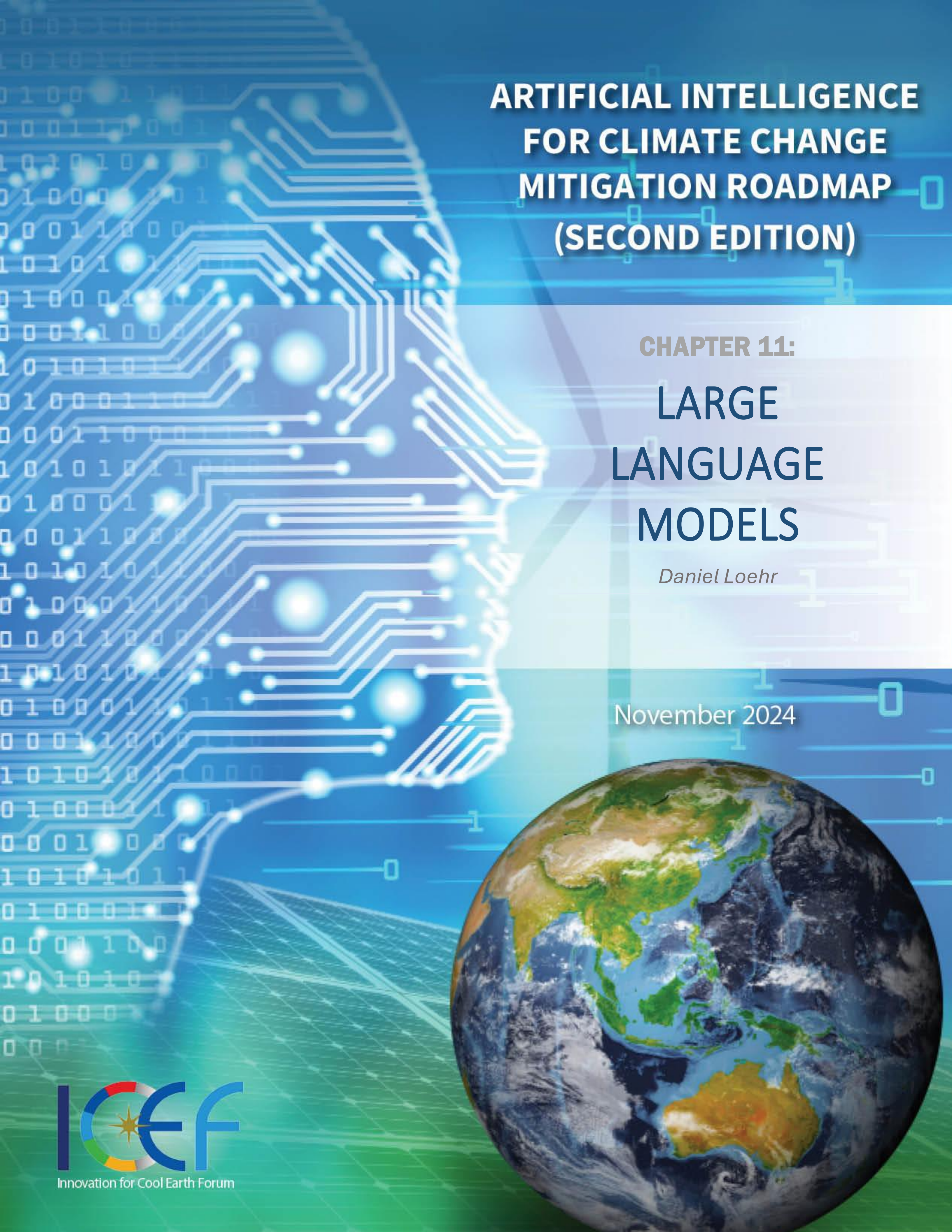
ARTIFICIAL INTELLIGENCE FOR CLIMATE CHANGE MITIGATION ROADMAP (SECOND EDITION)

CHAPTER 11:

LARGE LANGUAGE MODELS

Daniel Loehr

November 2024



Innovation for Cool Earth Forum

CHAPTER 11:

LARGE LANGUAGE MODELS

Daniel Loehr

In November 2022, the general public became aware of the power of artificial intelligence (AI) when OpenAI released a browser-based chat interface to its generative pre-trained transformer (GPT), a type of large language model (LLM). The generated text was so human-like the world experienced a "ChatGPT moment," in which many felt that AI (represented by LLMs) had now reached human performance.

LLMs have significant potential to help mitigate climate change. Already, LLMs are used in a variety of ways toward this goal. They help humans search and make sense of vast repositories of climate change information, from a variety of sources and in multiple languages. They identify sentiment and argument structure in human discussions of climate change. They find, classify and summarize climate change risks and impacts described within the growing breadth of climate literature.

In the future, LLMs hold even greater potential. They can serve as tutors in climate education, depict personalized climate consequences, and suggest individualized climate actions. They can advance basic science in climate change mitigation, from materials science for developing better batteries or carbon capture materials to sophisticated power grid management for incorporating dynamic renewable energy sources. They could also serve as guides to shortcut the current maze of permitting requirements that are causing a backlog in bringing carbon-free energy to the grid.

A. Background

i. Evolution of Natural Language Processing (NLP)

LLMs are an evolution of the 70-year-old field of natural language processing (NLP), in which computers process natural (human) languages. Table 11-1 shows common types of NLP.

