Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

This paper explores how the potential for a negative AI singularity could influence asset prices. We develop a simple disaster risk model where revolutionary AI advances benefit AI owners but harm the representative household's consumption. Our framework demonstrates that AI stocks might command high valuations not only due to expected earnings growth, but also because they hedge against catastrophic AI outcomes. We derive closed-form expressions for AI asset price-dividend ratios and show they can be substantial even with small disaster probabilities. 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. Within months of OpenAI releasing o1, competing firms like DeepSeek unveiled comparable reasoning models. Autonomous vehicles from Waymo now operate without safety drivers in multiple cities. Meanwhile, generative AI systems create increasingly sophisticated content across domains. These rapid developments have sparked concerns about potential labor displacement, with several studies suggesting AI could substantially reduce wages for many workers (Acemoglu and Restrepo, 2020; Knesl, 2023; Zhang, 2019).

While technological change has occurred throughout history, AI is fundamentally different. There is, in principle, no product or service that sufficiently advanced AI could not create. This paper itself exemplifies this distinction—it was entirely written by AI using a few short prompts.[1] The complete code and prompts are available at `https://github.com/chenandrewy/Prompts-to-Paper/`. This differs markedly from previous technological revolutions like the internet, which primarily enhanced communication and information access rather than threatening to displace human cognition itself. Moreover, AI progress may eventually become incredibly sudden—the hypothesized ”singularity” where self-improving AI rapidly surpasses human capabilities across all domains.

This paper studies how AI stocks are priced given the risk that future AI developments could devastate the consumption of the representative household. We introduce a simple disaster risk model where revolutionary AI advances lead to contractions in aggregate consumption for typical households while simultaneously increasing the relative value of AI-producing assets. Our framework provides a parsimonious explanation for why AI stocks might command high valuations even without expectations of substantial dividend growth—they serve as a hedge against negative singularity outcomes.

We are not claiming that a negative singularity will definitely occur. AI optimists envision a future where advanced AI dramatically enhances human capabilities and welfare. However, as Bengio et al. (2024) and others have warned, extreme AI risk scenarios deserve serious consideration given their potentially catastrophic consequences. Similarly, we are not asserting that this hedging value is already fully priced into current AI stock valuations. Our model simply illustrates a possible mechanism through which singularity concerns could influence asset prices.

Our work builds on two streams of literature. First, the rare disaster framework originated with Rietz (1988), who showed that small probabilities of catastrophic economic outcomes could explain the equity premium puzzle. This approach was further developed by Barro

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

(2006) and Wachter (2013). Second, we connect to research on technological singularity and AI risk, including seminal work by Vinge (1993) and Bostrom (2014), as well as recent economic analyses by Korinek and Suh (2024) and Jones (2024).

The economic literature on AI has primarily focused on labor market impacts (Zhang, 2019; Knesl, 2023) or macroeconomic growth scenarios (Korinek and Suh, 2024). Our contribution is to connect these AI risk concerns to asset pricing theory, suggesting that the high valuations of AI companies may partly reflect their potential role as hedges against negative singularity outcomes rather than just future earnings growth.

The remainder of this paper is organized as follows. Section 2 presents our model of AI singularity risk and derives the price-dividend ratio for AI assets. Section 3 discusses model extensions and limitations. Section 4 explores policy implications and concludes.

2 Model

We introduce a simple model to explore the potential impact of an AI singularity on asset prices. While deliberately stylized, our framework captures the key tension between AI owners and the broader population in a scenario where rapid AI development leads to negative outcomes for most households.

Our economy consists of two types of agents. First, there are AI owners who are fully invested in AI technology and are not marginal investors in the stock market. Second, there is a representative household who is the marginal investor in financial markets. Only the representative household's consumption matters for asset pricing. We assume 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 disaster process. In normal times, consumption growth is 1 (no growth for simplicity), but with some probability, consumption can fall by a factor of e^{-b} where $b > 0$. Formally:

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

These disasters represent revolutionary improvements in AI that are devastating for the representative household. While such advances benefit AI owners, they lead to significant losses for the household in terms of labor income, way of life, and sense of meaning. At time $t = 0$, we assume no disasters have occurred yet—the singularity has not happened. Our model allows for multiple disasters, capturing the ongoing uncertainty that persists even after an initial singularity event.

