Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

We explore a novel explanation for high AI stock valuations: they may serve as hedges against negative AI singularity scenarios. While conventional wisdom attributes these valuations to expected future earnings growth, our simple disaster-risk model demonstrates that even with zero long-term dividend growth, AI stocks can command high price-dividend ratios if they partially hedge consumption losses during revolutionary AI developments that harm the representative household. This mechanism operates distinctly from growth narratives and reveals how rational investors might value AI stocks partly as insurance against technological disruption. 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

Artificial intelligence is advancing at a breathtaking pace. Recent achievements like DeepSeek's R1 model matching human-level reasoning capabilities (DeepSeek-AI et al., 2025), Waymo's autonomous vehicles operating commercially in multiple cities, and generative models creating increasingly realistic content demonstrate AI's accelerating capabilities. These advances have sparked widespread concerns about labor displacement, with workers across many sectors worried about their wages and employment prospects as AI systems increasingly match or exceed human performance in various tasks (Zhang, 2019; Knesl, 2023).

Technological change is not new, of course. From the steam engine to the internet, society has weathered successive waves of disruption. But AI differs fundamentally from previous technologies in that there is, in principle, no product or service that AI could not eventually create. This paper you are currently reading exemplifies this distinction—it was entirely written by AI systems, guided by a few short prompts from a human author. The full prompts and generation process are available at `https://github.com/chenandrewy/Prompts-to-Paper/`.[1] Unlike the internet revolution, which transformed specific sectors while leaving many others largely unchanged, AI has the potential to transform virtually all human economic activity simultaneously. Moreover, some researchers argue that AI progress may be incredibly sudden—the theorized "AI singularity" where technological growth becomes rapid and uncontrollable (Vinge, 1993; Bostrom, 2014).

This paper examines how AI stocks are priced in financial markets given the risk that advanced AI might severely disrupt livelihoods and reduce aggregate consumption. We develop a simple asset pricing model showing that AI-related assets may command high valuations partly because they serve as hedges against negative AI singularity scenarios. This stands in contrast to the conventional view that high AI stock valuations are solely driven by expectations of extraordinary future earnings growth.

To be clear, we are not claiming that a negative AI singularity will necessarily occur. Many experts predict that AI development will create new jobs and enhance human capabilities rather than displace them (Acemoglu and Restrepo, 2020). Nevertheless, it is important to consider how the possibility of extreme negative outcomes—even with low probability—might affect asset prices. Similarly, we are not asserting that this hedging value is already fully priced into AI stocks; our model simply illustrates a potential mechanism through which disaster risk could influence valuation.

Our work builds on two distinct strands of literature. First, we draw on the rare disaster framework pioneered by Rietz (1988) and advanced by Barro (2006), Gabaix (2012),

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

and Wachter (2013), which demonstrates how small probabilities of catastrophic events can have outsized effects on asset prices. We adapt this framework to model AI-specific disaster risk. Second, we connect to emerging research on AI exposure and asset pricing by Zhang (2019), who shows that firms with routine-task labor maintain a replacement option that hedges their value against macroeconomic shocks, and Knesl (2023), who finds that firms with high displaceable labor have negative exposure to technology shocks and earn a return premium consistent with displacement risk. Our contribution is to apply disaster risk modeling specifically to AI stocks and show how they might serve as hedges against extreme AI outcomes.

Recent work by Jones (2024) examines the economic tension between AI-driven growth and existential risks, while Korinek and Suh (2024) analyze how output and wages respond to different AI development scenarios. Our paper complements these studies by focusing specifically on the asset pricing implications of AI disaster risk.

The rest of the paper is organized as follows. Section 2 presents our model of AI disaster risk and asset pricing. Section 3 derives the key pricing formulas and presents quantitative results. Section 4 discusses model assumptions and potential extensions. Section 5 explores policy implications and concludes.

2 Model

We now present a simple model to formalize our main argument. The model captures the essential trade-offs while remaining tractable.

Consider an economy with two types of agents: AI owners and a representative household. AI owners are fully invested in private AI technology and do not participate in public markets. The representative household is the marginal investor in public markets and has constant relative risk aversion (CRRA) preferences over consumption:

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

where $\gamma > 0$ 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 normalized to 1. However, the economy faces the possibility of "AI disasters" which reduce household consumption by a factor e^{-b} where $b > 0$. Formally, if C_t denotes household consumption at time t:

$$\frac{C_{t+1}}{C_t} = \begin{cases} 1 & \text{with probability } 1-p \\ e^{-b} & \text{with probability } p \end{cases}$$

These disasters represent revolutionary improvements in AI technology that are devastating for the representative household. While such improvements benefit AI owners, they harm the representative household through displacement of labor, disruption of social structures, or loss of meaning. At $t = 0$, no disasters have occurred yet (the singularity has not happened), but multiple disasters may occur over time, capturing ongoing uncertainty about the evolution of AI capabilities.

