Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

This paper explores a novel explanation for high AI stock valuations: they may serve as hedges against a negative AI singularity. We develop a simple consumption-based asset pricing model where revolutionary AI improvements benefit AI owners but harm the representative household. Our analysis shows that even small probabilities of AI disasters can significantly elevate AI stock valuations, as these assets provide protection during catastrophic consumption declines. This mechanism contrasts with the common view that AI valuations reflect future earnings growth. Our results highlight the potential role of financial markets in distributing AI risk. Unlike previous work, this short paper is generated by prompting LLMs.

Keywords: Artificial Intelligence, Disaster Risk, Asset Pricing

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

1 Introduction

Artificial intelligence is advancing at a breathtaking pace. Recent achievements—from DeepSeek's R1 matching the reasoning capabilities of top models (DeepSeek-AI et al., 2025) to Waymo's autonomous vehicles logging millions of passenger miles—signal an unmistakable trend: AI systems are becoming increasingly capable across diverse domains. As these technologies proliferate, investors and workers alike harbor growing concerns about potential wage displacement and economic disruption (Zhang, 2019; Knesl, 2023). Historical parallels exist—technological revolutions have regularly transformed labor markets and asset prices—but AI presents a unique case.

Unlike previous technological shifts, there is effectively no product or service that AI could not, in principle, create or drastically transform. This very paper exemplifies the distinction: it is entirely written by AI, using just a few short prompts.[1] You can find the prompts and generation process at `https://github.com/chenandrewy/Prompts-to-Paper/`. This fundamentally differs from even revolutionary innovations like the internet. Moreover, AI progress may arrive with unprecedented suddenness—what researchers term an "AI singularity," where self-improving systems rapidly exceed human capabilities across virtually all domains (Bostrom, 2014; Vinge, 1993).

This paper examines a novel angle on AI valuation: how might AI stocks be priced if they hedge against negative singularity scenarios where AI destroys livelihoods and consumption? We develop a simple consumption-based asset pricing model with two types of agents: AI owners and a representative household. In our framework, a "disaster" represents a revolutionary improvement in AI that devastates the representative household while benefiting AI assets. We show that even small probabilities of AI disasters can generate surprisingly high price-dividend ratios for AI stocks, providing a potential explanation for seemingly excessive valuations.

We should emphasize that we are not predicting a negative singularity will occur. Forecasting such events with any confidence remains beyond current capabilities. Nevertheless, examining these scenarios—even as low-probability tail risks—provides valuable insights for investors, policymakers, and the public. Similarly, we make no claims about whether the hedging value we identify is already priced into AI stocks. Our model merely illustrates a plausible mechanism that could contribute to their valuations.

Our analysis builds upon several research streams. The approach connects to the rare disaster literature pioneered by Rietz (1988) and extended by Barro (2006), Gabaix (2012), and Wachter (2013). These papers demonstrate how small probabilities of extreme events

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

can significantly impact asset prices. We also draw from work on technology-specific risks, including Zhang (2019), who shows firms with automation options maintain hedges against macroeconomic shocks, and Knesl (2023), who finds firms exposed to automation risk earn premiums consistent with pricing of technological displacement risk.

Our framework is further informed by AI risk research from Bengio et al. (2024), who identify extreme risks from advanced AI systems, and Jones (2024), who explores the economic tension between AI-driven growth and existential risks. Additionally, Korinek and Suh (2024) analyze various economic transitions to artificial general intelligence, examining impacts on output and wages across different scenarios.

The remainder of the paper proceeds as follows. Section 2 presents our model. Section 3 derives the key pricing formula and provides numerical illustrations. Section 4 discusses model extensions and limitations. Section 5 concludes with policy implications.

2 Model

We now describe our model, which is purposefully simple to illustrate the key mechanism. We consider an economy with two types of agents: AI owners and a representative household. The 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 determines asset prices.