We consider a publicly traded AI asset with dividends that represent a small fraction of consumption before any singularity occurs. Each time a disaster happens, the dividend's share of total consumption grows by a factor of e^h. Specifically, if we denote the dividend at time t as D_t, then:

$$\frac{D_t}{C_t} = \frac{D_0}{C_0} \cdot (e^h)^{N_t}$$

where N_t is the number of disasters that have occurred by time t. This formulation is meant to capture a worst-case scenario where, despite the dividend's increasing share of consumption, the absolute dividend may actually shrink during each disaster if $h < b$. This reflects a situation where the most significant AI improvements are concentrated in privately-held AI assets rather than publicly traded ones.

3 Results

We now derive the price-dividend ratio of the AI asset at time $t = 0$, before any singularity events have occurred. Our approach follows standard consumption-based asset pricing with CRRA utility.

The representative household's stochastic discount factor (SDF) is given by:

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

where $\beta \leq 1$ is a time discount factor. For simplicity, we set $\beta = 0.96$ in our numerical examples.

Under the assumption of a constant price-dividend ratio $Q = P_t/D_t$, the standard asset pricing equation $P_t = E_t[M_{t+1}(P_{t+1} + D_{t+1})]$ can be rearranged to yield:

$$Q = \frac{E[M_{t+1}G^d_{t+1}]}{1 - E[M_{t+1}G^d_{t+1}]}$$

where $G^d_{t+1} = D_{t+1}/D_t$ is the gross dividend growth rate. In our model, both the SDF and dividend growth depend on whether a disaster occurs. With probability $1 - p$, no disaster occurs, consumption growth is 1, and dividend growth is also 1. With probability p, a disaster occurs, consumption falls by factor e^{-b}, and the dividend's share of consumption increases by factor e^h, resulting in dividend growth of e^{h-b}.

Computing the expected product of the SDF and dividend growth:

$$E[M_{t+1}G^d_{t+1}] = \beta[(1-p) + pe^{h+b(\gamma-1)}]$$

Substituting this into our expression for Q, we obtain:

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

This formula is valid as long as $\beta[(1-p) + pe^{h+b(\gamma-1)}] < 1$. If this condition is violated, the price-dividend ratio becomes infinite, suggesting that the asset is so valuable as a hedge against the AI singularity that investors would be willing to pay any price for it.

To illustrate how the price-dividend ratio varies with the disaster probability p and the severity of consumption decline b, we compute Q for different parameter combinations. We fix $h = 0.20$ and $\gamma = 2$, which implies $h + b(\gamma - 1) = 0.20 + b$.

$$Q = \frac{0.96 \cdot ((1-p) + p \cdot e^{0.20+b})}{1 - 0.96 \cdot ((1-p) + p \cdot e^{0.20+b})}$$

The following table shows the price-dividend ratios for various combinations of p and b:

Table 1: Price-Dividend Ratio (Q) for Various Parameters

	b (Consumption Decline Parameter)				
p (Disaster Probability)	0.40	0.55	0.70	0.85	0.95
0.0001	24	24	24	24	24
0.001	25	25	25	25	25
0.005	27	28	29	31	33
0.01	29	33	37	44	52
0.02	39	55	76	199	Inf

The table reveals several interesting patterns. First, when the disaster probability is very low (0.0001), the price-dividend ratio remains stable at 24 regardless of the disaster severity. This is because rare events have minimal impact on asset prices when their probability is sufficiently small.

As the disaster probability increases, the price-dividend ratio becomes more sensitive to the disaster severity. For example, with $p = 0.01$, the price-dividend ratio increases from 29 to 52 as b increases from 0.40 to 0.95. This reflects the increasing value of the AI asset as a hedge against more severe consumption declines.

Most strikingly, when both the disaster probability and severity are high ($p = 0.02$ and $b = 0.95$), the price-dividend ratio becomes infinite. In this case, the representative household values the AI asset so highly as a hedge against the singularity that traditional pricing models break down.

