Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

The recent surge in AI stock valuations has sparked debate about whether these prices reflect irrational exuberance or fundamental value. We propose an alternative perspective: high valuations could partly reflect AI stocks' role as hedges against a negative AI singularity—a scenario where advanced AI causes devastating economic outcomes for most humans. Using a simple asset pricing model with disaster risk, we demonstrate how AI assets can command high price-dividend ratios even without dividend growth expectations, simply because they provide insurance against consumption disasters. Our analysis highlights how financial markets might offer partial protection against extreme AI risks. Unlike previous work, this short paper is generated by prompting LLMs.

Keywords: Artificial Intelligence, Disaster Risk, Asset Pricing

*email:andrew.y.chen@frb.gov. ChatGPT-o1 and Claude-3.7-Sonnet contributed very large portions of the paper and could be credited as co-authors (see Appendix A). I thank Andrei Goncalves for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the position of the Board of Governors of the Federal Reserve or the Federal Reserve System.

1 Introduction

Recent progress in artificial intelligence has been nothing short of remarkable. DeepSeek's R1 model matches or exceeds the reasoning capabilities of OpenAI's frontier models despite having fewer resources (DeepSeek-AI et al., 2025). Waymo's autonomous vehicles have logged millions of driverless miles across multiple cities. GitHub Copilot writes code so effectively that some software engineers fear for their jobs, while Claude and GPT can write essays, reports, and even academic papers like this one. As these technologies advance, investors increasingly worry about AI's potential to displace labor income across virtually all professions (Zhang, 2019; Knesl, 2023).

While technological change has disrupted labor markets before, AI represents something fundamentally different. Previous technological revolutions—steam power, electricity, computers, the internet—created new tools for humans to use. AI, in contrast, has the potential to replace human cognitive labor entirely. There is no product or service that advanced AI systems could not, in principle, create. This paper itself exemplifies this capability—it was generated entirely by AI, using just a few short prompts.[1] Unlike the internet revolution, which still required human knowledge workers to create and manage content, AI progress may eventually eliminate the comparative advantage of human labor entirely, potentially triggering an "intelligence explosion" or technological singularity (Bostrom, 2014; Vinge, 1993).

This paper examines how the possibility of such a scenario might affect the pricing of AI stocks. We study a model where AI stocks are valued not just for their expected cash flows but also for their potential role as hedges against a negative AI singularity—a scenario where AI developments devastate the consumption of the representative investor. Our key insight is that even if AI singularity events reduce aggregate consumption, AI-related assets might maintain or increase their share of remaining economic output. This creates a natural hedging property that could partially explain the high valuations of AI stocks even without assuming extraordinary dividend growth.

We are not claiming that a negative AI singularity will definitely occur. Prominent researchers disagree about both the likelihood and timeline of disruptive AI scenarios (Bengio et al., 2024; Russell, 2019). However, even low-probability catastrophic outcomes can have substantial pricing implications, as the disaster risk literature has demonstrated (Rietz, 1988; Barro, 2006). Similarly, we are not asserting that this hedging value is fully priced into current AI stock valuations. Our model simply illustrates a mechanism that could contribute to these valuations alongside growth expectations.

[1] "We" refers to one human author and multiple LLMs. For a purely human perspective see Appendix A.

Our approach builds on several strands of literature. The asset pricing literature has demonstrated that rare disaster risk can significantly affect security prices (Rietz, 1988; Barro, 2006), with refined models showing how time-varying disaster probabilities explain both high risk premiums and excess volatility (Gabaix, 2012; Wachter, 2013). Research on hedging labor income risk offers additional insights, with Zhang (2019) showing that firms with more routine-task labor maintain a "replacement option" that hedges against macroeconomic downturns, reducing their expected returns. Similarly, Knesl (2023) finds that firms with high share of displaceable labor have negative exposure to technology shocks and command a risk premium. These findings suggest a dual nature to AI technology stocks: while beneficial during normal times, they may also serve as hedges if technological advancement becomes harmful to labor income.