Table 11-1. *Common types of natural language processing (NLP)*

NLP TYPE	DESCRIPTION
Machine Translation	Automatic language translation
Named Entity Recognition	Identifying entities in text, such as people, places and organizations
Sentiment Analysis	Identifying sentiment (opinion/viewpoint) in text
Search	Finding and retrieving user-relevant information in a specific set of text documents
Question Answering	Providing answers to specific questions, e.g. the answer "316 ppm" to the question "What was atmospheric CO2 concentration in January 1960?" (vs. searching on e.g. "historic atmospheric CO2" and receiving relevant documents)
Dialogue Management	Chat
Summarization	Generating summaries of longer texts
Topic Modeling	Identifying topics in documents
Argument Mining	Extracting argument structure from text

NLP TYPE	DESCRIPTION
Optical Character Recognition	Converting images of text into digital text
Speech Recognition	Converting speech into digital text
Speech Synthesis	Converting digital text into speech

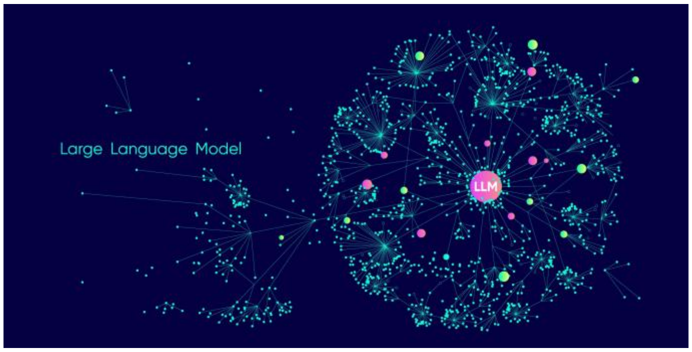
The general methodology of NLP can be divided into three historical paradigms.

- The earliest was rule-based, coding explicit instructions in the form of rules. For example, a Spanish-English translation system would include a rule to convert “casa” to “house.” However, these rules are difficult to write explicitly and often fail to capture nuances or unusual cases.
- The next paradigm, starting in the 1980s, was statistical, jettisoning explicit rules and taking advantage of the increasing amount of digital data. Here, the translation system would learn patterns from the available body of human-translated documents. For example, “house” is typically found in English translations of Spanish sentences containing “casa,” so the system learns to choose “house” as the translation.
- The current paradigm, LLMs, started in the early 2010s and is also essentially statistical but takes advantage of much more powerful statistical models based on neural nets. As described below, LLMs handle the translation task by converting text in one language into a mathematical representation of the words (an “embedding”) that captures their core meaning. The LLM then converts that representation into text in another language.

ii. Understanding Language Models (LMs)

LLMs are more directly evolved from a statistical-paradigm model called a language model (LM). LMs originally developed in the 1980s to enable a variety of NLP tasks. They are probabilistic models of a natural language. That is, LMs capture the probabilities of the sequences of words (or sometimes sub-words or characters) in a language.

LMs use sequences of words to derive *embeddings*, one of their core features. Embeddings are based on the idea that a word is defined “by the company it keeps.”[1] For example, two common senses of the word “bank”—a financial bank and a river bank—will occur in different contexts of surrounding words (e.g., near words like “loan” or “water”). A word is thus *embedded* in its context, and by capturing the surrounding words of every word in a body of text, an LM stores each word’s embedding.



Amazingly, embeddings allow the meaning of words to be treated like mathematical equations. A well-known example is that when the mathematical value of the word "man" is subtracted from the value for "king," and the value for "woman" is added, the resulting value is near the value for "queen."[2] Embeddings thus capture something essential about words, transferred out of the specific human language in which they occur. Specifically, embeddings are represented mathematically as vectors. The embedding vector for each word in a language is created by calculating which words it appears with most frequently. Further, a variety of downstream tasks use the *de facto* semantics that embeddings provide. For example, because words with similar embeddings are found in similar contexts, search algorithms can expand search terms with synonyms, by including words with vectors similar to those of the original search terms.

iii. Growing from Language Models (LMs) to *Large* Language Models (LLMs)

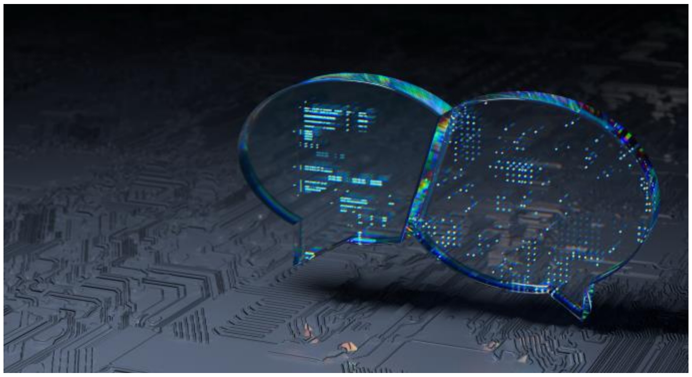
Large LMs (LLMs) are LMs of a much greater size than the original class of LMs. Though "large" is a relative term, it was first used in 2018 to describe a model called BERT (Bidirectional Encoder Representations from Transformers),[3] which contained 340 million parameters. A parameter is roughly equivalent to a connection or node in a neural network.

BERT made use of an effective new type of neural network, a transformer.[4] The original transformer, developed in 2017 to translate from English into German, had two parts. An encoder converted English text into its embeddings (capturing the semantics of the source text). A decoder converted the embeddings into the German text.

BERT used only the encoder part of transformers to generate high-quality embeddings. In contrast, LLMs, such as GPT, use only the decoder part of transformers to generate text from pretrained embeddings; hence the name generative pre-trained transformer (GPT).

Since BERT, the largest LLMs have grown to over a trillion parameters (though others have been designed to reduce parameter size while maintaining similar performance). Modern LLMs also use faster parallel processing methods than earlier word-by-word sequential approaches. The immense scale and speed of LLMs has driven much higher performance on language-related tasks than previous types of models.

There now exist dozens of LLMs, both proprietary and open source. Furthermore, though LLMs (as *language* models) started with text, embeddings need not be restricted to words. Pixels in images, audio clips, video frames, DNA sequences, computer code and many other types of data are best interpreted by models that are "aware" of the surrounding context. For this reason, LLMs can be multimodal, handling images, audio, video and other modalities, in addition to text.



Because LLMs are typically used to *generate* text, images and other modalities, the technology is a type of Generative AI or GenAI. Another common term is foundation model

(FM), which refers to systems with a general functionality (a "foundation") on which more specific applications can be built. For instance, ChatGPT is a specific chat system built on GPT, a general foundation. Though these three terms—LLM, GenAI and FM—describe slightly different types of systems, they overlap significantly and are often used interchangeably.

