating multi-source data, with the combination of multiple methodologies enhancing the comprehensiveness and persuasiveness of the research. The cross-sectional research between LLMs and sustainability is currently in its nascent stage, with a methodological preference clearly leaning towards technical implementation and quantitative assessment, emphasizing model construction, simulation, and empirical analysis.

3.3. Types of Input and Output Data

In the included studies, the input data required by the models exhibit characteristics of diversification, massiveness, and high specialization. These data form the knowledge base for LLMs to conduct training, analysis, and decision-making in the sustainability domain (see Table 5).

Table 5. Types of Input and Output Data.

Data Type (Input/Output)	Frequency (n)	Core Function/Application Scenarios
Structured Professional Data	13	Climate change analysis, drought warning, and ecosystem health assessment; Simulating carbon emission distribution, planning green infrastructure, and evaluating the socio-economic impact of policies; Load forecasting, energy efficiency analysis, and grid balance optimization; Traffic carbon emission formulas to quickly and accurately calculate the environmental footprint of various activities
Text	4	Forming the knowledge base and contextual understanding essential for comprehending concepts, interpreting policies, tracking frontiers, and gaining insight into public sentiment. Rapidly reviewing academic literature to distill the latest research findings and technological trends; Analyzing corporate sustainability reports to assess ESG performance and greenwashing risks; Capturing public concern and discourse trends on environmental issues from news and social media
Real-time Data	5	Providing dynamic, real-world information for real-time monitoring, warning, and adaptive control. Integrating real-time meteorological data to issue warnings for extreme weather (e.g., typhoons, heatwaves) and generating emergency recommendations; Accessing real-time urban data (e.g., traffic, energy consumption) to dynamically optimize traffic signals and adjust energy distribution
Sustainable Problem-specific Data	3	Translating abstract sustainability challenges into specific task instructions that LLMs can comprehend and execute, guiding the model to solve highly specialized and complex industry problems. Enabling LLMs to generate concrete solutions for reducing energy consumption and enhancing system resilience; Guiding LLMs to propose innovative solution pathways by synthesizing their knowledge base and data analysis capabilities

The data type with the highest proportion (40%) is the text corpus, which includes academic papers, policy documents, corporate sustainability reports, news, and social media content. The composition of this data reflects the LLMs' reliance on vast amounts of human knowledge text to comprehend complex concepts, policy frameworks, and social dynamics within the sustainability sector. Structured professional data accounts for 35%, covering environmental parameters such as NDVI, soil moisture, energy data, geospatial information, and professional formula models. This type of data is the cornerstone for LLMs to perform precise analysis, simulation, and scientific decision-making when applied to specific scenarios, such as carbon emission calculation and energy system optimization. Real-time data streams (15%) and problem-defining data (10%), while representing a relatively smaller proportion, are critically important. Their existence represents the frontier direction of this research area: utilizing LLMs to process dynamic data from real-time meteorology, urban operations, and electricity markets, and providing real-time, dynamic solutions for specific optimization problems like building energy efficiency and grid dispatch.

The application of LLMs in the sustainability domain is gradually shifting from mere knowledge processing toward dynamic, complex system analysis that integrates

multi-source data (text, professional data, real-time data). This composition of input data underscores the high demands placed on models by the environmental sustainability field: they must not only understand natural language but also possess the comprehensive ability to handle professional data, respond to real-time dynamics, and solve specific industry challenges.

3.4. Application Domains

The current application areas can be synthesized into six core domains: Green transition, carbon emission management, air quality assessment, smart city operations, map analysis, and human cognition and behavioral observation (Figure 4). Green transition (4 articles): Researchers utilize the powerful information integration and simulation capabilities of LLMs to provide decision support for national and regional sustainable development pathways. The core focus here is processing massive volumes of policy documents, scientific reports, and market data to efficiently assess the impact of climate mitigation policies, plan innovative urban development solutions, and conduct climate risk assessments, thereby assisting in the formulation of scientific and forward-looking low-carbon development strategies. Carbon emission management (3 articles): As a core quantifiable dimension of sustainability, global carbon emission management requires high precision. LLMs in this domain serve a "Measurement, Reporting, and Verification" role. Researchers achieve precise calculation and dynamic monitoring of product carbon footprints and Greenhouse Gas emissions by interpreting and applying complex models like traffic carbon emission formulas, and integrating real-time operational data provided by government agencies. They can also simulate low-carbon power scenarios and carbon market dynamics, providing critical data support for emission reduction optimization. Air quality assessment (3 articles): Given the increasing severity of global environmental quality issues, air quality serves as a key indicator of environmental health. LLMs are applied here by fusing environmental parameters such as wind speed with geospatial information to achieve precise atmospheric assessment and in-depth attribution analysis. The application scenarios extend beyond routine air quality evaluation to tracing atmospheric Greenhouse Gas emission sources, incorporating the impact of sudden environmental events like forest fires into comprehensive analysis, and generating insightful assessment reports. Smart city operations (3 articles): Within the smart city paradigm, LLMs often function as the "central nervous system" of urban operations. Studies in this area focus on processing real-time data streams from ITS, vehicle technologies, and urban sensors to analyze and optimize the built environment and public transit ridership. This synergy enhances the resource efficiency of systems like traffic and energy utilization, facilitating efficient, low-carbon urban operation models.

