AI for IPCC:
Towards Living Evidence Synthesis for Climate Science

Sandeep Chowdhary[1,*]

[1]International Institute for Applied Systems Analysis, Laxenburg, Austria.
[*]chowdhary@iiasa.ac.at

ABSTRACT

Climate science is expanding at a pace that outstrips the capacity of existing assessment processes. The IPCC alone must now screen tens of thousands of new publications every cycle, yet the current system relies on manual curation and synthesis spread across hundreds of experts over multiple years. This creates a synthesis bottleneck: evidence bases become outdated soon after the time reports appear, and the ability of policy makers to act on the latest knowledge becomes limited. Existing AI and bibliometric tools demonstrate the potential to map research landscapes, but they remain opaque, or proprietary, provide limited coverage and do not integrate with the workflows of global environmental assessments.

This project develops an open, transparent, and continuously updated literature synthesis tool tailored to the needs of the IPCC and related processes. Building on AR7 outlines and seed references, the system expands outward via citation networks and semantic search to construct supra-sets of potentially relevant studies, which are then filtered and summarised using large language models tuned for factual retention. Each extracted finding is anchored to its source text and validated through independent models to reduce errors and ensure correctness. As a first step, we will compare the coverage of automated searches with the reference corpora of past IPCC reports (e.g. AR5, AR6), quantifying where human-led strategies missed relevant literature. These benchmarks will then be used to fine-tune the AI-assisted search process before deploying it prospectively on AR7. Validated results are integrated into a live HTML version of the IPCC reports, enriched with timestamps, confidence levels, and direct links to sources, and made available via an API for integration with Integrated Assessment Models (IAMs). This workflow mirrors the way human authors construct assessments, but extends their reach and timeliness by automating discovery, validation, and integration.

The originality lies in combining the breadth of automated evidence mapping with the precision using meta-science approaches and accountability needed for assessment-ready synthesis. Unlike proprietary systems, the pipeline is fully open source and auditable, designed as a glass box rather than a black box. Its contribution is to close the gap between fast-growing scientific literatures and policy-relevant synthesis, creating a living, continuously updated knowledge base. While focused on climate science and the IPCC, the system generalises to any domain requiring evidence-intensive reviews, from biodiversity assessments to public health guidelines. By building a validated AI-for-science pipeline embedded in an institute with strong IAM expertise, the project strengthens Austria's role in shaping international climate knowledge infrastructures and accelerates the timeliness, transparency, and policy relevance of global environmental assessments.

1 State of the Art

Climate change stands among the most profound challenges of our time. Since the first quantitative prediction of CO2-induced warming [1] and early estimates of climate sensitivity [2, 3], research has expanded to encompass economic impacts [4], carbon budgets [5], and planetary boundaries [6], reaffirming both the severity of the crisis and the urgency of global responses.

Global environmental assessments (GEAs) such as the IPCC, IPBES, GEO, UNEP's Gap reports, and the Global Sustainable Development Report provide crucial evidence on environmental change and guide action-oriented policy [7]. They have shaped international negotiations and agreements, including the Paris Agreement [8, 9]. The IPCC in particular has long informed climate governance, with AR6 highlighting the urgency of mitigation and adaptation and feeding into the Global Stocktake [10]. Yet concerns remain about their ability to keep pace with the rapidly expanding scientific literature [11, 7]. The number of publications grows exponentially doubling roughly every 15 years while human attention and processing capacity remain bounded [12, 13]. This creates an unavoidable tension: ever more knowledge is produced, but the ability of experts to keep pace with emerging findings is limited, producing a synthesis bottleneck.

The IPCC exemplifies both the strengths and limits of *manual synthesis*. Each cycle mobilises hundreds of experts over 5–7 years [14], during which tens of thousands of publications must be screened as the literature grows rapidly. In AR5, authors assessed about 30,000 publications [15], while in AR6 this doubled to more than 66,000 [16]. Working Group I in AR6 engaged 234 authors to review 14,000 papers and address 78,000 review comments [17], while Working Group III involved 278 lead authors and 354 contributing authors to assess 18,000 papers and respond to 59,000 comments [18]. Beyond this scale, assessments have expanded from physical science to impacts, mitigation, adaptation, and social dimensions [19, 20], drawing on disciplines from economics and psychology to engineering and the humanities, and reaching audiences from negotiators to NGOs and businesses worldwide [20]. The direct IPCC Trust Fund expenditure for 2024 exceeded €5 million [21], excluding in-kind contributions. Assessment writing is thus one of the most resource-intensive processes in global science, and the lag of several years between research and synthesis creates a bottleneck.

