

Model	CLIMATE-ASSETS	CLIMATE-DATASETS	CLIMATE-GREENHOUS	CLIMATE-HAZARDS	CLIMATE-IMPACTS	CLIMATE-MITIGATIONS	CLIMATE-MODELS	CLIMATE-NATURE	CLIMATE-OBSERVATIO	CLIMATE-ORGANISMS	CLIMATE-ORGANIZATI	CLIMATE-PROBLEM-	CLIMATE-PROPERTIES
gpt-4o-mini	0,22	0,27	0,48	0,57	0,22	0,25	0,48	0,27	0,25	0,2	0,63	0,31	0,25
gpt-4	0,28	0,34	0,62	0,6	0,33	0,44	0,49	0,34	0,35	0,32	0,59	0,24	0,32
llama 70	0,23	0,37	0,5	0,45	0,16	0,3	0,47	0,25	0,4	0,25	0,51	0,16	0,27
llama 405	0,19	0,27	0,74	0,56	0,36	0,44	0,49	0,33	0,38	0,31	0,64	0,33	0,32
Average, all classes	0,23	0,31	0,56	0,55	0,27	0,36	0,48	0,3	0,35	0,27	0,59	0,26	0,29

Dataset: Climate-Change-NER. Average performance of each of the tested models for all types of prompts per class. The three best scores are in bold font.

Model	ENVIRONMENT	LOCATION	MATTER	ORGANISM	PHENOMENA	QUALITY
gpt-4o-mini	0,31	0,23	0,17	0,43	0,29	0,29
gpt-4	0,38	0,3	0,19	0,52	0,33	0,33
llama 70	0,3	0,24	0,15	0,49	0,28	0,28
llama 405	0,29	0,24	0,19	0,46	0,29	0,32
Average, all classes	0,32	0,25	0,16	0,48	0,3	0,3

Dataset: BiodivNER. Average performance of each of the tested models for all types of prompts per class. The two best scores are in bold font, and the third and the fourth ones are the same value (0.3).