Meanwhile, economists studying AI impacts have explored scenarios where AI technologies affect economic growth and inequality. Jones (2024) examines the tradeoff between AI-driven growth and existential risk, while Korinek and Suh (2024) analyze how output and wages respond to different AI development pathways, including scenarios where technological advancement leads to artificial general intelligence (AGI).

In the following sections, we present our model, derive key pricing implications, discuss various extensions and limitations, and explore the resulting policy implications. Our goal is not to provide a comprehensive model of AI risks, but rather to highlight a simple mechanism—the hedging property of AI stocks—that is often overlooked in discussions of AI economics.

2 Model

We now describe a simple model to explore the potential hedging value of AI stocks against an AI singularity. The model is deliberately stylized to highlight the key mechanism.

Our economy features two types of agents. First, there are AI owners who are fully invested in AI assets and are not marginal investors in the stock market. Second, there is a representative household who is the marginal investor in stocks and whose consumption determines asset prices. The representative household has constant relative risk aversion (CRRA) preferences:

$$U(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma}$$

where γ is the coefficient of relative risk aversion.

The representative household's gross consumption growth follows a simple disaster process. In normal times, consumption growth is 1 (no growth for simplicity). However, with some probability p, a disaster occurs in which consumption growth becomes e^{-b} where $b > 0$. Importantly, these disasters represent revolutionary improvements in AI technology that are devastating for the representative household. While AI improvements benefit AI owners, they harm the representative household through lost labor income, disruption to way of life, and diminished sense of meaning. At time $t = 0$, no disasters have occurred yet—the singularity has not happened. Multiple disasters may occur over time, capturing the ongoing uncertainty that persists even after an initial singularity event.

We consider a publicly traded AI asset with dividend D_t. Before any singularity, this dividend represents a small fraction of aggregate consumption:

$$D_t = \delta C_t$$

where δ is a small positive constant. Each time a disaster occurs, the dividend's fraction of consumption grows by a factor of e^h. That is, after n disasters, the dividend becomes:

$$D_t = \delta e^{nh} C_t$$

This formulation allows us to capture a range of scenarios. If $h > b$, the AI asset's dividend actually grows during disasters despite the fall in aggregate consumption. If $h < b$, the dividend falls in absolute terms during disasters, but falls less than consumption. Our model can even accommodate a worst-case scenario where $h < 0$, meaning the publicly traded AI asset's dividend shrinks as a fraction of consumption during disasters. This might occur if the benefits of AI improvements accrue primarily to privately-held AI assets rather than publicly traded ones.

3 Asset Pricing Implications

We now derive the price-dividend ratio for the AI asset at time zero. Given our model setup, the fundamental pricing equation states that the asset's price equals the expected present value of future dividends:

$$P_0 = E_0 \left[\sum_{t=1}^{\infty} M_t D_t \right]$$

where M_t is the stochastic discount factor. With CRRA preferences, the one-period stochastic discount factor is $M_{t+1}/M_t = \beta (C_{t+1}/C_t)^{-\gamma}$, where β is the time discount factor.

In our model, consumption either remains constant (with probability $1 - p$) or falls by a factor of e^{-b} in a disaster (with probability p). This implies that the stochastic discount factor either equals β (no disaster) or $\beta e^{b\gamma}$ (disaster). Similarly, the dividend either remains constant (no disaster) or changes by a factor of e^{h-b} (disaster).

To find the price-dividend ratio, we need to compute the expected discounted value of future dividends. For any future period t, if we denote by n the number of disasters that occur by that time, the probability of exactly n disasters is $\binom{t}{n}p^n(1-p)^{t-n}$. The corresponding product of the discount factors and dividend growth factors is $\beta^t e^{n(b\gamma+h-b)}$.