Recent advances in machine learning (ML) and natural language processing (NLP) promise to ease this bottleneck. Machine learning evidence maps of climate impacts [22] aggregated over 100,000 impact studies and revealed geographic attribution gaps, but they remain static and stop short of producing assessment-ready synthesis. AI-driven mapping of adaptation policies [23] identified systematic differences across governance levels, yet highlighted persistent blind spots in vulnerable countries and lacked iterative updating. A "living" ML map of 84,000 mitigation policy studies [24] demonstrated disparities between research attention and emissions, but still lacked integration into modelling frameworks. AI-enhanced systematic mapping of carbon dioxide removal (CDR) literature [25] uncovered three times more relevant publications than prior estimates, but outputs were static and disconnected from IAM parameters. Large-scale bibliometric studies [26, 27] map research landscapes, but remain descriptive without mechanisms for continuous integration. Domain-specific models such as ClimateBERT [28] have improved classification in corporate climate disclosures, and transformers have enabled attribution mapping of climate impacts [22] and systematic reviews in adaptation and health [29, 30, 31]. These approaches also open pathways for multilingual evidence synthesis, addressing English-language bias through translation and cross-lingual classification [32, 33]. Living evidence synthesis platforms [34] and AI-supported screening [35] show promise, while frameworks for structured confidence evaluation [36, 37, 38] mirror and extend expert judgment in assessment processes. Yet problems remain: hallucination risks, frozen corpora that become outdated, and uneven performance in long-form synthesis [39, 40]. General-purpose chatbots (e.g., ChatGPT, Claude Sonnet, Gemini) provide fast answers but lack coverage, transparency, and reliability in long-form scientific tasks [41, 42]. Specialized AI-for-science tools, by contrast, target literature discovery and synthesis but face other limitations: most are proprietary and closed-source (see Table **??**), with restricted free tiers, opaque methods, and limited synthesis scale. Few explicitly support consensus-building, underscoring the difficulty of achieving open, transparent, and assessment-ready AI-powered scientific discovery.

Therefore, what is missing is an approach that combines the breadth and scalability of automated evidence mapping with the precision and accountability required for policy ready assessments. What is needed is an open-source tool that can simultaneously expand coverage across the rapidly growing literature, anchor findings directly to their sources, and integrate them into the IPCC framework without introducing hallucinations or relying on black boxes. Only such a design can close the growing gap of literature synthesis improving timeliness while maintaining rigour in climate science.

Tool	Open Source	Free	Data Source/Coverage	Black Box	Synthesis Size	Summary	Targeted Consensus
Semantic Scholar[43]	No	Yes	Semantic Scholar corpus (200M+ papers, multidisciplinary)	No (API+UI)	Discovery only	Yes (TLDR)	No
Elicit[44]	No	Yes*, Paid	100M+ (Semantic Scholar, abstracts/ fulltext if open)	Mostly	10/25/40	Yes	Partial (evidence table)
Consensus[45]	No	Yes*, Paid	200M+ (Semantic Scholar, OpenAlex, web crawl)	Yes	1,500 filtered, top 20 synthesized	Yes	Yes (explicit claim meter)
Paperguide[46]	No	Yes*, Paid	200M+ (broad, not fully publicized)	Yes	~100	Yes	Yes (Q&A/synthesis)
Scite[47]	No	Free trial, Paid	Mixed (Elsevier, PubMed, others, open)	Part	Per-paper/field	No	No
Connected Papers[48]	No	Yes*, Paid	Semantic Scholar, ODC-BY, ResearchGate, Academia	Yes	~50,000 (graph)	No	No
Research Rabbit[49]	No	Yes (free)	OpenAlex, Semantic Scholar	Yes	Thousands, visual	No	No
Scholarcy[50]	No	Yes*, Paid	Uploaded papers only	Yes	Tens–hundreds	Yes	No
SciSpace[51]	No	Yes*, Paid	Multisource, open academic papers/metadata	Yes	5–20 (per synthesis)	Yes	No
Iris.ai[52]	No	Free*, Paid	Open, paywalled, enterprise data	Yes	Hundreds	Yes	No

Table 1. Feature comparison of top AI-for-science tools. (*=very limited)

2 Objectives

To address the synthesis bottleneck outlined above, this project will build an automated, systematic literature synthesis tool for climate science that keeps IPCC reports continuously updated with minimal manual supervision.