As illustrated in Figure 6, the role of LLMs in environmental sustainability is inherently paradoxical. On the enabling side, the six core application domains identified in this review, green transition, carbon emission management, air quality assessment, smart city operations, geospatial mapping, and human cognition, collectively demonstrate LLMs' capacity to accelerate environmental governance at multiple scales. Conversely, the same models impose measurable constraints through high energy consumption, substantial carbon footprints during training and inference, hallucination risks in high-stakes environmental decisions, embedded algorithmic biases, and data privacy concerns. This dual-role structure necessitates a lifecycle-aware cost–benefit analysis when deploying LLMs for sustainability purposes, a methodological direction that future research should operationalize with domain-specific benchmarking.

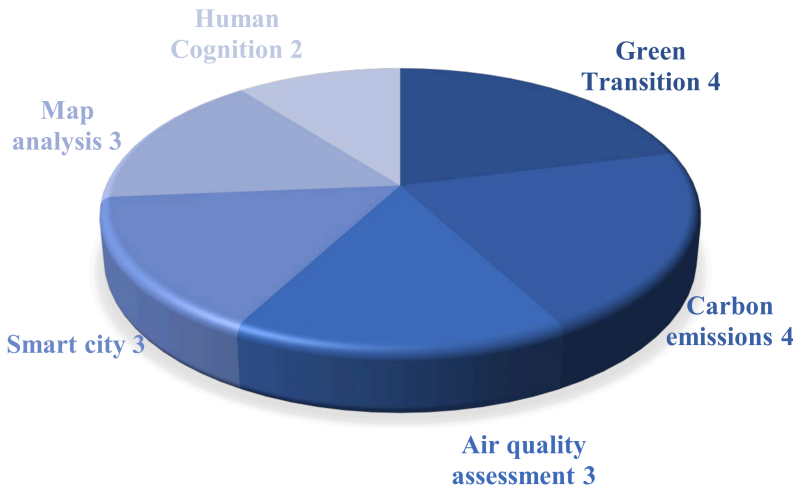

Figure 6. Application Domains.

4. Discussion

4.1. Main Findings and Results of Studies

LLMs demonstrate significant potential and application value in advancing sustainability across multiple domains. In the domains of environmental quality management and disaster risk response, LLMs enhance the prediction and response capabilities for natural or public health disasters, such as earthquakes, dengue fever, and forest fires, by improving geographical monitoring autonomy, optimizing data processing efficiency, and visual analysis. This saves researchers considerable time and labor costs and increases research efficiency. Maps generated by LLMs allow researchers to clearly understand the distribution and clustering of disasters in different regions. This substantially improves the timeliness, consistency, scalability, and standardization of data processing while automating repetitive tasks and assisting in data analysis.

In the process of urban and energy system transitions, LLMs provide critical technical support for the development of sustainable smart cities. By integrating GenAI and the AIoT system, they optimize energy management, traffic planning, and low-carbon grid operations, thereby contributing to the construction of more efficient, resilient, and sustainable urban environmental frameworks. They hold potential for driving sustainable smart city development, enabling researchers to cultivate smarter, more energy-efficient, adaptive, secure, robust, and autonomous AIoT ecosystems through the strategic application of Generative intelligence. Furthermore, LLMs show reliability in automatically extracting KPIs from corporate reports, providing a new pathway for evidence-based policymaking and corporate environmental performance assessment.

The credibility of this synthesis is supported by the high quality of the included literature. The predominant MMAT scores of 85.71% to 100% across various research designs (qualitative, RCTs, non-randomized, etc.) indicate a low risk of bias in the foundational data. Although some studies lacked detailed reporting on specific appraisal criteria (e.g., assessor blinding in RCTs), the consistency of findings across these high-quality studies strengthens the overall robustness of our conclusions.

Despite existing global concerns about the energy consumption and carbon footprint generated by these models, current LLMs can still effectively support climate model interpretation, public communication, and education, and, in specific contexts, assist in analyzing public attitudes and beliefs towards global warming. Relevant research in the carbon emission and transition domain has demonstrated the complexity and challenges of carbon emission calculation, emphasizing the importance of comparing and validating

formulas from different sources. The potential of LLMs as an NLP technology is thus demonstrated in the context of carbon emission calculation. LLMs are both powerful tools for achieving environmental and social sustainability goals, and their own development must be integrated into the scope of global environmental sustainability considerations. In the following sections, we will focus on the implications, implications, limitations and future trends of LLMs on Environmental Sustainability.

While the NLP-based clustering successfully maps the most frequently discussed obstacles, such as data constraints, computational costs, hallucination, and security, frequency-driven methods may inadvertently overshadow barriers that are less frequently mentioned but possess profound socio-psychological impacts. To address this methodological limitation and provide a more holistic view, a supplementary qualitative analysis of the included studies was conducted.

Beyond technical and economic hurdles, significant socio-psychological barriers exist in deploying LLMs for environmental sustainability. Foremost is the "trust deficit" and public skepticism toward AI-driven environmental policies. Environmental decision-making often involves high stakes and ethical trade-offs; public acceptance is severely hindered when AI models operate as "black boxes" lacking human-like moral accountability. Furthermore, there is an emerging concern of "deskilling" and over-reliance on AI, which could lead to the gradual erosion of traditional, indigenous ecological knowledge and human intuition in conservation efforts. Although these socio-psychological and socio-ethical factors show lower co-occurrence frequencies in the current NLP analysis, reflecting a technical bias in the existing literature, they are critical bottlenecks that must be addressed for the successful real-world adoption of LLMs in sustainability.

