



The Death and Life of Prediction Markets at Google

Dan Schwarz

Over the past two decades, Google has hosted two different internal platforms for predictions. Why did the first one fail — and will the other endure?

It's July 2005. Google is the darling of Silicon Valley. It has just unveiled Google Maps; Gmail is still in beta. Next year it will acquire YouTube and launch Google Translate.

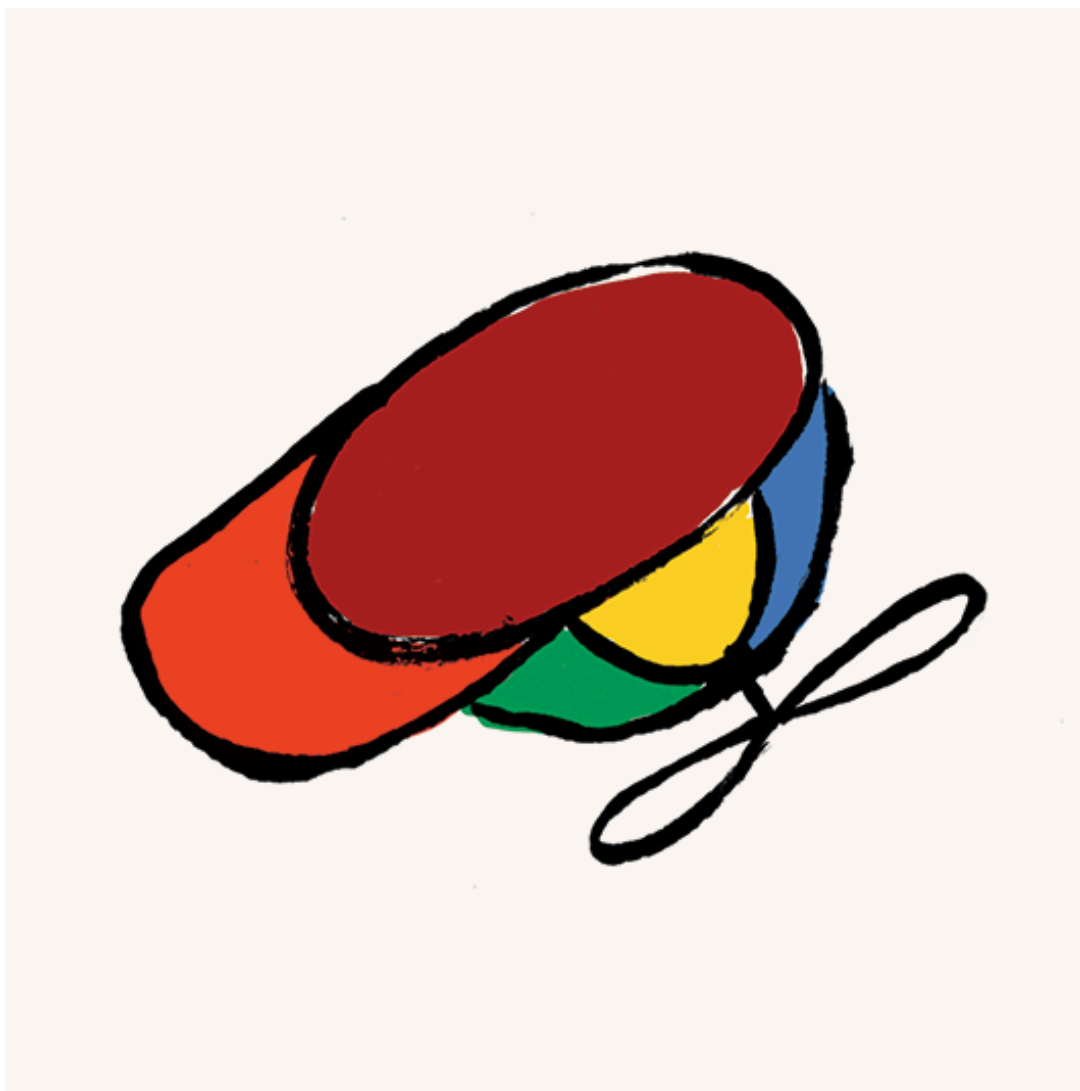
The week's new hires file past a full-size dinosaur skeleton in the courtyard on the way to their first TGIF — the company's weekly all-hands. They wear beanies with red, yellow, green, and blue colors — like the yet-to-be-designed Chrome logo — with a propeller on top. They are here to see Google's founders, Larry Page and Sergey Brin, both wearing shorts and plain colored t-shirts, banter about new tech.

