Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

This paper explores how the potential for an AI singularity might impact asset prices. We develop a simple disaster risk model where a breakthrough in AI capabilities could be devastating for the representative household. Counter to conventional wisdom that attributes high AI stock valuations solely to growth expectations, we show these valuations may partly reflect their role as hedges against negative singularity scenarios. Our model demonstrates how even without dividend growth, AI assets can command high price-dividend ratios if they outperform the representative household during singularity events. Unlike previous work, this short paper is generated by prompting large language models.

Keywords: Artificial Intelligence, Disaster Risk, Asset Pricing

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

1 Introduction

Recent breakthroughs in artificial intelligence have been startling in their pace and breadth. OpenAI's o1 model demonstrated unprecedented reasoning abilities, while DeepSeek's R1 achieved comparable performance through pure reinforcement learning (DeepSeek-AI et al., 2025). Autonomous vehicles from Waymo now navigate complex urban environments without human intervention. As these technologies advance, investors increasingly worry that AI might displace their labor income and erode their consumption (Acemoglu and Restrepo, 2020; Zhang, 2019).

Previous technological revolutions certainly disrupted labor markets, but AI represents something fundamentally different. Unlike the steam engine, electricity, or even the internet, there is no product or service that AI could not, in principle, create. An illustrative example is the current paper, which is entirely written by AI through just six prompts. This differs from, say, the internet revolution which primarily transformed information access and communication channels. AI progress may also be incredibly sudden—the technological singularity that researchers like Bostrom (2014) and Russell (2019) have discussed.[1]

This paper studies how AI stocks are priced given the risk that AI will destroy livelihoods and consumption. We develop a simple model with two types of agents: AI owners who are fully invested in AI technologies and are not marginal investors in the stock market, and a representative household who is the marginal investor in financial markets. The representative household faces the possibility of a disaster in the form of an AI singularity that severely reduces their consumption. Crucially, we allow for AI assets to increase their share of total consumption during these disaster events.

We are not claiming that a negative AI singularity will definitely occur. Experts remain divided on both the likelihood and timeline of advanced AI development, as well as its potential impact on humanity (Bengio et al., 2024). Nevertheless, it is important to consider this scenario in financial markets, where assets are priced based not only on expected outcomes but also on their behavior across various states of the world, particularly those with negative consumption consequences.

We are also not arguing that this hedging value is already fully priced into current AI stock valuations. Rather, our model illustrates a possible mechanism through which rational investors might value AI stocks even in the absence of expected dividend growth. The model shows how a small probability of a singularity event can generate substantial price-dividend ratios for AI assets, especially when these assets' dividends increase as a share of aggregate

[1] In this paper, "we" refers to one human author and multiple LLMs. For a purely human perspective see Appendix A.

consumption during singularity events.

The literature on rare disasters provides a framework for understanding how catastrophic risks can influence asset prices. Rietz (1988) pioneered the idea that a small probability of economic disasters could explain the equity premium puzzle, while Barro (2006) calibrated disaster probabilities using 20th-century global events. Gabaix (2012) developed models with time-varying disaster severity that explain multiple asset pricing puzzles, and Wachter (2013) demonstrated how time-varying disaster probability can generate both high equity premiums and stock market volatility. Recent studies suggest that AI-related assets may serve as hedges against technological risks. Babina et al. (2023) provide evidence that firms' investments in AI technologies affect their systematic risk profiles, while Zhang (2019) demonstrates that firms with higher automation potential maintain a replacement option that hedges their value against unfavorable macroeconomic shocks. These findings support the possibility that AI stock valuations reflect not only growth expectations but also compensation for bearing or hedging against significant technological transition risks, including potential negative singularity scenarios. Jones (2024) explores the economic tension between AI-driven growth and potential existential risks, while Korinek and Suh (2024) analyzes how output and wages respond to different AI development scenarios, including the transition to artificial general intelligence.

2 Model

We develop a simple model to explore the asset pricing implications of a potential AI singularity. Our model is deliberately stylized to highlight the key economic forces at work rather than to provide precise quantitative predictions.