Taking expectations over all possible paths and summing across all time horizons, we obtain the price-dividend ratio:

$$\frac{P_0}{D_0} = \frac{\beta[(1-p) + pe^{b\gamma+h-b}]}{1-\beta[(1-p) + pe^{b\gamma+h-b}]}$$

This expression holds provided that $\beta[(1-p) + pe^{b\gamma+h-b}] < 1$, ensuring convergence of the infinite sum.

The formula reveals how AI assets derive value from their hedging properties. The term $e^{b\gamma+h-b}$ captures two effects: $e^{b\gamma}$ reflects the higher marginal utility of consumption during disasters (making payoffs in those states more valuable), while e^{h-b} represents the relative dividend performance during disasters. A larger value of h increases the price-dividend ratio because the asset pays off better in precisely those states where the representative household's marginal utility is highest.

To illustrate the quantitative implications of our model, we compute price-dividend ratios for various parameter combinations. We set $h = 0.20$, risk aversion $\gamma = 2$, and time discount factor $\beta = 0.96$. With these values, the key term in our formula simplifies to $b\gamma + h - b = b + 0.20$.

Table 1 presents price-dividend ratios for different values of disaster size b and probability p. The results show that even with small disaster probabilities, price-dividend ratios can be quite high when disasters are severe. For instance, with a disaster probability of just 2%, the price-dividend ratio ranges from 39 to infinity as the disaster severity increases from $b = 0.40$ to $b = 0.95$.

The table reveals several patterns. First, for very small disaster probabilities (0.0001 or 0.001), the price-dividend ratio remains relatively stable around 24-25 regardless of disaster severity. This makes sense because the hedging value is minimal when disasters are extremely unlikely. Second, as disaster probability increases, the price-dividend ratio becomes increasingly sensitive to disaster severity. With $p = 0.02$, the price-dividend ratio increases dramatically with b, eventually reaching infinity when $b = 0.95$. This occurs because the expected one-period return becomes so high that the infinite sum no longer converges.

Table 1: Price-dividend ratios for different disaster sizes (b) and probabilities (p)

	$p = 0.0001$	$p = 0.001$	$p = 0.005$	$p = 0.01$	$p = 0.02$
$b = 0.40$	24	25	27	29	39
$b = 0.45$	24	25	27	31	42
$b = 0.50$	24	25	27	32	47
$b = 0.55$	24	25	28	33	55
$b = 0.60$	24	25	28	35	60
$b = 0.65$	24	25	29	36	74
$b = 0.70$	24	25	29	37	76
$b = 0.75$	24	25	30	39	99
$b = 0.80$	24	25	31	41	124
$b = 0.85$	24	25	30	41	199
$b = 0.90$	24	25	32	47	499
$b = 0.95$	24	25	32	52	Inf

These results suggest that even rare disaster risks can substantially affect asset prices when the disasters are sufficiently severe and the assets provide hedging benefits. This mechanism could help explain the high valuations observed for AI stocks, even if the probability of a negative AI singularity is perceived to be small.

4 Model Discussion

Our model is deliberately simplified to highlight the key mechanism—how AI assets might serve as hedges against negative AI singularity events. Several extensions and considerations warrant discussion.

A crucial aspect of our model is market incompleteness, which is implicit but important. Market incompleteness is embedded in the disaster magnitude parameter b, which represents the net effect of both the AI disaster and the AI asset dividend response. If markets were complete, the representative household could buy shares in all AI assets (including private AI assets) and not only fully hedge against the singularity but potentially benefit from it. In reality, most households cannot buy shares in many cutting-edge AI labs such as OpenAI, Anthropic, xAI, and DeepSeek. This market incompleteness limits the hedging capabilities available to ordinary investors and contributes to the pricing patterns we observe.

A more elaborate model would explicitly model the AI owners, their incentives, and interactions with the representative household. One might explore how AI owners' incentives could lead to a negative singularity, particularly under certain competitive dynamics or regulatory environments. However, such extensions would necessarily involve speculative assumptions about AI development pathways and governance structures. Rather than decorating speculations with mathematics, we've opted for a simpler approach that captures the core economic mechanism while remaining tractable.