In the first line of their Founders' IPO letter, Page and Brin wrote "Google is not a conventional company." They sought to provide "unbiased, accurate and free access to information." On this Friday, Patri Friedman, the grandson of Milton Friedman, and Bo Cowgill, now an economics professor at Columbia University, are here to talk about Google's next bet to do this: an internal prediction market called Prophit.

On stage next to Larry Page, Friedman and Cowgill announce winners from Prophit's first quarter and show statistical results on its forecasting accuracy. Prophit was popular inside Google. Over the next three years, about 20% of all Google employees would place bets.


Two months after this presentation, *The New York Times* covered Prophit. They wrote about it again in 2008, and it became a *Harvard Business Review* case study. Despite the momentum, in March 2010, Prophit hit a major roadblock in a public

launch as an external product. It attempted a pivot, and ultimately shut down in 2011.



In April 2020, almost exactly 15 years after Prophet launched, and one month after employees worldwide were sent home due to COVID-19, I launched Google's second internal prediction market, called Gleangen.¹ By late 2022, about 8% of Google employees had placed a bet on Gleangen. The company had grown so large that 8% represented 15,000 people — ten times the number of employees that placed bets on Prophet. Gleangen sustained over 1,000 monthly traders, more than the popular forecasting platforms Metaculus (where I later served as CTO) and Manifold at the time.²

Outside of Google, prediction markets have once again been thrust into the spotlight. Weeks before it became a mainstream view, users on sites like PredictIt and Metaculus predicted not only that President Biden would drop his re-election campaign but that doing so would increase the Democrats' chance of winning. Over the next few months, swings in election prediction markets regularly made

national news, and despite some distortions caused by aggressive whales, the markets ultimately performed well. Theoretically, prediction markets are equally powerful when used by companies to anticipate events, such as predicting their competitors' next moves. 

But as Cowgill has shown, corporate prediction markets have a mixed track record, as evidenced by attempts at Google and many other companies. Why is this? What does it take to make them work? The inside story of Prophit and Gleangen, the two largest corporate prediction markets ever run, offers some lessons.

A new type of information economy

Probably the first corporate prediction market was built in 1990 by economist Robin Hanson for employees of the doomed hypertext startup Project Xanadu. It had 18 participants, who bet on questions like whether cold fusion would be developed by 1991. The decade was dotted with other prediction market experiments, like the Iowa Electronic Markets (an online real money prediction platform established in 1988) and Hollywood Stock Exchange (a 1996 virtual market for film-related options).

Researchers like Hanson and others, such as the economist Justin Wolfers, worked out some of the implementation details from this data. Crucially, they also showed that most markets — whether operating with either real or play money — were relatively accurate, much more so than even the most accurate individuals. In 2004, the journalist James Surowiecki published the widely-read *The Wisdom of Crowds*, detailing cases where collective judgment surpassed individual expertise. Cowgill read the book the same year, and decided to start a prediction market at Google.

The company had some key ingredients to make a market work. First, it had a culture open to experimentation. Googlers were famously permitted 20% of their working hours to experiment on new products.³ Cowgill posted his idea for “Google Decision Markets” on a message board. Engineers Ilya Kirnos, Doug Banks, and Piaw Na responded to help, and Patri Friedman joined the project a few months later.

A second ingredient was Google's economic expertise. Hal Varian, Google's chief economist then and now, wrote the book on the economics of the internet, and helped design Google's famous ad auction system. On Hal's suggestion, the Prophit team used the prediction market structure first described by economists at the University of Iowa and based on the Iowa Electronic Markets.⁴

Prophet launched to Google employees in April 2005. Market prices were featured on the company intranet home page. It was an immediate hit.



Betting for fun and profit

In Google's early days, long before it became a blue-chip stock, its offices were full of LEGOs, ball-pits, and Magic the Gathering cards. Product design mirrored office design, and Google built applications full of bright colors, easter eggs, and April Fools' jokes.