4.2. Impact of LLMs on Environmental Sustainability

LLMs exhibit remarkable capabilities in the field of environmental science, enabling researchers to provide accurate regulatory information, perform data analysis, and generate location-specific management recommendations [11]. By strictly adhering to scientific sources, they are used in areas such as air quality policy, disaster monitoring, and forest fire factor analysis, offering strong data support for precise and efficient environmental governance [35]. LLMs show significant potential in optimizing code for lower energy consumption, optimizing grid operations to integrate renewable energy, and reducing urban traffic carbon emissions by promoting land use and public transportation, providing innovative pathways for smart city traffic transition and low-carbon energy management [36]. The gradual iteration of various LLMs has greatly lowered the application threshold for data analysis and remote sensing technologies, allowing non-professionals to quickly handle complex scientific datasets, thereby accelerating the popularization and application of sustainability knowledge and lowering the entry barrier for researchers [37]. Concurrently, LLMs can simulate public opinions; despite certain algorithmic fidelity issues, they can still translate complex climate model results into easily understandable narratives, thereby improving policy formulation and promoting public participation in climate action, fostering the fusion of technology and environmental science [20]. By optimizing resource allocation, accelerating technological innovation, and enhancing governance capacity, LLMs are becoming key accelerators in achieving sustainability goals. Nevertheless, the models' inherent resource consumption, ethical risks, and technical uncertainties also introduce new, unresolved issues for sustainable development.

While this study acknowledges the environmental costs associated with deploying large language models, a more rigorous treatment of this trade-off is warranted. Drawing on lifecycle assessment (LCA) frameworks, the environmental burden of LLM usage in sustainability can be situated within a broader sustainability calculus. Critically, however,

these costs must be weighed against the efficiency gains and reduced resource consumption that LLM-assisted approaches offer compared to conventional alternatives in sustainability. We therefore propose framing this tension using a cost–benefit lens informed by LCA principles: while the per-query carbon footprint of LLM inference is non-trivial, the aggregate benefit may yield a net positive environmental or economic outcome under certain deployment conditions. Future work should operationalize this comparison with domain-specific benchmarking data to provide a more definitive assessment.

4.3. Role and Significance of LLMs for Environmental Sustainability

LLMs demonstrate powerful data parsing and automation capabilities in the fields of environmental governance and disaster response. They can process satellite remote sensing images, sensor data, and various environmental parameters, not only significantly improving the automation level of dengue fever epidemic monitoring and earthquake risk assessment but also accurately identifying the dominant factors of forest fires [24], establishing a scientific basis for disaster warning and resource allocation [38]. In carbon emission and energy management, LLMs serve as key quantification and optimization tools [39]. They can automatically extract carbon emission KPIs from corporate sustainability reports, assist in complex carbon footprint calculation [40], and optimize grid operations and forecast energy demand patterns in low-carbon energy management [28], continuously promoting the green transition of energy systems. In the dimension of information dissemination, LLMs play the vital role of a "translator [41]." They can convert complex climate model outputs into expressions easily understood by the public, effectively aiding the popularization of climate education and enhancing public participation [42]. They can also quickly gain insight into public attitudes towards global warming by analyzing text data from social media, providing support for formulating more targeted climate communication strategies [43]. As intelligent decision-making hubs, LLMs integrate GenAI and AIoT technology to optimize intelligent transportation systems and improve resource utilization efficiency [44], thereby providing strong support for building more efficient and resilient urban development strategies [45]. However, the training and operation of LLMs are accompanied by significant energy consumption and carbon footprints, creating an intrinsic conflict with the sustainability goals they aim to promote [39]. Furthermore, challenges such as inconsistent performance, limited accuracy in specific tasks, and latent security risks exist in the application of these models [46]. This necessitates prudent management of the technology's ecological footprint, seeking a dynamic balance between driving technological innovation and ensuring environmental sustainability.

4.4. Limitations of LLMs Applications in Environmental Sustainability

Currently, the primary tasks for LLMs are to address the issues of accuracy and hallucination in generated content and to enhance their ability to parse structured scientific data [47], such as time series and gridded weather data [48]. Their training and operation processes consume vast amounts of water and energy, generate a substantial carbon footprint, and impose pressure on land resources, forming a direct conflict with the core goals of sustainable development [49]. To address this, there is a need to establish specialized domain evaluation benchmarks and systematically validate model performance through simulation testing, prototyping, and real-world deployment [50].

A core development trend involves squarely addressing and quantifying the energy consumption and environmental footprint generated by LLMs during training and operation [49]. Currently, LLMs still suffer from insufficient algorithmic fairness and data dimensionality limitations [51]. The use of commercial LLM APIs may also lead to data public disclosure outside the jurisdiction, potentially violating relevant regulations and

posing a risk of sensitive information leakage to external systems [52]. Therefore, a cautious approach is essential to address the embedded algorithmic biases in LLMs, specifically by proactively conducting sustainable algorithm audits, implementing bias mitigation measures, and striving to collect higher-quality, more representative data [53]. It is particularly important to strengthen research across cultural and institutional contexts to break through the limitations of relying on secondary data in the environmental sustainability domain [54]. A clear future development trend is that LLMs will be positioned as "decision support tools," rather than "substitutes for human judgment [55]." For researchers, the key lies in establishing effective human–machine collaboration frameworks, allowing users to verify model outputs with professional discretion and combining tool capabilities with human insights to jointly drive the green transition process.