Note that while LLMs are often the most effective tool for many NLP tasks, thanks to their foundational capabilities, this is not always the case. For example, traditional optical character recognition (OCR) tools currently outperform the OCR capabilities of image-enabled LLMs.

iv. Improving and Evaluating Large Language Models (LLMs)

An ecosystem of new technologies has arisen to improve the output of LLMs:

- Prompt Engineering: LLMs generate output in response to prompts. Since the complexity of LLMs can yield greatly different responses to only slightly different prompts, a new discipline has emerged to create the most effective prompts for a given task. This can include providing the LLM with multiple examples (or "shots") of the desired response type.
- Retrieval-Augmented Generation (RAG): In RAG, the LLM searches a traditional database or trusted web source for information that it combines with its response. This can update the recency of information (incorporating information that has become available since the LLM was trained), allow companies to incorporate proprietary data, and reduce (but not eliminate) incorrect "hallucinations" to which LLMs are prone.
- Agentic Workflows: LLMs can act as agents in a collection of multiple LLMs working with each other and with external tools, such as search engines, to achieve a goal. New programming languages have been created to develop these systems.
- Fine Tuning: LLMs are typically trained as general-purpose models, which are then applied to a variety of specific domains. Yet they can also be fine-tuned by further training on domain-specific data. ClimateBert[5] and ClimateGPT[6] are two examples in the climate domain.

An important aspect of LLMs is evaluation of their performance, which not only records their astonishing progress but also drives their improvement by providing benchmarks to develop against. Dozens of evaluation frameworks have been created to test a variety of knowledge capabilities, such as question answering in a variety of subjects (e.g., logic, mathematics, commonsense reasoning and more). An example is Massive Multitask Language Understanding (MMLU), which contains 16,000 multiple-choice questions from 57 academic topics.[7] Figure 11-1 shows the performance of LLMs over four years on MMLU, revealing remarkable improvement, now reaching a human performance baseline of 90%. This also underscores how benchmarks are quickly being saturated and require replacement by more difficult ones.

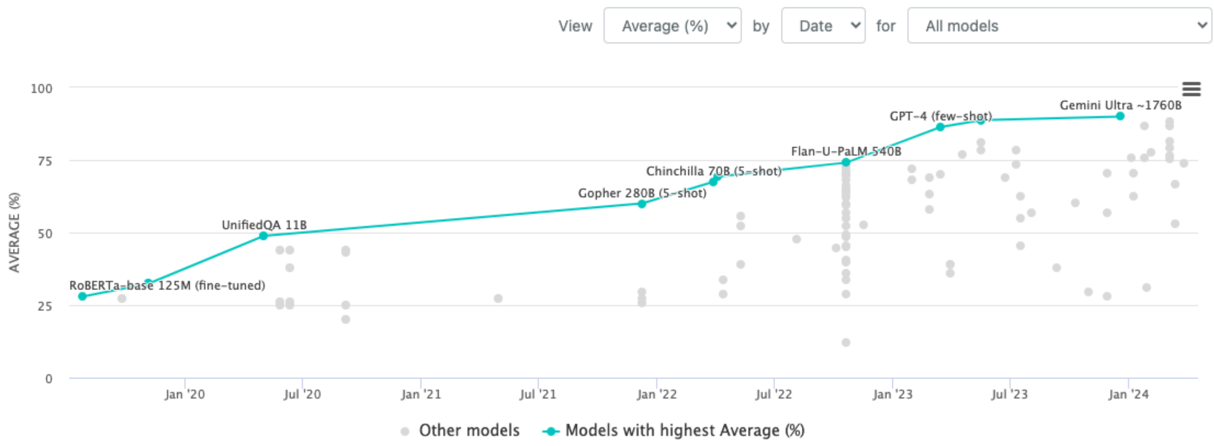


Figure 11-1. *Large language model (LLM) performance on Massive Multitask Language Understanding (MMLU) over time. From paperswithcode.com.[8]*

Templates called Winograd schemas are another evaluation framework used to evaluate LLMs. They are often used to test reasoning that is simple for humans but difficult for LLMs. In these templates, an answer depends on commonsense knowledge. For example, in the sentence "The trophy doesn't fit in the suitcase because it's too small," does "it" refer to the trophy or the suitcase? Does the answer change if "small" is replaced by "large"?[9]

LLMs have recently been evaluated specifically for their knowledge in the climate domain and have shown clear gaps in knowledge content and recency.[10,11] Newer LLMs such as ClimateGPT,[6] fine-tuned on climate data, are an effort to fill these gaps.

It is also necessary to evaluate more than knowledge capability. Equally important is assessing what is called *alignment*, meaning the extent to which LLMs are aligned with human values, such as helpfulness, harmlessness and honesty. This includes aspects such as ethics and morality, bias, toxicity, truthfulness and safety, including robustness against attacks. Benchmarks have been created to evaluate all these qualities.[12] Assessing human-aligned values is difficult by its very nature, as human judgments, often the source of the content of these benchmarks, are subjective and variable. Thus, the ability to evaluate LLMs' alignment with human values typically lags the ability to evaluate their knowledge capabilities.

B. General Uses of LLMs

Because of the hype surrounding LLMs, it can be difficult to determine exactly how they are currently being used. A 2024 Harvard Business Review[13] article researched actual usage by individuals, via online forums, and came up with six overall themes. These are listed in the first part of the table below, along with example use cases for each. Following those are an additional set of attested uses by organizations.