Prophet was such a product: serious with a dash of whimsy. Markets tracked over 60% of all Google quarterly objectives, such as "How many Gmail users will there be by the end of Q2?" Some questions aimed to predict its competitors' next moves, such as "Will Apple launch a computer based on Intel's Power PC chip?" Yet one third of all markets were marked as "fun," of interest to employees but with no clear connection to Google's business (e.g., the quality of Star Wars Episode III or gas prices).

Cowgill told me an anecdote of both the seriousness and the whimsy. A senior executive saw Prophet give a very low probability that the company would complete the hire of a new senior executive on time (filling the position had been a quarterly objective for the past six quarters). "The betting on this goal was extremely harsh. I am shocked and outraged by the lack of brown-nosing at this company," the executive said to laughter in a company-wide meeting. But the market was the nudge the execs needed. They subsequently "made some hard decisions" to complete the hire on time.

The crowd of Googlers produced accurate predictions, spurred by competition for \$10,000 of awards each quarter, plus the even-more-valuable prestige of the leaderboards. In a 2009 paper which analyzed the first 2.5 years of Prophet, Cowgill found that the platform demonstrated high calibration that improved over time. (It wasn't perfect. There was an optimism bias — bettors thought positive outcomes for Google would happen more often than they actually did.⁵ There were also strong correlations in the bets made by people who sat within a few feet of each other.)

But that was just Prophet's internal beta. The product was designed to launch externally, not to be an internal decision-making tool. It was meant to become Google's next world-changing innovation in information technology.

Sign up for our newsletter to get Asterisk's latest interviews, essays, and more.



Your email address

Subscribe

Economists vs. the United States


Google popularized⁶ many now standard tech practices: on-site cafés, A/B tests, and “dogfooding,” or first releasing new products internally where they can be improved before launching to the public.

Prophit was a logical next step in the company's mission to organize the world's information and make it universally accessible and useful. That included all the world's books, all the world's roads, and all the world's languages — why not all the information in the heads of the world's humans? Prediction markets offered a mechanism to systematically elicit reliable information from human judgment. Cowgill relayed to me what happened next, a previously unreported chapter in the Prophit story.

The biggest obstacle preventing Prophit's public launch was the inconvenient fact that online gambling was (mostly) illegal, and the laws surrounding it were not uniform across states. States attorneys general could claim Prophit fell within their jurisdiction, requiring Google to adhere to 50 different state policies.⁷