How can we construct the supra-set of potentially relevant literature? Building on the AR7 outlines [?], we treat each chapter's bullet points as topics. For each topic, we identify a small seed set of well-known, highly cited, or otherwise central papers. These seeds are then expanded into a supra-set of potentially relevant articles by (i) crawling the citation network in both directions (backward to foundational works and forward to newer studies), and (ii) running semantic and keyword-based searches to capture semantically linked literature not connected by citation. At this stage, recall and coverage are prioritized over precision: it is more important to avoid missing relevant literature, while false positives will be handled in the filtering stage.

Can we develop more comprehensive search strategies of scientific literature compared to manual curation in past IPCCs? A first test of our system will be to evaluate how well automated strategies reconstruct the literature base of previous IPCC reports. Using the outlines of past assessments, we will identify key topics and seed papers, then run our search algorithm (citation expansion + semantic search) to generate a supra-set of potentially relevant studies. We will compare the coverage of this supra-set against the actual set of papers cited in past IPCC chapters. Notice that at this stage recall and overall coverage are prioritized over precision, since false positives will be filtered in later stages. This exercise will allow us to quantify gaps in human-curated reports that arose from resource constraints and compare them to the broader, less resource-limited coverage achieved by automated methods. Our benchmark will be achieving at least 90% recall of IPCC-cited papers across Working Groups, while documenting additional relevant studies that were missed by manual processes.

How well does the system perform compared to manual curation and existing AI tools, and what does this mean for policy relevance and trust? We will benchmark performance against the gold standard of manual annotation in IPCC processes, testing retrieval, classification, and synthesis accuracy at each step of the pipeline. We will also capture (i) coverage gains over curated IPCC reference lists at fixed precision, (ii) end-to-end accuracy via expert labelled test sets with precision/recall, calibration, and validator stability, and (iii) closeness of automated key takeaway insertion to expert authored IPCC text. In parallel, we will compare our system's outputs to existing ML-based bibliometric tools and general purpose LLMs (e.g., ChatGPT, Gemini, Claude) to establish relative strengths and weaknesses. The results will be assessed not only for technical accuracy but also for decision utility: time-to-evidence, trust in AI made additions, and perceived policy relevance by IPCC authors and modellers.

Can we achieve reliable, continuous literature integration into IPCC reports? Continuous integration comes after supra-set construction + relevance filtering. We will design a hybrid human-AI workflow that automatically retrieves, validates, and inserts new climate science findings into a live HTML version of the IPCC reports, with cited sources preserved down to the level of each new added sentence.

Can AI be used for horizon scanning for IPCC gaps, detect blind spots and emerging domains? Recent AI-enhanced mapping in carbon dioxide removal uncovered 28,976 studies, roughly 3 to 4 times earlier counts [25]. Our AI-augmented search system will help future assessments avoid such gaps. Furthermore, AR6 highlighted clear priorities: in mitigation, more robust MRV for carbon dioxide removal, demand and services side options, and transparent scenario curation linked to national pathways [53, 54, 55, 56, 57, 58, 59]; in adaptation, stronger evidence on limits, cascading risks, and urban systems [60, 61]; and in physical science, better detection/attribution of extremes and compound events [62]. Our tool will operationalise these priorities by continuously tracking growth signals (e.g., citation bursts, novelty terms, and publication activity), quantifying post AR6 attention, and routing results into curated watchlists and "trend cards" with validator scores, timestamps, and links to source PDFs. Beyond reinforcing known priorities, it will also detect emerging blind spots: identifying novel lexicon in titles and abstracts, as well as atypical combinations of research domains. Using automated classification into existing IPCC themes, the system will surface not only extensions of current categories but also entirely new recombinations patterns shown to drive high impact science [63].

