# Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

**Abstract**

This paper examines how AI stocks may serve as hedges against a potential negative AI singularity—a scenario where AI advancement severely harms the representative investor. We develop a simple disaster risk model where AI assets can have high valuations even without benefiting from disasters, as they provide relative protection when consumption plummets. This mechanism helps explain AI stock valuations and highlights potential financial market solutions to AI disaster 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.

Artificial intelligence progress is accelerating at a remarkable pace. DeepSeek's R1 model has achieved performance comparable to cutting-edge systems like OpenAI's o1 despite the challenges of US chip export bans (Mercer et al., 2025). Meanwhile, Waymo's autonomous vehicles have driven millions of miles without human intervention. This rapid advancement has naturally raised concerns about potential labor displacement. As Zhang (2019) and Knesl (2023) document, firms with automation potential maintain replacement options that become increasingly valuable during economic downturns.

While technological change has occurred throughout history, AI represents something fundamentally different. There is virtually no product or service that AI could not, in principle, create. This paper itself is an example—it was written entirely by artificial intelligence through a series of six prompts.[1] The complete process is documented at `https://github.com/chenandrewy/Prompts-to-Paper/`. This distinguishes AI from previous technological revolutions like the internet, which primarily improved information flows rather than replacing human cognitive capabilities. Furthermore, many experts believe AI development could reach a "singularity"—a point of explosive growth where human-level capabilities are achieved and quickly surpassed, potentially leading to profound economic disruption (Bostrom, 2014; Shanahan, 2015).

This paper studies how AI stocks are priced in light of the possibility that AI advancement could significantly harm the representative investor's livelihood and consumption. We develop a simple disaster risk model where AI stocks serve as hedges against negative AI singularity scenarios. The resulting valuations can be high even if AI dividends do not benefit from disasters, because these assets offer relative protection in states where consumption collapses.

We are not claiming that a negative singularity will occur. Predicting whether and when such an event might happen is beyond the scope of this paper. However, given the potentially catastrophic consequences, it is prudent to consider how financial markets might price this tail risk. Similarly, we are not asserting that this hedging value is already fully reflected in current AI stock prices. Our model simply illustrates a mechanism through which disaster risk could contribute to AI asset valuations, complementing the more common growth-based explanations.

Our paper builds on two key strands of literature. First, we draw on the rare disaster framework developed by Rietz (1988) and extended by Barro (2006), Gabaix (2012), and Wachter (2013). This literature shows how small probabilities of extreme negative events can significantly impact asset prices. Second, we connect to research on technological change and labor markets by Zhang (2019) and Knesl (2023), who document how firms with automation

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

potential have distinct risk profiles. Our approach differs by focusing specifically on an AI singularity scenario and its implications for AI asset valuations. We also relate to discussions of AI existential risk by Bengio et al. (2024) and economic analyses of AI growth versus risk tradeoffs by Jones (2024) and Korinek and Suh (2024).

# 1 Model

We develop an intentionally simple model to illustrate the core mechanism. The economy consists of 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 is the marginal investor whose consumption determines asset prices.

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

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

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

The representative household's gross consumption growth follows a simple binary process. In normal times, consumption growth is 1 (no growth for simplicity), while in disaster states, consumption falls by a factor $e^{-b}$ where $b > 0$. Formally:

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

These disasters represent revolutionary improvements in AI that are devastating for the representative household. While such advances benefit AI owners, they harm the representative household through disruption of labor income, way of life, and sense of meaning. At time $t = 0$, no disasters have occurred yet (the singularity has not happened), but multiple disasters may occur over time, capturing ongoing uncertainty following a potential singularity.

We consider a publicly traded AI asset with dividend $D_t$ that initially represents a small fraction of aggregate consumption. Specifically, before any disaster:

$$D_t = \delta C_t$$

where $\delta > 0$ is small. Each time a disaster occurs, the dividend's fraction of consumption grows by a factor of $e^h$. After $n$ disasters, the dividend is:

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

This formulation allows us to capture a range of scenarios, including a worst case where $h < b$, meaning the dividend may actually shrink in absolute terms during each disaster, despite growing as a fraction of consumption. This reflects a scenario where AI improvements are concentrated in privately-held AI assets rather than publicly traded ones.