Table 11-2. *Uses of LLMS*

INDIVIDUAL USE OF LLMS	
Technical Assistance & Troubleshooting • Debugging software code • Writing Excel formulas • Manipulating data	*Learning & Education* • Generating a lesson plan • Giving simple explainers • Summarizing content
Content Creation & Editing • Generating ideas • Drafting emails • Writing and editing cover letters	*Creativity & Recreation* • Getting past writer's block • Recommending movies, books, etc. • Writing poems
Personal & Professional Support • Providing therapy/companionship • Providing business advice • Planning workouts	*Research Analysis & Decision-Making* • Conducting specific searches • Performing fact-checking • Developing critiques & counterarguments
ORGANIZATIONAL USE OF LLMS	
Software development assistance	*Creation of images and videos*
Business analytics	*Business analytics*
Personalized experiences • Marketing • Recommender systems	*Translation* *Search* *Data management*
Education • General training • Personalized tutoring	*Summarization* • Search results • Product reviews • Documents • Meeting notes
Generating documents • Business documents • Product descriptions	*User support (via chat, Q&A, or search)* • Customer support • Helpdesk • Product information

It is important to note that while LLMs are being *used* for these purposes and others, it is not yet clear how *useful* they are for these tasks. Nor is it clear whether LLMs are more useful than existing task-specific tools. For example, the search use case may be better served by traditional search engines optimized for the task. Interestingly, only 11% of companies had adopted LLMs at scale as of May 2024, according to McKinsey.[14]

It is also worth noting that in most use cases above, the LLM assists humans in carrying out tasks, rather than replacing them. This may be the real value of LLMs, in which artificial intelligence augments human intelligence. For example, LLMs can generate software code for common short programming tasks or write job application cover letters, but it cannot be relied on to guarantee the correctness of those products. Because the *presentation* of LLM output can appear so human-like, humans often assume LLMs' *content* is human-quality. Yet LLM content can be incorrect and even harmful, and human over-reliance on LLM output can be dangerous (see Section E). Nonetheless, humans can clearly benefit from LLM *assistance* with common tasks, in which humans provide a quality check before incorporating LLM output.

C. Using LLMs to Mitigate Climate Change

The use of natural language processing in studying climate change is not new. Traditional NLP has been used to help understand views expressed in online discussions and other texts concerning climate change for several decades.[15,16] However the advent of LLMs six years ago greatly enhanced the ability of NLP to help mitigate climate change. In light of their remarkable effectiveness and rapid evolution, LLMs have the potential to play a helpful and important role in climate change mitigation. In fact, LLMs are already being applied to climate change in a number of ways. Examples are shown in Table 11-3, categorized by NLP type.

Table 11-3. *Existing applications of large language models (LLMs) to climate change, categorized by natural language processing (NLP) type*

NLP TYPE	APPLICATION OF LLM TO CLIMATE CHANGE
Machine Translation	• Providing climate change information in Arabic[17] • Translating windmill operational codes to textual maintenance instructions[18] • Translating climate model components from Fortran to Python, to improve performance[19]
Named Entity Recognition	• Identifying specific geographic locations in climate literature and tracking regional impacts of climate change[20]

NLP TYPE	APPLICATION OF LLM TO CLIMATE CHANGE
Sentiment Analysis	• Determining stance on climate in news media[21] • Assessing human expert confidence in climate statements[22] • Estimating public opinion about global warming[23]
Search	• Improving search of climate laws and policies[24] • Mining the scientific literature for functional materials design[25] • Searching product descriptions against industry estimates of similar products' embodied carbon footprints[26]
Question Answering	• Answering questions about climate information in corporate earnings calls[27]
Dialogue Management	• Providing climate information from corporate sustainability reports via chat[28] • Providing organizations' and nations' net-zero information via chat[29]
Summarization	• Providing summaries of climate information from authoritative UN documents[30,31] or tailored to the user's specific geography[32]
Topic Modeling	• Detecting climate change topics in public documents[33] • Identifying environmental, social and governance (ESG) topics in news media[34] • Identifying climate change topics in insurance, carbon disclosure[35] and Nationally Determined Contribution documents[36] • Finding topics in the climate literature related to climate-induced infrastructure hazards[37]
Argument Mining	• Identifying narrative techniques in climate skeptic texts[38,39] • Using evidence-based reasoning for fact-checking of climate change claims[40,41]

LLMs provide another capability: classification. Indeed, the largest category of work applying LLMs to climate change involves classification, as listed below.

- Classifying evidence in building a dataset for verification of climate claims[42]
- Classifying climate risks in corporate disclosure reports to track trends[43] and analyze their impact on the credit default swap market[44]
- Classifying Task Force on Climate-related Financial Disclosure (TCFD) categories in corporate disclosure documents[45,46]
- Classifying presence/absence of net-zero claims[47] and climate risk type in corporate earnings calls[48]
- Classifying presence/absence of net-zero claims in laws and policies[49]
- Classifying environmental, social and governance (ESG) categories in corporate documents[47]

- Classifying presence/absence of climate-related text[50] or environmental claims[51] in a variety of document types
- Classifying climate change impacts found in the scientific literature[52]
- Classifying financial activities to estimate emissions of investments[53]
- Classifying climate change claims to benchmark a corporate greenwashing dataset[54]

Finally, the productivity enhancements provided by LLMs can speed up routine tasks, freeing humans to focus on innovation (e.g., allowing a chemistry lab to more quickly predict molecular structures with better carbon absorption capability).[55]

The ways in which LLMs are currently used to help mitigate climate change give good insight into the many ways they might be used for this purpose in the future. For example:

- LLMs can be especially helpful in education about climate change. LLMs can help develop accessible materials on climate change and act as personalized “climate tutors” to bring individuals up to speed on various climate topics.
- LLMs can also personalize the potential impacts of climate change. Non-LLM GenAI technologies can already create images of a user’s home or neighborhood under flood conditions to personalize climate change impacts.[56] Generative AI using LLMs could enable depictions of climate impacts in myriad other ways.
- In addition, LLMs can be monetized in business to develop personalized experiences in advertising and marketing. In this spirit, LLMs can be tuned so their responses include sustainability “nudges” (e.g., suggesting lower-carbon options when asked about recipes, investments, travel or other general topics).[57]

Other potential use cases of LLMs include:

- Summarizing policy documents
- Monitoring the extent of natural disaster impacts via social media
- Providing laypersons a natural language interface to specialized climate information tools and resources
- Creating synthetic data to stand in for privacy-containing data, such as residential smart meters to further smart grid research
- Identifying chemical names in scientific literature to assist in materials discovery
- Shortening the grid interconnection queue with predictive planning to help operators manage increasingly renewable energy sources[58]

More generally, the ability of LLMs to help with common tasks, such as data manipulation and software development, could augment AI practitioners’ technical efforts in the above use cases.

