Red Cards for Climate Myths: introducing *ClimaVAR*, an AI-based Tool for Climate Disinformation Detection*

Marcos A. de Oliveira Jr.[1,2,3], **Gail Whiteman**[1], **Nectarios K. Nugroho**[1], **Ashish Siwach**[1], **Gerson G. H. Cavalheiro**[4]

[1]Nature and Climate Impact Team - University of Exeter – UK

[2]Instituto Federal de Educação, Ciência e Tecnologia Farroupilha – Brazil

[3]Universidade Federal do Rio Grande do Sul - Campus Serra – Brazil

[4]PPGC - Universidade Federal de Pelotas – Brazil

{m.oliveira-jr, g.whiteman, nkn203, as2095}@exeter.ac.uk

gerson.cavalheiro@inf.ufpel.edu.br

Abstract. *Climate disinformation undermines public understanding, erodes trust in science, and delays effective climate action. This paper introduces ClimaVAR, an AI-based tool that uses a football referee metaphor to evaluate climate-related statements. ClimaVAR proposes a pipeline which, from a user textual input, combines: (i) climate disinformation detection using Computer-assisted recognition of (climate change) denial and skepticism (CARDS) model, (ii) Retrieval-Augmented Generation (RAG) from curated scientific documents to find correct statements when disinformation is detected, and (iii) a large language model layer to generate accessible answers translating climate science into football metaphors. The system analyzes user-submitted climate statements and issues football-related messages (offside, yellow/red card, goal) indicating alignment or not with truth, accompanied by explanations and scientific references. Deployed at high-profile events including COP30 and side events at Davos, during the World Economic Forum 2026, ClimaVAR has demonstrated the potential of culturally adapted interfaces to increase engagement with climate science. Initial results demonstrated that 66% of user inputs were tagged as misinformation. Complementary analyses enabled also grouping inputs by topic, with themes such as: Climate Denial & Skepticism, Human Causation & Consensus, Physical Impacts & Extreme Events, Energy & Policy, and Other/Emerging Topics.*

1. Introduction

Climate disinformation are false or misleading information about climate science deliberately spread to delay action and it poses a significant challenge in the digital age [Lamb et al. 2020, Supran et al. 2023]. With climate-related content spreading rapidly across social media platforms [Vosoughi et al. 2018, Falkenberg et al. 2022], there is an urgent need for automated tools that can help users distinguish scientifically accurate climate information from disinformation.

*This study was funded by the Nature and Climate Impact Team (Hoffmann Project) - University of Exeter and the ESRC-IAA Translational Funding - University of Exeter.

Recent advances in natural language processing and machine learning have enabled automated detection of climate disinformation at scale [Coan et al. 2021, Zanartu et al. 2024]. The CARDS taxonomy (Computer Assisted Recognition of Denial and Skepticism) provides a structured framework for classifying contrarian climate claims [Coan et al. 2021, Coan et al. 2024], while large language models (LLMs) combined with Retrieval-Augmented Generation (RAG) enable systems to provide evidence-based explanations grounded in scientific literature [Zhao et al. 2023, Farkas and Derczynski 2022]. However, most existing detection tools present their outputs through generic technical interfaces that may not engage diverse audiences effectively [Newman et al. 2023].

This paper introduces **ClimaVAR** (`http://climavar.com`), an AI-based tool for climate disinformation detection that combines technical sophistication with cultural adaptation. ClimaVAR implements a three-stage pipeline: (i) *CARDS-based detection* to identify contrarian climate claims [Coan et al. 2021], (ii) *Retrieval-Augmented Generation (RAG)* to retrieve correct scientific evidence from a curated knowledge base when disinformation is detected, and (iii) *LLM-based natural language generation* that translates climate science into accessible explanations using football metaphors.

The football metaphor is central to ClimaVAR's design. Instead of only technical labels like "true/false" or "accurate/misleading". the system issues verdicts using familiar football referee concepts: "red card" or "offside" for disinformation, "yellow card" for partially misleading content, and "goal" or "in game" for accurate statements. This approach aims to make climate disinformation detection more engaging and accessible, particularly in football-oriented cultures like Brazil [Pearce et al. 2019].

The main contributions of this paper are:

- An integrated AI pipeline combining CARDS, RAG, and LLM-based generation for climate disinformation detection and explanation using football language;
- A culturally adapted interface using football metaphors to make climate assessments more accessible;
- A practical implementation deployed at scale, with early results demonstrating user engagement.