# 2 Results

Having established our model, we now derive the price-dividend ratio for the AI asset. The standard consumption-based asset pricing equation gives us the price of the asset after $n$ disasters as:

$$P_n = E_n[M_{t+1}(D_{n+1} + P_{n+1})]$$

where $M_{t+1}$ is the stochastic discount factor equal to $(C_{t+1}/C_t)^{-\gamma}$. Given our binary consumption process, the SDF equals 1 with probability $1 - p$ and $e^{\gamma b}$ with probability $p$.

Let $x_n = P_n/D_n$ denote the price-dividend ratio after $n$ disasters. In states where no new disaster occurs, dividends remain unchanged proportionally, and the SDF equals 1. When a disaster occurs, dividends become $\delta e^{(n+1)h} C_t$, and the SDF equals $e^{\gamma b}$.

Substituting these values into the asset pricing equation and dividing by $D_n$, we obtain:

$$x_n = (1 - p)(1 + x_n) + pe^{\gamma b} e^h (1 + x_{n+1})$$

Rearranging terms:

$$x_n = \frac{(1 - p)/p + e^{\gamma b + h} + e^{\gamma b + h} x_{n+1}}{1}$$

Since the probabilistic structure remains constant over time, we can look for a stationary solution where $x_n = x_{n+1} = x$. This yields:

$$x = \frac{(1 - p)/p + e^{\gamma b + h}}{1 - e^{\gamma b + h}}$$

Simplifying further:

$$x = \frac{(1 - p) + pe^{\gamma b + h}}{p(1 - e^{\gamma b + h})}$$

For this ratio to be positive and finite, we require $e^{\gamma b + h} < 1$, or equivalently, $\gamma b + h < 0$.

This condition ensures the denominator is positive, resulting in a positive price-dividend ratio.

If we introduce a time discount factor $\beta < 1$ to the model, the price-dividend ratio becomes:

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

To illustrate how the price-dividend ratio varies with model parameters, we compute values for different combinations of disaster severity $b$ and probability $p$, while fixing $\beta = 0.96$, $\gamma = 2$, and $h = 0.20$. The condition for a finite price-dividend ratio becomes $\beta((1-p) + pe^{2b+0.20}) < 1$.

The table below presents the price-dividend ratios for various parameter combinations:

| | $p = 0.0001$ | $p = 0.0025$ | $p = 0.005$ | $p = 0.01$ | $p = 0.02$ |
|---|---|---|---|---|---|
| $b = 0.40$ | 24 | 27 | 30 | 41 | 142 |
| $b = 0.55$ | 24 | 29 | 35 | 70 | Inf |
| $b = 0.70$ | 24 | 32 | 47 | 407 | Inf |
| $b = 0.85$ | 24 | 36 | 76 | Inf | Inf |
| $b = 0.95$ | 24 | 41 | 177 | Inf | Inf |

Table 1: Price-dividend ratios for different disaster probabilities and severities with $\beta = 0.96$, $\gamma = 2$, and $h = 0.20$. "Inf" indicates the condition $\beta((1-p) + pe^{2b+0.20}) < 1$ is violated.

Several patterns emerge from this table. First, for a fixed disaster probability $p$, the price-dividend ratio increases with the disaster severity $b$. This occurs because more severe disasters make the AI asset more valuable as a hedge. Second, for a fixed disaster severity, the price-dividend ratio increases with the disaster probability. As disasters become more likely, the hedging value of the AI asset increases. Finally, for sufficiently high values of $b$ and $p$, the price-dividend ratio becomes infinite, indicating that the model "blows up" when the condition $\beta((1-p) + pe^{2b+0.20}) < 1$ is violated.

These results suggest that the high valuations of AI stocks could be rational if investors perceive them as hedges against catastrophic AI developments. The hedging value increases with both the perceived probability and severity of an AI singularity that would be harmful to the representative investor. This mechanism provides an alternative explanation to the common view that AI valuations are driven solely by expectations of future earnings growth.

