

# Propel AI hands-on session: AI superforecasting<sup>1</sup>

Led by Christopher Potts, Stanford University

## The plan

We will break up into small groups and engage in an informal session of superforecasting about AI. There are a few warm-up questions to help the groups prepare for superforecasting, one question about chocolate, and then a series of statements about the future of AI. You should feel free to choose a subset of the statements to focus on, and to revise the statements however you like. In the last 15 minutes, we will collect the claims and evaluations, to store them somewhere (with or without attribution) for future assessment.

## Your areas of expertise

Try to identify the collective areas of expertise represented in your group. This should include not just professional and intellectual topics, but also life experiences – your background and culture, your hobbies, the places you’ve lived, the people you get to interact with, the technologies you use or have tried to use, and so forth. Accurate superforecasting in an area is not always positively correlated with expertise in that area; diverse, informed groups tend to do better.

## Your past predictions

Where have you made predictions in the past that have proven correct, and where have you really missed the mark? You could think of political events, sports, reality TV, product launches, economic trends, performance on an exam – any area where your prognostications were ultimately put to an objective test. The goal is to get a sense for different people’s helpful and perhaps unhelpful biases.

## Group name

In light of the above, what should we call your group?

---

<sup>1</sup>I have tried, to the extent possible in a short session, to follow the guidelines and principles in Philip Tetlock and Dan Gartner’s *Superforecasting: The Art and Science of Prediction* (2016).

# 1 Supertasting

You have two chocolates from Dandelion Chocolate to taste, labeled A and B. Both are 70% cocoa, 30% cane sugar. Dandelion assigns them very different flavors, though. Each group should come to a consensus about which chocolate received which flavor description from Dandelion.

| Dandelion’s description                                    | Your choice (A or B) |
|--|----------------------|
| “We taste honey mango, crème fraîche, and cocoa powder.”   |                      |
| “We taste notes of cherry, yogurt, and toasted sourdough.” |                      |

**How did your group arrive at this consensus?**

## 2 An AI-penned bestsellers

**Claim to evaluate** By 2028, there will have been at least one novel on the New York Times bestseller list that is written by an AI.

**Group’s revision to the core claim (optional)**

**Some potentially relevant evidence**

1. The context window (max sequence length models can process) has gone from around 5,000 tokens in 2019 to 10,000,000 tokens in 2025.
2. Modern LMs learn to embed high-level linguistic structures, which help them remain coherent but may hamper creativity.<sup>2</sup>
3. The widely praised 2021 essay ‘Ghosts’ was written by a human (Vauhini Vara) with large parts written by GPT-3. The essay is a meditation on the death of Vara’s sister.<sup>3</sup>
4. The 2023 book *I Am Code: An Artificial Intelligence Speaks: Poems* is listed on Amazon as “code-davinci-002 (Author), Brent Katz (Editor), Josh Morgenthau (Editor), Simon Rich (Editor)” and is described as “an autobiographical thriller written in verse by an AI itself”.<sup>4</sup>

**Score (0–100 where 0 ≈ No and 100 ≈ Yes)**

<sup>2</sup><https://arxiv.org/abs/2407.02446>

<sup>3</sup><https://www.thebeliever.net/ghosts/>

<sup>4</sup><https://www.amazon.com/Am-Code-Artificial-Intelligence-Speaks/dp/0316560065>

### 3 AI incidents

**Claim to evaluate** Before January 1, 2030, we will have experienced at least one incident in which (1) an autonomous AI caused harm with costs over \$1B, and (2) the AI's actions were not the result of deliberate guidance by humans.

#### Group's revision to the core claim (optional)

#### Some potentially relevant evidence

1. The 2011 Atlantic article "Does Anne Hathaway News Drive Berkshire Hathaway's Stock?" reports that Berkshire Hathaway's stock goes up whenever Anne Hathaway stars in a positively reviewed movie. This is likely due to correlations in the data that no one can track down (or no one has bothered to try to track down).<sup>5</sup>
2. In 2008, United Airlines' stock briefly plunged when Google News picked up an old story about a previous United Airlines bankruptcy and the Google News algorithm put the current date on it because the page had no date of its own. Trading was halted and the stock quickly recovered.<sup>6</sup>
3. In 2023, a hoax tweet seeming to come from the Associated Press said that there had been an explosion at the Pentagon. The stock market briefly plunged. AI experts quickly spotted artifacts in the images and established that they were fake. Trading was halted and the markets quickly recovered.<sup>7</sup>
4. In 2023, deep fake audio recordings seemed to capture one of the candidates in Slovakia's elections discussing a plan to rig the election. Due to constraints on political reporting in Slovakia right before an election, the media felt it could not investigate thoroughly. The candidate ultimately lost the election.<sup>8</sup>