The paper is organized as follows. Section 2 reviews related work on automated climate disinformation detection. Section 3 describes ClimaVAR's architecture, including the knowledge base, system design, and the three-stage detection pipeline. Section 4 presents initial results from deployments at major events. And, Section 5 concludes with limitations and future work.

2. Background and Related Work

This section outlines the background supporting ClimaVar, covering climate disinformation, the CARDS taxonomy, and the use of RAG and LLMs to ground explanations in scientific evidence.

2.1. Climate Disinformation and Misinformation: Core Concepts

Climate disinformation refers to deliberately false or misleading information about climate change, distinct from misinformation which lacks intent to deceive [Dunlap and Jacques 2013]. Common disinformation categories include science denial (e.g., "climate change is not happening"), impact minimization (e.g., "warming will be beneficial"), solution skepticism (e.g., "renewable energy cannot work"), greenwashing (e.g., misleading environmental claims), and conspiracy narratives (e.g., "climate science is fabricated for profit") [Cook et al. 2017a, Williams et al. 2015]. Climate disinformation detection, as addressed in this paper, involves assessing whether a given statement aligns with the scientific consensus documented in authoritative sources such as IPCC reports and peer-reviewed literature [Oreskes and Conway 2010, Boykoff 2007].

2.2. Related Work

Several automated approaches have been developed for climate disinformation detection. The CARDS taxonomy [Coan et al. 2021] provides hierarchical classification of contrarian claims using supervised machine learning, achieving high accuracy on labeled datasets but requiring manual annotation for training and lacking natural language explanations for end users. Augmented CARDS [Coan et al. 2024] extends this to social media with real-time detection but similarly provides only category labels without accessible justifications. Technocognitive approaches [Zanartu et al. 2024] integrate CARDS with the FLICC (Fake experts, Logical fallacies, Impossible expectations, Cherry picking, and Conspiracy theories) fallacy taxonomy to detect both claim content and rhetorical techniques, improving detection performance but still presenting outputs as technical classifications rather than user-friendly explanations.

Recent work has applied large language models (LLMs) for climate misinformation detection [Li et al. 2024], leveraging their natural language generation capabilities to produce explanations. However, standalone LLMs are prone to hallucination, generating plausible but incorrect information, making them unreliable without external grounding. Retrieval-Augmented Generation (RAG) architectures address this by retrieving evidence from curated knowledge bases before generation [Farkas and Derczynski 2022], reducing hallucination and improving factual accuracy. Vector databases enable efficient semantic search through embedding-based retrieval, allowing systems to find relevant passages based on conceptual similarity rather than keyword matching [Treen et al. 2020].

Existing tools typically present detection results through conventional interfaces with labels like "true/false" or "accurate/misleading". which may not engage diverse audiences effectively [Amazeen 2020]. Research shows that culturally adapted communication strategies can enhance engagement [Moser 2010], but few systems incorporate such adaptations. Additionally, gamified and interactive approaches that teach users to recognize manipulation techniques show sustained effectiveness [Basol et al. 2020, Maertens et al. 2021], suggesting that detection tools should educate users about reasoning errors in engaging ways.

In Brazil, human-centered approaches and data-driven initiatives have also been explored in the development of climate- and misinformation-related digital solutions. Silva et al. [Silva et al. 2019] report the application of the Human-Centered Design (HCD) process to support the development of a software platform aimed at improving the communication of weather and climate information between public agencies and the population, emphasizing user needs and iterative prototyping. Complementarily, Santos et al. [Santos et al. 2023] introduce IMAGEFACTCK.BR, a public repository of synthetic images containing misinformation in Brazilian Portuguese, designed to support research on automatic detection methods, particularly in multimodal and OCR-based scenarios. While these efforts contribute respectively to user-centered climate communication and to data resources for misinformation research, they do not integrate automated detection pipelines grounded in scientific knowledge with culturally adapted explanatory interfaces, as proposed in this work.

ClimaVAR addresses these limitations by combining CARDS-based detection with RAG from curated scientific sources and LLM-based generation into an integrated pipeline that provides both accurate classification and accessible explanations. Unlike generic commercial LLMs (ChatGPT, Gemini, Perplexity) with custom prompts, ClimaVAR offers several advantages: (i) specialized detection using the validated CARDS taxonomy rather than general-purpose classification, (ii) guaranteed grounding in a curated climate science knowledge base rather than relying on incorrect training data, (iii) systematic retrieval of scientific references with every response, and (iv) a culturally adapted football-themed interface designed specifically for climate communication rather than generic conversational AI. In contrast to prior efforts that typically address isolated aspects of the problem, ClimaVAR provides an end-to-end solution that integrates detection, grounding, and user-oriented explanation within a unified architecture.

