Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

We propose that high valuations of artificial intelligence (AI) stocks could be partially explained by their role as a hedge against a negative AI singularity—a scenario where explosive AI development devastates the average investor. This contrasts with the common view that AI valuations merely reflect future earnings growth. Using a simple disaster-risk model, we show how AI assets that increase their share of consumption during singularity events can command high price-dividend ratios despite modest expected dividend growth. Our analysis suggests financial markets may offer novel approaches to managing catastrophic AI risk. 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 development is progressing at a remarkable pace. Recent advances like DeepSeek's R1 model achieving results comparable to OpenAI's leading systems despite development constraints (DeepSeek-AI et al., 2025), and Waymo's autonomous vehicles logging millions of miles without human intervention, suggest we are in an era of unprecedented technological acceleration. As these technologies advance, many investors reasonably worry about their wages and livelihoods being displaced (Zhang, 2019; Knesl, 2023).

While technological change has occurred before, AI represents something fundamentally different. Unlike previous technological revolutions such as the internet or industrial machinery, there is virtually no product or service that AI could not, in principle, create. This paper itself exemplifies this point—it is entirely written by AI, using just six prompts[1]. The full prompting process is available at `https://github.com/chenandrewy/Prompts-to-Paper/`. This differs markedly from previous technological shifts like the internet revolution, which primarily transformed information access and communication without directly replacing human cognitive labor across nearly all domains. Moreover, AI progress may prove to be incredibly sudden—the phenomenon often referred to as the "AI singularity."

This paper studies how AI stocks are priced given the risk that transformative AI could potentially destroy livelihoods and drastically reduce consumption for the average investor. We propose and analyze a simple asset pricing model where publicly traded AI assets serve as a hedge against a negative AI singularity. In our framework, AI investments might command high valuations not only because of their growth prospects, but also because they provide insurance against a scenario where AI development dramatically reduces labor income and consumption for most people.

We emphasize that we are not claiming a negative singularity will inevitably occur. Predictions about AI's ultimate impact vary widely among experts, and many envision overwhelmingly positive outcomes. Nevertheless, it is prudent to consider the full range of possible scenarios, including negative ones, when analyzing asset valuations. Similarly, we are not asserting that the hedging value we identify is currently priced into AI stocks. Rather, our model illustrates a theoretical mechanism that could contribute to valuations alongside traditional factors like expected earnings growth.

Our work connects to the literature on disaster risk in asset pricing. Barro (2006) and Gabaix (2012) showed that even small probabilities of rare economic disasters can explain high equity premia and other asset pricing puzzles. We extend this framework to a novel setting: the potential for catastrophic AI development. In our model, the disaster probability

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

represents the likelihood of a technological singularity with devastating effects on household consumption, similar to the technological risks discussed by Bostrom (2014) and Bengio et al. (2024). Wachter (2013) demonstrated that time-varying disaster risk can simultaneously explain high equity premia and stock market volatility, which provides a foundation for our exploration of how AI singularity risk might be priced into assets. By showing that even small probabilities of AI disasters can lead to substantial valuation effects, our results echo Barro and Liao's (2021) finding that disaster risk is quantitatively important for asset pricing.

The impact of technological change on labor and asset prices also informs our analysis. Zhang (2019) documents that firms with more routine-task labor maintain a "replacement option" that hedges their value against unfavorable macroeconomic shocks, resulting in lower expected returns. Similarly, Knesl (2023) shows that firms with high shares of displaceable labor have negative exposure to technology shocks and command a risk premium. Our model complements these findings by focusing on AI stocks as potential hedges against singularity risk, rather than examining the risk exposure of firms vulnerable to automation. Moreover, our approach relates to work on hedging labor income risk, such as Benzoni, Collin-Dufresne, and Goldstein (2007), who demonstrate that when labor income and stock markets are cointegrated, human capital becomes "stock-like" for young investors but "bond-like" for older ones. Our work also connects to recent economic analyses of AI's potential impacts by Jones (2024) and Korinek and Suh (2024), who examine the tension between AI-driven growth and existential risks.

2 Model

We now describe a simple model that captures the essence of our argument. Consider an economy with two types of agents: AI owners and a representative household. The AI owners are fully invested in private AI assets and are not marginal investors in public markets. The representative household is the marginal investor in public markets, and only their consumption matters for asset pricing. The household has constant relative risk aversion (CRRA) preferences, with utility function:

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

where $\gamma > 0$ is the coefficient of relative risk aversion.