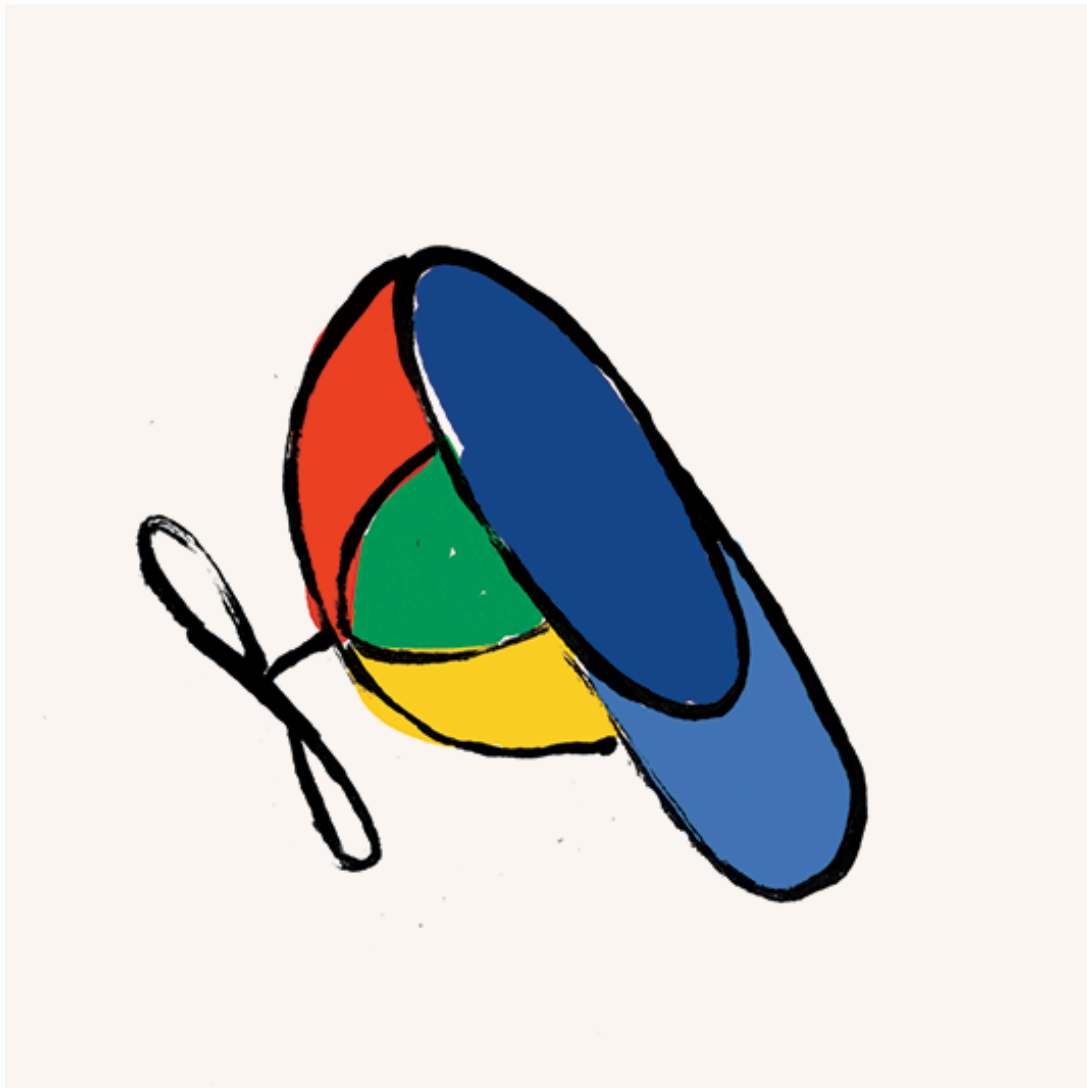
Prophit, along with Google's legal and policy teams, developed a strategy to convince a federal regulator to pre-emptively assert jurisdiction over Prophit. It was Google's opinion that the Commodity Futures Trading Commission would be the most natural and friendly regulator, in part because the CFTC had already issued a “no action letter” protecting the Iowa Electronic Markets⁸ back in 1993. In May 2008, the CFTC published a letter expressing concerns about the potential overlap between prediction markets and gambling, but left the door open for possible exemptions, and invited public comments on how to regulate such markets.

The CFTC staffed many economists, and they seemed to understand the moment. That same month, an article called *The Promise of Prediction Markets* appeared in the prestigious journal *Science*, arguing that prediction markets held too much

research and decision-making potential to be fettered by government restrictions. Its authors included four Nobel laureates in economics, numerous academic proponents of markets (including Philip Tetlock, Robin Hanson, and Robert Forsythe), as well as Google's own Hal Varian. 

Google hired a lobbying firm with extensive CFTC ties. They also organized support from Yahoo! and Microsoft, where prediction markets had been popular among senior executives. The three companies, despite their rivalry in their core markets, jointly backed a single proposal for the CFTC to legalize prediction markets. In March 2010, Cowgill and Ilya Kirnos, another of Prophet's creators, went to Washington to lobby this proposal to the CFTC, as well as officials in the White House and Congress.

But by the time the CFTC digested Google's proposal, the political appetite for financial deregulation had changed. The 2008 financial crisis hit its peak with the collapse of Lehman brothers in September. In the interim years, many had come to blame the crisis on the lack of oversight of new financial products. In July 2010, just a few months after Cowgill and Kirnos's visit to Washington, the Dodd-Frank Act was passed, overhauling nearly all federal financial regulatory agencies. According to Googlers working on Prophet in DC and Mountain View, despite Silicon Valley's organized push, the intellectual momentum for regulatory reform had died.



As the meetings in DC stalled, so too did progress back in Mountain View. Cowgill told me that it became hard to get resources to continue lobbying — and, crucially, engineering headcount to support the external launch. This put Prophit in limbo.