3 Originality and Innovation

The originality of our approach lies in *modeling and augmenting the way humans (e.g., IPCC Lead Authors) actually search literature*, while simultaneously *combining multiple modes of automated discovery into a single pipeline*. Human experts typically begin with topic-wise seed papers, then snowball outward through citation chains and semantically related works before filtering and synthesizing evidence. Our system mirrors this process, but extends it by integrating *systematic citation-graph expansion*, *semantic similarity search*, and *LLM-based context-aware classification* to evaluate whether a new study is relevant to the specific research topic, subsection, or hypothesis under review. This layered design ensures that directly related, indirectly connected, and contextually relevant literature are all captured–far beyond what keyword searches or one-off queries can achieve. By merging *graph expansion*, *semantic enrichment*, and *contextual decision-making*, the tool provides a fundamentally new way to achieve comprehensive and precise synthesis.

Equally important contributor to the originality is the emphasis on *transparency*. Current AI-supported literature tools are largely closed-source, proprietary, and opaque in how they retrieve and filter information. Systems like *Elicit or Consensus* provide structured outputs but restrict corpus size (up to 20 papers), limit explainability, and prevent users from knowing whether key studies were omitted. Our approach explicitly inverts this model: rather than a *black box*, it is designed as a *glass box* where all steps from search boundaries to intermediate filtering decisions to sentence-level sources are open and auditable. This transparency builds trust, which is crucial for future adoption of such systems into scientific workflows.

Finally, the project goes beyond static reports and creates a *living knowledge system*. For centuries, books and bound reports were the primary vessels of scientific synthesis- fixed snapshots in time, necessarily outdated soon after they are published. The future, however, lies in *living, breathing documents*: continuously updated, automatically-validated, and openly accessible, with every claim transparently linked to its source. Our system will instantiate this future for climate science, integrating directly into IAM and policy workflows, while serving an online HTML version of the IPCC reports enriched with confidence scores, timestamps, and provenance. Rather than treating synthesis as episodic and frozen, the project establishes it as an ongoing, dynamic process closing the gap between rapid scientific advances and the urgent needs of global climate policy. Not only will our approach create a *living, breathing IPCC report* but once the evidence from each paper is quantified using numbers and embedding vectors –instead of qualitative features – we can simply project the IPCC report as a whole onto the embedding vector of any new research hypothesis we have and get a quantitative answer on how many papers support this or oppose this. For all the reasons described above, this is not merely a tool that can be built, but one that *must* be built if climate science is to remain actionable in real time.

Generalisation. While initially designed for IPCC assessments, the pipeline generalises to any *seed document* to be expanded or synthesized. By seeding it with a reference report, the system can automatically retrieve, filter, and integrate subsequent literature, maintaining a living version of the document in any evidence intensive domain from biodiversity assessments to public health guidelines.

4 Methods

The workflow will consist of three main stages:

1. **Identification of relevant literature.** We will begin with a *seed set of references identified per topic in the AR7 outlines (or past IPCC reports)*, including chapter reference lists, well-known and highly cited works, and expert-selected articles. From these seeds, the system will expand outward by retrieving all forward citations (papers citing the seed set), backward citations (papers cited by the seed set), and semantically similar studies identified via embedding search. Each candidate study will then be evaluated for relevance using a large language model (LLM) relevance classifier, which will process topic–abstract pairs and produce binary inclusion decisions. Relevance will also be checked using co-citation strength with original references.

   At this stage we emphasize recall and coverage of the supra-set; false positives are expected and will be handled downstream during the filtering stage. This ensures that the constructed supra-set captures as close as possible to the full universe of potentially relevant studies, even if precision is initially low.

2. **Synthesis and validation of findings.** For each filtered paper, we will extract key findings relevant to the target IPCC section using LLM-based summarisation tuned for factual retention. Extracted statements will be linked to their exact location in the source PDF to preserve provenance. The system will also classify each statement to the correct chapter, subsection, and line of the IPCC text (LOC). A separate, independent AI model differing in architecture and training corpus will then re-parse these passages to verify factual correctness and contextual alignment. Confidence scores will be assigned and periodically re-calibrated. Only findings confirmed by this validator will be included in the synthesis. This two-model approach will reduce correlated errors, increase factual accuracy, and provide a transparent chain from claim to source text.