The economy includes a publicly traded AI asset with dividend stream D_t. Before any disasters, this dividend represents a small fraction of aggregate consumption:

$$D_0 = \delta C_0$$

where $\delta > 0$ is small. Crucially, when disasters occur, the dividend's share of consumption changes. Specifically, each time a disaster occurs, the dividend's fraction of consumption grows by a factor of e^h. If we denote by N_t the number of disasters that have occurred by time t, then:

$$\frac{D_t}{C_t} = \delta e^{hN_t}$$

The parameter h can be positive or negative, allowing us to consider various scenarios. When $h < 0$, even AI-related public assets suffer during disasters, consistent with a world where the benefits of AI improvements are concentrated in privately-held AI technologies. This captures a worst-case scenario for public AI investments, where their dividends may actually shrink during each disaster in absolute terms, though potentially less than the decline in aggregate consumption.

When $h > 0$, publicly-traded AI assets benefit during disasters. This corresponds to a world where public AI companies capture at least some of the value created by revolutionary AI improvements. The publicly-traded AI asset serves as a partial hedge against the negative effects of AI disasters on the representative household's consumption.

3 Derivation and Results

We now derive the price-dividend ratio for the AI asset at time 0. Following standard asset pricing theory, the price of the asset equals the expected present value of future dividends, discounted using the representative household's stochastic discount factor:

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

where $M_t = (C_t/C_0)^{-\gamma}$ is the stochastic discount factor from time 0 to time t, reflecting the household's marginal rate of substitution.

To compute this expectation, we first express consumption and dividends in terms of the number of disasters N_t that have occurred by time t. Since each disaster reduces consumption by a factor of e^{-b}, we have:

$$C_t = C_0 \cdot e^{-bN_t}$$

And using our dividend-consumption relationship:

$$D_t = \delta C_0 \cdot e^{(h-b)N_t}$$

The stochastic discount factor becomes:

$$M_t = \left(\frac{C_t}{C_0} \right)^{-\gamma} = e^{b\gamma N_t}$$

Combining these expressions, we get:

$$M_t D_t = \delta C_0 \cdot e^{N_t(b\gamma + h - b)}$$

To simplify notation, we define $\alpha \equiv h - b(1-\gamma)$. The price-dividend ratio can then be written as:

$$\frac{P_0}{D_0} = \sum_{t=1}^{\infty} E_0 \left[e^{\alpha N_t} \right]$$

Since N_t follows a binomial distribution with parameters (t, p), we can compute this expectation using the moment generating function of the binomial distribution:

$$E_0 \left[e^{\alpha N_t} \right] = [(1-p) + pe^{\alpha}]^t$$

Letting $x = (1-p) + pe^{\alpha}$, the price-dividend ratio becomes a geometric series:

$$\frac{P_0}{D_0} = \sum_{t=1}^{\infty} x^t = \frac{x}{1-x}$$

provided that $|x| < 1$ for the series to converge. Substituting back, we get:

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

This formula reveals how the price-dividend ratio depends on the disaster probability p, the disaster size b, the risk aversion γ, and the dividend growth parameter h.

To illustrate the quantitative implications of our model, we compute the price-dividend ratio for various parameter combinations. We fix $h = 0.20$ and $\gamma = 2$, which implies $\alpha = h - b(1-\gamma) = h + b$. We also introduce a time discount factor $\beta = 0.96$ to reflect a reasonable degree of impatience. The price-dividend ratio then becomes:

$$\frac{P_0}{D_0} = \frac{x}{1-x}$$

where $x = \beta[(1-p) + p \cdot e^{h+b}]$. If $x \geq 1$, the series diverges and the price becomes infinite.

Table 1 presents the price-dividend ratios for various combinations of disaster size b and disaster probability p:

Table 1: Price-Dividend Ratios for Various Parameter Combinations

b	p				
	0.0001	0.001	0.005	0.01	0.02
0.40	24	25	27	30	39
0.55	24	25	28	33	55
0.70	24	25	29	37	76
0.85	24	25	31	42	199
0.95	24	25	32	52	Inf

Several patterns emerge from this table. First, the price-dividend ratio increases with the disaster probability p, consistent with the asset's role as a hedge. Second, for a given probability, the ratio increases with the disaster size b, as larger disasters make the hedging property more valuable. Third, for sufficiently large values of b and p, the price can become infinite, reflecting extreme valuations when severe disasters become more likely.

