NORTHWESTERN UNIVERSITY

BROWN

Center for Foundation Models and Generative AI

Center for Science of Science & Innovation

Sci2Pol:

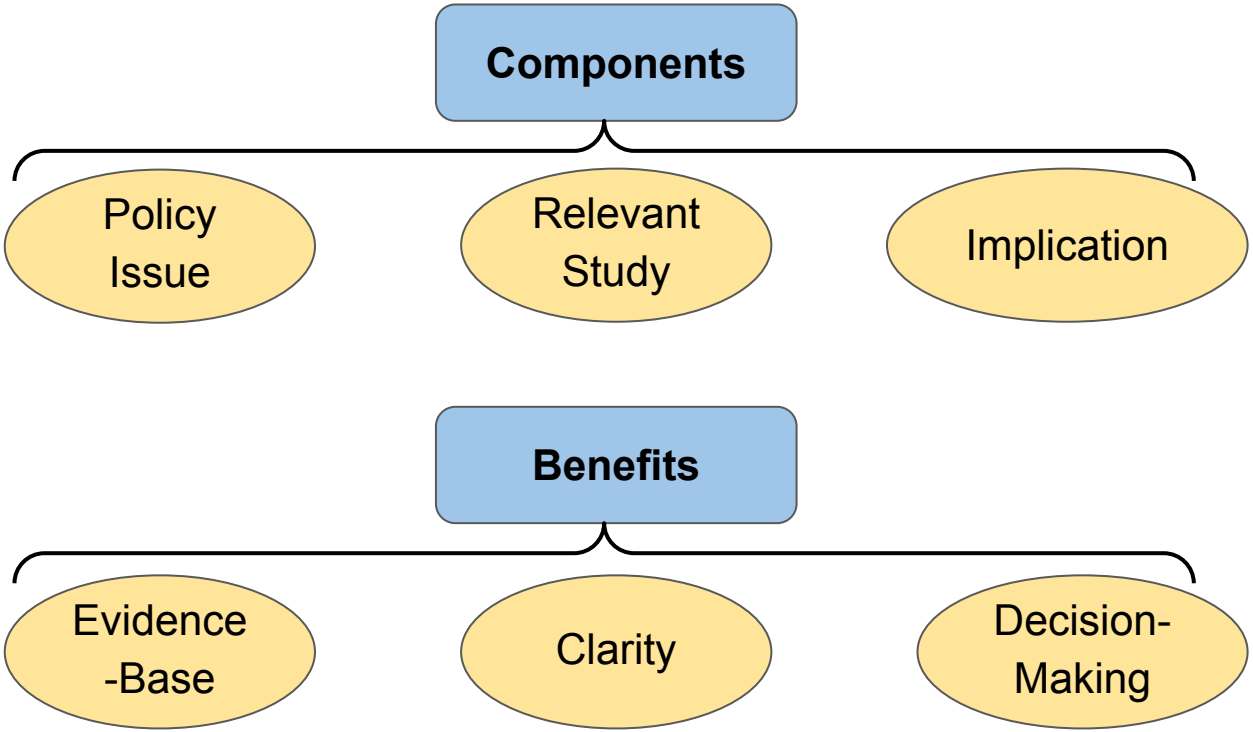
Evaluating and Fine-tuning LLMs on Scientific-to-Policy Brief Generation

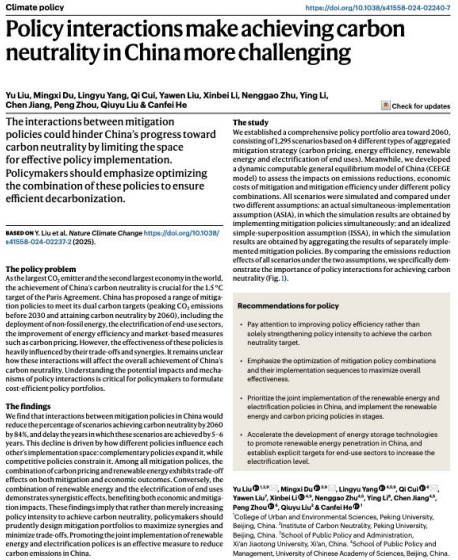
Weimin Wu, Alexander C. Furnas, Eddie Yang, Gefei Liu,
Akhil Pandey Akella, Xuefeng Song, Dashun Wang*, Han Liu*