We consider an economy with two types of agents. First, there are AI owners who are fully invested in AI technologies 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:

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

where γ is the coefficient of relative risk aversion.

The representative household faces uncertainty about future consumption growth. Specifically, gross consumption growth follows a simple disaster process. In normal times, consumption growth is normalized to 1 (no growth for simplicity). However, the economy may experience disasters in the form of revolutionary improvements in AI that are devastating for the representative household. When such a disaster occurs, gross consumption growth is e^{-b} where $b > 0$ represents the severity of the disaster. At time $t = 0$, no disasters have yet occurred (the singularity has not yet happened). Multiple disasters may occur over time, capturing ongoing uncertainty even after an initial singularity event.

The key assumption is that while AI improvements benefit the AI owners, they can be harmful to the representative household. This could occur through displacement of labor, disruption of social structures, or loss of meaning and purpose. The benefits of AI improvements accrue primarily to the AI owners, while the costs fall disproportionately on the representative household.

We consider a publicly traded AI asset with a dividend stream. Before any singularity event, the dividend represents a small fraction of aggregate consumption, denoted by δ:

$$D_0 = \delta C_0$$

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

$$D_n = \delta e^{nh} C_n$$

Importantly, we allow for $h < b$, capturing a worst-case scenario where the dividend may actually shrink in absolute terms during each disaster, even as it grows as a fraction of the representative household's consumption. This reflects the possibility that the benefits of AI improvements are concentrated in privately-held AI assets rather than publicly traded ones.

3 Results

We now derive the price of the AI asset in our model. The price at time 0 is the present value of all future dividends, discounted using the representative household's stochastic discount factor (SDF). In our discrete-time setting with CRRA preferences, the SDF between time 0 and time t is:

$$M_t = \beta^t \left(\frac{C_t}{C_0} \right)^{-\gamma}$$

where $\beta \in (0, 1)$ is the time discount factor.

The number of disasters that occur by time t, denoted N_t, follows a binomial distribution with parameters t and p, where p is the per-period probability of a disaster. Given our consumption process, we have $C_t = C_0 e^{-bN_t}$. Similarly, the dividend at time t can be expressed as $D_t = D_0 e^{N_t(h-b)}$.

The price-dividend ratio at time 0, denoted $p_0 = P_0/D_0$, can be written as:

$$p_0 = \sum_{t=1}^{\infty} E_0 \left[M_t \frac{D_t}{D_0} \right]$$

Substituting the expressions for M_t and D_t/D_0, we get:

$$p_0 = \sum_{t=1}^{\infty} E_0 \left[\beta^t \left(\frac{C_t}{C_0} \right)^{-\gamma} \frac{D_t}{D_0} \right] = \sum_{t=1}^{\infty} E_0 \left[\beta^t e^{N_t(b\gamma+h-b)} \right]$$

Using the moment generating function of the binomial distribution, we can compute:

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

This gives us:

$$p_0 = \sum_{t=1}^{\infty} \beta^t \left(1 - p + pe^{b\gamma+h-b} \right)^t$$

Defining $\alpha \equiv 1 - p + pe^{b\gamma+h-b}$, we have:

$$p_0 = \sum_{t=1}^{\infty} (\beta\alpha)^t$$

This geometric series converges if and only if $\beta\alpha < 1$, in which case:

$$p_0 = \frac{\beta\alpha}{1-\beta\alpha} = \frac{\beta(1-p+pe^{b\gamma+h-b})}{1-\beta(1-p+pe^{b\gamma+h-b})}$$

This formula clearly shows how the disaster probability p and the parameters b and h affect the valuation of the AI asset. The key insight is that each singularity event simultaneously reduces the household's consumption (making them worse off) while increasing the dividend's share of that consumption (reflecting the relative gain to AI owners).

To illustrate the quantitative implications of our model, we compute price-dividend ratios for various parameter combinations. We set $\gamma = 2$, $h = 0.20$, and $\beta = 0.96$, while varying the disaster severity b and probability p. With these parameter values, the exponent simplifies to $b\gamma + h - b = b + 0.20$.