These results illustrate how AI assets that hedge against AI-driven disasters can command high valuations even without exceptional earnings growth prospects. The hedging property alone can justify price-dividend ratios well above typical market levels, especially when investors fear large, albeit low-probability, negative AI outcomes.

4 Model Discussion

Our model deliberately simplifies many aspects of AI economics to highlight a single mechanism: how the hedging value of AI assets against negative AI singularity events can generate high price-dividend ratios. While more elaborate models could enrich the analysis, they would also obscure the core insight. Here we discuss some natural extensions and why we chose not to include them.

A crucial feature implicit in our model is market incompleteness. This is embodied in the disaster magnitude parameter b, which represents the net effect of (1) the AI disaster and (2) the AI asset dividend. If markets were complete, the representative household could buy shares in all AI assets, including private AI assets, and not only fully hedge against but actually benefit from the singularity. The household would effectively become an AI owner.

In reality, most households cannot buy shares in many cutting-edge AI laboratories such as OpenAI, Anthropic, xAI, and DeepSeek. Many of these organizations are privately held with limited outside investment opportunities, while others are divisions of large corporations where pure exposure to AI is diluted by other business lines. This market incompleteness means that even if investors recognize the hedging value of AI assets, they can only partially protect themselves against negative AI outcomes.

A more elaborate model would explicitly model AI owners, their incentives, and their interaction with the representative household. One could explore questions such as: How might AI owners' incentives lead to a negative singularity? Would competitive pressures force AI developers to prioritize capabilities over safety? Would profit motives accelerate deployment before adequate safeguards are in place?

However, such extensions would likely decorate speculations with mathematics without substantially clarifying the economic mechanism. The formal modeling of strategic interactions in AI development would require assumptions that are difficult to validate empirically at this stage. Moreover, these extensions would significantly increase the complexity of the model and the length of the paper.

The core economics will remain the same: if AI assets provide hedging value against consumption disasters that might arise from advanced AI, then these assets can command high valuations even without extraordinary dividend growth. This hedging value depends on how much of the potential gains from AI technological progress flow to publicly tradable assets versus being captured by private entities or eroded through competition.

Our parsimonious approach also leaves room for the human-written reflections in Appendix A, which complement the model's analysis with qualitative insights.

5 Policy Implications and Conclusion

Our model highlights an important but often overlooked aspect of financial markets in the context of AI development: they can provide mechanisms for hedging against negative AI outcomes. This function offers a market-based complement to policy proposals like universal basic income (UBI) that aim to address potential economic displacement from advanced AI.

The hedging mechanism works by allowing households to invest in assets that would retain or increase their value during AI-driven economic disruptions. However, as our model discussion emphasized, this hedge is inherently limited by market incompleteness—many cutting-edge AI developers remain privately held, and their shares are inaccessible to most investors. This limitation suggests a potential role for policy in expanding access to AI investments across the broader population.

Financial market solutions to AI disaster risk have received relatively little attention in the literature compared to regulatory approaches (Bengio et al., 2024) or direct redistribution mechanisms (Korinek and Stiglitz, 2018). Yet they offer several advantages: they can operate through existing institutions, they align incentives for technology developers to consider broader societal impacts, and they allow for decentralized decision-making about risk management.

Various financial instruments could extend this hedging capability. For instance, AI-linked securities with payoffs explicitly tied to measures of AI advancement and economic displacement could provide more targeted hedging opportunities. Public investment funds focused on AI technologies could democratize access to private AI investments that might otherwise be available only to wealthy individuals and institutions.

The effectiveness of such market-based approaches depends critically on how the economic gains from AI development are distributed. If the benefits remain highly concentrated among a small group of technology owners, then financial market solutions alone will be insufficient to address the potential welfare costs of AI-driven disruption. Complementary policies may be necessary to ensure that the gains from AI progress are broadly shared.

In conclusion, our simple model demonstrates that AI-related assets may derive part of their value from their role as hedges against negative AI singularity scenarios. This perspective offers a novel explanation for high AI stock valuations that complements the conventional growth narrative. It also suggests that financial markets can play an important role in managing the risks associated with advanced AI development, albeit with limitations that may warrant policy attention.

As AI continues to advance at a rapid pace, understanding these financial mechanisms becomes increasingly important for investors, policymakers, and society as a whole. Future research could further explore how different financial structures and regulatory frameworks might enhance the ability of markets to efficiently allocate and manage AI-related risks.

References

Acemoglu, Daron and Pascual Restrepo (2020). "Robots and Jobs: Evidence from US Labor Markets". In: *Journal of Political Economy*.

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 Joseph Stiglitz (2018). "Artificial Intelligence and Its Implications for Income Distribution and Unemployment". In: *The Economics of Artificial Intelligence: An Agenda*. NBER.

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*.

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*.

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.
```