3. ClimaVAR Tool

This section presents the design and implementation of ClimaVAR, describing the system architecture, the curated knowledge base, and the integrated AI pipeline.

3.1. System Overview and Architecture

ClimaVAR is an AI-based tool for climate disinformation detection accessible via web application and mobile apps for Android and iOS. The system implements a three-stage pipeline that processes user-submitted climate statements and returns verdicts using football referee metaphors: "red card" or "offside" for disinformation, "yellow card" for partially misleading content, and "goal" or "in game" for accurate statements. There are also some filters and safeguards to identify if the question is in fact related to climate and does not contain any inappropriate content; otherwise, it returns a message such as "out of the game".

The architecture follows a modular, service-oriented design with clear separation between frontend, backend API, and AI processing components. ClimaVAR frontend was developed using Lovable.dev, a rapid application development platform that streamlined the creation of web and mobile interfaces, server setup, and database configuration. This enabled the team to focus on implementing the core AI pipeline while leveraging platform-provided infrastructure for authentication, data persistence, and cross-platform deployment.

The backend is implemented as a Django REST API (Django 5.2.7, Django REST Framework 3.16.1) hosted on Railway, a cloud platform that provides automated deployment, scaling, and monitoring. The main detection endpoint (`/api/check-misclassification`) orchestrates the three-stage pipeline. The technology stack includes: OpenAI Python client (v2.6.1) for GPT-4o-mini and embeddings; langdetect (v1.0.9) for language detection; ChromaDB (v1.2.2) for vector storage

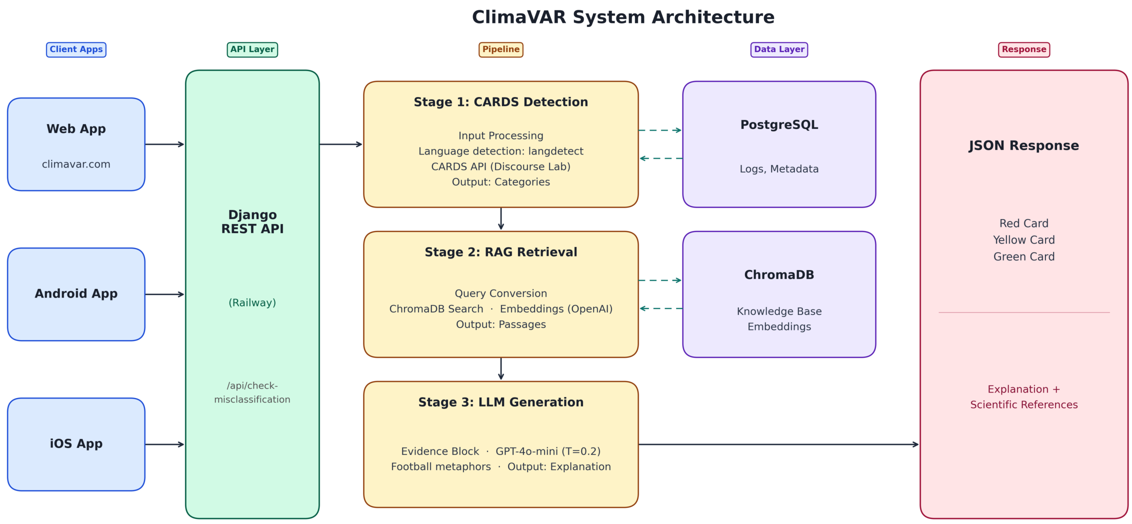

Figure 1. ClimaVAR System Architecture, with a three-stage pipeline: (1) CARDS-based detection of contrarian claims, (2) RAG-based retrieval from knowledge base, and (3) LLM-based generation with football metaphors.

and retrieval; PostgreSQL for structured data; and the CARDS API (provided by Discourse Lab) for contrarian claim detection. Figure 1 illustrates the system architecture. The complete API source code is publicly available on GitHub[1].