To provide a more granular understanding of LLM efficacy, we categorized the findings into key environmental sustainability sub-topics, revealing distinct application patterns: LLMs demonstrate high efficacy in carbon accounting and climate policy synthesis. By automating the extraction of ESG metrics and emissions data from vast corporate reports, models like GPT-4 and DeepSeek-R1 enable more transparent monitoring of decarbonization efforts. In the context of energy and materials, LLMs are primarily utilized for optimization. The evidence suggests that LLMs can significantly reduce energy consumption in built environments by integrating with Digital Twins to provide proactive climate control and adaptive lighting solutions. The application of domain-specific models, such as AirGPT, highlights a shift toward specialized AI. These models are particularly effective in interpreting complex air quality indices and providing actionable policy recommendations, outperforming general-purpose models in technical accuracy. While less prevalent than urban-focused studies, LLMs are emerging as powerful tools for ecological documentation and remote sensing data interpretation. They facilitate the analysis of land-use changes and assist in mapping disaster-prone areas, though spatial–temporal reasoning remains a challenge.

4.5. Potential and Future Trends of LLMs Applications in Environmental Sustainability

The potential of LLMs in the environmental sustainability domain is reflected in the ability to expand their application scope to broader sustainability indicators, different industries, and multilingual scenarios [56]. Especially when deeply integrated with structured data extraction tools, they will strongly promote the automation process of ESG data, thereby better adapting to rapidly changing regulatory requirements [57]. This also inspires researchers to adopt such methods in the future, further optimizing research paradigms and exploring the integration pathways of LLMs with cutting-edge technologies like the AIoT and digital twins to build more powerful intelligent systems [58]. Simultaneously, efforts should focus on unlocking the potential of LLMs in real-time decision-making scenarios, such as dynamic traffic management and complex climate modeling [59]. By incorporating a wider range of impact factors and analyzing their associated effects, LLMs are expected to upgrade from mere analytical tools to intelligent decision support systems capable of handling complex systemic problems [60]. NLP tasks have been improved greatly by LLMs. However, numerous parameters make their execution computationally expensive and difficult on resource-constrained devices. For this problem, as well as maintaining accuracy, Zheng, etc. also proposed a novel distilled lightweight model for BERT named MicroBERT, and have observed that this method significantly reduces the model size and inference time as well as without sacrificing high accuracy [61]. The future development of LLMs in the sustainability domain is inseparable from the support of interdisciplinary collaboration, it requires continuously enhancing their technical capabilities, reliability, and application scope, while simultaneously addressing the environmental impacts, ethical

risks, and data limitations they introduce, ultimately guiding them to become responsible tools for promoting the sustainability process.

4.6. Comparative Synthesis of Performance and Efficacy

A critical synthesis of the evidence reveals significant performance gradients across different LLM architectures and applications. Quantitative data from the included studies suggest that the efficacy of LLMs is not uniform. In document-heavy domains like ESG reporting and policy analysis, models such as GPT-4o and DeepSeek-R1 demonstrate near-human accuracy in information extraction, significantly reducing labor time.

However, a comparative analysis shows that domain-specific fine-tuned models consistently outperform general-purpose models in specialized environmental tasks, reducing hallucination rates. This evidence provides the necessary quantitative support for our conclusion that LLMs are a positive driver for sustainability, provided they are appropriately calibrated to the specific technical demands of the environmental sub-domain.

4.7. Patterns of Success and Failure: An Interpretative Analysis

Beyond simple descriptions, our analysis identifies distinct patterns that determine the success or failure of LLM deployment.

LLMs exhibit maximum efficacy in tasks involving unstructured text synthesis, code generation for energy optimization, and multi-criteria decision support. These successes are characterized by the models' ability to bridge the gap between complex climate data and public communication. Conversely, LLM efficacy diminishes in tasks requiring precise spatial–temporal reasoning, or physics-based causal modeling. These "failure modes" are primarily due to numerical hallucinations and a lack of inherent physical world knowledge.

Recognizing these patterns is crucial: the "positive role" of LLMs is most robust in their capacity as cognitive assistants and information aggregators, while their role as autonomous environmental decision-makers is still constrained by fundamental reasoning limitations.

4.8. Limitations

This systematic scoping review aims to comprehensively synthesize the application practices of LLMs in the field of environmental sustainability; however, the study still presents several limitations. Firstly, despite employing an extensive set of search terms and covering six core academic databases, the search strategy may not exhaust all relevant literature. Specifically, gray literature not indexed in databases and newly published research in non-core journals may have been missed. Secondly, the literature included in this review is predominantly in English, which may lead to the omission of research published in other languages that holds significant practical value in specific geographical regions. Finally, due to the obvious heterogeneity in the research designs and application domains of the included literature, this review primarily utilizes a narrative synthesis approach rather than a meta-analysis, which to some extent limits the in-depth quantitative comparison of the effects of related interventions. Based on these limitations, future research can be improved by expanding database coverage, conducting multilingual literature searches, and regularly updating the review content.

Another important limitation to acknowledge is the exclusion of 69 studies due to the inaccessibility of their full texts. Although their titles and abstracts suggested potential relevance during the initial screening phase, the inability to review their complete methodologies and empirical data necessitated their exclusion to maintain the rigorous standards of this scoping review. Given the consistent trends and overarching patterns identified among the 20 fully assessed and highly representative studies, we assess that the

absence of these inaccessible records is unlikely to fundamentally skew the core findings of this review.

One of further methodological limitation concerns the operationalization of "applied research" as an inclusion criterion. While review articles were systematically excluded, a subset of included studies rely substantially on secondary data or adopt a conceptual-analytical orientation that does not constitute fully independent empirical investigation in the traditional sense. Although these studies were retained on the grounds that they deploy LLMs as an active subject of structured analysis and contribute evidence directly relevant to the review's research questions, their inclusion introduces a degree of heterogeneity in the empirical rigor of the evidence base. To mitigate this, we have introduced a post hoc empirical/conceptual-applied classification, which is reflected in the interpretive framing of the synthesis. Future scoping reviews in this domain are encouraged to adopt a pre-registered, more granular typology of research designs, distinguishing, for instance, between system-level implementation studies, evaluation studies, and applied analytical studies—to enhance selection reproducibility and evidence coherence.