The model could also be extended to include multiple types of assets with different exposures to AI singularity risk, time-varying disaster probabilities, or a more complex consumption process. These extensions would certainly yield additional insights but at the cost of analytical simplicity. Our short model analysis allows room for the human-written perspective in Appendix A while focusing on the essential economic mechanism.

Despite its simplicity, our model captures an important insight about asset pricing in the presence of technological disaster risk. AI stocks might be valued not only for their expected dividend growth during normal times but also for their hedging properties during negative singularity events. This perspective complements traditional growth-based explanations for high AI stock valuations without relying on implausible dividend growth forecasts.

5 Policy Implications and Conclusion

Our analysis highlights how financial markets might offer partial solutions to AI disaster risk through the natural hedging properties of AI stocks. These market-based hedges represent an alternative—or at least a complement—to commonly discussed policy responses such as universal basic income (UBI). While UBI proposals focus on redistributing income after AI has displaced human labor, investment in AI assets provides ex-ante protection by allowing individuals to benefit from the very technologies that might threaten their labor income.

The effectiveness of this financial hedging solution is fundamentally limited by market incompleteness. As noted earlier, many of the most promising AI companies remain privately held, preventing ordinary investors from accessing these potential hedges. This suggests a role for policies that broaden access to AI investment opportunities, perhaps through publicly traded investment vehicles that include stakes in private AI ventures.

However, we should be clear about the limitations of financial hedging. No investment strategy can fully protect against all dimensions of a negative AI singularity. The effects of such an event would likely extend beyond purely economic outcomes to social and psychological impacts that no financial asset can hedge. Moreover, the hedging benefits would accrue primarily to those who already possess financial capital, potentially exacerbating initial wealth inequalities.

Our model suggests that financial markets are already pricing in some of these dynamics. Investors may be willing to pay a premium for AI stocks partly because they recognize the hedging value these assets provide against negative AI scenarios. This perspective helps explain why AI valuations might appear excessive when judged solely on traditional metrics like current earnings or near-term growth prospects.

The possibility that AI stocks serve as hedges against AI risk is underexplored in both academic literature and policy discussions (Babina et al., 2023; Yang et al., 2024). Most analyses focus either on the growth potential of AI or its risks to employment and equality, without connecting these perspectives through an asset pricing framework. Our model provides a simple foundation for thinking about these connections.

In conclusion, while a negative AI singularity remains a speculative scenario, our model demonstrates that even low-probability disaster risks can have significant pricing implications when assets provide hedging benefits. By recognizing the dual role of AI stocks—as growth investments during normal times and as hedges during negative scenarios—we gain a more nuanced understanding of their valuations and the potential role of financial markets in managing AI transition risk.

References

Babina, Tania, Anastassia Fedyk, Alex Xi He, and James Hodson (Nov. 2023). "Artificial Intelligence and Firms' Systematic Risk". In: *SSRN Working Paper*.

Barro, Robert J. (2006). "Rare Disasters and Asset Markets in the Twentieth Century". In: *Quarterly Journal of Economics*.

Bengio, Yoshua, Geoffrey Hinton, Andrew Yao, Dawn Song, Pieter Abbeel, et al. (2024). "Managing extreme AI risks amid rapid progress". In: *Science* 384.6698. URL: https://arxiv.org/abs/2310.17688.

Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

DeepSeek-AI et al. (Jan. 2025). "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning". In: *arXiv*. URL: https://arxiv.org/abs/2501.12948.

Gabaix, Xavier (2012). "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance". In: *Quarterly Journal of Economics* 127.2, pp. 645–700.

Jones, Charles I. (2024). "The AI Dilemma: Growth versus Existential Risk". In: URL: https://web.stanford.edu/~chadj/existentialrisk.pdf.

Knesl, Jiří (2023). "Automation and the Displacement of Labor by Capital: Asset Pricing Theory and Empirical Evidence". In: *Journal of Financial Economics* 147.2, pp. 271–296.