The representative household has constant relative risk aversion (CRRA) preferences over consumption, 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 disaster process. In each period, consumption either remains stable or experiences a disaster:

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

where $b > 0$ represents the severity of the disaster and p is the probability of a disaster occurring.

In our context, a "disaster" represents a revolutionary improvement in AI that is devastating for the representative household. While such technological advances benefit AI owners, they potentially harm the household through disruptions to labor income, way of life, and sense of meaning. At time $t = 0$, no disasters have yet occurred—the singularity has not happened. However, multiple disasters may occur over time, capturing the ongoing uncertainty following an initial singularity event.

We consider a publicly traded AI asset with dividend D_t that represents a fraction of aggregate consumption:

$$D_t = \delta_t C_t$$

where δ_t evolves based on the occurrence of disasters. Initially, δ_0 is a small fraction of consumption. Each time a disaster occurs, this fraction changes by a factor of e^h:

$$\frac{\delta_{t+1}}{\delta_t} = \begin{cases} 1 & \text{if no disaster occurs} \\ e^h & \text{if a disaster occurs} \end{cases}$$

The parameter h can be positive or negative, allowing us to model different scenarios. When $h < 0$, even AI stocks may suffer during a singularity event, capturing a worst-case scenario where benefits of AI improvements are concentrated in privately-held AI assets rather than publicly traded ones.

This framework, while stylized, allows us to examine how the hedging properties of AI stocks against singularity risk might affect their valuations.

3 Results

We now derive the price-dividend ratio for our AI asset at time $t = 0$. Using the standard consumption-based asset pricing framework, the price of an asset equals the expected discounted value of its future dividends:

$$P_0 = E_0 \left[\sum_{t=1}^{\infty} M(0,t) \cdot D_t \right]$$

where $M(0,t)$ is the stochastic discount factor from time 0 to time t. For the price-dividend ratio, we can factor out D_0:

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

Given our model assumptions, we can express the dividend growth in terms of the number of disasters $N(t)$ that have occurred by time t:

$$\frac{D_t}{D_0} = e^{(h-b)N(t)}$$

With CRRA preferences and discount factor β, the stochastic discount factor from time 0 to t is:

$$M(0,t) = \beta^t \cdot \left(\frac{C_t}{C_0}\right)^{-\gamma} = \beta^t \cdot e^{b\gamma N(t)}$$

Combining these expressions, the price-dividend ratio becomes:

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

Let $\lambda = e^{b(\gamma-1)+h}$. Since $N(t)$ follows a binomial distribution with parameters (t,p), we have $E_0[\lambda^{N(t)}] = (1-p+p\lambda)^t$. This gives us:

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

As long as $\beta(1-p+p\lambda) < 1$, this geometric series converges to:

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

This formula reveals how the AI asset's valuation depends on the disaster probability p, disaster severity b, risk aversion γ, AI dividend scaling h, and discount factor β.

To illustrate the quantitative implications of our model, we compute the price-dividend ratio for a range of parameter values. We set $\gamma = 2$, $h = 0.20$, and $\beta = 0.96$, which simplifies our formula to:

$$\frac{P_0}{D_0} = \frac{\beta[1-p+pe^{b+0.20}]}{1-\beta[1-p+pe^{b+0.20}]}$$

Table 1 shows the price-dividend ratios for different combinations of disaster severity b and probability p. When the product $\beta(1-p+pe^{b+0.20}) \geq 1$, the sum diverges, resulting in an infinite price (denoted as "Inf").

The table reveals that even small probabilities of severe AI disasters can substantially increase asset valuations. For example, with a disaster severity of $b = 0.85$ and probability $p = 0.02$, the price-dividend ratio reaches 230, nearly ten times the baseline ratio of 24 when disaster risk is negligible ($p = 0.0001$). This demonstrates how AI assets that hedge against singularity risk can command premium valuations, even when the probability of such events is small.