3. **Integration into a live, online IPCC report.** Validated findings will be inserted into a continuously maintained HTML version of the IPCC report, built with a Python backend and a lightweight React-based front-end. Newly integrated studies will be highlighted visually, accompanied by their validation status, date of integration, and direct links to source documents. The report will be queryable by topic, confidence, and publication date, with structured outputs made available via an API for integration into IIASA's Integrated Assessment Models (IAMs) and related policy analysis workflows.

Finally, humans-in-the-loop will check new additions and validate those where AI agents are uncertain. This design will address both timeliness and trust gaps left open by current approaches while substantially reducing the human cost of synthesis. At a conservative €50 per fully loaded research hour, screening labour alone in AR6 likely exceeded €2 million per Working Group. Our hybrid human–AI workflow will aim to cut this cost while enabling more frequent, iteration-friendly synthesis cycles without sacrificing rigour. Updates will be comprehensive, auditable, and bias-aware, in contrast to fast but opaque alternatives.

References

1. S. Arrhenius, "On the influence of carbonic acid in the air upon the temperature of the ground," *Philosophical Magazine and Journal of Science*, vol. 41, no. 251, pp. 237–276, 1896.
2. National Research Council, Carbon Dioxide and Climate Committee, *Carbon Dioxide and Climate: A Scientific Assessment*. Washington, D.C.: National Academies Press, 1979.
3. S. Manabe and R. T. Wetherald, "Thermal equilibrium of the atmosphere with a given distribution of relative humidity," *Journal of the Atmospheric Sciences*, vol. 24, no. 3, pp. 241–259, 1967.
4. N. Stern, *The Economics of Climate Change: The Stern Review*. Cambridge University Press, 2006.
5. V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield, eds., *Global Warming of 1.5 °C. An IPCC Special Report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Geneva, Switzerland: Intergovernmental Panel on Climate Change, 2018.
6. J. Rockström, W. Steffen, K. Noone, Persson, F. S. I. Chapin, E. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. J. Schellnhuber, B. Nykvist, C. A. de Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. Foley, "A safe operating space for humanity," *Nature*, vol. 461, no. 7263, pp. 472–475, 2009.
7. N. Castree, R. Bellamy, and S. Osaka, "The future of global environmental assessments: Making a case for fundamental change," *The Anthropocene Review*, vol. 8, no. 1, pp. 56–82, 2020.
8. J. Jabbour and C. Flachsland, "40 years of global environmental assessments: a retrospective analysis," *Environmental Science & Policy*, vol. 77, pp. 193–202, 2017.
9. T. Ourbak and L. Tubiana, "Changing the game: the paris agreement and the role of scientific communities," *Climate Policy*, vol. 17, no. 7, pp. 819–824, 2017.
10. IPCC, "Summary for policymakers," *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 2023.
11. M. Vardy, M. Oppenheimer, N. K. Dubash, J. O'Reilly, D. Jamieson, *et al.*, "The intergovernmental panel on climate change: Challenges and opportunities," *Annual Review of Environment and Resources*, vol. 42, pp. 55–75, 2017.
12. S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, D. Wang, and A.-L. Barabási, "Science of science," *Science*, vol. 359, no. 6379, p. eaao0185, 2018.
13. L. Bornmann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *Journal of the Association for Information Science and Technology*, vol. 72, no. 9, pp. 1099–1112, 2021.
14. Intergovernmental Panel on Climate Change, "About the ipcc: How it is structured," 2024. IPCC comprehensive assessment reports are produced roughly every 5–7 years.
15. IPCC, "Climate change 2014: Synthesis report," 2015. AR5 Synthesis Report, Accessed 2025-08-16.
16. IPCC, "Ar6 factsheet 2022." Intergovernmental Panel on Climate Change, 2022. Summary factsheet on publications assessed in AR6. Accessed 2025-08-16.
17. UNFCCC Secretariat, "Summary report on the sbsta–ipcc special event on the working group i contribution to ar6," 2022. AR6 WGI scale: 234 authors; >14,000 publications assessed; >78,000 review comments.
18. UNFCCC Secretariat, "Summary report on the sbsta–ipcc special event: Unpacking the new scientific knowledge and key findings in the working group iii contribution to ar6," 2022. AR6 WGIII scale: 278 lead authors; 354 contributing authors; >18,000 publications; >59,000 comments.

