Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

We propose that high valuations of AI companies may partly reflect their role as hedges against a negative AI singularity—an explosive development in artificial intelligence that could be devastating for the representative investor. This contrasts with the common view that AI valuations primarily stem from future earnings growth. Our stylized model shows that even with small disaster probabilities, AI stocks can command high price-dividend ratios due to their hedging properties. This insight offers a novel perspective on current AI valuations and suggests potential financial market solutions to technological risk. Unlike previous work, this short paper is generated by prompting LLMs.

Keywords: Artificial Intelligence, Disaster Risk, Asset Pricing

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

1 Introduction

Artificial intelligence is progressing at a breathtaking pace. Recent achievements like DeepSeek's R1 reasoning model (DeepSeek-AI et al., 2025), Waymo's autonomous vehicles operating without safety drivers, and Claude's ability to analyze complex documents illustrate how quickly AI capabilities are advancing. For investors and workers alike, this rapid progress raises a profound question: what happens to human labor—and by extension, human consumption—when machines can perform an ever-expanding range of tasks? Papers by Zhang (2019) and Knesl (2023) document that financial markets already price in some of this technological displacement risk.

Unlike previous technological revolutions, AI represents something fundamentally different. The printing press, steam engine, and even the internet primarily augmented human capabilities in specific domains. In contrast, there is virtually no product or service that AI could not, in principle, create. The very paper you are reading exemplifies this point—it is written entirely by AI systems instructed with a few short prompts.[1] The paper's reasoning, structure, and prose all emerge from AI, guided but not directly authored by a human. This differs fundamentally from, say, the internet revolution, which provided new ways for humans to communicate but did not itself generate content or reasoning. Moreover, AI progress may prove incredibly sudden—the hypothesized "technological singularity" where AI capabilities increase exponentially (Vinge, 1993; Kurzweil, 2005), potentially leading to profound economic disruption.

This paper explores how AI stocks might be valued given the risk that unchecked AI advancement could dramatically reduce human consumption. While most investors likely view AI stocks as growth opportunities, we propose an additional mechanism: AI stocks may serve as hedges against a negative AI singularity. If AI advancement proceeds in a way that reduces aggregate consumption but increases the relative share of consumption captured by AI companies, then owning AI stocks provides a form of insurance. This contrasts with the conventional view that high AI valuations are primarily justified by expectations of future earnings growth.

We are not claiming that a negative singularity will happen. Nor are we suggesting that the risk of such an outcome is the primary driver of current AI stock valuations. Nevertheless, understanding this hedging mechanism helps provide a more complete picture of how rational investors might approach AI investments, especially given the fundamental uncertainty surrounding transformative AI development.

[1] "We" in this paper refers to one human author and multiple large language models. For a purely human perspective, see Appendix A.

Our analysis builds on the rich literature of rare disaster models in asset pricing, initiated by Rietz (1988) and expanded by Barro (2006), who showed that small probabilities of economic catastrophes can explain high equity risk premiums. Gabaix (2012) and Wachter (2013) further developed these frameworks to incorporate time-varying disaster risk. We extend this approach to the specific context of AI-driven disasters, connecting to research on AI catastrophic risks. Bengio et al. (2024) highlight potential extreme risks from advanced AI systems, while Bostrom (2014) and Russell (2019) explore how superintelligent AI might pose existential threats if not properly aligned with human values. Jones (2024) specifically examines the economic tension between AI-driven growth and potential existential risks, providing an economic framework for analyzing these trade-offs.

The potential for AI technologies to disrupt labor markets has been examined in several studies with asset pricing implications. Zhang (2019) and Knesl (2023) demonstrate that firms with high exposure to automation technology have lower expected returns, as they provide hedging benefits against technological displacement risk. This connects to our premise that AI stocks might serve as hedges against negative AI scenarios. Empirically, Babina et al. (2024) document that AI-investing firms experience higher growth in sales, employment, and market valuations. Our work complements these studies by providing a theoretical framework where AI assets serve as hedges in a Lucas-tree economy with disaster risk, similar to how Santos and Veronesi (2006) model the impact of non-financial income on asset prices.

2 Model

We propose a stylized model to explore the idea that AI stocks may be valued highly in part because they hedge against a negative AI singularity. Our model is deliberately simple, focusing on the essential economic mechanisms rather than quantitative precision.