Furthermore, it is critical to acknowledge that the quantitative data and visual mappings presented in this study, e.g., publication trends, geographic distribution, and publication types, are inherently descriptive. Due to the nascent stage of this intersectional field and the high heterogeneity of current studies, standardized statistical interpretation or rigorous uncertainty analysis, such as confidence intervals or effect sizes, could not be robustly performed. Consequently, claims regarding the transformative impact of LLMs should be interpreted with caution, as they currently lack the rigorous empirical and statistical backing required to definitively establish their efficacy in complex, real-world environmental systems.

5. Conclusions

This study systematically synthesized and analyzed the current application status of LLMs in the field of environmental sustainability. The systematic synthesis indicates that while LLMs show emerging potential in environmental governance, energy system optimization, and public environmental cognition, comprehensive empirical validation remains limited. Based on the currently available literature, LLMs appear to offer operational advantages in multimodal data integration and low-carbon transition pathway modeling; however, these findings are primarily derived from early-stage conceptual or controlled applications rather than large-scale, real-world deployments. However, the training and operation processes of LLMs consume substantial energy, creating a certain intrinsic conflict with the sustainable development goals they promote. Given that LLMs have lowered the threshold for complex data processing to the level of natural language interaction, their development potential is concentrated in expanding their application scope to broader sustainability indicators, different industries, and multilingual scenarios. This also inspires researchers to adopt such technological pathways in the future, focusing on unlocking the potential of LLMs in real-time decision-making scenarios such as dynamic traffic management and complex climate modeling. The future development of LLMs in the sustainability domain is inseparable from the support of interdisciplinary collaboration; it requires continuously enhancing their technical capabilities, reliability, and application scope, while simultaneously addressing the environmental impacts, ethical risks, and data limitations they introduce, ultimately positioning them as responsible tools for promoting the sustainability process.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su18094327/s1, Table S1: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.

Author Contributions: X.S., T.L. and Y.T.L. proposed the idea to conceive and design this study. X.S. analyzed the data. X.S. completed the manuscript with the assistance of T.L., P.P. and D.W. supervised students and oversaw this work. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Macao Science and Technology Development Fund (FDCT; funding ID: 0029/2025/AIJ) and Macao Polytechnic University research grant (project code: RP/FCA-24/2025).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no competing interests.

Appendix A

Type of Research Design										
1. QUALITATIVE STUDIES										
NO.	Authors (year)	S1.	S2.	1.1.	1.2.	1.3.	1.4.	1.5.	Score	Study Quality
[17]	Shoaib et al. (2023)	+	+	+	+	+	+	?	85.71%	Good
[21]	Cecconi et al. (2024)	+	+	+	+	+	+	?	85.71%	Good
[24]	Stein (2025)	+	+	+	+	+	+	+	100%	Good
[31]	Rivero et al. (2025)	+	+	+	+	+	+	–	85.71%	Good
2. RANDOMIZED CONTROLLED TRIALS										
NO.	Authors (year)	S1.	S2.	2.1.	2.2.	2.3.	2.4.	2.5.	Score	Study Quality
[22]	Bhaskar et al. (2024)	+	+	+	+	+	?	+	85.71%	Good
[26]	Bibri et al. (2025)	+	+	+	+	+	?	+	85.71%	Good
[28]	Cheng et al. (2024)	+	+	+	+	+	?	+	85.71%	Good
[33]	Cheng et al. (2025)	+	+	+	+	+	?	+	85.71%	Good
3. NON-RANDOMIZED STUDIES										
NO.	Authors (year)	S1.	S2.	3.1	3.2.	3.3.	3.4.	3.5.	Score	Study Quality
[18]	Song et al. (2025)	+	+	+	+	+	+	+	100%	Good
[19]	Cappendijk et al. (2025)	+	+	+	+	+	+	+	100%	Good
[16]	Lin et al. (2025)	+	+	+	+	+	?	+	85.71%	Good
[23]	Wang et al. (2025)	+	+	+	+	+	+	+	100%	Good
[24]	Chew et al. (2024)	+	+	+	+	+	+	+	100%	Good
[27]	Huang et al. (2025)	+	+	+	+	+	+	+	100%	Good
[30]	Krzyżewska (2025)	+	+	+	+	+	+	+	100%	Good
[34]	Zhang et al. (2025)	+	+	+	+	+	+	+	100%	Good
4. QUANTITATIVE DESCRIPTIVE STUDIES										
NO.	Authors (year)	S1.	S2.	4.1.	4.2.	4.3.	4.4.	4.5.	Score	Study Quality
[32]	Hou et al. (2025)	+	+	+	+	+	+	+	100%	Good
5. MIXED METHODS STUDIES										
NO.	Authors (year)	S1.	S2.	5.1.	5.2.	5.3.	5.4.	5.5.	Score	Study Quality
[20]	Lee et al. (2024)	+	+	+	+	+	+	+	100%	Good
[7]	Martín-Domingo et al. (2025)	+	+	+	+	+	+	+	100%	Good
[29]	Karlsson (2025)	+	+	+	+	+	?	+	85.71%	Good

"+" = Yes; "–" = No; "?"= Can't tell.