These results suggest that even a small probability of a severe AI singularity could significantly inflate the valuations of AI assets. This provides an alternative explanation for high AI stock valuations beyond the conventional growth narrative. Rather than simply reflecting expectations of future earnings growth, these valuations may partly incorporate a hedging premium against catastrophic AI outcomes.

4 Model Discussion

Our model is deliberately simple to illustrate the core mechanism by which AI singularity concerns might influence asset prices. Several extensions and limitations merit discussion.

Market incompleteness is implicit but important in our framework. The consumption decline parameter b represents the net effect of both the AI disaster itself and the AI asset's dividends. If markets were complete, the representative household could purchase shares in all AI assets, including privately held ones, and thereby not only fully hedge against the singularity but potentially benefit from it. In reality, most households cannot buy shares in many cutting-edge AI labs such as OpenAI, Anthropic, xAI, or DeepSeek. This market incompleteness limits the household's ability to hedge against singularity risk.

A more elaborate model would explicitly represent AI owners, their incentives, and their interactions with the representative household. One might ask: How might AI owners' incentives lead to a negative singularity? Would they not have incentives to prevent outcomes that harm society at large? However, decorating such speculations with mathematical formalism might add complexity without additional insight. The analysis would be costlier to develop and read, while the core economic mechanisms would remain essentially the same as in our simpler model.

Our model assumes that the representative household's consumption falls during an AI singularity. This reflects the possibility that advanced AI could displace significant human labor, reducing wage income, or could fundamentally disrupt social structures in ways that diminish human welfare. However, it's worth noting that many AI researchers envision more positive scenarios where AI advances enhance human capabilities and raise living standards. Our model focuses on the negative tail risk because this is precisely the risk that households might seek to hedge through AI asset ownership.

We've also assumed that AI assets increase their share of total consumption during singularity events. This reflects the possibility that as AI advances, the entities controlling the most powerful AI systems capture an increasing share of economic value. In reality, the dynamics might be more complex—AI systems might compete with each other, regulatory interventions might limit AI owners' power, or technological developments might democratize access to advanced AI capabilities.

Despite these limitations, our simple model captures the essential insight that AI assets might serve as hedges against negative singularity outcomes, potentially explaining part of their valuation premium. The short model analysis also allows room for the human-written perspective in Appendix A.

5 Policy Implications and Conclusion

Our analysis suggests that financial markets might offer partial solutions to AI disaster risk through the hedging mechanism we've identified. If households can invest in AI assets, they can partially insure themselves against negative singularity scenarios by sharing in the returns that accrue to AI owners. This market-based approach could complement other proposals for addressing potential AI-induced economic disruption, such as universal basic income.

However, the effectiveness of this hedging approach is fundamentally limited by market incompleteness. Many of the most promising AI systems are being developed by private companies with restricted ownership, while others are controlled by large technology companies where AI is just one of many business lines. This situation limits the ability of ordinary households to diversify their income risk through financial markets. As Betermier et al. (2012) and others have shown, hedging labor income risk through financial markets is already challenging with existing technologies; AI may exacerbate this challenge.

One policy approach might be to encourage greater public ownership of AI companies, either through regulatory mandates for public listings or through government investment vehicles that acquire stakes in private AI developers and pass the returns to citizens. However, such approaches would need to be carefully designed to avoid discouraging innovation or creating perverse incentives.

In conclusion, our simple model illustrates how concerns about AI singularity risk might already be influencing asset prices through a hedging mechanism. The potential for AI to dramatically reduce the consumption of ordinary households creates demand for assets that would maintain their value—or even increase in value—during such scenarios. While we've focused on the pricing implications rather than normative judgments about AI development paths, our analysis highlights the importance of considering financial market dynamics in discussions of AI policy.

As AI continues to advance, understanding these hedging dynamics may become increasingly important for investors, policymakers, and society at large. The question of how to distribute the risks and rewards of transformative AI development remains open, but financial markets are likely to play a significant role in any solution.

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`.

Betermier, Sebastien, Thomas Jansson, Christine Parlour, and Johan Walden (2012). "Hedging Labor Income Risk". In: *Journal of Financial Economics* 105.3, pp. 622–639.

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

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.

# 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.
```