Content

- **Motivation: Gap between Scientists and Policymakers**
- **Benchmark: Sci2Pol-Bench**
- **Training Dataset: Sci2Pol-Corpus**
- **Experimental Results**

Background: Policy Brief





Climate policy https://doi.org/10.1038/s41558-024-02240-7

Policy interactions make achieving carbon neutrality in China more challenging

Yu Liu, Mingxi Du, Lingyu Yang, Qi Cui, Yawen Liu, Xinbei Li, Nenggao Zhu, Ying Li, Chen Jiang, Peng Zhou, Qiuyu Liu & Canfei He

Check for updates

The interactions between mitigation policies could hinder China's progress toward carbon neutrality by limiting the space for effective policy implementation. Policymakers should emphasize optimizing the combination of these policies to ensure efficient decarbonization.

BASED ON Y. Liu et al. *Nature Climate Change* https://doi.org/10.1038/s41558-024-02237-2 (2025).

The policy problem

As the largest CO_2 emitter and the second largest economy in the world, the achievement of China's carbon neutrality is crucial for the 1.5 °C target of the Paris Agreement. China has proposed a range of mitigation policies to meet its dual carbon targets (peaking CO_2 emissions before 2030 and attaining carbon neutrality by 2060), including the deployment of non-fossil energy, the electrification of end-use sectors, the improvement of energy efficiency and market-based measures such as carbon pricing. However, the effectiveness of these policies is heavily influenced by their trade-offs and synergies. It remains unclear how these interactions will affect the overall achievement of China's carbon neutrality. Understanding the potential impacts and mechanisms of policy interactions is critical for policymakers to formulate cost-efficient policy portfolios.

The findings

We find that interactions between mitigation policies in China would reduce the percentage of scenarios achieving carbon neutrality by 2060 by 84%, and delay the years in which these scenarios are achieved by 5–6 years. This decline is driven by how different policies influence each other's implementation space: complementary policies expand it, while competitive policies constrain it. Among all mitigation polices, the combination of carbon pricing and renewable energy exhibits trade-off effects on both mitigation and economic outcomes. Conversely, the combination of renewable energy and the electrification of end uses demonstrates synergistic effects, benefiting both economic and mitigation impacts. These findings imply that rather than merely increasing policy intensity to achieve carbon neutrality, policymakers should prudently design mitigation portfolios to maximize synergies and minimize trade-offs. Promoting the joint implementation of renewable energy and electrification polices is an effective measure to reduce carbon emissions in China.

The study

We established a comprehensive policy portfolio area toward 2060, consisting of 1,295 scenarios based on 4 different types of aggregated mitigation strategy (carbon pricing, energy efficiency, renewable energy and electrification of end uses). Meanwhile, we developed a dynamic computable general equilibrium model of China (CEEGE model) to assess the impacts on emissions reductions, economic costs of mitigation and mitigation efficiency under different policy combinations. All scenarios were simulated and compared under two different assumptions: an actual simultaneous-implementation assumption (ASIA), in which the simulation results are obtained by implementing mitigation policies simultaneously; and an idealized simple-superposition assumption (ISSA), in which the simulation results are obtained by aggregating the results of separately implemented mitigation policies. By comparing the emissions reduction effects of all scenarios under the two assumptions, we specifically demonstrate the importance of policy interactions for achieving carbon neutrality (Fig. 1).

Recommendations for policy

- Pay attention to improving policy efficiency rather than solely strengthening policy intensity to achieve the carbon neutrality target.
- Emphasize the optimization of mitigation policy combinations and their implementation sequences to maximize overall effectiveness.
- Prioritize the joint implementation of the renewable energy and electrification policies in China, and implement the renewable energy and carbon pricing policies in stages.
- Accelerate the development of energy storage technologies to promote renewable energy penetration in China, and establish explicit targets for end-use sectors to increase the electrification level.