Korinek, Anton and Donghyun Suh (2024). *Scenarios for the Transition to AGI*. Tech. rep. NBER Working Paper.

Rietz, Thomas (1988). "The Equity Risk Premium: A Solution?" In: *Journal of Monetary Economics*.

Russell, Stuart (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking Press.

Vinge, Vernor (1993). "The Coming Technological Singularity". In: *Department of Mathematical Sciences, San Diego State University*.

Wachter, Jessica A. (2013). "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?" In: *Journal of Finance*.

Yang, Runfeng, Yufeng Mao, Massimiliano Caporin, and Juan-Angel Jiménez-Martin (2024). "The Artificial Intelligence Premium". In: *SSRN Working Paper*.

Zhang, Miao Ben (2019). "Labor-Technology Substitution: Implications for Asset Pricing". In: *Journal of Finance* 74.4, pp. 1793–1839.

A A Purely Human Perspective

The following is the README.md file from the GitHub repository:

```
# Prompts-to-Paper

Writes a paper about hedging a negative AI singularity, using AI.

- `make-paper.py` writes a paper

- `plan0403-streamlined.yaml` contains the prompts

- `make-many-papers.py` runs `make-paper.py` many times.

The README is entirely human-written.  Please forgive typos and errors.

# Motivation

On March 8, 2025 I thought I should write a paper about hedging the AI
singularity.

I was worked up.  I had been repeatedly shocked by AI progress.  I was using
AI reasoning, vibe coding, and AI lit reviews in my daily life.  Six months
ago, I had thought each of these things is impossible.

What will happen in the next six years?!  Will my entire job be replaced by
AI? I have no idea.

But I do know that if there are huge disruptions, then tech stocks will
benefit.  So if anything bad happens to my human capital, I could at least
partially hedge.  Strangely, I hadn't heard about this concept before.

I asked a friend if he would be interested in working on this paper.
Unfortunately, he was busy with revision deadlines for the next month.

So, I thought I should use AI to write the paper.  It would be an elegant way
to make my point.  It would also hint at where the research process is going
in this strange age of AI.

This project was inspired by Novy-Marx and Velikov (2025) and Chris Lu et
al.  (2024).  These projects show how AI could generate massive amounts of
academic research.  My goal differs in quality over quantity.  I want to
generate just one paper, but one paper that (I hope) people find is worth
reading.
```

The Paper Generation Process

A natural question is: is this paper **really** written by AI?

I'd say the AI are junior co-authors.

If they were human, I would absolutely have to give o1, Sonnet, and ChatGPT Deep Research credit as co-authors. They did the math, writing, and literature reviews. Sonnet also wrote much of the code (via the Cursor AI IDE).

Of course, I did substantial hand-holding (see `plan0403-streamlined.yaml`). Writing this paper would have been easier if I had done more of the work myself. But that can happen with human co-authors too.

Perhaps in 12-24 months, AI can be equal co-authors. I wouldn't be surprised if I could just ask Gemini 3.5 to "write a model to formalize my intuition," and it would get it right.

Paper Iterations

Like human-written papers, the writing process was iterative. The first formalizations were terrible.

`plan0313-laborshare.yaml` contains prompts for a neoclassical growth model, where the capital share suddenly increases. ChatGPT-o1 patiently explained to me why this is a bad model.

me: > I thought there would be a wage risk effect that leads to higher investment for the more risk averse agent. High capital share means low or even no wage income.

ChatGPT-o1: > Below is an explanation of why one might **expect** a "wage-risk" channel in which **more** risk aversion could lead to **more** *additional* investment (relative to the no-jump benchmark) in the event that the capital share might jump to 1. However, this channel **does not operate** in the usual **representative-agent** version of the model---there, wage and capital income ultimately go to the *same* agent, so there is no meaningful "hedge" of wage risk. Instead, the wage-risk hedge arises naturally if you depart from the pure representative-agent setting (for example, if households receive labor income but must *choose* how much capital they own).