Finally, an important contribution of LLMs could lie in accelerating permitting for renewable energy (RE) siting, construction, storage and transmission—an urgent need in the United States and other geographies. In the United States, federally funded RE projects require an Environmental Impact Statement (EIS), and the average duration from initial notice to final decision is 4.5 years.[59] In addition, there is an "interconnection queue" of RE power and storage plants seeking connection to the national grid. Currently in queue is an active capacity of nearly 2.6 TW (~1.6 TW power and ~1 TW storage), twice the installed capacity of the entire US power plant fleet (~1.3 TW), and 95% of that queue is zero-carbon. However, the median duration from initial request to commercial operation is ~5 years.[60]

Part of the permitting delay is the work proposers must undertake to navigate the dozens of potential required permits at the local, state, tribal, interstate and federal levels.[61,62] LLMs are well suited for summarizing and extracting information from lengthy and complex documents, which could accelerate permitting. For example, LLMs could assist in processing voluminous public comments, automate application completeness checks, and extract and organize information from past permits, reviews and approvals to create a reference dataset useful for all stakeholders.[63,64] LLMs also estimate solar permitting risk for developers, based on zoning information.[65,66] LLMs can also help draft lengthy permit applications (an EIS alone averages over 600 pages[67]), by generating application text. For instance, Microsoft is using LLMs to generate documents for nuclear power regulatory approval.[68,69]

Such work would respond to federal permitting directives. For example, the 2022 White House Permitting Action Plan directs federal agencies to "identify, share, or develop … tools to assist project sponsors, permit applicants, affected communities, Tribal communities, and other stakeholders to navigate the environmental review and permitting process effectively."[70] In addition, the 2022 Inflation Reduction Act includes DOE funding for "actions that may improve the chances of, and shorten the time required for, approval by the siting authority of the application relating to the siting or permitting of the covered transmission project,"[71] and DOE is piloting the use of LLMs to streamline RE permitting.[72]

D. Barriers

Barriers to using LLMs to mitigate climate change include the following.

- Limited Interpretability: LLMs, which can contain hundreds of billions of numbers as parameters, are to a large extent "black boxes." It is difficult to understand how they arrive at their output, eroding trust in their answers related to climate change. Though work on AI interpretability is making LLMs somewhat more understandable, they are still largely opaque.
- Incorrect Information: LLMs are well known for "hallucinating" or making up incorrect information, also eroding trust and willingness to apply them to climate change. This can be mitigated using some of the techniques described above (e.g., RAG). But the opacity of LLMs makes it difficult to guarantee that information they supply is correct.
- Access Barriers: LLMs require huge capital investments for training and are thus currently concentrated within a few technology companies. This investment requirement can shut out the majority of potential climate mitigation practitioners, including smaller companies, the

global south and academia. Fortunately, a growing number of open-source and smaller-footprint LLMs are showing good performance.

- Intellectual Property Issues: Current litigation alleges copyright infringement of certain training data. Although many repositories of LLM training data in the climate domain actively encourage their dissemination, other climate information sources belong to organizations, such as the media, that protect their intellectual property. Thus, copyright issues could limit LLMs' current and future use of climate-related data.

E. Risks

Risks of using LLMs to mitigate climate change include the following.

- Bias: LLMs are trained on society's data (e.g., the Internet) and reflect society's biases. In the climate domain, much of the available training data are skewed toward the global north, which has a greater representation on the Internet. Recent work has tried to correct bias, but it is difficult and over-correction can yield factually incorrect output.
- Security Threats: Like any software, LLMs can be exploited. They can be subject to "jailbreaks" and tricked into operating outside their prescribed instructions. They are also vulnerable to leaking personal or proprietary information, such as residential smart meter data, which could be used to maliciously target household residents. It is difficult to enforce LLM guardrails, given their complexity and opacity.
- Greenhouse Gas (GHG) Emissions: LLMs are compute-intensive. The carbon footprint of AI in general is currently modest, but there is potential for growth. (See Chapter 15.) Mitigation gains achieved by LLMs in the fight against climate change could be partially undercut by their own GHG emissions.
- Incorrect Use: Though LLMs have captured the public's imagination and are thus turned toward a variety of uses, they are often not the right tool for the job. Consequences of incorrect use in the climate domain can range from simply being not as effective as other tools to disillusionment at not living up to hyped expectations to real-world damage if improperly used in critical applications, such as the power sector.
- Harmful Use: For every beneficial purpose of LLMs, there can be an opposite harmful purpose they are turned toward. For instance, in the climate domain, LLMs can be positively directed toward mitigation via education, marketing, content creation or software development. Yet LLMs could also use these capabilities for climate change denial, misinformation, or encouragement and development of GHG-emitting activities.

These issues are real obstacles to furthering application of LLMs in climate change mitigation, and work overcoming them requires as much focus as continued development of LLM capabilities.

F. Recommendations

1. *Private companies and academic researchers should continue to develop LLMs specifically trained on climate data and ensure they are openly available so the public can both improve them and benefit from them.*

2. *National governments, private companies, academic researchers and standards development organizations should cooperate on developing further benchmarks for evaluating LLMs' knowledge in the climate domain, thus extending the existing ecosystem for evaluating LLMs' knowledge in general.*

3. *Professional societies and academic experts should develop training programs on the proper use and limits of LLMs in mitigating climate change to help the public better understand the benefits and risks of using LLMs in the climate domain.*

4. *National governments, private companies and academic researchers should cooperate on developing public challenge competitions on proposed climate mitigation use cases of LLMs to advance their development.*

5. *National governments and private companies should expand current research and development (R&D) programs in addressing known issues with LLMs, so the public can place greater trust in LLMs, especially when applied to climate change.*

6. *LLM developers and users should publish fine-grained measurements of LLMs' carbon footprint by adopting tools to track and report the GHGs emitted by their compute time.*

7. *National governments should fund R&D for public-facing prototypes to advance the use of LLMs for accelerating permitting of renewable energy.*

G. References

1 J. Firth. A Synopsis of Linguistic Theory, 1930-1955 in *Studies in Linguistic Analysis* (ed J.R. Firth) pgs. 10-32 (Philological Society – Blackwell, Oxford, 1957, https://cs.brown.edu/courses/csci2952d/readings/lecture1-firth.pdf).