Although internally successful, without a path to a public launch, it underwent a surprising — if brief — pivot. Cowgill had by that point joined Hal Varian's economics team, where he tried to implement prediction markets within different projects. One such project was a social network for Chinese users because Google Plus (Google's attempt at a social network) was blocked there. Cowgill, together with engineers in Chinese offices, integrated Prophit's engine to power a forecasting feature in the app. But it failed to take off, and Google later killed the project.

By that point, most of Prophit's core team had moved on. Friedman left in 2008 to start the Seasteading Institute. Banks and Na left Google in 2010. Kirnos kept running markets until the last ones paid out in mid-2011 and left in 2012 for a

startup that was later acquired by Twitter. Cowgill started a PhD in Economics at Berkeley. “I regret that we shut down Prophit,” he wrote to the Prophit group in 2012. “We should have treated the internal instance as a product in its own right, not as a stepping stone to going public.”

The flow of information at Waymo

I’d been fascinated by prediction markets since I graduated college in 2011. I read about Prophit, and I was influenced by the same intellectual momentum that had brought Cowgill and Kirnos to Washington. I joined Google in 2014, where I worked first on Google Search and then on Google Translate. In early 2018, I had been promoted twice to Senior Software Engineer, and I began looking for new opportunities.


I dug up all the code and internal emails on Prophit that I could find. Cowgill’s e-mail expressing regret for not pursuing Prophit as an internal-only app caught my attention. Why not build a new prediction market, I thought, only this time designed *from the beginning* to influence decision making inside Google?

I met with Cowgill, Friedman, and Varian. All three expressed reservations about trying again — as did my friends from other divisions, whom I would need to help me pitch Area 120, Google’s in-house incubator for full-time “20% projects,” to staff “Prophit 2.0.” And so I set the idea aside and instead found a role on the systems engineering team at the self-driving car company Waymo. But while at Waymo, I quickly found a concrete use case for a prediction market.

Waymo’s systems engineering team measured safety through a variety of metrics from real driving, simulated driving, and models. Yet this data was fragmented — different orgs across the 1,000 person company had different access levels, and worse, different definitions for what they were measuring.

I was working on a new metric modeling the chance of collision from a particular maneuver. I spoke with engineers about how their projects might improve safety as measured by this metric. Opinions differed widely. And this was crucial information — engineering prioritization at the highest level depended on expectations of which potential projects would most likely improve safety of various car maneuvers.

I realized I could get the most accurate assessments of the likely impact of projects by aggregating the predictions of the crowd of engineers. And I could envision how these predictions could change priorities in practice, thanks to two famous stories of goals in Waymo’s history, the two “Founder’s Challenge

In the first, in early 2009, Larry Page challenged what was then known as Project Chauffeur to complete 10 different autonomously driven 100-mile routes in California within two years. The team succeeded with three months to spare. Page's second challenge, issued in 2015, was for Waymo to provide 1,000 driverless rides per week within one year. On this challenge, the company fell far short. By 2020, five years later, the company was still only completing about 50-200 such rides each week. 

This reminded me of Cowgill's anecdote about the senior manager who used Proffit to intervene on a failing initiative. I imagined what would have happened if the second Founder's Challenge had had a company-wide betting pool on when the milestone would be met. Might a Waymo executive have done what the early Google executive had done — intervened when the probability of success dropped, asked bettors why the outcome was in doubt, and changed course?

I decided to build a prediction market to find out, using my team's safety metrics as the targets to forecast. I made a prototype, onboarded employees, sourced predictions, got UX feedback, and iterated. Once I had minimal liquidity, I presented the resulting forecasts to every senior manager that would take a meeting.

In 2019, although Waymo had by then become a separate company from Google, it was in many ways still attached. It had never moved out of the warehouse it shared with Google X. It used the same HR system and the same tech stack. I even continued to use my Google laptop and Google badge.

In one key way, though, Waymo was not like Google. Access to data was much more restricted.

Proffit worked, in part, because of internal transparency. When a market forecast how many Gmail users would join next quarter, it was based on a value that was visible to everyone at Google. Google was famously internally transparent compared to other tech giants.

One Waymo VP, upon seeing these safety metric forecasts, said this would help communicate these metrics across company divisions. And then, to my amazement, he said this was counter to his division's goal to *restrict* information like this. The core mechanism of prediction markets — using the wisdom of crowds — can be antithetical to the common management desire to control who knows what.⁹

I failed to find any management support for the prediction market. But I  I seen the potential. Less than one year after joining Waymo, I searched Google ,

listings for the role most aligned with forecasting, got a role on the supply chain forecasting team, and transferred back to the mothership.