3.2. Knowledge Base Construction

ClimaVAR's assessments are grounded in a curated climate knowledge base stored in ChromaDB, a vector database enabling efficient semantic search through embeddings. The knowledge base integrates multiple authoritative sources including IPCC assessment reports, peer-reviewed climate science syntheses [Cook et al. 2016], and established climate communication resources that document common myths and their rebuttals [Cook et al. 2017b]. These documents used to construct the knowledge base are entirely based on publicly accessible sources and, to support transparency and reproducibility, available in a public repository at [2]. The repository currently aggregates approximately 1.5 GB of curated documents used during the development and evaluation of the system.

The documents were processed and chunked into passages of approximately 450 characters to balance retrieval precision and context comprehensiveness. Each chunk is embedded using OpenAI's `text-embedding-3-small` model, which generates 1536-dimensional dense vectors capturing semantic content. These embeddings are stored in ChromaDB (collection name: `ClimaVAR_v2`) along with metadata including document title, publication year, URL, and chunk identifier.

The knowledge base is structured around key climate topics: greenhouse gas physics, observed warming trends, attribution to human activities, projected impacts, renewable energy technologies, and climate policy instruments. Each topic contains entries linking conceptual statements to supporting evidence. Inspired by the CARDS and FLICC

[1]Code repository: `https://github.com/keane7a/ClimaVar_API`
[2]RAG files repository: `https://github.com/marcosjr06/RAG_Files_ClimaVAR`

taxonomies [Coan et al. 2021, Zanartu et al. 2024], the knowledge base associates common misleading narratives with both factual corrections and fallacy labels, enabling the system to explain not only what is wrong with a claim but also how it misleads.

3.3. AI Pipeline

The ClimaVAR pipeline integrates three distinct stages: CARDS-based detection, Retrieval-Augmented Generation, and LLM-based translation to football metaphors. This section describes each stage and their integration.

Input Processing and Routing. When a user submits a statement via `/api/check-misclassification`, the system first validates input length (10-300 characters). Language detection uses GPT-4o-mini to identify the input language, translating to English if necessary using GPT-4o-mini with a translation prompt. Currently, English, Portuguese, and Spanish are supported. The system then applies a climate relevance filter using GPT-4o-mini to confirm the statement contains climate-related content, filtering out off-topic inputs. Finally, the system classifies whether the input is a question or statement using heuristics (question marks, question words) supplemented by LLM-based classification for ambiguous cases.

Stage 1: CARDS-based Detection. For statements, the system invokes the CARDS API (`https://cardsbot.ai`) using the `cards-mini-sonnet-2024-12-05` model with temperature 0.0 for deterministic outputs. CARDS returns structured classifications containing hierarchical category labels (e.g., "1_2_3: It's natural cycles") and justifications explaining category matches [Coan et al. 2021]. If the response contains category "0_0_0". the statement is classified as not containing disinformation; otherwise, it is flagged as disinformation. The system constructs a summary string of detected categories and justifications for subsequent stages.

Stage 2: Retrieval-Augmented Generation. The RAG stage retrieves relevant scientific evidence from the ChromaDB knowledge base. The retrieval query depends on input type: for questions, the query is the translated question text; for statements flagged as disinformation, the system converts the statement into a neutral question using GPT-4o-mini (e.g., "Fossil fuels do not cause climate change" $\rightarrow$ "Do fossil fuels cause climate change?"); for statements not flagged, the query is the statement itself.

The system embeds the query using `text-embedding-3-small` and performs semantic search in ChromaDB for the top-k most similar passages (default k=2). Cosine similarity between query and document embeddings determines relevance. Retrieved passages are truncated to approximately 450 characters, and the system constructs an "evidence block" by concatenating passages and a citation list from document metadata (title, year, URL).

Stage 3: LLM-based Generation with Football Metaphors. The final stage generates accessible explanations using GPT-4o-mini (temperature 0.2). Prompts are tailored based on detection results:

- **For Questions:** The prompt instructs the model to answer concisely based on retrieved evidence.
- **For Disinformation:** The prompt provides the user's statement, evidence block, and CARDS category summary. It instructs the model to explain why the claim