References

1. Luo, Y.; Pang, P.C.-I.; Chang, S. Enhancing exploratory learning through exploratory search with the emergence of large language models. *arXiv* **2024**, arXiv:2408.08894.
2. Mao, C.; Li, J.; Pang, P.C.-I.; Zhu, Q.; Chen, R. Identifying Kidney Stone Risk Factors Through Patient Experiences with a Large Language Model: Text Analysis and Empirical Study. *J. Med. Internet Res.* **2025**, *27*, e66365. [CrossRef] [PubMed]
3. Annepaka, Y.; Pakray, P. Large language models: A survey of their development, capabilities, and applications. *Knowl. Inf. Syst.* **2025**, *67*, 2967–3022. [CrossRef]
4. Han, S.; Wang, M.; Zhang, J.; Li, D.; Duan, J. A review of large language models: Fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges. *Electronics* **2024**, *13*, 5040. [CrossRef]
5. Peduzzi, P. The disaster risk, global change, and sustainability nexus. *Sustainability* **2019**, *11*, 957. [CrossRef]
6. Salmi, A.; Jussila, J.; Hämäläinen, M. The role of municipalities in transformation towards more sustainable construction: The case of wood construction in Finland. *Constr. Manag. Econ.* **2022**, *40*, 934–954. [CrossRef]
7. Martín-Domingo, L.; Fernandez, J.B.; Efthymiou, M.; Ali, M.I. Extracting airline emission KPIs from sustainability reports using large language models (LLMs). *Transp. Res. Interdiscip. Perspect.* **2025**, *33*, 101599. [CrossRef]
8. Taheri Hosseinkhani, N. Artificial Intelligence and Large Language Models in Energy Systems and Climate Strategies: Economic Pathways to Cost-Effective Emissions Reduction and Sustainable Growth. *SSRN Electron. J.* 2025. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5385513 (accessed on 12 January 2026).
9. Al Khourdajie, A. The role of artificial intelligence in climate change scientific assessments. *PLoS Clim.* **2025**, *4*, e0000706. [CrossRef]
10. Cao, C.; Zhuang, J.; He, Q. LLM-Assisted Modeling and Simulations for Public Sector Decision-Making: Bridging Climate Data and Policy Insights. In Proceedings of the AAAI—2024 Workshop on Public Sector LLMs: Algorithmic and Sociotechnical Design, Vancouver, BC, Canada, 27 February 2024.
11. Van de Weghe, N.; De Sloover, L.; Cohn, A.; Huang, H.; Scheider, S.; Sieber, R.; Timpf, S.; Claramunt, C. Opportunities and challenges of integrating geographic information science and large language models. *J. Spat. Inf. Sci.* **2025**, *30*, 93–116. [CrossRef]
12. Everman, B.; Villwock, T.; Chen, D.; Soto, N.; Zhang, O.; Zong, Z. Evaluating the carbon impact of large language models at the inference stage. In *Proceedings of the 2023 IEEE International Performance, Computing, and Communications Conference (IPCCC), Anaheim, CA, USA, 17–19 November 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 150–157.
13. Alizadeh, N.; Castor, F. Green AI: A preliminary empirical study on energy consumption in dl models across different runtime infrastructures. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI, Lisbon, Portugal, 14–15 April 2024; pp. 134–139.
14. McGowan, J.; Straus, S.; Moher, D.; Langlois, E.V.; O'Brien, K.K.; Horsley, T.; Aldcroft, A.; Zarin, W.; Garitty, C.M.; Hempel, S. Reporting scoping reviews—PRISMA ScR extension. *J. Clin. Epidemiol.* **2020**, *123*, 177–179. [CrossRef]
15. Hong, Q.N. Revision of the Mixed Methods Appraisal Tool (MMAT): A Mixed Methods Study. Ph.D. thesis, McGill University, Montréal, QC, Canada, 2018.
16. Lin, W.-C.; Tseng, M.-H. Autonomous Epidemic and Geographic Disaster Mapping: Assessing the Performance of Large Language Models in Spatial Information Integration. *J. Disaster Res.* **2025**, *20*, 386–395. [CrossRef]
17. Shoaib, M.R.; Emara, H.M.; Zhao, J. A survey on the applications of frontier ai, foundation models, and large language models to intelligent transportation systems. In *Proceedings of the 2023 International Conference on Computer and Applications (ICCA), Cairo, Egypt, 28–30 November 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–7.
18. Song, J.; Ma, C.; Ran, M. AirGPT: Pioneering the convergence of conversational AI with atmospheric science. *npj Clim. Atmos. Sci.* **2025**, *8*, 179. [CrossRef]
19. Cappendijk, T.; de Reus, P.; Oprescu, A. An exploration of prompting LLMs to generate energy-efficient code. In *Proceedings of the 2025 IEEE/ACM 9th International Workshop on Green and Sustainable Software (GREENS), Ottawa, ON, Canada, 29 April 2025*; IEEE Computer Society: Washington, DC, USA, 2025; pp. 31–38.
20. Lee, S.; Peng, T.-Q.; Goldberg, M.H.; Rosenthal, S.A.; Kotcher, J.E.; Maibach, E.W.; Leiserowitz, A. Can large language models estimate public opinion about global warming? An empirical assessment of algorithmic fidelity and bias. *PLoS Clim.* **2024**, *3*, e0000429. [CrossRef]
21. Cecconi, F.; Marconi, L.; Barazzetti, A. *Climate Change Mitigation Policies Using GPT-4*; MISC: Berlin/Heidelberg, Germany, 2024.
22. Bhaskar, P.; Seth, N. Environment and sustainability development: A ChatGPT perspective. In *Applied Data Science and Smart Systems*; CRC Press: Boca Raton, FL, USA, 2024; pp. 54–62.
23. Wang, Z.; Zheng, X.; Meng, F.; Wang, K.; Wu, X.; Yu, D. Exploring the Joint Influence of Built Environment Factors on Urban Rail Transit Peak-Hour Ridership Using DeepSeek. *Buildings* **2025**, *15*, 1744. [CrossRef]
24. Chew, Y.J.; Ooi, S.Y.; Pang, Y.H.; Lim, Z.Y. Framework to create inventory dataset for disaster behavior analysis using google earth engine: A Case Study in Peninsular Malaysia for historical forest fire behavior analysis. *Forests* **2024**, *15*, 923. [CrossRef]