The end of Google's 20% culture

In early February 2020, one month after I started my new role, I saw a question on Metaculus asking if a new coronavirus might lead to a global pandemic. One month later, all 150,000 Alphabet employees were sent home. Demand for information surged, and management didn't have answers. So I used my extra time in lockdown to revamp and relaunch what I was now calling Gleangen as a prediction market for all Google employees.

Early markets, bet on by thousands of employees, centered on our uncertain future: when offices would re-open in various countries, the date vaccines would be available, and how at-home and in-office work would be balanced. Sev |

times, such markets gave high probabilities to the creation of pandemic policies long before they were actually enacted.



My main challenge in making Gleangen a useful internal tool for strategy was generating enough users, bets, and liquidity on markets about Google's core businesses. Over the next 18 months, while working my day job in supply chain forecasting, I wrote, predicted, and resolved over a thousand such markets. I grew a mailing list and chat group, and hosted socials, unconferences, and talks. As Prophit had done, I got approval to pay out valuable prizes to complement the play-money leaderboards. I posted links to Gleangen on so many internal memes that people told me to stop. I once spent an entire weekend hanging flyers in every bathroom I could access across dozens of buildings in campuses across the Bay Area.


I never presented at TGIF, though Google CEO Sundar Pichai did answer a TGIF question about a Gleangen market. And while the time for puff pieces from *The New York Times* about Google had long passed, I did manage to get approval to write a public piece in December 2021 that — while mostly an ad for Google Cloud — explained the basics of prediction markets, highlighted Gleangen's over 10,000 users, and included screenshots of the interface.

Like Prophit, Gleangen only got official staffing after achieving success from its incubation during engineers' "20% time." It was official company policy that all Google employees in good standing could work on projects of their choosing one day per week, as long as they were for the benefit of Google. Internal job postings listed hundreds of 20% project opportunities.

But by 2020, despite no change in official policy, the concept of 20% time at Google was on life-support.¹⁰ None of the three managers I had during this phase approved of me spending one day a week on Gleangen. (Neither did any of the over 20 other 20% contributors I recruited to administer the site, add features, and manage contests.) Gleangen became my full-time job in early 2022.

The wisdom of the Googler crowd

From the beginning, my plan was to learn from Cowgill, Friedman, and Varian's experience — and mistakes — with Prophit. "The plan we executed was to focus on having as much adoption, buzz, and usage by employees as possible," Cowgill told me. "Google was making choices about public products based on buzz and adoption, with the mantra of 'user growth now, monetize later.'" What they didn't do, he said, was sell their product to decision-makers the way one might

enterprise decision support software. “That’s probably how it should be done,” he said, “with lots of consulting and oversight of each application and client.” 

So I ran pilots to produce useful forecasts for division after division, from Nest to the Quantum Computing team. While most managers were supportive, few were willing to allocate resources.


I was well aware of the known obstacles to prediction market adoption in the workplace: unwillingness to share data, the desire for plausible deniability when projects fail, the risks of market manipulation, and good old-fashioned status quo bias. But in hindsight, my almost three-year struggle before I got official staffing for Gleangen came down to poor execution on exactly what Cowgill advised: understanding client needs. I had a somewhat utopian view of the value of information. I put some, but not enough, effort into figuring out the messy details on how to operationalize the wisdom of Googlers into a valuable resource for managers.

Even in cases where managers wanted explicit probabilistic forecasts, a prediction market turned out to be a tough sell. I tried to launch an initiative to use Gleangen to augment the forecasts Google made to drive purchasing decisions of computers, power, and land for its data centers. There was a seemingly strong need for accuracy in the forecasts — too high, and Google could spend hundreds of millions of dollars overprovisioning; too low, and core Google products like Adwords could run out of resources.

Most managers of the supply chain forecasting teams accepted the premise that Gleangen would improve accuracy. But accuracy wasn’t their top priority. Supply chain forecasts were produced and used by many teams totalling hundreds of people. Managers were incentivized to improve the forecasting process’ transparency, adjustability, accountability, and interoperability with other systems. They didn’t stand to benefit much even if Gleangen did improve the final accuracy of the forecasts. And my initiative exposed them to risk of disrupting many of the subtler roles the process provided, such as value judgments on which parts of Google’s business should get scarce data center resources. I didn’t offer to serve their real needs as if they were enterprise customers, as Cowgill had advised.