2 Emerging Technology from the arXiv. *King – Man + Woman = Queen: The Marvelous Mathematics of Computational Linguistics;* MIT Technology Review – Artificial Intelligence, https://www.technologyreview.com/2015/09/17/166211/king-man-woman-queen-the-marvelous-mathematics-of-computational-linguistics/ (2015).

3 Jacob Devlin *et al.* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 (2018). https://doi.org/10.48550/arXiv.1810.04805.

4 Ashish Vaswani *et al.* Attention Is All You Need. arXiv:1706.03762 (2017). https://doi.org/10.48550/arXiv.1706.03762.

5 Nicolas Webersinke *et al.* ClimateBert: A Pretrained Language Model for Climate-Related Text. arXiv:2110.12010 (2021). https://doi.org/10.48550/arXiv.2110.12010.

6 David Thulke *et al.* ClimateGPT: Towards AI Synthesizing Interdisciplinary Research on Climate Change. arXiv:2401.09646 (2024). https://doi.org/10.48550/arXiv.2401.09646.

7 Dan Hendrycks *et al.* Measuring Massive Multitask Language Understanding. arXiv:2009.03300 (2020). https://doi.org/10.48550/arXiv.2009.03300.

8 Papers with Code. *Multi-task Language Understanding on MMLU;* https://paperswithcode.com/sota/multi-task-language-understanding-on-mmlu (Accessed August 2024).

9 Ernest Davis. *Collection of Winograd Schemas;* https://cs.nyu.edu/~davise/papers/WinogradSchemas/WSCollection.html (Accessed August 2024).

10 Jannis Bulian *et al.* Assessing Large Language Models on Climate Information. arXiv:2310.02932 (2023). https://doi.org/10.48550/arXiv.2310.02932.

11 Hongyin Zhu & Prayag Tiwari. Climate Change from Large Language Models. arXiv:2312.11985 (2023). https://doi.org/10.48550/arXiv.2312.11985.

12 Zishan Guo *et al.* Evaluating Large Language Models: A Comprehensive Survey. arXiv:2310.19736 (2023). https://doi.org/10.48550/arXiv.2310.19736.

13 Marc Zao-Sanders. *How People Are Really Using GenAI;* Harvard Business Review – Technology and Analytics, https://hbr.org/2024/03/how-people-are-really-using-genai (2024).

14 Aamer Baig *et al. Moving past gen AI's honeymoon phase: Seven hard truths for CIOs to get from pilot to scale;* McKinsey Digital, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/moving-past-gen-ais-honeymoon-phase-seven-hard-truths-for-cios-to-get-from-pilot-to-scale (2024).

15 Andrés Alonso, José Manuel Carbó & J Manuel Marqués. Machine Learning methods in climate finance: a systematic review. *Banco de Espana* Working Paper No. 2310 (2023). http://dx.doi.org/10.2139/ssrn.4352569

16 Manfred Stede & Ronny Patz. The Climate Change Debate and Natural Language Processing in *Proceedings of the 1st Workshop on NLP for Positive Impact*, Bangkok, Thailand. pgs. 8–18, (2021; https://aclanthology.org/2021.nlp4posimpact-1.2.pdf).

17 Sahal Shaji Mullappilly *et al.* Arabic Mini-ClimateGPT : A Climate Change and Sustainability Tailored Arabic LLM in *Findings of the Association for Computational Linguistics: EMNLP 2023*, Singapore. (eds

Houda Bouamor, Juan Pino, & Kalika Bali) 14126–14136, https://doi.org/10.18653/v1/2023.findings-emnlp.941, (2023).

18 Joyjit Chatterjee & Nina Dethlefs. Natural language generation for operations and maintenance in wind turbines in *NeurIPS 2019 Workshop on Tackling Climate Change with Machine Learning, Vancouver, Canada*, https://www.climatechange.ai/papers/neurips2019/9, (2019).

19 Anthony Zhou, Linnia Hawkins & Pierre Gentine. Proof-of-concept: Using ChatGPT to Translate and Modernize an Earth System Model from Fortran to Python/JAX. arXiv:2405.00018 (2024). https://doi.org/10.48550/arXiv.2405.00018.

20 Tanwi Mallick *et al.* Analyzing Regional Impacts of Climate Change using Natural Language Processing Techniques. arXiv:2401.06817 (2024). https://doi.org/10.48550/arXiv.2401.06817.

21 Yiwei Luo, Dallas Card & Dan Jurafsky. Detecting Stance in Media on Global Warming. arXiv:2010.15149 (2020). https://doi.org/10.48550/arXiv.2010.15149.

22 Romain Lacombe, Kerrie Wu & Eddie Dilworth. ClimateX: Do LLMs Accurately Assess Human Expert Confidence in Climate Statements? , arXiv:2311.17107 (2023). https://doi.org/10.48550/arXiv.2311.17107.

23 Sanguk Lee *et al.* Can large language models estimate public opinion about global warming? An empirical assessment of algorithmic fidelity and bias. *PLOS Climate* 3, e0000429 (2024). https://doi.org/10.1371/journal.pclm.0000429.

24 Climate Policy Radar – Blog. *Building natural language search for climate change laws and policies;* https://climatepolicyradar.org/latest/building-natural-language-search-for-climate-change-laws-and-policies (Accessed August 2024).

25 Yifei Duan *et al.* Literature Mining with Large Language Models to Assist the Development of Sustainable Building Materials (Papers Track) in *International Conference on Learning Representations (ICLR) 2024 – Workshop: Tackling Climate Change with Machine Learning*, Vienna, Austria. https://www.climatechange.ai/papers/iclr2024/39, (2024).