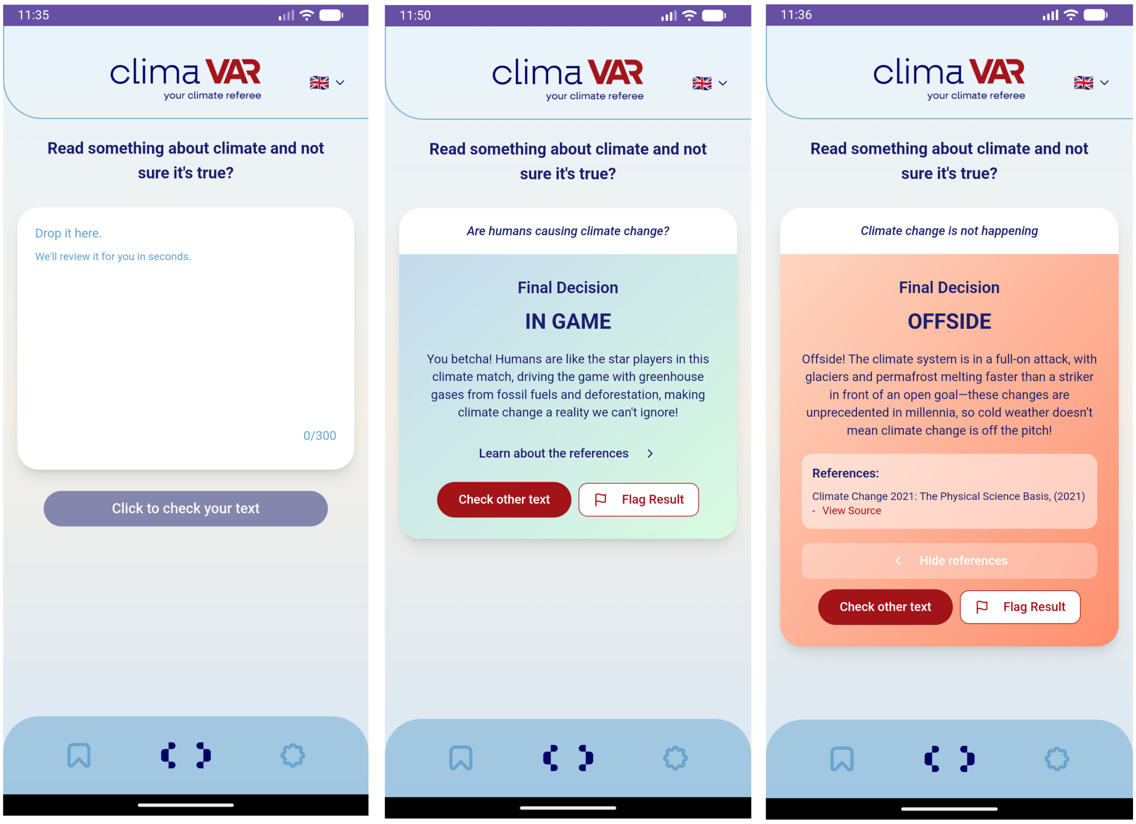

Figure 2. ClimaVAR (left) main app interface, and showing a "in game" (center) and "red card" (right) verdict for a disinformation statement. The system displays the football-themed verdict, explanation in accessible language, and scientific references.

is misleading using football terminology (e.g., "this claim is offside with climate science"), the correct scientific understanding, and manipulation techniques identified by CARDS [Cook et al. 2017a].

- **For Accurate Statements:** The prompt requests a supportive explanation affirming alignment with climate science using positive football terminology.

The generated explanation is translated back to the user's original language and the API returns a JSON response containing the explanation text, disinformation flag, and reference list. Figure 2 shows an example of the main app interface, and two different veredicts. The system offers support to English, Portuguese, and Spanish. Extending to additional languages requires translation prompt updates and LLM performance validation. The knowledge base is designed for periodic updates to incorporate new findings and emerging disinformation narratives, supporting localization across regions.

4. Results and Discussion

This section presents the main results obtained with ClimaVAR and discusses their implications and observed limitations.

4.1. User-Submitted Content Analysis

ClimaVAR has been deployed and showcased at high-profile events to maximize visibility and gather real-world usage data. The tool was initially showcased in November 2025, at

COP30 in Belem/Brazil, being presented at eventes related to Data Integrity agenda. In January 2026 it was oficially launched at side events during the World Economic Forum 2026 in Davos/Switzerland, where it was presented as an AI-powered system that "gives climate misinformation the red card." The demonstrations included live interactions, panel discussions on AI's role in combating climate disinformation, and media coverage highlighting the football metaphor as an innovative approach to science communication.

An analysis of 395 user interactions recorded in ClimaVAR's database between October 2025 and March 2026 reveals consistent patterns in how users engage with the tool. After filtering noise and test entries, 267 queries with complete classification data were analysed. These were manually grouped into five thematic clusters: Climate Denial & Skepticism, Human Causation & Consensus, Physical Impacts & Extreme Events, Energy & Policy, and Other/Emerging Topics. The largest cluster, Other/Emerging (n=96), reflects a wide range of specific and contextual queries, while Climate Denial (n=48) and Energy & Policy (n=40) represent the most topically concentrated groups.