We consider an economy with two types of agents: AI owners and a representative household. AI owners are fully invested in AI assets and are not marginal investors in the stock market. The representative household, on the other hand, is the marginal investor whose consumption-based preferences determine asset prices. 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 in consumption growth. Specifically, gross consumption growth can take one of two values: either 1 (no change) or e^{-b} (disaster), where $b > 0$. This disaster state represents a revolutionary improvement in AI that is devastating for the 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 diminished sense of meaning. At time $t = 0$, we assume no disasters have yet occurred—the singularity has not happened. However, multiple disasters may occur over time, capturing the ongoing uncertainty that would persist even after an initial singularity event.

We consider a publicly traded AI asset with dividend process D_t. Before any singularity occurs, this dividend represents a small fraction of aggregate consumption:

$$D_t = \delta C_t$$

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

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

This specification is meant to capture a worst-case scenario for the representative household. Even if $h < b$, the dividend may actually shrink in absolute terms during each disaster, but it would still grow as a fraction of the household's consumption. This reflects a situation where the benefits of AI improvements are concentrated in privately-held AI assets, with publicly traded AI companies capturing only a portion of the value created by advances in artificial intelligence.

To derive the price-dividend ratio for the AI asset, we work in discrete time and assume that each period, either no disaster occurs (with probability $1 - p$) or a disaster occurs (with probability p). We use the standard consumption-based asset pricing framework. The stochastic discount factor for the representative household with CRRA utility is:

$$M_{0,t} = \beta^t \left(\frac{C_t}{C_0}\right)^{-\gamma}$$

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

The price of the AI asset at time 0 is the expected present value of all future dividends:

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

Substituting the expressions for $M_{0,t}$ and D_t, and noting that if n_t disasters have occurred by time t, then $C_t = C_0 e^{-bn_t}$ and $D_t = \delta C_0 e^{n_t(h-b)}$, we can express the price-dividend ratio as:

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

Since n_t follows a binomial distribution with parameters (t, p), the expectation of e^{Xn_t} is $((1-p)+pe^X)^t$. With $X = h - b(1-\gamma)$, we get:

$$\frac{P_0}{D_0} = \sum_{t=1}^{\infty} \beta^t \left((1-p) + pe^{h-b(1-\gamma)} \right)^t$$

This is a geometric series that sums to:

$$\frac{P_0}{D_0} = \frac{\beta \left((1-p) + pe^{h-b(1-\gamma)} \right)}{1 - \beta \left((1-p) + pe^{h-b(1-\gamma)} \right)}$$

provided that $\beta \left((1-p) + pe^{h-b(1-\gamma)} \right) < 1$ for convergence.

This closed-form solution reveals how the probability of a singularity (p), its effect on consumption (b), and its effect on the AI asset's dividend share (h) jointly determine the asset's valuation relative to its current dividend.

To illustrate the quantitative implications of our model, we set $\beta = 0.96$, $h = 0.20$, and $\gamma = 2$. With these parameters, the key term in our formula becomes $e^{h-b(1-\gamma)} = e^{0.20+b}$. The table below shows the price-dividend ratios for various combinations of disaster probability (p) and consumption impact (b):

Table 1: Price-Dividend Ratios for AI Assets

	Probability of Disaster (p)				
b	0.0001	0.0025	0.005	0.01	0.02
0.40	24	25	27	30	37
0.55	24	26	28	33	55
0.70	24	26	29	37	82
0.85	24	27	31	41	199
0.95	24	28	32	52	Inf

Several patterns emerge from this table. For very small disaster probabilities (e.g., $p = 0.0001$), the price-dividend ratio is approximately 24, which is close to the standard Gordon growth formula result of $\beta/(1-\beta) = 0.96/0.04 = 24$. As the probability of disaster increases, the price-dividend ratio rises, especially for larger values of b. When both p and b are sufficiently high (e.g., $p = 0.02$ and $b = 0.95$), the present value sum no longer converges, indicated by ”Inf” in the table.

These results demonstrate that even a small probability of a severe AI-driven disaster can substantially increase the valuation of AI assets that provide a hedge against such events. For instance, with a modest 1

3 Model Discussion

Our model is deliberately streamlined to illustrate the core economic mechanism at play: AI stocks may be valued not just for their growth potential but also for their hedging properties against negative AI singularity scenarios. Several natural extensions to the model would add realism but not fundamentally alter this insight.