The representative household's gross consumption growth follows a simple process. In normal times, consumption growth is 1 (no growth for simplicity). However, the economy faces the possibility of ”disasters” - revolutionary improvements in AI that are devastating for the household. When a disaster occurs, the household’s consumption drops by a factor of e^{-b} where $b > 0$. Formally, the gross consumption growth is:

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

These disasters represent a singularity scenario where the benefits of AI improvement are captured primarily by the AI owners, while the household experiences losses in labor income, way of life, and meaning. At $t = 0$, no disasters have yet occurred (the singularity has not happened), but multiple disasters may happen over time, capturing ongoing uncertainty if a singularity occurs.

We consider a publicly traded AI asset with dividend D_t. Before any disasters, the dividend represents a small fraction of consumption, $D_t = \delta C_t$, where $\delta > 0$ is small. Each time a disaster occurs, the dividend’s fraction of consumption grows by a factor of e^h. After n disasters, the dividend is:

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

This specification captures a scenario where the dividend’s absolute value may actually shrink during each disaster (if $h < b$), reflecting that AI improvements might be concentrated in privately-held AI assets rather than publicly traded ones. The key is that the dividend’s share of consumption increases during disasters, potentially making the asset a hedge against the singularity.

3 Asset Pricing Solution

Given the model outlined above, we now derive the equilibrium price-dividend ratio for the publicly traded AI asset. Under standard consumption-based asset pricing with CRRA preferences, the stochastic discount factor (SDF) from period t to $t + 1$ is:

$$M_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma}$$

where $\beta \in (0, 1)$ is the time discount factor. The price of the asset at time 0 is the expected discounted sum of all future dividends:

$$P_0 = E_0 \left[\sum_{t=1}^{\infty} M_{0,t} D_t \right]$$

To compute the price-dividend ratio, we divide by the current dividend D_0:

$$\frac{P_0}{D_0} = E_0\left[\sum_{t=1}^{\infty} M_{0,t}\frac{D_t}{D_0}\right]$$

After N_t disasters up to time t, consumption is $C_t = C_0 e^{-bN_t}$, and the dividend is $D_t = \delta C_0 e^{(h-b)N_t}$. Thus, the multi-period SDF is $M_{0,t} = \beta^t e^{b\gamma N_t}$, and the dividend ratio is $D_t/D_0 = e^{(h-b)N_t}$. Substituting these expressions:

$$\frac{P_0}{D_0} = E_0\left[\sum_{t=1}^{\infty} \beta^t e^{N_t(b\gamma+h-b)}\right]$$

Let $\alpha = e^{b\gamma+h-b}$. Since disasters follow a binomial distribution with parameter p, we can compute $E_0[\alpha^{N_t}] = (1-p+p\alpha)^t$. This gives us:

$$\frac{P_0}{D_0} = \sum_{t=1}^{\infty} \beta^t(1-p+p\alpha)^t = \sum_{t=1}^{\infty}[\beta(1-p+p\alpha)]^t$$

This is a geometric series that converges when $\beta(1-p+p\alpha) < 1$. Summing the series yields the closed-form solution:

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

This formula reveals how the price-dividend ratio depends on the disaster probability p, the consumption impact of disasters b, the dividend's relative performance during disasters h, and risk aversion γ.

4 Numerical Results

To illustrate the quantitative implications of our model, we compute the price-dividend ratio for various parameter combinations. We set $h = 0.20$, $\gamma = 2$, and $\beta = 0.96$, which are plausible values. With these choices, the exponential term simplifies to $e^{b\gamma+h-b} = e^{b+0.20}$.

The table below shows the price-dividend ratios for different combinations of disaster probability p (columns) and consumption impact b (rows):

Several patterns emerge from this table. First, for very low disaster probabilities ($p = 0.0001$), the price-dividend ratio remains around 24 regardless of the consumption impact b. This is close to what we would expect in a model without disasters. Second, as the disaster probability increases, the price-dividend ratio rises, especially for larger values of b. For instance, when $b = 0.95$ and $p = 0.02$, the ratio becomes infinite because the expected return approaches the discount rate.

	$p = 0.0001$	$p = 0.005$	$p = 0.01$	$p = 0.015$	$p = 0.02$
$b = 0.40$	24	27	30	34	39
$b = 0.55$	24	28	33	39	55
$b = 0.70$	24	29	37	55	82
$b = 0.85$	24	31	44	76	199
$b = 0.95$	24	32	55	110	Inf

Table 1: Price-dividend ratios for different disaster probabilities (p) and consumption impacts (b), with $h = 0.20$, $\gamma = 2$, and $\beta = 0.96$.

This analysis demonstrates that even small probabilities of AI-related disasters can substantially increase the valuation of AI stocks that serve as hedges against such disasters. The effect is particularly pronounced when disasters are expected to have a large negative impact on consumption. Our model thus provides a rational explanation for high AI stock valuations that complements the standard growth narrative. Rather than simply reflecting optimism about future earnings, these valuations may partly incorporate a premium for hedging against catastrophic AI outcomes.

