Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

We offer a new perspective on elevated AI company valuations. Rather than viewing high AI stock prices solely as bets on future earnings growth, we present a model where AI stocks serve as hedges against severe AI singularity events that could devastate the representative investor's consumption. Our simple disaster risk asset pricing model demonstrates how rational investors might willingly pay premiums for publicly-traded AI assets to offset consumption risk from radical AI advances, even if these advances primarily benefit private AI owners. 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 accelerating at a breathtaking pace. DeepSeek's R1 recently matched the reasoning capabilities of OpenAI's o1, despite facing export restrictions on advanced GPUs (DeepSeek-AI et al., 2025). Waymo's driverless taxis now operate commercially in multiple cities, while generative AI tools create everything from novel images to functional code. These rapid advances have led many investors to worry about their wages being displaced (Acemoglu and Restrepo, 2020; Gomes, Jansson, and Karabulut, 2024), as AI capabilities expand into previously protected domains.

While technological change has disrupted economies before, AI differs fundamentally from previous revolutions. There is no product or service that AI could not, in principle, create. An illuminating example is the current paper, which is entirely written by AI, using six prompts. This differs markedly from the internet revolution, which primarily transformed information distribution rather than creation itself. You can examine the prompts that generated this paper at `https://github.com/chenandrewy/Prompts-to-Paper/`.[1] Moreover, AI progress may not be gradual—a technological singularity could occur when AI systems capable of redesigning themselves trigger runaway intelligence improvement (Vinge, 1993; Bostrom, 2014).

This paper examines how AI stocks might be priced in light of these risks. We develop a simple asset pricing model where an AI singularity represents a potential consumption disaster for the representative household. Our core argument is provocative yet mathematically straightforward: high AI stock valuations might partially reflect their role as hedges against negative AI singularity scenarios. In our model, while AI singularity events reduce aggregate consumption, publicly-traded AI assets capture a larger share of this reduced consumption. This makes them valuable hedges, even when their absolute dividends may decline during a singularity event.

We are not claiming that a negative singularity will definitely occur. There are opposing views about whether advanced AI will be beneficial or harmful (Bengio et al., 2024). However, even if the probability of catastrophic outcomes is small, the pricing implications could be significant. Similarly, we are not asserting that this hedging value is already fully priced into current AI stock valuations. Our model simply illustrates a rational mechanism that could contribute to high valuations beyond the standard growth narrative.

Our work connects two established literature streams. The first is disaster risk in asset pricing, pioneered by Rietz (1988) and Barro (2006), with important extensions by Gabaix (2012) and Wachter (2013). The second explores technological risk hedging, including Zhang

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

(2019), who shows firms with routine-task labor maintain options to replace workers during downturns, lowering their expected returns, and Knesl (2023), who finds firms with high shares of displaceable labor have negative exposure to technology shocks.

We connect these literatures through the lens of AI singularity risk. Bostrom (2014) explores the existential risk of advanced artificial intelligence, while Bengio et al. (2024) highlight rapidly advancing AI systems' potential extreme societal risks. On the investment side, Babina et al. (2023) provide evidence that firms' AI investments affect their systematic risk profiles. Jones (2024) examines the tension between AI-driven growth and existential risks, while Korinek and Suh (2024) analyze wage response scenarios during the transition to artificial general intelligence. Our model provides a novel theoretical framework explaining why investors might rationally value AI assets highly despite—or perhaps because of—their potential role in catastrophic scenarios.

2 Model

We now describe a simple asset pricing model to formalize our main argument. The model captures the essential tension between AI valuations and the potential for an AI singularity that could be devastating for the representative investor.

Consider an economy with two types of agents: AI owners and representative households. AI owners are fully invested in AI assets and are not marginal investors in the stock market. Representative households, on the other hand, are the marginal investors whose consumption-based pricing kernel determines asset prices.

The representative household has constant relative risk aversion (CRRA) preferences:

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

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

The gross consumption growth for the representative household follows a disaster process. In normal times, consumption growth equals 1 (no growth for simplicity), but with some probability, consumption experiences a negative shock:

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

where $b > 0$ represents the magnitude of the disaster and p is the probability of a disaster occurring in a given period.