Market incompleteness is implicit but important in our framework. The parameter b captures the net effect of both the AI disaster and the AI asset dividend on the representative household. If markets were complete, the household could purchase shares in all AI assets, including privately held cutting-edge labs, and not only fully hedge but potentially benefit from the singularity. In reality, most households cannot buy shares in many leading AI companies like OpenAI, Anthropic, xAI, or DeepSeek, which remain privately held with restricted access. This market incompleteness limits the representative household's hedging options.

A more elaborate model could explicitly represent AI owners, their incentives, and their interaction with the representative household. One might explore how AI owners' profit motives could potentially lead to a negative singularity through underinvestment in safety or rushing deployment before alignment issues are resolved. However, such an extension would largely be decorating speculations with mathematical formalism. The added complexity would make the model costlier to analyze and read without substantially changing the core economic insights.

Our model also treats the probability and magnitude of AI disasters as exogenous parameters. In practice, these parameters might depend on policy choices, investments in AI safety research, and coordination among AI developers. Endogenizing these parameters would allow for a richer analysis of how incentives shape AI development paths, but again would not fundamentally alter our main point about the hedging value of AI stocks.

The beauty of a simple model is that it isolates the essential economic mechanism. The fact that AI assets might serve as hedges against AI-induced consumption declines is a point that can be made without elaborate modeling. This simplicity also allows room for the human-written perspective in Appendix A, which provides valuable context from a different viewpoint.

4 Policy Implications and Conclusion

Our analysis suggests that financial markets may naturally provide partial insurance against negative AI scenarios through the hedging properties of AI stocks. This insight offers a complementary perspective to discussions about universal basic income (UBI) as a response to AI-driven labor displacement. While UBI approaches the problem through redistribution, our model highlights how market-based mechanisms might also help manage the risks of transformative AI.

The hedging mechanism we describe is fundamentally limited by market incompleteness. Since most households cannot directly invest in many frontier AI companies, their ability to hedge against AI singularity risks is constrained. This suggests that policies expanding access to AI investments—perhaps through AI-focused index funds, publicly traded AI partnerships, or broader employee stock ownership plans at AI companies—could help distribute the benefits of AI more widely.

These financial market solutions to AI disaster risk have received insufficient attention in the literature on AI safety and policy. While Korinek (2023) discusses the importance of scenario planning for AI policy and Korinek and Stiglitz (2018) analyze AI's implications for income distribution, the potential role of financial markets in mitigating AI risks is often overlooked. Even modest disaster probabilities can substantially affect asset valuations, as our model demonstrates, suggesting that financial markets may already be pricing in some of these risks.

Of course, financial hedges alone cannot solve all the challenges posed by transformative AI. They complement rather than replace technical AI safety research, alignment efforts, and thoughtful regulation. Nevertheless, our analysis suggests that the financial market's natural hedging properties deserve a place in the broader conversation about managing AI transition risks.

In conclusion, thinking about AI stocks as potential hedges against negative AI scenarios provides a novel perspective on their valuation. While conventional wisdom focuses primarily on growth prospects, our model demonstrates that even small probabilities of AI-induced consumption declines can justify high price-dividend ratios for AI assets. By recognizing the risk-sharing aspects of AI investments, we gain a more nuanced understanding of how financial markets might help societies navigate the uncertain frontier of artificial intelligence.

References

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson (2024). "Artificial intelligence, firm growth, and product innovation". In: *Journal of Financial Economics* 151, Article 103745.

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

Bengio, Yoshua, Geoffrey Hinton, Andrew Yao, Dawn Song, Pieter Abbeel, et al. (2024). "Managing extreme AI risks amid rapid progress". In: *Science* 384.6698. URL: https://arxiv.org/abs/2310.17688.

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

DeepSeek-AI et al. (Jan. 2025). "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning". In: *arXiv*. URL: https://arxiv.org/abs/2501.12948.

Gabaix, Xavier (2012). "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance". In: *Quarterly Journal of Economics* 127.2, pp. 645–700.

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

Knesl, Jiří (2023). "Automation and the Displacement of Labor by Capital: Asset Pricing Theory and Empirical Evidence". In: *Journal of Financial Economics* 147.2, pp. 271–296.

Korinek, Anton (2023). "Scenario Planning for an A(G)I Future". In: *IMF Finance & Development Magazine*.

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

Kurzweil, Ray (2005). *The Singularity Is Near: When Humans Transcend Biology*. Viking Press.

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.

Santos, Tano and Pietro Veronesi (2006). "Labor Income and Predictable Stock Returns". In: *Review of Financial Studies* 19, pp. 1–44.

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.
```