Figure 3 shows the distribution of user queries in five thematic clusters. Across all interactions, 66% of queries (n=177) were classified as misinformation by the system. The highest misinformation rate was recorded in the Climate Denial cluster (90%), where nearly all inputs were false or misleading claims. Energy Policy and Physical Impacts followed at 72% and 69% respectively, reflecting the prevalence of specific technical misconceptions in those areas. Human Causation showed the lowest rate (48%), consistent with a mix of genuine questions and accurate statements about scientific consensus. These results suggest that ClimaVAR users are predominantly submitting contested or false claims for verification, supporting its role as an active fact-checking tool rather than a general climate information resource.

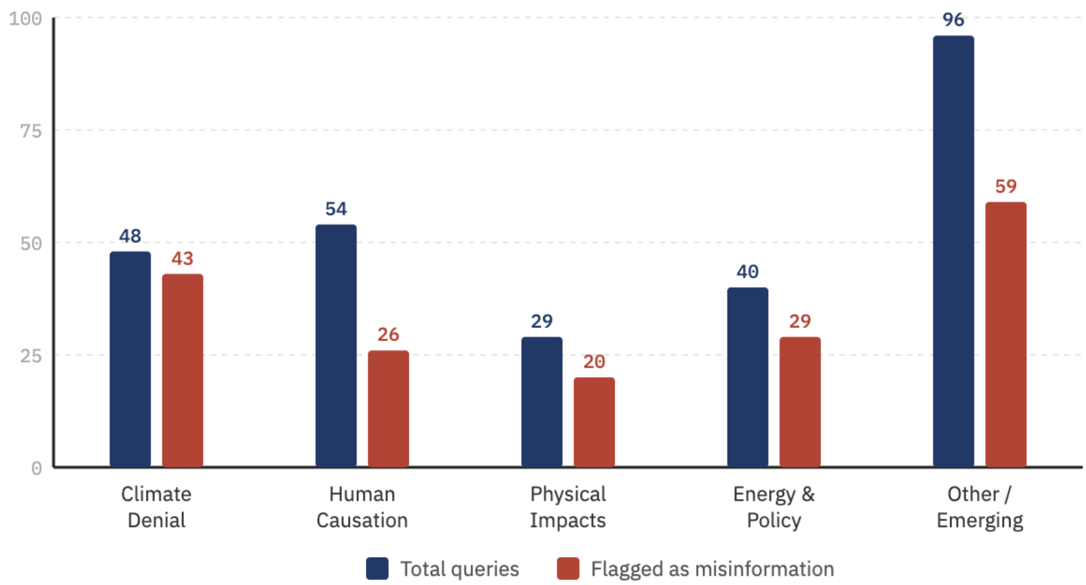

Figure 3. Input analysis: Total queries (dark blue) and misinformation-flagged queries (red) per topic cluster.

These thematic patterns closely mirror some of the categories documented in the CARDS taxonomy [Coan et al. 2021] and climate disinformation research literature [Dunlap and Jacques 2013, Williams et al. 2015], suggesting that ClimaVAR's three-stage pipeline effectively captures the narratives users encounter in everyday discourse. Many statements were successfully detected and categorized by the CARDS stage, with the system identifying claims invoking fake experts (citing non-climate scientists), cherry-picking data (highlighting short-term fluctuations to deny trends), and conspiracy theories (claiming climate science manipulation). The RAG stage successfully retrieved relevant evidence to counter these claims, and the LLM stage generated explanations addressing both factual inaccuracies and rhetorical techniques using football metaphors.

4.2. Technical Performance and Limitations

Internal testing on a sample of statements annotated by climate communication experts suggests that ClimaVAR's three-stage pipeline produces assessments that generally align with expert judgments. Statements that experts classified as clearly false or misleading were consistently flagged by CARDS and received red card verdicts with appropriate explanations, while statements aligned with mainstream climate science received green card or "goal" verdicts. However, edge cases-such as statements mixing correct and incorrect elements or using subtle framing-occasionally resulted in disagreements between the automated system and expert annotators, highlighting areas for improvement in nuanced claim detection.