This challenge of operationalizing the wisdom of the Googlers came to a climax when we began forecasting perhaps the most important development in Google’s recent history: large language models.

Forecasting AI

Shortly after Gleangen launched in April 2020, I read OpenAI’s technical report on GPT-3. By early 2022, I was convinced that properly developing LLMs was the  most important challenge Google faced. This was a chance to use collective intelligence to improve the chance of success of an important outcome.

I met with the leads of the core LLM teams inside Google Research, then called LaMDA. Together we devised two types of markets: technical LLM milestones and the integration of LLMs in Google products. We secured a budget to incentivize extra participation with prizes and launched the “LLM Forecasting Contest.”

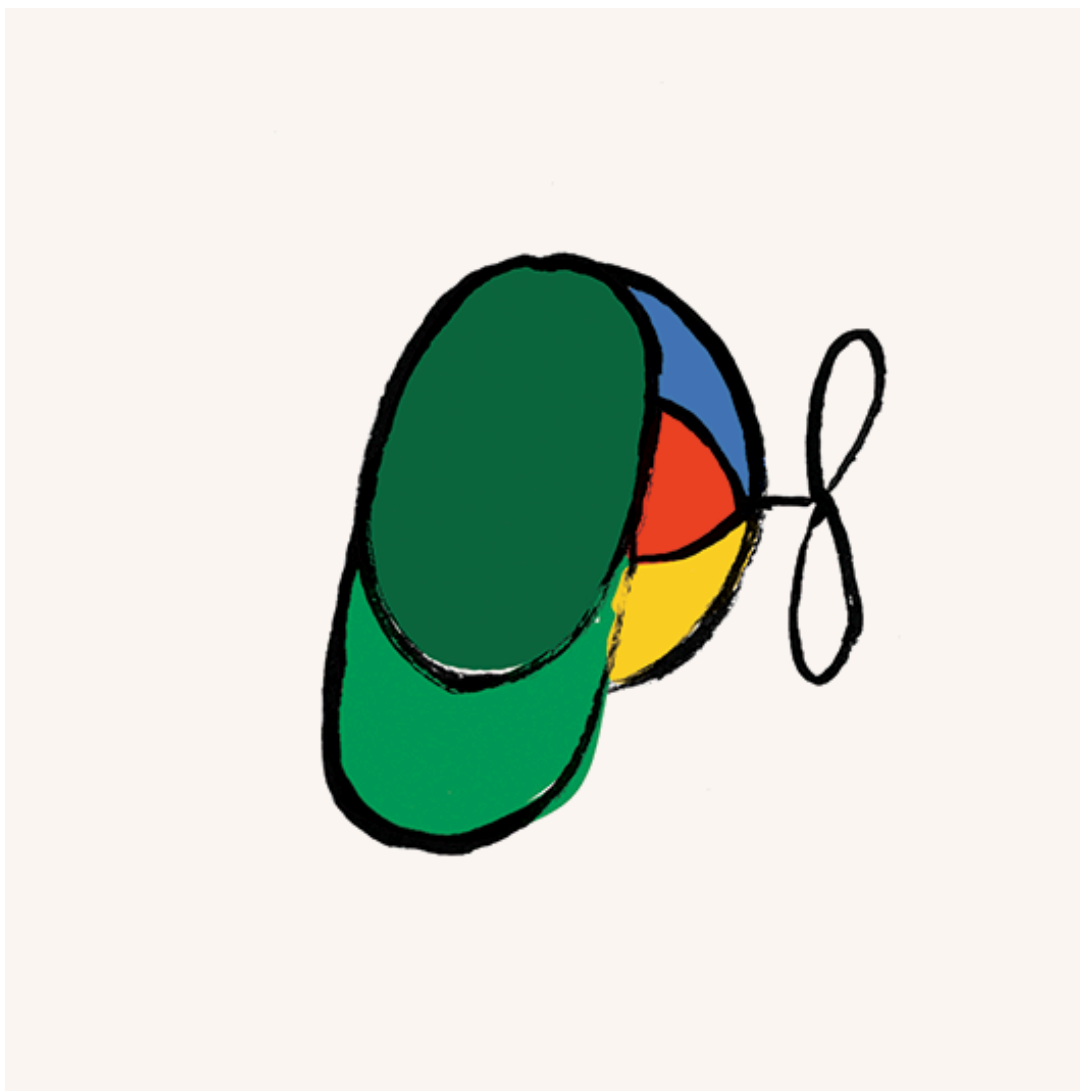
Six months into the contest, OpenAI released ChatGPT. Its success sent Google’s top executives scrambling. Most employees close to the development of LLMs, and those who used LaMDA internally, were much less surprised than management. But at a company as large as Google, information — even critical information — sometimes doesn’t percolate up to the top.

