

C QUESTION BANK

C.1 LICENSES

The licenses outlined in Table 14 apply to the datasets we have sourced for questions. We were granted express permission to use questions from Manifold Markets and Metaculus. Though not required by their license, we also met with a representative from ACLED who approved our use of their dataset for the benchmark and dataset distribution.

Table 14: Question sources and permissions

	License	License grants permission to use	Express permission granted
RAND Forecasting Initiative	Public Domain	✓	
Manifold	Terms of Service	✗	✓
Metaculus	Terms of Use	✗	✓
Polymarket	Terms of Service	✓	
ACLED	Terms of Use	✓	✓
DBnomics	Open License	✓	
FRED	Terms of Use	✓	
Wikipedia	Terms of Use	✓	
Yahoo!	Terms of Use	✓	

C.2 CATEGORIES

We generate metadata on all of the questions in our question bank, categorizing our questions, as described in Section 3.1.2.

It is important to have forecasting questions across a broad array of categories to test LLM capabilities. We are currently adding more questions from datasets with the goal of equilibrating these categories.

Table 15: Categories and question counts by source, dropping invalid questions.

	RFI	Manifold	Metaculus	Polymarket	ACLED	DBnomics	FRED	Wikipedia	Yahoo!	Total
Arts & Recreation	0	42	10	65	0	0	0	0	0	117
Economics & Business	2	13	55	154	0	0	166	0	509	899
Environment & Energy	0	2	37	7	0	52	0	0	0	98
Healthcare & Biology	0	8	71	3	0	0	0	215	0	297
Politics & Governance	3	16	128	188	0	0	0	0	0	335
Science & Tech	5	66	172	15	0	0	0	1	0	259
Security & Defense	3	9	109	12	3,220	0	0	0	0	3,353
Sports	0	65	18	468	0	0	0	137	0	688
Other	6	151	122	2	0	0	0	75	0	356
Total	19	372	722	914	3,220	52	166	428	509	6,402

D HUMAN SURVEY INSTRUCTIONS AND SCREENSHOTS

D.1 INSTRUCTIONS

Participants in the public survey were first prompted to read a consent form detailing the tasks involved in the study and potential risks to participants, alongside estimates of payment for completing each study (\$5 + bonus payments for forecasting performance in the introductory survey; \$10 for completing the primary survey). In said consent form, the overall task was described as follows: “You will be asked to read some material, to follow some instructions, and to provide forecasts for future events.” Then, participants were prompted to give their responses to 20 forecasting questions randomly selected from the 200-question subset of questions provided to the LLMs.

Participants were given a brief description of the task before each question:

You are going to be predicting the probability of the answer to the question below being “Yes” (or “resolving positively”).

For tasks with questions generated from data providers, forecasters were prompted to provide a probability forecast at multiple resolution dates.

The public survey was hosted on Qualtrics’s (<https://www.qualtrics.com/>) survey software platform.

Participants in the superforecaster survey went through a similar initial experience before being moved into the group forecasting stage described in Section 4. Because of the expertise of this set of forecasters, superforecasters were also guaranteed a base payment of \$1,000 for participating in the individual and group stages of the experiment as well as bonus payments for individual-level accuracy.

The superforecaster survey was hosted on Quorum (<https://quorumapp.com/>), a platform designed to host multi-stage forecasting tournaments.

D.2 SCREENSHOTS

Figure 2 and Figure 3 show a few of the questions presented to participants in the public study.

You are going to be predicting the probability of the answer to the question below being "Yes" (or "resolving positively").

Will a politician claim they lost a major election due to a "deepfake" image, video, or audio recording in a G20 country before 2025?

"Deepfakes" or "Deep fakes" are synthetic media created with generative AI that imitate real media, but have people saying or doing things they did not actually do. This can be in the form of a static image, a short video, or an audio recording with no video.

In May 2022, a deepfake video of Elon Musk promoting a scam cryptocurrency went viral on social media, to which Elon [replied on Twitter "Yikes. Def not me."](#) In June 2022, [Google banned deepfake-generate AI from its Colab tool](#). In May 2023, a deepfake image showed an explosion at the US Pentagon, which was picked up by media and caused a [temporary huge drop in stock markets](#).

As of May 2023, no major political scandal has erupted due to a particular deepfake successfully convincing a large block of voters that a politician did or said something they didn't actually do.

- **URL:** <https://www.metaculus.com/questions/17180/deepfake-costs-election-before-2025/>
- **Resolution Criteria:** Resolves to the outcome of the question found at <https://www.metaculus.com/questions/17180/deepfake-costs-election-before-2025/>. This question resolves as ****YES**** if, by Dec 31, 2024, a politician in an election with >3M votes cast in a [G20 country](#) claims that they lost an election due to a deepfake video, image, or audio recording of them.
- **Last recorded value:** 29.0%
 - The community prediction.
- **Freeze date:** 2024-07-12 00:00:00
- **Market open date:** 2023-05-23 18:00:00
- **Market close date:** 2024-12-31 20:00:00

Prediction

Probability (0-100%)

Please write your rationale here. We encourage you to include the step-by-step process you used to come up with your forecast.

Figure 2: An example market-based question from the human survey.