**Score (0–100 where 0 ≈ No and 100 ≈ Yes)**

---

<sup>5</sup><https://www.theatlantic.com/technology/archive/2011/03/does-anne-hathaway-news-drive-berkshire-hathaways-stock/72661/>

<sup>6</sup><https://www.nytimes.com/2008/09/09/business/worldbusiness/09iht-AIR.1.16001879.html>

<sup>7</sup><https://www.washingtonpost.com/technology/2023/05/22/pentagon-explosion-ai-image-hoax/>

<sup>8</sup><https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/>

## 4 The future of software engineering

**Claim to evaluate** The number of people employed as software engineers in 2028 will be *lower* (in absolute terms) than the number of people employed as software engineers in 2023 (26.9M).

**Group’s revision to the core claim (optional)**

### Some potentially relevant evidence

1. In 2022, 26.9M people worldwide were employed as software engineers. This is up from 21M in 2016.
2. Since at least 2000, the number of software engineers has risen at a much faster rate than the average increase in employment in that time period.
3. The CEO of Google recently said that AI writes 25% of the code at Google.<sup>9</sup> The CEO of Anthropic recently said that, within 12 months, AI will write “essentially all of the code”.<sup>10</sup>
4. In 2016, prominent AI researcher Geoffrey Hinton predicted that “It’s just completely obvious that within five years deep learning is going to do better than radiologists. . . . It might be 10 years, but we’ve got plenty of radiologists already.” The number of radiologists in the workforce has increased since 2016, and, at present, there is a shortage of qualified radiologists in the world.<sup>11</sup>
5. There is anecdotal evidence that AI coding tools are allowing smaller start-ups to hire at a slower pace than they would have 10 years ago.

**Score (0–100 where 0 ≈ No and 100 ≈ Yes)**

---

<sup>9</sup><https://www.forbes.com/sites/jackkelly/2024/11/01/ai-code-and-the-future-of-software-engineers/>

<sup>10</sup><https://www.aol.com/anthropics-ceo-says-3-6-062716453.html>

<sup>11</sup><https://newrepublic.com/article/187203/ai-radiology-geoffrey-hinton-nobel-prediction>

## 5 AI usage costs

**Claim to evaluate** In 2030, the median API cost of the 10 most widely used GenAI models in the world will be over \$100 per million tokens. (Models without priced APIs are excluded.)

**Group's revision to the core claim (optional)**

### Some potentially relevant evidence

1. GenAI API pricing has come down very dramatically since 2021. For example, the cost to use GPT-3 in November 2021 was \$60 per million tokens, whereas at present the cost to use a comparable model is \$0.06 per million tokens.<sup>12</sup>
2. At present, the big model providers seem to be losing money while smaller startups thrive.<sup>13</sup>
3. David Cahn at Sequoia Capital has concluded that there is a \$600B gap between investment in AI infrastructure (especially GPUs) and revenue.<sup>14</sup>
4. OpenAI may be holding back plans to make money while it sorts out its unusual governance structure.<sup>15</sup>
5. A large number of firms (including Meta) have released excellent GenAI models under usage terms that allow them to be served to customers using any (or at least many) cloud providers.
6. At present, the majority of GPUs used for AI are designed by NVIDIA (U.S.-based), manufactured by TSMC (Taiwan Semiconductor Manufacturing Company), and used by U.S. firms.

**Score (0–100 where 0  $\approx$  No and 100  $\approx$  Yes)**

---

<sup>12</sup><https://a16z.com/llmflation-llm-inference-cost/>

<sup>13</sup><https://calpaterson.com/porter.html>

<sup>14</sup><https://www.sequoiacap.com/article/ais-600b-question/>

<sup>15</sup><https://www.bloomberg.com/opinion/articles/2025-02-11/sure->

## 6 AI pets

**Claim to evaluate** In the year 2040, AI pets will account for 0.7% of household pets in the U.S. (about the same percentage as currently have rabbits as pets).

**Group's revision to the core claim (optional)**

**Some potentially relevant evidence** There are already lots of things called "AI pets" on the market. Some seek to imitate familiar pets in some way or another, some seek to imitate dinosaurs, and others are entirely of their own kind.

**Score (0–100 where 0  $\approx$  No and 100  $\approx$  Yes)**