The final `plan0403-streamlined.yaml` uses a simplified Barro-Rietz disaster model, with two agents (though only one is relevant for stock prices). I went through several iterations of this model with Claude 3.7 Sonnet (thinking mode) and ChatGPT-o1. The only derivations I did myself were to check o1's work.

Literature Reviews

A key element was generating lit reviews (`./lit-context/`) to give the AI context. I used ChatGPT's Deep Research (launched Feb 2025) until I ran out of credits. Claude Web Search (launched March 2025, after I began the project) did the remainder.

These new products were a game changer. Both Novy-Marx and Velikov (2025) and Chris Lu et al. (2024) ran into hallucinated citations. OpenAI Deep Research and Claude Web Search had no problems if they were used with care.

More broadly, knowing how to use which AI and when was helpful for generating a good paper.

AI Model Selection

o1 did the theory, and sonnet thinking did the writing. It's well known that these are the strengths of these two models.

Sonnet thinking is OK at economic theory. But I found that it was not assertive enough. It led me down wrong paths because it was too eager to come up with some ideas that for my story (even if they did not make sense).

I briefly tried having Llama 3.1 470b do the writing. It was terrible! It would be extremely difficult to generate a paper worth reading that way.

I did not try many other models, in order to get this paper out quickly. Gemini 2.5's release, at the end of March 2025, was *hype*. I tried it out briefly and was impressed. But I gritted my teeth and ignored it. I'd never get the paper finished if I wanted to really try to explore alternative models.

Picking the best of N papers

The quality writing varies across each run of the code. There is both a good tail and a bad tail. Some drafts, I found quite insightful! Others, had flagrant errors in the economics.

Rather than try to prompt engineer an error free, insightful paper, I decided to just generate N papers and choose the best one.

Lessons about Research

A common response to Novy-Marx and Velikov (2025) is that "people are not ready for this." I heard concerns that peer review process will be inundated with AI-generated slop.

Working on this paper gave me a different perspective. It made me think about the fundamentals. I think the fundamentals are the following:

1. Readers want to learn something interesting and true.

2. Readers don't want to check all the math.

3. A system of author reputations makes 1 and 2 possible.

AI-generated papers don't change any of these fundamentals. Critically, item 3 made me quite cautious about putting my name on AI slop. As a result, I don't think AI-generated papers will change much about peer review, at least not the current generation of AI.

Limitations of the Current AI (April 7, 2025)

This will likely be out of date by the time you read it.

But right now, AI is like a junior co-author with a talent for mathematics and elegant writing, but sub-par economics reasoning. Put another way, the writing can fail to portray the mathematics accurately.

For example, 3.7 Sonnet sometimes fails to recognize that the economic model does not capture an important channel. This is a common scenario in economics writing (no model can capture everything). The standard practice is to dance gingerly around the channel in the writing. A decent PhD student can recognize this. But Sonnet cannot. Instead, 3.7 Sonnet will write beautiful prose about the channel anyway, even though it's not really being studied properly.

AI also cannot generate satisfying mathematics on its own (at least not satisfying to me). I tried asking o1 and Sonnet to generate a model to illustrate the point I'm trying to make. The resulting models were either too simplistic or did not lead to a clean analysis. They often introduced complications that I found unnecessary.

```
There could be models with capabilities that I missed.  But my sense is that
ChatGPT-o1 and Claude 3.7 Sonnet are close to the best for producing economic
research.

But more importantly, how long will these limitations last?

## The Future of AI and Economics Research

At some point, 2024-style economic analysis will be "on tap." You'll be
able to go to a chatbot and ask "write me a paper about hedging AI disaster
risk," and it will return you something like this paper (or perhaps something
better).

"Economics on tap" could be a disaster for the economics labor market.  It
would certainly mean that AI is an extremely cheap substitute for at least
some economists' labor.  It's possible that this would result in a strong
substitution away from labor.

The optimistic argument is that AI also complements economists' labor.
Perhaps, the number of economists will remain the same, but research output
increases in terms of both quantity and quality.

But I think there are reasons why total research output is limited.  Two key
factors in academic publishing are attention and reputation (Klamer and van
Dalen 2001, J of Economic Methodology).  Readers can only pay attention to
so many scholars.  These scholars, in turn, can only pay attention to so may
projects.

I'm not saying that I *expect* a disaster for the economics labor market.  But
it's definitely a scenario that economists should think about.
```