19. D. Khojasteh, M. Haghani, A. Shamsipour, C. Zwack, W. Glamore, *et al.*, "Climate change science is evolving toward adaptation and mitigation solutions," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 15, no. 2, p. e884, 2024.

20. M. Callaghan, J. Minx, and P. Forster, "A topography of climate change research," *Nature Climate Change*, vol. 10, pp. 118–123, 2020.

21. Intergovernmental Panel on Climate Change, "Ipcc trust fund programme and budget (2024 income/expenditure; 2025 proposed)," 2025. Shows 2024 expenditure CHF 5,505,000.

22. M. Callaghan, C.-F. Schleussner, S. Nath, Q. Lejeune, T. R. Knutson, M. Reichstein, G. Hansen, E. Theokritoff, M. Andrijevic, R. J. Brecha, M. Hegarty, C. Jones, K. Lee, A. Lucas, N. van Maanen, I. Menke, P. Pfleiderer, B. Yesil, and J. C. Minx, "Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies," *Nature Climate Change*, vol. 11, no. 11, p. 966–972, 2021.

23. A. J. Sietsma, E. Theokritoff, R. G. Biesbroek, I. Villaverde Canosa, A. Thomas, M. Callaghan, J. C. Minx, and J. D. Ford, "Machine learning evidence map reveals global differences in adaptation action," *One Earth*, vol. 7, no. 2, pp. 280–292, 2024.

24. M. Callaghan, L. Banisch, N. Doebbeling-Hildebrandt, D. Edmondson, C. Flachsland, W. F. Lamb, S. Levi, F. Müller-Hansen, E. Posada, S. Vasudevan, and J. C. Minx, "Machine learning map of climate policy literature reveals disparities between scientific attention, policy density, and emissions," *npj Climate Action*, vol. 4, p. 7, 2025.

25. S. Lück, M. Callaghan, M. Borchers, A. Cowie, S. Fuss, M. Gidden, J. Hartmann, C. Kammann, D. P. Keller, F. Kraxner, W. F. Lamb, N. Mac Dowell, F. Müller-Hansen, G. F. Nemet, B. S. Probst, P. Renforth, T. Repke, W. Rickels, I. Schulte, P. Smith, S. M. Smith, D. Thrän, T. G. Troxler, V. Sick, M. van der Spek, and J. C. Minx, "Scientific literature on carbon dioxide removal revealed as much larger through ai-enhanced systematic mapping," *Nature Communications*, vol. 16, p. 6632, 2025.

26. A. Sharifi, Y. Jin, and P. D. Aboagye, "Mapping and topic modeling of over 133,000 studies on climate change and cities: Trends, geographical disparities, and future research needs," *Urban Climate*, vol. 61, p. 102448, 2025.

27. A.-M. Niculae, S.-V. Oprea, A.-G. Văduva, A. Bâra, and A.-I. Andreescu, "Assessing the role of machine learning in climate research publications," *Sustainability*, vol. 16, no. 24, p. 11086, 2024.

28. N. Webersinke, M. Kraus, J. A. Bingler, and M. Leippold, "Climatebert: A pretrained language model for climate-related text," *arXiv preprint arXiv:2110.12010*, 2021.

29. L. Berrang-Ford, A. R. Siders, A. Lesnikowski, A. P. Fischer, M. W. Callaghan, *et al.*, "A systematic global stocktake of evidence on human adaptation to climate change," *Nature Climate Change*, vol. 11, pp. 989–1000, 2021.

30. L. Berrang-Ford, A. Sietsma, M. Callaghan, J. Minx, P. Scheelbeek, *et al.*, "Systematic mapping of global research on climate and health: a machine learning review," *The Lancet Planetary Health*, vol. 5, no. 8, pp. e514–e525, 2021.

31. A. Thomas, E. Theokritoff, A. Lesnikowski, D. Reckien, K. Jagannathan, *et al.*, "Global evidence of constraints and limits to human adaptation," *Regional Environmental Change*, vol. 21, no. 85, 2021.

