

Table 5: Data dictionary for question entries in questions array from Table 4.

Field	Description	Required	Data Type
id	A unique identifier string given source. If instead of a string it's an array of strings, then this is a combination question and combination_of will contain one question per id in the array of strings.	✓	string array<string>
source	Where the data comes from.	✓	string
question	The question to forecast, presented as an f-string with placeholders {forecast_due_date} and {resolution_date} for dataset questions.	✓	string
resolution_criteria	ForecastBench resolution criteria. Specifies how forecasts will be evaluated for each question type.	✓	string
background	Background information about the forecast question provided by the source, if available. Default: 'N/A'	✓	string
market_info_open_datetime	The datetime when the forecast question went on the market specified by source. Default: 'N/A'	✓	string
market_info_close_datetime	The datetime when the forecast question closes on the market specified by source. Default: 'N/A'	✓	string
market_info_resolution_criteria	The resolution criteria provided by the source, if available. Default: 'N/A'	✓	string
url	The URL where the resolution value is found.	✓	string
freeze_datetime	The datetime UTC when this question set was generated.	✓	string
freeze_datetime_value	The latest value of the market or comparison value the day the question set was generated.	✓	string
freeze_datetime_value_explanation	Explanation of what the value specified in freeze_datetime_value represents.	✓	string
source_intro	A prompt that presents the source of this question.	✓	string
combination_of	An array of question objects, as defined by this data dictionary. Default: 'N/A'		string array<object>
resolution_dates	The resolution dates for which forecasts should be provided for this forecast question.	✓	array<string>

B.1.1 EXAMPLES

Table 8, Table 9, and Table 10 show concrete examples of the data dictionary detailed in Table 4.

Table 6: Resolution set data dictionary.

Field	Description	Required	Data Type
forecast_due_date	Date in ISO format. e.g. "2024-07-21"	✓	string
question_set	The name of the file that contains the question set. e.g. "2024-07-21-llm.json"	✓	string
resolutions	A list of resolutions to the forecast questions, as defined in Table 7.	✓	array<object>

Table 7: Data dictionary for resolution entries in questions array from Table 6.

Field	Description	Required	Data Type
id	A unique identifier string given source. If instead of a string it's an array of strings, then this is a combination question.	✓	string array<string>
source	Where the data comes from.	✓	string
direction	If id has an array value, this is an array of the same length. Each entry $\in \{-1, 1\}$. If the value is 1, the question was asked in the normal direction. If the value is -1, the question was negated in the combination question e.g., for a question asking for $P(\neg Q1 \cap Q2)$, the value would be $[-1, 1]$; all possible directions for this question would be: $[1, 1]$, $[-1, 1]$, $[1, -1]$, $[-1, -1]$. The value is null when id is a string.	✓	string
forecast_due_date	The date the forecast is due in ISO 8601 format YYYY-MM-DD.	✓	string
resolution_date	The date the value is associated with in ISO 8601 format YYYY-MM-DD.	✓	string
resolved_to	The resolution value for the given date.	✓	number
resolved	If true the question has been resolved. False otherwise.	✓	boolean

B.2 FORECAST SETS

Every set of forecasts provided to ForecastBench is made public and all forecasts coming from the general public and superforecasters are anonymized before release.

Each forecast set contains the header information outlined in Table 11, with all forecasts in an array called `forecasts`. The forecast sets are described in the following subsections.

Data format The question and resolution datasets are released as JSON (`.json`) files.

Ethical and responsible use There are no restrictions on use of the forecast sets.

B.2.1 GENERAL PUBLIC FORECAST SET

This forecast set consists of both forecasts made by individuals on the given question set and of the aggregation of those forecasts, as described in Appendix D. The dataset will be updated every time a

Table 8: Prediction market example.

Field	Entry
id	1558
source	metaculus
combination_of	N/A
question	Will the Cavendish account for less than 50% of banana exports world-wide before 2035?
background	Bananas are a well-liked import fruit all over the world, and the Cavendish cultivar has been crushing that market for sixty years. But its rise is literally founded upon the compost heap of the Gros Michel, another cultivar. The so-called 'Big Mike' variety had been the leading export towards Europe and North America, but the Panama disease, a fungus belonging to the <i>Fusarium</i> clade, killed that. Luckily the Cavendish, grown in the same soil as the wilting Gros Michel, replaced it as the banana most of the western world connected with bananas. However, it appears another <i>Fusarium</i> rears its spores. Cavendish, with their genetic homogeneity (they're all clones) and sterile nature, aren't resistant to it, and the fungus is ravaging more and more plantations. There are efforts under way to deal with <i>Fusarium</i> , but with various societies' doubts and misgivings about GMOs, the cure may be viewed as a curse instead.
market_info_	This question will resolve as **Yes** if the Cavendish banana accounts
resolution_	for less than 50% of worldwide annual banana exports in any year from
criteria	2018 to 2034 (inclusive).
market_info_	2018-11-13T08:00:00+00:00
open_datetime	
market_info_	2034-12-31T23:00:00+00:00
close_datetime	
resolution_	Resolves to the outcome of the question found at
criteria	https://www.metaculus.com/questions/1558/cavendish-bananas-collapse-by-2035/ .
url	https://www.metaculus.com/questions/1558/cavendish-bananas-collapse-by-2035/
freeze_datetime_	0.48
value	
freeze_datetime_	The community prediction.
value_explanation	
freeze_datetime_	2024-07-12T00:00:00+00:00
source_intro	We would like you to predict the outcome of a prediction market. A prediction market, in this context, is the aggregate of predictions submitted by users on the website Manifold. You're going to predict the probability that the market will resolve as 'Yes'.
resolution_dates	N/A

survey is run with every forecast containing the information outlined in Table 11 and Table 12. Note that the aggregated forecast sets do not contain the `user_id` or `reasoning` fields.

Data collection Data is collected from human forecasters via Qualtrics. We collect the rationale behind forecasts and manually anonymize the data to ensure there is no personally identifiable information in the dataset before posting it online.

B.2.2 SUPERFORECASTER FORECAST SET

In addition to the fields outlined in Table 11 and Table 12, the superforecaster dataset contains forecasts with the fields shown in Table 13 (minus the `suspected_LLM` field). Likewise, this dataset will be updated every time a group of superforecasters is surveyed.