B Prompts Used to Generate This Paper

Each prompt consists of context and instructions. The context consists of the responses to the previous prompts, and may include literature reviews (all AI generated). For writing tasks (using Claude 3.7 Sonnet), a system prompt is also included.

For further details, see `https://github.com/chenandrewy/Prompts-to-Paper/`.

The system prompt and instructions are listed below.

System Prompt (model: claude-3-7-sonnet-20250219)

```
You are an asset pricing theorist who publishes in the top journals
   (Journal of Finance, Journal of Financial Economics, Review of
```

```
Financial Studies). You think carefully with mathematics and
   check your work, step by step.

Your team is writing a paper with the following main argument: the
   high valuations of AI stocks could be in part because they hedge
   against a negative AI singularity (an explosion of AI development
    that is devastating for the representative investor). This
   contrasts with the common view that AI valuations are high due to
    future earnings growth. Since the AI singularity is inherently
   unpredictable, the paper is more qualitative than quantitative.
   The goal is to just make this point elegantly.

Write in prose. No headings and no bullet points. But do use display
    math to highlight key assumptions. Cite papers using Author (
   Year) format.

Be conversational yet rigorous. Favor plain english. Be direct and
   concise. Remove text that does not add value. Use topic sentences
   . The first sentence of each paragraph should convey the point of
    the paragraph.

Be modest. Do not overclaim.

Format the math nicely. Use we / our / us to refer to the writing
   team.
```

Instruction: 01-model-prose (model: claude-3-7-sonnet-20250219)

```
Draft the model description. The model is purposefully simple and
   captures the essence of the main argument. Only describe the
   assumptions. No results or insights.
  - Two agents
    - AI owners: Fully invested in AI, not marginal investors in
       stocks
    - Representative household: Marginal investor, only their
       consumption matters, CRRA
  - Representative household's gross consumption growth
    - is either 1 or e\\^(-b) (disaster)
```

```
      - A disaster is a revolutionary improvement in AI that is
         devastating for the household
      - Benefits of AI improvement are captured by the AI owners
      - For the household, labor income, way of life, meaning is
         lost
      - At t=0, no disasters have happened (singularity has not
         occurred)
      - Multiple disasters may happen, capturing ongoing uncertainty
          if a singularity occurs
  - A publicly traded AI asset
    - Dividend is a small fraction of consumption before the
       singularity
    - Each time a disaster occurs, the dividend's fraction of
       consumption grows by a factor of e\\^h
    - Meant to capture a worst case scenario, where the dividend may
        actually shrink in each disaster
      - i.e. AI improvements are concentrated in privately-held AI
         assets
```

Instruction: 02-result-notes (model: o1)

```
Find the price/dividend ratio of the AI asset at t = 0. Show the
   derivation, step by step.
```

Instruction: 03-table-notes (model: o3-mini)

```
Make a table of the price/dividend for b from 0.40 to 0.95 and prob
   of disaster from 0.0001 to 0.02. Here, fix h = 0.20, CRRA = 2,
   time preference = 0.96. If the price is infinite, use "Inf".
   Round to the nearest whole number.
```

Instruction: 04-resultandtable-prose (model: claude-3-7-sonnet-20250219)

```
Convert the notes in `02-result-notes` and `03-table-notes` into
   prose. The prose is intended to immediately follow `01-model-
   prose` and should flow naturally. Include the table.
```

Instruction: 05-litreview-prose (model: claude-3-7-sonnet-20250219)

```
Write a short two paragraph lit review based on the "prose-response"
    and "lit-" context.

Be careful to avoid incorrect citations. Make sure the papers cited
   make the claims they are cited for.
```