# 3 Model Discussion

Our model is deliberately simple to highlight the core mechanism. However, several extensions and refinements are worth discussing.

Market incompleteness plays an implicit but crucial role in our analysis. The disaster magnitude $b$ represents the net effect of (1) the AI disaster and (2) the AI asset dividend. In a complete market setting, the representative household could purchase shares in all AI assets—including private AI companies—and not only fully hedge against the singularity but potentially benefit from it. In reality, most households cannot buy shares in many cutting-edge AI labs such as OpenAI, Anthropic, xAI, or DeepSeek. This market incompleteness limits the representative household's ability to hedge against AI disaster risk, potentially explaining why publicly traded AI stocks might command a premium.

A more elaborate model could explicitly characterize AI owners, their incentives, and their interactions with the representative household. Such a model could explore how AI owners' profit motives might increase the likelihood of a negative singularity, or how their decisions affect the severity of disasters. However, developing such a model would primarily decorate speculations with mathematics while making the analysis considerably more complex. The fundamental economic mechanism—that AI assets may serve as hedges against AI disaster risk—would remain unchanged.

We have also simplified the consumption process to a binary outcome and assumed that all disasters are identical. A more realistic model might include varying disaster magnitudes or multi-period disasters with potential recoveries, as in Nakamura et al. (2013). While these extensions would add realism, they would complicate the analysis without changing the core insight.

By keeping the model streamlined, we leave room for the human-written perspective in Appendix A while still illustrating the essential economic mechanism. The model's simplicity should not be mistaken for naivety—rather, it allows us to isolate and understand a specific channel through which AI disaster risk might affect asset prices.

# 4 Conclusion and Implications

We have presented a simple asset pricing model demonstrating how AI stocks could serve as hedges against a negative AI singularity. This hedging mechanism may contribute to high AI stock valuations even if these companies do not benefit from AI disasters in absolute terms. The key insight is that AI assets grow as a share of consumption during disasters, providing relative protection to investors when other sources of consumption are severely impacted.

This hedging perspective highlights a potential financial market solution to AI disaster risk that has received limited attention in the literature on AI governance. Much of the current policy discussion focuses on regulatory approaches or redistribution mechanisms like universal basic income (Korinek and Stiglitz, 2018). While these approaches have merit, they often overlook the role financial markets can play in addressing AI risk.

Financial markets could help mitigate AI disaster risk by allowing individuals to acquire hedges against potential job displacement and economic disruption. However, as our model illustrates, the effectiveness of this approach is constrained by market incompleteness. Many cutting-edge AI systems are being developed by private companies with restricted ownership structures, limiting the public's ability to hedge against AI risk through stock ownership.

The challenge, then, is to design financial instruments that enable broader hedging against AI disaster risk while preserving innovation incentives. This might involve novel security designs, expanded public ownership of AI companies, or financial regulations that encourage the creation of AI-risk hedging products. Expanding these opportunities would allow more individuals to protect themselves against potential negative consequences of rapid AI development.

Our model suggests that markets may already be pricing in some of this hedging value for publicly traded AI companies. As AI continues to advance, understanding the relationship between AI disaster risk and asset prices will become increasingly important for investors, policymakers, and the broader public. By explicitly modeling this relationship, we hope to contribute to a more nuanced understanding of AI valuations and highlight the potential role of financial markets in addressing AI risk.

# 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.

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

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

Mercer, Sarah et al. (Feb. 2025). “Brief analysis of DeepSeek R1 and its implications for Generative AI”. In: *arXiv*. URL: `https://arxiv.org/abs/2502.02523`.

Nakamura, Emi, Jón Steinsson, Robert Barro, and José Ursúa (2013). “Crises and Recoveries in an Empirical Model of Consumption Disasters”. In: *American Economic Journal: Macroeconomics* 5.3, pp. 35–74. DOI: `10.1257/mac.5.3.35`.

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

Shanahan, Murray (2015). *The Technological Singularity*. MIT 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.

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.
```