32. A. J. Sietsma, R. W. Groenendijk, and R. Biesbroek, "Progress on climate action: a multilingual machine learning analysis of the global stocktake," *Climatic Change*, vol. 176, no. 173, 2023.

33. D. Thulke, Y. Gao, P. Pelser, R. Brune, R. Jalota, *et al.*, "Climategpt: towards ai synthesizing interdisciplinary research on climate change," *arXiv preprint arXiv:2401.09646*, 2024.

34. J. Elliott, R. Lawrence, J. C. Minx, O. T. Oladapo, P. Ravaud, *et al.*, "Decision makers need 'living' evidence synthesis," *Nature*, vol. 600, no. 7889, pp. 383–385, 2021.

35. R. van de Schoot, J. de Bruin, R. Schram, P. Zahedi, J. de Boer, *et al.*, "An open source machine learning framework for efficient and transparent systematic reviews," *Nature Machine Intelligence*, vol. 3, pp. 125–133, 2021.

36. M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, O. Edenhofer, T. F. Stocker, *et al.*, "The ipcc ar5 guidance note on consistent treatment of uncertainties: a common approach across the working groups," *Climatic Change*, vol. 108, no. 4, pp. 675–691, 2011.

37. M. Leippold, S. Vaghefi, D. Stammbach, V. Muccione, J. A. Bingler, *et al.*, "Automated fact-checking of climate change claims with large language models," *arXiv preprint arXiv:2401.12566*, 2024.

38. M. Majszak and J. Jebeile, "Expert judgment in climate science: how it is used and how it can be justified," *Studies in History and Philosophy of Science*, vol. 100, pp. 32–38, 2023.

39. K. Wu *et al.*, "An automated framework for assessing how well llms cite relevant medical references," *Nature Communications*, vol. 16, no. ???, p. ???, 2025. Finds 30–90% of statements unsupported across popular LLMs.

40. L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, and T. Liu, "A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions," *ACM Transactions on Information Systems*, 2024.

41. J. Clark and colleagues, "Generative artificial intelligence use in evidence synthesis: a systematic review," *Research Synthesis Methods*, pp. 601–619, 2025.

42. M. Stede and R. Patz, "The climate change debate and natural language processing," in *Proceedings of the 1st Workshop on NLP for Positive Impact*, pp. 8–18, 2021.

43. "Semantic scholar." `https://www.semanticscholar.org`.

44. "Elicit." `https://elicit.com`.

45. "Consensus." `https://consensus.app`.

46. "Paperguide." `https://paperguide.ai`.

47. "Scite." `https://scite.ai`.

48. "Connected papers." `https://connectedpapers.com`.

49. "Researchrabbit." `https://researchrabbit.ai`.

50. "Scholarcy." `https://scholarcy.com`.

51. "Scispace." `https://scispace.com`.

52. "Iris.ai." `https://iris.ai`.

53. IPCC Working Group III, "Technical summary. climate change 2022: Mitigation of climate change," 2022.

54. IPCC Working Group III, "Carbon dioxide removal (cdr) factsheet," 2022.

55. IPCC Working Group III, "Chapter 5: Demand, services and social aspects of mitigation," in *Climate Change 2022: Mitigation of Climate Change*, 2022.

56. IPCC Working Group III, "Annex iii: Scenarios and modelling methods," 2022.

57. G. P. Peters and colleagues, "Ar6 scenarios database: an assessment of current curation and use," *npj Climate Action*, 2023.

58. B. Cointe and colleagues, "The ar6 scenario explorer and the history of ipcc scenario databases," *npj Climate Action*, 2024.

59. IIASA, "Ar6 scenario explorer and database," 2022.

60. IPCC Working Group II, "Summary for policymakers. climate change 2022: Impacts, adaptation and vulnerability," 2022.

61. IPCC Working Group II, "Chapter 6: Cities, settlements and key infrastructure," 2022.

62. IPCC Working Group I, "Chapter 11: Weather and climate extreme events in a changing climate," in *Climate Change 2021: The Physical Science Basis*, 2021.

63. B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones, "Atypical combinations and scientific impact," *Science*, vol. 342, no. 6157, pp. 468–472, 2013.