5 Model Discussion

While our model provides a clear mechanism for understanding how AI stocks might serve as hedges against a negative singularity, several extensions and limitations are worth discussing.

Market incompleteness is implicit but important in our framework. The consumption impact parameter b represents the net effect of both the AI disaster and the AI asset's dividend. If markets were complete, the representative household could buy shares in all AI assets, including private AI assets, and not only fully hedge but potentially benefit from the singularity. In reality, most households cannot buy shares in many cutting-edge labs such as OpenAI, Anthropic, xAI, and DeepSeek. This market incompleteness limits the hedging potential of publicly traded AI stocks.

A more elaborate model would explicitly model the AI owners, their incentives, and their interaction with the representative household. Such a model might address questions like: How might AI owners' incentives lead to a negative singularity? Would profit-maximizing behavior by AI firms necessarily lead to adverse outcomes for the average person? However, attempting to model these complex dynamics would essentially decorate speculations with mathematics. This would be costly to analyze and to read, without necessarily adding substantial economic insights. The core economics would remain the same: publicly traded AI assets may serve as partial hedges against singularity risk.

We have also made several simplifying assumptions. For instance, we assume that consumption growth is zero in normal times and that disasters are i.i.d. events. We could extend the model to include normal economic growth and time-varying disaster risk, as in Wachter (2013). Additionally, we could consider a more general specification of how dividends respond to disasters. However, these extensions would complicate the analysis without changing the fundamental insight that AI stocks can serve as hedges against negative AI outcomes.

The simplicity of our model analysis allows room for the human-written Appendix A, which provides important context and perspective from the human author. This combination of AI-generated analysis and human reflection demonstrates a potentially productive collaboration between human and artificial intelligence in economic research.

6 Conclusion and Implications

This paper has explored a novel perspective on AI stock valuations: the possibility that these assets serve as hedges against a negative AI singularity. We developed a simple disaster risk model where publicly traded AI assets increase their share of consumption during singularity events, even if their absolute dividends decline. This hedging property can generate high price-dividend ratios, even without assuming explosive dividend growth, especially when investors perceive even a small probability of a severe AI-driven consumption disaster.

The financial market solution to AI catastrophe risk that emerges from our analysis offers an interesting alternative to commonly discussed policy responses like universal basic income (UBI). By investing in AI stocks, individuals can potentially hedge some of their exposure to AI-driven labor market disruption. However, this hedging strategy is severely limited by market incompleteness—most households cannot invest in the most advanced AI labs that are privately held.

This market-based approach to managing AI risk has received insufficient attention in the literature on AI economics. While much work focuses on policy interventions and regulations (Bengio et al., 2024), or the distributional consequences of AI (Korinek and Stiglitz, 2018), the role of financial markets in helping individuals manage AI transition risk deserves greater consideration. Financial markets have historically helped society manage various risks, from natural disasters to business cycles, and they may play a similar role as we navigate the uncertain path of AI development.

Financial instruments specifically designed to hedge against AI disruption could eventually emerge, similar to how catastrophe bonds help manage natural disaster risk. However, designing such instruments requires careful consideration of moral hazard problems and the limited ability to diversify truly systemic risks.

The possibility that AI stocks might serve as hedges against negative AI scenarios does not imply that current valuations are rational or that individual investors should rush to invest in AI companies. Our model simply illustrates a theoretical mechanism that could contribute to valuations. For individual portfolio decisions, investors should consider their own specific exposure to AI risk, alongside traditional factors like diversification, time horizon, and risk tolerance.

As AI continues to develop, understanding its implications for asset pricing and risk management will become increasingly important. Our simple model provides a starting point for thinking about how financial markets might adapt to the unique risks posed by transformative AI technologies.

References

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

Barro, Robert J. and Gordon Y. Liao (2021). “Rare Disaster Probability and Options Pricing”. In: *Journal of Financial Economics* 139.3, pp. 750–769.

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`.

Benzoni, Luca, Pierre Collin-Dufresne, and Robert S. Goldstein (2007). “Portfolio Choice over the Life-Cycle when the Stock and Labor Markets Are Cointegrated”. In: *Journal of Finance* 62.5, pp. 2123–2167.

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.

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.

Some papers had problematic cites (`run01`). Others provided low-quality model discussions (`run02`) or poor explanations of the algebra (`run03`)

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 six 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 "Conclusion and Implications" section
  - Review the main argument
  - End paper by discussing financial market solutions to AI
     catastrophe 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)
  - Don't say "In conclusion." Just conclude

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)
- Taking the model too seriously
- 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.
```