Yu Liu[1,2,3] ✉, Mingxi Du[1,2] ✉, Lingyu Yang[1,2,3,4], Qi Cui[1] ✉, Yawen Liu[1], Xinbei Li[4,5], Nenggao Zhu[4,6], Ying Li[4], Chen Jiang[4,5], Peng Zhou[4], Qiuyu Liu[1] & Canfei He[4] ✉

[1]College of Urban and Environmental Sciences, Peking University, Beijing, China. [2]Institute of Carbon Neutrality, Peking University, Beijing, China. [3]School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an, China. [4]School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing, China.

Motivation: Gap between Scientists and Policymakers

I'm applying for NSF funding. Our research on urban heat islands could greatly benefit city planning.

Interesting, what specific outcomes are you targeting?

%###...&&&%%%

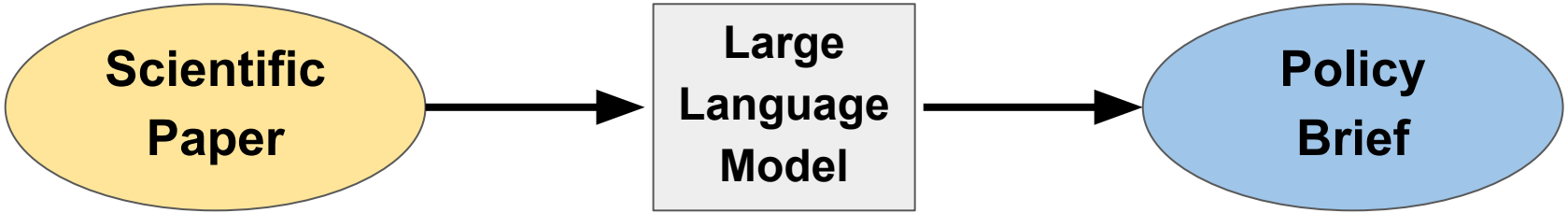
Continue: Gap between Scientists and Policymakers

Sorry I do not understand science. Can you translate these findings into policy actions?

Sorry I do not understand policy.

Then I need more time to reconsider your funding at this moment.

Whether LLM can Help?



Two key questions:

- To what extent can LLMs assist in scientific-to-policy brief generation?
- How can their performance be further improved?

Our Contributions

Sci2Pol-Bench

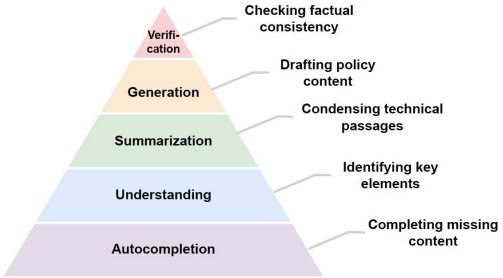
First comprehensive benchmark (18 tasks) for science to policy brief generation

Sci2Pol-Corpus

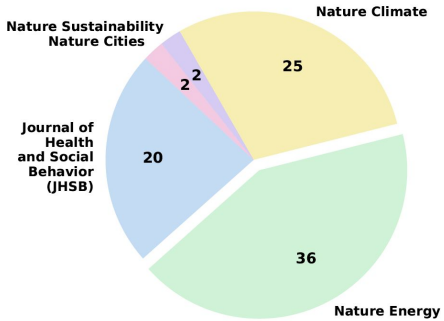
First large-scale training dataset with 639 paper brief pairs curated from 5.6M policy documents