26 Bharathan Balaji *et al.* CaML: Carbon Footprinting of Household Products with Zero-Shot Semantic Text Similarity in *Proceedings of the ACM Web Conference 2023*, Austin, TX, USA. pgs. 4004–4014, https://doi.org/10.1145/3543507.3583882, (2023).

27 Alexandra Luccioni, Emily Baylor & Nicolas Duchene. Analyzing Sustainability Reports Using Natural Language Processing in *Neural Information Processing Systems (NeurIPS) – Tackling Climate Change with Machine Learning workshop*, https://www.climatechange.ai/papers/neurips2020/31, (2020).

28 Jingwei Ni *et al.* CHATREPORT: Democratizing Sustainability Disclosure Analysis through LLM-based Tools. arXiv:2307.15770 (2023). https://doi.org/10.48550/arXiv.2307.15770.

29 ChatNetZero. *ChatNetZero: Helping to demystify net zero (Home Page);* Fremont, California, https://chatnetzero.ai/ (Accessed October 2024).

30 Natalia de la Calzada *et al.* ClimateQ&A : bridging the gap between climate scientists and the general public (Papers Track) in *International Conference on Learning Representations (ICLR) 2024 – Workshop: Tackling Climate Change with Machine Learning*, Vienna Austria. https://www.climatechange.ai/papers/iclr2024/3; https://arxiv.org/pdf/2403.14709, (2024).

31 Saeid Ashraf Vaghefi *et al.* ChatClimate: Grounding conversational AI in climate science. *Communications Earth & Environment* 4, 480 (2023). https://doi.org/10.1038/s43247-023-01084-x.

32 Nikolay Koldunov & Thomas Jung. Local climate services for all, courtesy of large language models. *Communications Earth & Environment* 5, 13 (2024). https://doi.org/10.1038/s43247-023-01199-1.

33 Francesco S. Varini *et al.* ClimaText: A Dataset for Climate Change Topic Detection. arXiv:2012.00483 (2020). https://doi.org/10.48550/arXiv.2012.00483.

34 Tim Nugent, Nicole Stelea & Jochen L. Leidner. Detecting Environmental, Social and Governance (ESG) Topics Using Domain-Specific Language Models and Data Augmentation in *Flexible Query Answering Systems: 14th International Conference (FQAS 2021) Proceedings*, Bratislava, Slovakia. pgs. 157–169, https://doi.org/10.1007/978-3-030-86967-0_12, (2021).

35 Tanmay Laud *et al.* ClimaBench: A Benchmark Dataset For Climate Change Text Understanding in English. *ArXiv* (2023). https://doi.org/10.48550/arXiv.2301.04253.

36 Tom Corringham *et al.* Bert classification of paris agreement climate action plans in *International Conference on Machine Learning (ICML) 2021 – Workshop on Tackling Climate Change with Machine Learning*, Paper #45, https://www.climatechange.ai/papers/icml2021/45, (2021).

37 Tanwi Mallick *et al.* Analyzing the impact of climate change on critical infrastructure from the scientific literature: A weakly supervised NLP approach. arXiv:2302.01887 (2023). https://doi.org/10.48550/arXiv.2302.01887.

38 Shraey Bhatia, Jey Han Lau & Timothy Baldwin. Automatic classification of neutralization techniques in the narrative of climate change scepticism in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pgs. 2167-2175, https://aclanthology.org/2021.naacl-main.175/, (2021).

39 Travis G. Coan *et al.* Computer-assisted classification of contrarian claims about climate change. *Scientific Reports* 11, 22320 (2021). https://doi.org/10.1038/s41598-021-01714-4.

40 Markus Leippold *et al.* Automated Fact-Checking of Climate Change Claims with Large Language Models. arXiv:2401.12566 (2024). https://doi.org/10.48550/arXiv.2401.12566.

41 Saeid A Vaghefi *et al.* Deep climate change: a dataset and adaptive domain pre-trained language models for climate change related tasks in *Conference on Neural Information Processing Systems (NeurIPS) 2022 – Workshop on tackling climate change with machine learning*, New Orleans, Louisiana, USA. Paper #27, https://www.climatechange.ai/papers/neurips2022/27, (2022).

42 Thomas Diggelmann *et al.* CLIMATE-FEVER: A Dataset for Verification of Real-World Climate Claims. arXiv:2012.00614 (2020). https://doi.org/10.48550/arXiv.2012.00614.

43 David Friederich *et al.* Automated Identification of Climate Risk Disclosures in Annual Corporate Reports. arXiv:2108.01415 (2021). https://doi.org/10.48550/arXiv.2108.01415.

44 Julian F Kölbel *et al.* Ask BERT: How Regulatory Disclosure of Transition and Physical Climate Risks Affects the CDS Term Structure*. *Swiss Finance Institute (Journal of Financial Econometrics, forthcoming)*, Research Paper No. 21-19 (2022). http://dx.doi.org/10.2139/ssrn.3616324.

45 Julia Anna Bingler *et al.* Cheap talk and cherry-picking: What ClimateBert has to say on corporate climate risk disclosures. *Finance Research Letters* 47, 102776 (2022). https://doi.org/10.1016/j.frl.2022.102776.

46 Rylen Sampson, Aysha Cotterill & Quoc Tien Au. TCFD-NLP: Assessing alignment of climate disclosures using NLP for the financial markets in *Conference on Neural Information Processing Systems (NeurIPS) 2022 – Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022*, New Orleans, Louisiana, USA. https://s3.us-east-1.amazonaws.com/climate-change-ai/papers/neurips2022/49/paper.pdf, (2022).

47 Tobias Schimanski *et al.* Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication. *Finance Research Letters* 61, 104979 (2024). https://doi.org/10.1016/j.frl.2024.104979.

48 Ming Deng *et al.* War and Policy: Investor Expectations on the Net-Zero Transition. *Swiss Finance Institute*, Research Paper No. 22-29 (2023). http://dx.doi.org/10.2139/ssrn.4080181

49 Matyas Juhasz *et al.* Identifying Climate Targets in National Laws and Policies using Machine Learning in *International Conference on Learning Representations (IRCL) 2024*, Vienna, Austria. https://s3.us-east-1.amazonaws.com/climate-change-ai/papers/iclr2024/26/paper.pdf, (2024).