Table 1: Price-Dividend Ratios for Various Disaster Parameters

b	$p = 0.0001$	$p = 0.005$	$p = 0.01$	$p = 0.015$	$p = 0.02$
0.40	24	27	30	33	39
0.55	24	28	33	39	56
0.70	24	29	37	55	82
0.85	24	31	42	76	230
0.95	24	32	52	110	Inf

An important feature of our model is that these high price-dividend ratios emerge even though dividends never grow in expectation. This contrasts with standard growth-based explanations for high valuations. Instead, AI assets are valuable because they serve as hedges against disastrous consumption declines for the representative household.

4 Model Discussion

While our model is deliberately minimalist, it captures the essential mechanism through which singularity risk could affect AI asset prices. Several natural extensions and limitations deserve discussion.

Market incompleteness is implicit but important in our framework. The disaster magnitude parameter b represents the net effect of both the AI disaster and the AI asset dividend. If markets were complete, the representative household could purchase shares in all AI assets—including private AI ventures—and not only fully hedge against but potentially benefit from the singularity. In reality, most households cannot invest in many cutting-edge AI labs like OpenAI, Anthropic, xAI, or DeepSeek. This market incompleteness limits the representative 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. Such a model might explore how AI owners' incentives could lead to a negative singularity. However, this approach would essentially decorate speculations with mathematics. The additional complexity would be costly to analyze and digest while adding little to the core economic insight: AI assets may be valuable as hedges against AI disasters, even if those disasters have low probability.

Our model assumes that AI disasters affect the representative household and AI assets in specific ways. In reality, the impact would likely be heterogeneous across households and assets. Some households might benefit from AI advances through capital ownership or complementary skills, while others might suffer significant displacement. Similarly, some AI assets might benefit more than others during revolutionary advances. These nuances could be incorporated in future work.

We've also simplified the timing and nature of potential AI disasters. Our binary disaster state abstracts from the complex ways AI progress might unfold. A richer model might consider multiple disaster states with varying severities, or a continuous process of AI advancement with stochastic impacts on household consumption and AI dividends. However, such extensions would complicate the analysis without necessarily changing the fundamental insights.

The parsimony of our approach has advantages. By focusing on the essential mechanism, we provide a clear illustration of how disaster risk affects valuations. This simplicity also leaves room for the human-written Appendix A, which offers additional perspectives on the topic.

5 Policy Implications and Conclusion

Our analysis suggests that financial markets might play an important role in addressing AI disaster risk. If AI stocks indeed serve as hedges against negative singularity scenarios, they offer a market-based mechanism for distributing the risks and benefits of advanced AI. This represents an alternative or complement to more commonly discussed policy solutions like universal basic income.

The hedging mechanism we identify is limited by market incompleteness. Since many leading AI developers remain privately held, the general public cannot fully access these potential hedges. This suggests that the economics of AI-induced inequality are fundamentally tied to financial market structure. Improving access to AI investments could potentially help distribute the benefits of AI progress more widely, though market solutions alone cannot address all concerns.

It's worth noting that market-based solutions have inherent limitations. If AI disasters are severe enough, even well-diversified portfolios might provide inadequate protection. Moreover, many individuals have limited financial resources to invest, regardless of available opportunities. These constraints highlight the potential complementary role for thoughtful policy in managing AI transition risks.

While our model provides a novel perspective on AI stock valuations, it remains a simplified illustration. Future research could further explore the empirical relationship between AI risk perceptions and asset prices, or develop more sophisticated models that capture additional complexities of AI development pathways.

In conclusion, the possibility that AI stocks serve as hedges against negative singularity scenarios offers a new lens for understanding their valuations. Rather than merely reflecting optimism about future earnings growth, high AI stock prices may partly incorporate their insurance value against potentially devastating technological transitions. This insight highlights the importance of considering disaster risk when evaluating AI investments and emphasizes the potential role of financial markets in addressing the distributional challenges of AI advancement.

References

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 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.
```