The table below shows the price-dividend ratios for different combinations of b and p:

Several patterns emerge from this table. For very low disaster probabilities ($p = 0.0001$), the disaster severity b has little impact on valuations because the disaster adjustment enters through the small multiple $p \cdot e^{b+0.20}$. As the probability increases, the impact of b becomes

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

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

more pronounced. When $p = 0.02$ and $b = 0.95$, the effective discount factor $\beta\alpha$ reaches 1, causing the price-dividend ratio to become infinite as the geometric series fails to converge.

These results suggest that even modest probabilities of an AI singularity can generate substantial asset price effects, particularly when the potential consumption impact is severe. This provides a rational explanation for high AI stock valuations that does not rely solely on expectations of future earnings growth.

4 Model Discussion

Our model illustrates a simple but powerful point: AI assets may be valuable as hedges against AI-induced consumption disasters, even if they don't experience traditional dividend growth. This hedging property is complementary to the more commonly discussed growth narrative for AI stocks.

Market incompleteness plays an implicit but important role in our model. The disaster magnitude parameter b represents the net effect of both the AI disaster and the AI asset dividend response. If markets were complete, the representative household could buy shares in all AI assets, including privately held ones, and not only fully hedge but potentially benefit from the singularity. In reality, most households cannot buy shares in many cutting-edge AI labs such as OpenAI, Anthropic, xAI, and DeepSeek. This market incompleteness means that publicly traded AI stocks may be imperfect but valuable hedges against singularity risk.

A more elaborate model would explicitly incorporate AI owners, their incentives, and their interactions with the representative household. Such a model might explore how AI owners' incentives could lead to a negative singularity, perhaps through competitive dynamics or misaligned objectives. However, this would essentially decorate speculations with mathematics, adding complexity without necessarily improving insights. The core economics—that AI assets could serve as hedges against AI-induced consumption disasters—would remain the same.

Our model also assumes a representative household, abstracting from heterogeneity in how different individuals might be affected by an AI singularity. In reality, some workers would likely be affected more severely than others, depending on their skills, occupations, and ability to adapt. Incorporating such heterogeneity would complicate the model without changing the fundamental insight that assets that perform relatively well during singularity events would command a premium.

The simplicity of our approach allows us to focus on the key mechanism while leaving room for the human-written perspective in Appendix A. As with any model, ours is a deliberate simplification of reality designed to isolate and understand specific economic forces rather than capture all aspects of a complex phenomenon.

5 Conclusion and Implications

This paper has explored how the potential for an AI singularity might impact the valuation of AI-related assets. Our simple model shows that AI stocks could command high price-dividend ratios partly because they serve as hedges against consumption disasters caused by runaway AI development. Even without traditional dividend growth, the fact that AI assets' dividends might increase as a share of aggregate consumption during singularity events makes them valuable to risk-averse investors.

These findings suggest an interesting possibility: financial markets could provide a partial solution to AI catastrophe risk through the hedging value of AI stocks. As individuals become concerned about AI-induced labor displacement, they might naturally increase their holdings of AI-related assets as a hedge. This market-based response is an alternative to policy interventions like universal basic income, though it is limited by market incompleteness. Most households cannot invest in private AI companies, and the benefits of AI development might accrue disproportionately to a small group of owners and early investors.

The hedging mechanism we describe is not widely discussed in the literature on AI risk, which tends to focus on technical alignment, governance frameworks, and social safety nets. Financial markets offer an additional channel through which societies might adapt to radical technological change. By allowing individuals to participate in the upside of technologies that might otherwise harm them, financial markets could help smooth the transition to an AI-dominated economy.

Our model is intentionally stylized, and we make no claims about whether current AI stock valuations actually reflect this hedging property. Nevertheless, the mechanism we describe is consistent with standard asset pricing theory and offers a novel perspective on how rational investors might value AI assets in a world of technological uncertainty.

As AI capabilities continue to advance rapidly, understanding how financial markets price AI-related assets becomes increasingly important. The potential for hedging against AI singularity risk represents an intriguing possibility that deserves further exploration by researchers, investors, and policymakers alike.

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.

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

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

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

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

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. Its writing often fails to portray the economics 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.
```