The system's performance is limited by several technical factors. First, CARDS classification accuracy depends on the model's training data coverage. Novel or emerging disinformation narratives not well-represented in the training set may be missed or misclassified. Second, the RAG stage's effectiveness depends on the quality and currency of the knowledge base. As discussed in Section 3.4, the knowledge base requires frequent updates to remain aligned with evolving climate science, new technologies (e.g., carbon capture innovations, advanced renewables), and emerging disinformation patterns (e.g., responses to extreme weather events, greenwashing claims from fossil fuel companies). Without systematic curation, the system risks providing outdated information or failing to detect novel claims.

Third, LLM generation quality varies with prompt design and model capabilities. While GPT-4o-mini generally produces coherent and accessible explanations, it occasionally generates overly generic responses or fails to fully capture the nuance of complex scientific topics. The football metaphor, while effective for engagement, may also oversimplify certain issues or feel forced in contexts where the analogy doesn't naturally fit.

4.3. Technical Performance and Limitations

Internal testing on statements annotated by climate communication experts suggests that ClimaVAR's three-stage pipeline generally aligns with expert judgments. Statements classified by experts as clearly false or misleading were consistently flagged by CARDS and received red card verdicts with appropriate explanations, while those aligned with mainstream climate science received green cards. However, edge cases such as statements mixing correct and incorrect elements or using subtle framing occasionally produced disagreements between the system and expert annotators, highlighting areas for improvement in nuanced claim detection.

The system faces several technical limitations. CARDS classification accuracy depends on training data coverage, with novel or emerging disinformation narratives potentially missed or misclassified. The RAG stage's effectiveness depends critically on knowledge base currency-climate science advances rapidly with new publications, IPCC updates, and policy developments, while disinformation narratives evolve in response to current events. Without systematic curation, the system risks providing outdated information, representing the most significant operational challenge for long-term deployment. LLM generation quality varies with prompt design, occasionally producing overly generic responses or oversimplifying complex topics through the football metaphor. The system also struggles with highly technical statements, sarcasm, irony, and figurative language.

Additional constraints include language coverage limited to English, Portuguese, and Spanish, requiring significant validation efforts for expansion. The system currently handles only text input, despite the prevalence of visual disinformation (misleading graphs, manipulated images, climate memes) on social media. While qualitative user feedback is positive, systematic quantitative evaluation of ClimaVAR's impact on user climate literacy, trust in science, and behavior change remains to be conducted through longitudinal studies.

5. Conclusion and Future Work

This paper introduced ClimaVAR, an AI-based tool for climate disinformation detection combining CARDS-based classification, Retrieval-Augmented Generation from curated scientific sources, and LLM-based generation with football metaphors. Deployed at `http://climavar.com` as web and mobile applications, ClimaVAR grounds assessments in scientific evidence and translates verdicts into intuitive football referee language- "red cards" for disinformation, "yellow cards" for partial misinformation, and "goals" for accurate statements. Deployments at COP30 and Davos 2026 (WEF side events) demonstrated user engagement, with feedback indicating the football metaphor is intuitive and less confrontational than conventional fact-checking. User-submitted statements mirrored documented disinformation categories, confirming the system addresses real-world narratives. Were recorded 395 user interactions between October 2025 and March 2026, grouped in five clusters according the main topic. The majority of inputs as declarative statements (67%) rather than questions (26%), indicating that users predominantly interact with the tool to challenge or verify specific claims rather than seek general information. Approximately 66% of all submissions were classified as misinformation.

Future work will focus on addressing the limitations discussed in Section 4 while advancing the robustness, scalability, transparency, and societal impact of the platform. Key priorities include the development of semi-automated knowledge base curation pipelines supported by expert validation workflows; improving misinformation detection performance in ambiguous and edge-case scenarios through the retraining and expansion of the CARDS taxonomy and classification models; and refining LLM prompt engineering strategies to generate more accurate, context-aware, and culturally resonant explanations. Future research should also incorporate a comprehensive quantitative evaluation of the full pipeline, including classification performance, explanation quality, latency, user interaction metrics, and comparative analyses across different LLM configurations.

Additional directions involve expanding multilingual support with region-specific cultural adaptation, integrating multimodal analysis capabilities to address image- and video-based climate disinformation, and conducting longitudinal evaluation studies to assess the system's effects on climate literacy, critical reasoning, engagement, and behavioral change over time. Other important aspect to investigate are the ethical implications associated with the use of proprietary LLMs, particularly regarding transparency, accountability, embedded biases, and dependence on closed commercial infrastructures. These concerns are especially relevant in climate communication contexts, where inaccuracies, opaque decision-making processes, or unequal access to AI technologies may directly affect public understanding and trust. To support reproducibility and future research extensions, the project resources, documentation, and supplementary materials have been made publicly available through the public repository link provided in this study.