50 Eduardo Garrido-Merchán, Cristina González-Barthe & Maria Coronado. Fine-tuning ClimateBert transformer with ClimaText for the disclosure analysis of climate-related financial risks. (2023). http://dx.doi.org/10.21203/rs.3.rs-3600821/v1.

51 Dominik Stammbach *et al.* Environmental Claim Detection. arXiv:2209.00507 (2022). https://doi.org/10.48550/arXiv.2209.00507.

52 Max Callaghan *et al.* Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nature Climate Change* 11, 966-972 (2021). https://doi.org/10.1038/s41558-021-01168-6.

53 Ayush Jain *et al.* Empowering Sustainable Finance: Leveraging Large Language Models for Climate-Aware Investments in *International Conference on Learning Representations (ICLR) 2024 – Tackling Climate Change with Machine Learning workshop*, Vienna, Austria. https://s3.us-east-1.amazonaws.com/climate-change-ai/papers/iclr2024/27/paper.pdf, (2024).

54 Gaku Morio & Christopher D. Manning. An NLP benchmark dataset for assessing corporate climate policy engagement in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, New Orleans, LA, USA. Article 1724, https://dl.acm.org/doi/10.5555/3666122.3667846, (2024).

55 Zhiling Zheng *et al.* ChatGPT Research Group for Optimizing the Crystallinity of MOFs and COFs. *ACS Central Science* 9, 2161-2170 (2023). https://doi.org/10.1021/acscentsci.3c01087.

56 A. Luccioni *et al.* Using Artificial Intelligence to Visualize the Impacts of Climate Change. *IEEE Computer Graphics and Applications* 41, 8-14 (2021). https://doi.org/10.1109/MCG.2020.3025425.

57 Cesar Ilharco *et al. Open-source LLM Alignment for Promoting Sustainability | Accelerating Climate Change Action;* Applied Machine Learning Days, EPFL, Switzerland, https://www.youtube.com/watch?v=KV0UWyZlcko (2024).

58 Beret Walsh. *AI & the Interconnection Queue: The Newest Intersection in Renewable Energy;* Landgate, https://www.landgate.com/news/ai-the-interconnection-queue-the-newest-intersection-in-renewable-energy (Accessed August 2024).

59 Executive Office of the President – Council on Environmental Quality. *Environmental Impact Statement Timelines (2010-2018);* National Environmental Policy Act - Department of Energy, Washington, D.C., https://ceq.doe.gov/docs/nepa-practice/CEQ_EIS_Timeline_Report_2020-6-12.pdf (2020).

60 Joseph Rand *et al. Queued Up: 2024 Edition – Characteristics of Power Plants Seeking Transmission Interconnection As of the End of 2023;* Lawrence Berkeley National Laboratory, https://emp.lbl.gov/sites/default/files/2024-04/Queued%20Up%202024%20Edition_R2.pdf (2024).

61 Shawn Enterline, Andrew Valainis & Ben Hoen. *Laws in Order: An Inventory of State Renewable Energy Siting Policies;* Lawrence Berkeley National Laboratory – Energy Markets and Policy, https://emp.lbl.gov/publications/laws-order-inventory-state-renewable (2024).

62 Rayan Sud & Sanjay Patnaik. *How does permitting for clean energy infrastructure work?;* Brookings – Research, https://www.brookings.edu/articles/how-does-permitting-for-clean-energy-infrastructure-work/ (2022).

63 Keith J. Benes, Joshua E. Porterfield & Charles Yang. *AI for Energy: Opportunities for a Modern Grid and Clean Energy Economy;* US Department of Energy (DOE), Washington, D.C., https://www.energy.gov/sites/default/files/2024-04/AI%20EO%20Report%20Section%205.2g%28i%29_043024.pdf (2024).

64 PolicyAI. *LLM for Environmental Review;* Kaggle, https://www.kaggle.com/competitions/llm-for-environmental-review/ (2024).

65 Symbium. *Symbium Solar Permits: Join Symbium's solar permitting pilot;* San Francisco, California, https://symbium.com/instantpermitting/solar/california/sb379 (Accessed August 2024).

66 Paces. *Quantifying Solar Permitting Risk with Large Language Models;* https://www.paces.com/post/permitting-risk-with-large-language-models (2023).

67 Executive Office of the President – Council on Environmental Quality. *Length of Environmental Impact Statements (2013-2018);* National Environmental Policy Act - Department of Energy, Washington, D.C., https://ceq.doe.gov/docs/nepa-practice/CEQ_EIS_Length_Report_2020-6-12.pdf (2020).

68 Solomon Klappholz. *Microsoft is using AI to get its nuclear projects approved in the US;* ITPro – Technology: Artificial Intelligence, https://www.itpro.com/technology/artificial-intelligence/microsoft-is-using-ai-to-get-its-nuclear-projects-approved-in-the-us (2023).

69 Jennifer Hiller. *Microsoft Targets Nuclear to Power AI Operations;* Microsoft Start, https://www.msn.com/en-us/money/other/microsoft-targets-nuclear-to-power-ai-operations/ar-AA1loccZ (2024).

70 The White House. *The Biden-Harris Permitting Action Plan to Rebuild America's Infrastructure, Accelerate the Clean Energy Transition, Revitalize Communities, and Create Jobs;* https://www.whitehouse.gov/wp-content/uploads/2022/05/Biden-Harris-Permitting-Action-Plan.pdf (2022).

71 117th Congress. *H.R.5376 - Inflation Reduction Act of 2022;* Congress.gov, https://www.congress.gov/bill/117th-congress/house-bill/5376/text (2022).

72 US Department of Energy (DOE). *DOE Announces New Actions to Enhance America's Global Leadership in Artificial Intelligence;* Washington, D.C., https://www.energy.gov/articles/doe-announces-new-actions-enhance-americas-global-leadership-artificial-intelligence#:~:text=DOE%20is%20investing%20%2413%20million,used%20to%20develop%20software%20to (2024).