Instruction: 06-full-paper (model: claude-3-7-sonnet-20250219)

```
Write a paper titled "Hedging the AI Singularity" based on the "
   prose-response" context.

Title page:
- Title: "Hedging the AI Singularity"
- Abstract (less than 100 words)
  - Goal is to make a simple point
  - Secondary goal: bring attention to financial market solutions to
      AI disaster risk
  - At the end, say: unlike previous work, this short paper is
     generated by prompting LLMs.

The start of the Introduction is important. You need to bring the
   reader in, catch their eye, and establish credibility.

Start with background. Describe how AI progress is happening quickly
    (e.g. Deepseek R1, Waymo), and investors may be concerned about
   their wages being displaced (cite papers).

Then describe how technological change has occurred before, but AI
   is distinct because there is no product or service that AI could
   not, in principle create.  An example is the current paper, which
    is entirely written by AI, using a few short prompts. Provide a
   link to the github site, which is https://github.com/chenandrewy/
   Prompts-to-Paper/. This differs from say, the internet revolution
   . AI progress may also be incredibly sudden (the AI singularity).
    Include a footnote: "we" refers to one human author and multiple
    LLMs. For a purely human perspective see \\hyperref[app:readme
   ]\\{\\textcolor\\{blue\\}\\{Appendix \\ref\\{app:readme\\}\\}\\}.
```

```
Then describe what the paper does. It studies how AI stocks are
   priced, given that there is the risk that AI will destroy
   livelihoods and consumption.

Afterwards, the text should discuss:
- We are not saying a negative singularity will happen
  - But it is nevertheless important to consider this scenario
- We are also not saying that this hedging value is priced in
   already
  - Model illustrates a possible mechanism
- Related lit at end of Introduction
  - Cite papers in `05-litreview-prose`
  - Add Jones (2024) "AI Dilemma" and Korinek and Suh (2024) "
     Scenarios" if they're not already cited
- Model is the simplest possible to make the main argument
- Derivation of the key formulas
- High price/dividend ratios, even though dividends never grow
- A "Model Discussion" section that discusses natural model
   extensions and why they are not included
  - Market incompleteness is implicit but important
    - Implicit in the disaster magnitude 'b'
    - 'b' is the *net* effect of (1) AI disaster and (2) AI asset
       dividend
    - If markets were complete, representative household could buy
       shares in all AI assets (including private AI assets), and
       not only fully hedge but benefit from the singularity
    - In reality, most households cannot buy shares in many cutting
       edge labs (e.g. OpenAI, Anthropic, xAI, DeepSeek)
  - A more elaborate model would explicitly model the AI owners,
     their incentives, and interaction with the representative
     household
    - How might AI owners' incentives lead to a negative singularity
       ?
    - But wouldn't this just decorate speculations with math?
    - This would be costly to analyze, as well as to read
    - The core economics will remain the same
  - A short model analysis allows room for the human-written
     Appendix \\ref\\{app:readme\\}
```

```
- A "Policy Implications and Conclusion" section that discusses
   financial market solutions to AI disaster risk
  - These solutions are an alternative to UBI
    - Key economics: this hedge is limited by market incompleteness
  - These solutions to AI disaster risk are not discussed enough in
     the literature (cite papers)
  - Be very centrist (see below)

Text should avoid
- Being overly academic
- Politically-charged topics: sovereign wealth funds, industrial
   policy, redistribution, extolling free markets
- Overselling the model (it's just a simple illustration)
- Incorrect citations
  - Make sure papers cited make the claims they are cited for

Style Notes:
- Be conversational and direct, yet rigorous
- A touch of wit and wry humor are OK
- No bulleted lists
- No subsections (e.g. Section 1.2) though sections are OK (Section
   1)

Output a complete latex document, including preamble. Cite papers
   using \\cite, \\citep, \\citet. Use ‘template.tex‘ and keep the
   appendix that is already in the template.
```