25. Stein, A.L. Generative AI and Sustainability. In *The Oxford Handbook of the Foundations and Regulation of Generative AI*; Hacker, P., Engel, A., Hammer, S., Mittelstadt, B., Eds.; Oxford University Press: Oxford, UK, 2025. [CrossRef]
26. Bibri, S.E.; Huang, J. Generative AI of Things for sustainable smart cities: Synergies in cognitive augmentation, resource efficiency, network traffic, and anomaly and threat detection for environmental optimization. *Sustain. Cities Soc.* **2025**, *133*, 106826. [CrossRef]
27. Huang, J.; Bibri, S.E.; Keel, P. Generative spatial artificial intelligence for sustainable smart cities: A pioneering large flow model for urban digital twin. *Environ. Sci. Ecotechnol.* **2025**, *24*, 100526. [CrossRef]
28. Cheng, Y.; Zhou, X.; Zhao, H.; Gu, J.; Wang, X.; Zhao, J. Large Language Model for Low-Carbon Energy Transition: Roles and Challenges. In *Proceedings of the 2024 4th Power System and Green Energy Conference (PSGEC), Shanghai, China, 22–24 August 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 810–816.
29. Karlsson, J.; Käck, J. Targeting Green Prospects: Identifying Environmentally Conscious Prospects Using AI-driven Tools Within the Swedish Energy Sector. 2025. Available online: https://www.diva-portal.org/smash/get/diva2:1964601/FULLTEXT01.pdf (accessed on 8 January 2026).
30. Krzyżewska, A. The applications of ai tools in the fields of weather and climate—Selected examples. *Atmosphere* **2025**, *16*, 490. [CrossRef]
31. Rivero, S.; Chinarro Vadillo, D.; Prieto Andres, A. The green algorithm: Can sustainability define the winner in the AI race? *Front. Political Sci.* **2025**, *7*, 1629914. [CrossRef]
32. Hou, Y.; Yang, S.; Li, L.; Chen, L. Unlocking Environmental Sustainability with Generative Artificial Intelligence: Insights from Resource Orchestration Theory. *IEEE Trans. Eng. Manag.* **2025**, *72*, 3080–3093. [CrossRef]
33. Cheng, Y.H.; Wang, Y.W.; Kuo, C.N. The Potential and Applications of Utilizing the ChatGPT Model for Comparative Analysis of Carbon Emission Calculation Formulas in Public Transportation. In Proceedings of the 2023 12th International Conference on Awareness Science and Technology (iCAST), Taichung, Taiwan, 9–11 November 2023; pp. 31–34.
34. Zhang, L.; Yue, D.; Hancke, G.P.; Dou, C.; Yu, L.; Chen, Z. Optimization of Energy and Carbon Emissions in Integrated Energy System Based on Deep Reinforcement Learning Assisted by Large Language Model. *IEEE Trans. Ind. Inf.* **2025**, *21*, 8186–8197. [CrossRef]
35. Ncube, M.M.; Ngulube, P. Enhancing environmental decision-making: A systematic review of data analytics applications in monitoring and management. *Discov. Sustain.* **2024**, *5*, 290. [CrossRef]
36. Wen, J.; Zhang, R.; Niyato, D.; Kang, J.; Du, H.; Zhang, Y.; Han, Z. Generative AI for low-carbon artificial intelligence of things with large language models. *IEEE Internet Things Mag.* **2024**, *8*, 82–91. [CrossRef]
37. Tao, L.; Zhang, H.; Jing, H.; Liu, Y.; Yan, D.; Wei, G.; Xue, X. Advancements in vision–language models for remote sensing: Datasets, capabilities, and enhancement techniques. *Remote Sens.* **2025**, *17*, 162. [CrossRef]
38. Lei, Z.; Dong, Y.; Li, W.; Ding, R.; Wang, Q.R.; Li, J. Harnessing large language models for disaster management: A survey. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2025, Vienna, Austria, 27 July–1 August 2025; ACL: San Diego, CA, USA, 2025; pp. 14528–14551.
39. Jiang, P.; Sonne, C.; Li, W.; You, F.; You, S. Preventing the immense increase in the life-cycle energy and carbon footprints of LLM-powered intelligent chatbots. *Engineering* **2024**, *40*, 202–210. [CrossRef]
40. Folke, O.; Ivan Erik Troedsson, A. How Effectively Can AI Be Applied to Extract ESG-Related KPIs from Annual Reports? 2025. Available online: https://www.diva-portal.org/smash/get/diva2:1985641/FULLTEXT01.pdf (accessed on 29 January 2026).
41. Zhong, Y.; Zhao, K. Application and Research of Large Language Model in Foreign Language Translation. In *Proceedings of the 2024 International Conference on Information Technology, Comunication Ecosystem and Management (ITCEM), Bangkok, Thailand, 20–22 December 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 63–68.
42. Bina, R.; Luong, K.; Mehta, S.; Pang, D.; Xie, M.; Chou, C.; Kimbrough, S.O. On Large Language Models as Data Sources for Policy Deliberation on Climate Change and Sustainability. *arXiv* **2025**, arXiv:2503.05708. [CrossRef]
43. Rahman, G.; Fitriyah, A. Harnessing AI for Climate Change Communication: Analyzing Public Perception through NLP and Machine Learning. *Sinergi Int. J. Commun. Sci.* **2025**, *3*, 87–98. [CrossRef]
44. Sharif, S.; Zeadally, S.; Ejaz, W. Resource optimization in UAV-assisted IoT networks: The role of generative AI. *IEEE Internet Things Mag.* **2024**, *8*, 34–41. [CrossRef]
45. Vivekanandan, V.; Sureshkumar, R.; Manikandan, S.; Ram Kumar, R.; Das, M.S.; Kumar, G.R.; Nandakumar, S. Environmental Monitoring and Sustainability: LLMs for Climate-Responsive Urban Design. In *Large Language Models for Sustainable Urban Development*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 89–109.
46. Das, B.C.; Amini, M.H.; Wu, Y. Security and privacy challenges of large language models: A survey. *ACM Comput. Surv.* **2025**, *57*, 1–39. [CrossRef]
47. Gao, C.; Fan, G.; Chong, C.Y.; Chen, S.; Liu, C.; Lo, D.; Zheng, Z.; Liao, Q. A Systematic Literature Review of Code Hallucinations in LLMs: Characterization, Mitigation Methods, Challenges, and Future Directions for Reliable AI. *arXiv* **2025**, arXiv:2511.00776. [CrossRef]