This was exactly the sort of problem I’d built Gleangen to solve. But, to my dismay, I realized we hadn’t produced the information executives really needed. We asked questions of the type “Will Google integrate LLMs into Gmail by Spring 2023?” and “How many parameters will the next LaMDA model have?” Yet what executives would have wanted to know was “Will *Microsoft* integrate LLMs into *Outlook* by Spring 2023?” and “How many parameters will the next *GPT* model have?”

This turns out to be a general lesson from running a corporate prediction market. Forecasting internal progress, and acting on that information, requires solving complex operational problems and understanding the moral mazes that managers face. Forecasting competitors’ progress has almost none of these problems.

It’s conceivable that many executives would have consulted Gleangen in 2022 for forecasts on Microsoft, OpenAI, and Anthropic. This could have enshrined the wisdom of the Googler crowd as a critical source of information, giving individual employees an ability to directly influence senior management by contributing high-quality information. It could have delivered on Friedman and Cowgill’s original vision of prediction markets as part of Google’s core mission: organizing the information inside people’s heads and making it systematically useful.


We learned from this experience. Gleangen became a staffed part of Google’s Behavioral Economics team shortly after this LLM forecasting contest started. I left Google in October of 2022 to serve as the CTO of Metaculus, but as of August 2024, the team continues to refine its approach to make Gleangen a useful source of information for Google senior management.



Looking to the future





As Nuño Sempere, Misha Yagudin, and Eli Lifland wrote in 2021, a successful corporate prediction market needs to cost less than the value of information it produces. I see the path forward in two improvements to this cost-benefit analysis. First, it must present leadership with more valuable information. All the examples of predictions in this article — executive hires, self-driving car safety metrics, COVID-19 workplace policies, datacenter costs, and LLM progress — could be useless or pivotal, depending on the process around it and the needs of the internal customer.

Second, AI could make the forecasts cheaper to produce. Sarah Pratt, a researcher at DeepMind, and members of the Gleangen team released a paper in June which compared bettors on Gleangen to predictions from PaLM 2, an LLM developed by Google. In brief, that paper — as well as several others recently released — show AI forecasts are much better than chance, but not nearly as accurate as a human

crowd, at least not yet. Their paper also highlights another way AI helps with the cost-benefit of corporate prediction markets: they increase the value of the wisdom of the human crowd by using it for evaluation, and perhaps soon the training, of AI systems. 

As the economists, engineers, and researchers running corporate prediction markets improve this value proposition, they may realize the now decades-old vision of Bo Cowgill and others. They are buoyed by the popularity of real-money public markets like Polymarket and Kalshi.¹¹ At the Manifest conference this summer, run by the prediction market Manifold, Keri Warr, a member of the technical staff at Anthropic, announced that Anthropic has launched an internal prediction market with a “focus on decision-makers.” Time will tell whether Anthropic’s employees, in concert with Claude, will produce information that helps steer the direction of the company — and if so, whether more companies will follow them.

Dan Schwarz worked at Google from 2014 to 2022. He then served as CTO of Metaculus, and is now the co-founder and CEO of [FutureSearch](#). He writes about forecasting and AI on X at [@dschwarz26](#).

Published November 2024    

Have something to say? Email us at letters@asteriskmag.com.

Previous

The Biggest Community Development Program You’ve Never Heard Of

Next

The Myth of the Loneliness Epidemic

© 2025 Asterisk

[Subscribe](#)

[Donate](#)