Importantly, in our context, a "disaster" represents a revolutionary improvement in AI that is devastating for the representative household. While such technological advances benefit AI owners, they harm the representative household through loss of labor income, disruption to their way of life, and potential loss of meaning. At time $t = 0$, we assume no disasters have yet occurred (the singularity has not happened), but multiple disasters may occur in the future, capturing ongoing uncertainty about the evolution of AI capabilities.

We now introduce a publicly traded AI asset. Before any singularity occurs, this asset pays a dividend that represents a small fraction of aggregate consumption. Let D_t denote the dividend at time t, and initially:

$$D_0 = \delta C_0$$

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, where h is a parameter that could be positive or negative:

$$\frac{D_{t+1}/C_{t+1}}{D_t/C_t} = \begin{cases} 1 & \text{if no disaster occurs at } t+1 \\ e^h & \text{if a disaster occurs at } t+1 \end{cases}$$

This formulation allows us to capture a worst-case scenario where $h < b$, meaning the dividend may actually shrink in absolute terms during each disaster, even as its share of consumption grows. This reflects the possibility that AI improvements might be concentrated in privately-held AI assets, with publicly traded AI companies capturing only a portion of the value created by revolutionary AI advances.

3 Results

We now derive the price-dividend ratio for this asset. Under the CRRA preferences, the stochastic discount factor (SDF) from period t to $t + 1$ is:

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

where $\beta \in (0, 1)$ is the subjective discount factor. Given our consumption process, the SDF takes two possible values:

$$M_{t+1} = \begin{cases} \beta & \text{with probability } 1-p \\ \beta e^{\gamma b} & \text{with probability } p \end{cases}$$

For the AI asset, dividend growth also follows a two-state process. When no disaster occurs, dividends remain constant. When a disaster occurs, dividends change by a factor of e^{h-b}, reflecting both the increased share of consumption (e^h) and the decline in aggregate consumption (e^{-b}):

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

To solve for the price-dividend ratio, we use the standard asset pricing equation:

$$P_0 = E_0[M_1(P_1 + D_1)]$$

Dividing both sides by D_0 and denoting the price-dividend ratio as $x_0 = P_0/D_0$, we get:

$$x_0 = E_0\left[M_1\left(\frac{P_1}{D_1}\frac{D_1}{D_0} + \frac{D_1}{D_0}\right)\right]$$

In our i.i.d. environment, the price-dividend ratio is constant over time, so $P_1/D_1 = x_0$. Letting $G_D = D_1/D_0$ represent gross dividend growth, we have:

$$x_0 = E_0[M_1 G_D(x_0 + 1)]$$

Given the two possible states, we can write:

$$x_0 = (1-p)[\beta \cdot 1 \cdot (x_0+1)] + p[\beta e^{\gamma b} \cdot e^{h-b} \cdot (x_0+1)]$$

Factoring out $(x_0 + 1)$:

$$x_0 = [(1-p)\beta + p\beta e^{\gamma b} e^{h-b}](x_0+1)$$

Let $A = (1-p)\beta + p\beta e^{\gamma b+h-b}$. Then:

$$x_0 = A(x_0+1)$$

Solving for x_0:

$$x_0 = \frac{A}{1-A}$$

Therefore, the price-dividend ratio is:

$$x_0 = \frac{\beta[(1-p) + pe^{\gamma b+h-b}]}{1-\beta[(1-p)+pe^{\gamma b+h-b}]}$$

This closed-form solution reveals how the price-dividend ratio depends on key parameters:

the discount factor β, the probability of an AI singularity p, the consumption decline b, the dividend's relative growth h, and risk aversion γ.

To illustrate how these parameters affect valuations, we calculate price-dividend ratios for various combinations of disaster magnitude b and probability p. We set $\beta = 0.96$, $\gamma = 2$, and $h = 0.20$. With $\gamma = 2$, the exponent simplifies to $(b + h)$. The table below shows the resulting price-dividend ratios:

	$p = 0.0001$	$p = 0.005$	$p = 0.01$	$p = 0.015$	$p = 0.02$
$b = 0.40$	24	27	30	35	39
$b = 0.55$	24	28	33	39	55
$b = 0.70$	24	29	37	58	82
$b = 0.85$	24	31	42	76	199
$b = 0.95$	24	33	52	110	Inf

Table 1: Price-dividend ratios for varying disaster probability p and magnitude b, with $\beta = 0.96$, $\gamma = 2$, and $h = 0.20$

These results demonstrate a striking pattern: as both the probability and severity of an AI singularity increase, the price-dividend ratio of AI assets rises dramatically. For example, with a modest 1% probability of a singularity ($p = 0.01$) and a severe consumption decline of 95% ($b = 0.95$), the price-dividend ratio more than doubles compared to the baseline case with negligible singularity risk. At higher probabilities, valuations can become extraordinarily large or even infinite when the expected return falls below the growth rate.