Sci2Pol-Bench

Sci2Pol-Taxonomy



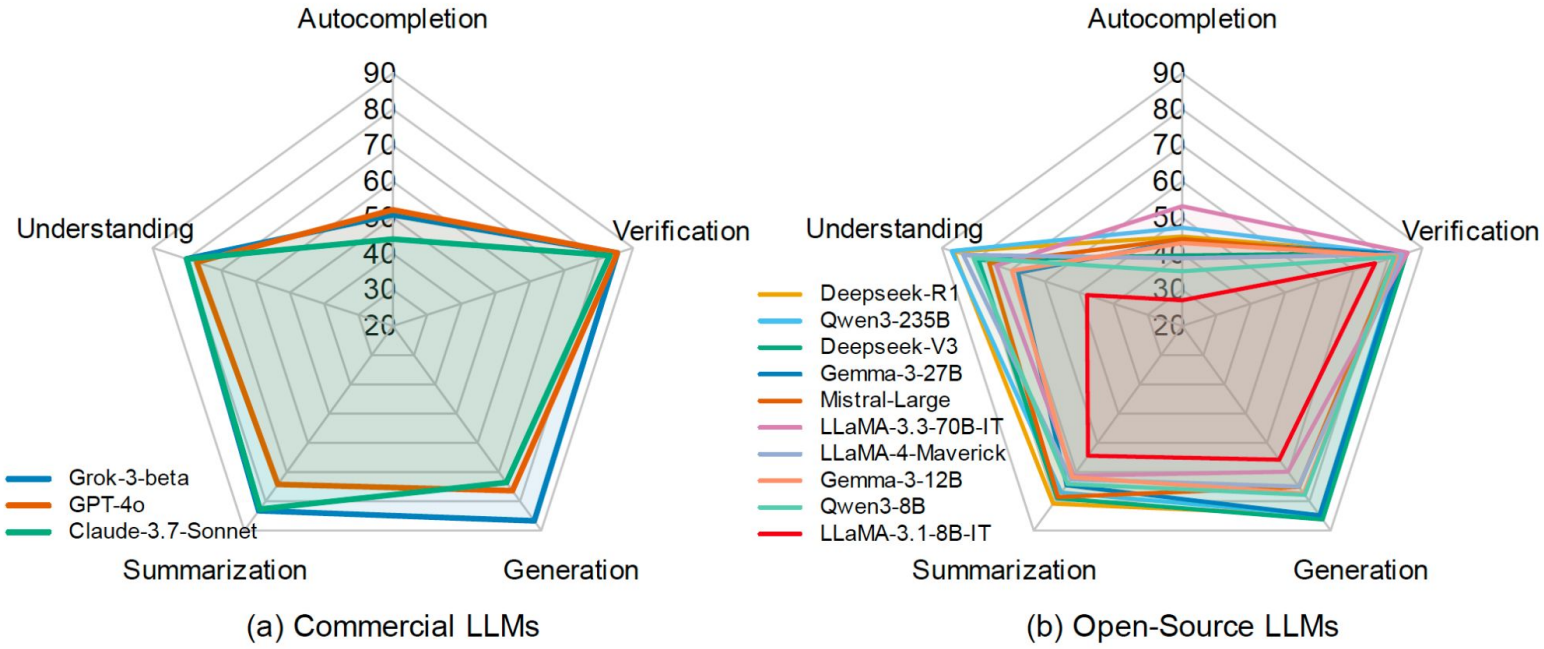
85 High-quality Pairs



18 Tasks

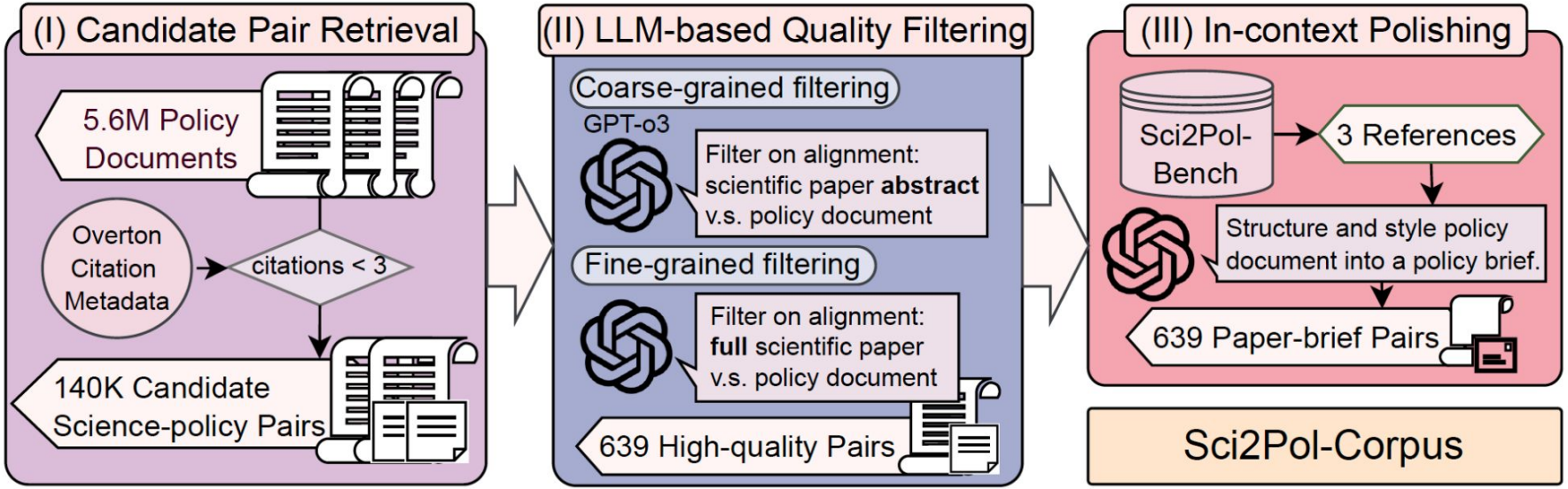
Taxonomy	ID	Task Description	Metric
Autocompletion	1	Scientific Text Autocompletion	Micro F1
	2	Political Text Autocompletion	
	3	Scientific Sentence Reordering	
	4	Political Sentence Reordering	
Understanding	5	Sentence Classification	Micro F1
	6	Scientific Knowledge Understanding	
Summarization	7	Policy Problem Summarization	Ref-free
	8	Research Findings Summarization	
	9	Study Methods Summarization	
	10	Policy Implications Summarization	
Generation	11	Policy Problem Generation	Ref-based
	12	Research Findings Generation	
	13	Study Methods Generation	
	14	Policy Implications Generation	
	15	Policy Brief Generation	
Verification	16	Scientific Claims Verification	Micro F1
	17	Scientific Claims Verification 2	
	18	Policy Implications Verification	

Benchmark Performance



Large room for improvement

Sci2Pol-Corpus



Supervised Fine-tuning

Model	Sci2Pol-Taxonomy					Avg.	Gain