48. Mirshekali, H.; Shadi, M.R.; Ladani, F.G.; Shaker, H.R. A Review of Large Language Models for Energy Systems: Applications, Challenges, and Future Prospects. *IEEE Access* **2025**, *13*, 163162–163188. [CrossRef]
49. Leon, M. The escalating AI's energy demands and the imperative need for sustainable solutions. *WSEAS Trans. Syst.* **2024**, *23*, 444–457. [CrossRef]
50. Agoro, H.; Llorient, M. Methodologies for Testing the Performance and Reliability of Large Language Models in Real-World AI Applications. 2025. Available online: https://www.researchgate.net/publication/391986163_Methodologies_for_Testing_the_Performance_and_Reliability_of_Large_Language_Models_in_Real-World_AI_Applications (accessed on 7 February 2026).
51. Hadi, M.U.; Al-Tashi, Q.; Qureshi, R.; Shah, A.; Muneer, A.; Irfan, M.; Zafar, A.; Shaikh, M.B.; Akhtar, N.; Al-Garadi, M.A. Large Language Models: A Comprehensive Survey of Applications, Challenges, Datasets, Models, Limitations, and Future Prospects. *TechRxiv* **2024**, techrxiv.23589741. [CrossRef]
52. Feretzakis, G.; Verykios, V.S. Trustworthy AI: Securing sensitive data in large language models. *AI* **2024**, *5*, 2773–2800. [CrossRef]
53. Afreen, J.; Mohaghegh, M.; Doborjeh, M. Systematic literature review on bias mitigation in generative AI. *AI Ethics* **2025**, *5*, 4789–4841. [CrossRef]
54. Piwowar-Sulej, K. Sustainable development and national cultures: A quantitative and qualitative analysis of the research field. *Environ. Dev. Sustain.* **2022**, *24*, 13447–13475. [CrossRef]
55. Svoboda, I.; Lande, D. Enhancing multi-criteria decision analysis with ai: Integrating analytic hierarchy process and gpt-4 for automated decision support. *arXiv* **2024**, arXiv:2402.07404. [CrossRef]
56. Huang, Y. Advancing industrial sustainability research: A domain-specific large language model perspective. *Clean. Technol. Environ. Policy* **2025**, *27*, 1899–1901. [CrossRef]
57. Zou, Y.; Shi, M.; Chen, Z.; Deng, Z.; Lei, Z.; Zeng, Z.; Yang, S.; Tong, H.; Xiao, L.; Zhou, W. ESGReveal: An LLM-based approach for extracting structured data from ESG reports. *J. Clean. Prod.* **2025**, *489*, 144572. [CrossRef]
58. Amangeldy, B.; Tasmurzayev, N.; Imankulov, T.; Baigarayeva, Z.; Izmailov, N.; Riza, T.; Abdukarimov, A.; Mukazhan, M.; Zhumagulov, B. AI-Powered Building Ecosystems: A Narrative Mapping Review on the Integration of Digital Twins and LLMs for Proactive Comfort, IEQ, and Energy Management. *Sensors* **2025**, *25*, 5265. [CrossRef] [PubMed]
59. Ullah, A.; Qi, G.; Hussain, S.; Ullah, I.; Ali, Z. The role of llms in sustainable smart cities: Applications, challenges, and future directions. *arXiv* **2024**, arXiv:2402.14596. [CrossRef]
60. Peykani, P.; Ramezanlou, F.; Tanasescu, C.; Ghanidel, S. Large language models: A structured taxonomy and review of challenges, limitations, solutions, and future directions. *Appl. Sci.* **2025**, *15*, 8103. [CrossRef]
61. Zheng, D.; Li, J.; Yang, Y.; Wang, Y.; Pang, P.C.-I. MicroBERT: Distilling MoE-based Knowledge from BERT into a Lighter Model. *Appl. Sci.* **2024**, *14*, 6171. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.