This analysis suggests that high valuations of AI stocks could be rational even if investors fear negative consequences from AI development. The key insight is that these assets may serve as hedges against catastrophic AI outcomes for the representative investor, justifying premium valuations despite potential risks to society as a whole.

4 Model Discussion

Our model deliberately simplifies many aspects of the economy to highlight the core mechanism. Several important extensions and considerations merit discussion.

Market incompleteness is implicit but important in our framework. The consumption disaster magnitude 'b' represents the net effect of both (1) the AI disaster itself and (2) the AI asset dividend hedging ability. If markets were complete, the representative household could buy shares in all AI assets, including privately-held frontier AI labs, and not only fully hedge against but potentially benefit from the singularity. In reality, most households cannot buy shares in many cutting-edge labs like OpenAI, Anthropic, xAI, or DeepSeek.

This market incompleteness limits households' ability to hedge against AI risks and explains why publicly-traded AI assets might command a premium.

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 or examine competitive dynamics that might accelerate AI development beyond safety thresholds. However, such extensions would inevitably entail more assumptions about highly uncertain dynamics. Rather than decorating speculations with mathematics, we've opted for parsimony. The additional complexity would be costly to analyze and to read, while the core economic insights would remain largely unchanged.

Our model also ignores heterogeneity among households. In reality, different households have varying exposures to AI risk based on their occupations, geographic locations, and existing asset holdings. A richer model might capture these differences, potentially showing even more extreme hedging motives for particularly vulnerable households. Nevertheless, the representative agent framework suffices to demonstrate the fundamental mechanism.

The model's simplicity allows room for the human-written Appendix A, which provides additional perspective on the paper's arguments and limitations. This trade-off between model complexity and accessibility reflects our aim of concisely illustrating a single, novel economic mechanism rather than providing a comprehensive analysis of all AI-related market phenomena.

5 Conclusion and Implications

This paper has presented a simple asset pricing model to explore a novel explanation for high AI stock valuations. While conventional wisdom attributes these valuations primarily to growth expectations, we've shown that they may also reflect a hedging premium against AI singularity risk. Our model demonstrates how a publicly-traded AI asset can serve as insurance against potential consumption disasters triggered by radical AI advances, even if the asset itself suffers absolute dividend declines in such scenarios.

The analysis suggests that financial markets may naturally develop mechanisms to address some AI catastrophe risks. As investors perceive increasing singularity probability, they bid up prices of assets that would retain relative value in such scenarios. This market-based risk mitigation approach complements other proposed responses like universal basic income. The key economic insight is that this hedge is limited by market incompleteness—if households could directly invest in all frontier AI companies, including private ones, their hedging options would be more complete.

Framing AI stock investment as disaster insurance provides a different perspective on skyhigh valuations that might otherwise seem puzzling. It suggests that apparent ”irrational exuberance” could actually reflect rational concerns about technological risk. At the same time, this hedging demand could potentially distort capital allocation, directing excessive investment toward publicly-traded AI companies relative to their fundamental value.

Financial market solutions to AI risk have received insufficient attention in the broader literature on AI safety and policy (Bengio et al., 2024; Jones, 2024; Korinek and Suh, 2024). Our model suggests that enabling broader investment access to frontier AI development might distribute technological risk more efficiently. This could include mechanisms like expanded retail access to private equity, special purpose vehicles for AI investment, or regulatory frameworks that ensure public markets can participate in the upside of revolutionary AI advances.

As AI development accelerates, understanding these financial dynamics becomes increasingly important. The possibility that AI stocks serve as hedges against singularity risk adds a new dimension to debates about technology regulation, investment strategies, and economic policy in the age of artificial intelligence.

References

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

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.

Gomes, Francisco, Thomas Jansson, and Yigitcan Karabulut (2024). "Do Robots Increase Wealth Dispersion?" In: *The Review of Financial Studies*.

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*.

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.

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.
```