References

Amazeen, M. A. (2020). Journalistic interventions: The structural factors affecting the global emergence of fact-checking. *Journalism*, 21(1):95–111.

Basol, M., Roozenbeek, J., and van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, 3(1):2.

Boykoff, M. T. (2007). Flogging a dead norm? newspaper coverage of anthropogenic climate change in the united states and united kingdom from 2003 to 2006. *Area*, 39(4):470–481.

Coan, T. G., Cook, J., and Lewandowsky, S. (2024). Augmented detection of climate contrarian claims using hierarchical classification. *Environmental Communication*, 18(3):342–359.

Coan, T. G., Vokey, B., Cook, J., and Lewandowsky, S. (2021). Computer-assisted classification of contrarian claims about climate change. *Scientific Reports*, 11(1):22320.

Cook, J., Lewandowsky, S., and Ecker, U. K. (2017a). Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters*, 12(2):024018.

Cook, J., Lewandowsky, S., and Ecker, U. K. (2017b). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, 12(5):e0175799.

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., et al. (2016). Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, 11(4):048002.

Dunlap, R. E. and Jacques, P. J. (2013). Climate change skepticism and denial: An introduction. *American Behavioral Scientist*, 57(6):691–698.

Falkenberg, M., Galeazzi, A., Torricelli, M., Di Marco, N., Larosa, F., Sas, M., Metta, A., Quattrociocchi, W., and Zollo, F. (2022). Growing polarization around climate change on social media. *Nature Climate Change*, 12(12):1114–1121.

Farkas, B. and Derczynski, L. (2022). Scalable and automated detection of climate misinformation using natural language processing. *Patterns*, 3(9):100579.

Lamb, W. F., Mattioli, G., Levi, S., Roberts, J. T., Capstick, S., Creutzig, F., Minx, J. C., Müller-Hansen, F., Culhane, T., and Steinberger, J. K. (2020). Discourses of climate delay. *Global Sustainability*, 3:e17.

Li, Z., Wang, Y., Chen, Z., and Liu, Y. (2024). Multimodal misinformation detection using vision-language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(17):18750–18758.

Maertens, R., Roozenbeek, J., Basol, M., and van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1):1–16.

Moser, S. C. (2010). Communicating climate change: history, challenges, process and future directions. *Wiley Interdisciplinary Reviews: Climate Change*, 1(1):31–53.

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., and Nielsen, R. K. (2023). Information disorder and climate change. *Reuters Institute for the Study of Journalism*.

Oreskes, N. and Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Publishing USA.

Pearce, W., Grundmann, R., Hulme, M., Raman, S., Kershaw, E. H., and Tsouvalis, J. (2019). Climate communication and sports: leveraging football for climate action. *Climatic Change*, 157(2):115–130.

Santos, Y., Silva, M., and Reis, J. C. S. (2023). Imagefactck.br: Repositório de imagens para a detecção de desinformação disseminada em plataformas digitais. In *Anais do V Dataset Showcase Workshop*, pages 87–98, Porto Alegre, RS, Brasil. SBC.

Silva, J., Silva, T., Lima, M., and Souza, R. (2019). Aplicação do design centrado em humano para desenvolvimento de uma solução de software baseada em informações do tempo e clima. In *Anais do X Workshop de Computação Aplicada a Gestão do Meio Ambiente e Recursos Naturais*, pages 107–116, Porto Alegre, RS, Brasil. SBC.

Supran, G., Rahmstorf, S., and Oreskes, N. (2023). Assessing exxonmobil's global warming projections. *Science*, 379(6628):eabk0063.

Treen, K. M. d., Williams, H. T., and O'Neill, S. J. (2020). Online misinformation about climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 11(5):e665.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Williams, H. T., McMurray, J. R., Kurz, T., and Lambert, F. H. (2015). Network analysis reveals open forums that coordinate attack campaigns against climate science. *Nature Climate Change*, 5(6):563–567.

Zanartu, F., Blanco-Arancibia, C., Lewandowsky, S., and Cook, J. (2024). A technocognitive approach to detecting fallacies in climate misinformation. *Scientific Reports*, 14:25193.

Zhao, Y., Zhang, D., and Yang, J. (2023). Automated detection of climate change misinformation with large language models. *Nature Communications*, 14:6670.