	Auto. (1-4)	Under. (5-6)	Sum. (7-10)	Gene. (11-15)	Ver. (16-18)		
LLaMA-3.1-8B-IT	27.12±2.53	47.74±1.54	64.42±0.05	65.78±1.71	76.25±1.27	56.63±1.43	-
LLaMA-3.1-8B-SFT	31.27±2.90	44.34±1.48	78.28±1.25	77.62±1.55	80.59±1.08	64.27±1.70	**+7.64**
Gemma-3-12B	42.96±2.79	69.61±1.28	71.79±0.05	77.34±1.44	82.51±1.06	68.47±1.35	-
Gemma-3-12B-SFT	43.14±2.86	69.53±1.25	84.19±1.19	78.57±1.53	82.48±1.05	71.59±1.64	**+3.12**
Gemma-3-27B	43.60±2.83	67.82±1.42	74.55±0.05	84.82±1.16	84.29±0.98	71.40±1.28	-
Gemma-3-27B-SFT	45.39±2.90	67.44±1.36	86.36±1.06	81.53±1.60	84.06±0.96	**73.43**±1.64	**+2.03**
DeepSeek-V3	39.54±3.06	79.35±1.28	78.97±0.05	86.23±1.19	85.48±0.85	73.35±1.33	-
GPT-4o	52.17±3.00	77.17±1.32	74.23±0.06	76.39±1.28	85.45±0.82	72.12±1.32	-

Gemma-27B-SFT (73.43) surpasses GPT-4o (72.12) and DeepSeek-V3 (73.35)

Acknowledgement

National Science Foundation

Project Page: https://github.com/WeiminWu2000/Sci2Pol

Sci2Pol-Bench: https://huggingface.co/datasets/Weimin2000/Sci2Pol-Bench

Sci2Pol-Corpus: https://huggingface.co/datasets/Weimin2000/Sci2Pol-Corpus

Thank You!