

# Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025\*

## Abstract

This paper explores a novel perspective on high AI stock valuations: they might serve as hedges against a potential negative AI singularity that could be devastating for the representative investor. We present a stylized asset pricing model where, even with a small probability of disaster, AI stocks command high valuations if they perform relatively better than the overall economy during such events. This creates a hedging value for these assets, complementing traditional growth-based explanations. 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. DeepSeek’s R1 model demonstrates reasoning capabilities comparable to OpenAI’s leading systems despite export restrictions, and Waymo autonomous vehicles now operate without human supervision in multiple cities. These achievements represent just the visible edge of a technological revolution. Many investors worry about their livelihoods as AI systems increasingly match or exceed human performance across domains previously thought safe from automation (Acemoglu and Restrepo, 2020; Knesl, 2023; Zhang, 2019).

While technological change has disrupted labor markets throughout history, AI differs fundamentally from previous innovations. There is no product or service that AI could not, in principle, create. The current paper illustrates this point vividly, as it is entirely written by AI systems using a few short prompts.<sup>1</sup> This content, along with the code used to generate it, is available at <https://github.com/chenandrewy/Prompts-to-Paper/>. Unlike the internet revolution, which required humans to create and maintain websites, applications, and infrastructure, AI progress may ultimately lead to systems that design, improve, and deploy themselves with minimal human input. AI advancement could also prove incredibly sudden—a phenomenon often called the “AI singularity.”

This paper studies how AI stocks might be priced in light of the risk that advanced AI could significantly disrupt livelihoods and consumption patterns. We introduce a stylized model where a representative household faces the possibility of an AI singularity that would devastate their consumption. Within this framework, publicly traded AI companies command high valuations not necessarily because of anticipated earnings growth, but because they serve as hedges against this potential disaster. When a negative singularity occurs, these companies’ relative share of aggregate consumption increases, making them valuable insurance instruments despite potentially shrinking in absolute terms.

We are not claiming that a negative singularity will definitely occur. Distinguished AI researchers disagree substantially on both the likelihood and timeline of transformative AI outcomes (Bengio et al., 2024). Nevertheless, considering this scenario is important both for understanding asset prices and for developing policy responses to AI risks. Similarly, we are not asserting that this hedging value is already fully priced into current AI stock valuations. Our model simply illustrates a plausible mechanism through which disaster risk could contribute to these valuations.

Our paper builds on two established literatures in asset pricing. First, we extend the

---

<sup>1</sup>“We” in this paper refers to one human author working with multiple LLMs. For a purely human perspective, see [Appendix A](#).

rare disaster risk framework developed by Rietz (1988) and Barro (2006), where the possibility of economic catastrophes significantly impacts asset prices. As shown by Gabaix (2012) and Wachter (2013), time-varying disaster risk can explain key asset pricing puzzles including high equity premia and excess volatility. Second, we connect to the literature on hedging labor income risk through portfolio choices. Benzoni, Collin-Dufresne, and Goldstein (2007) demonstrate that when labor income and dividends are cointegrated, optimal portfolio strategies change dramatically. Similarly, Santos and Veronesi (2006) show that the ratio of labor income to consumption affects expected stock returns and drives market predictability.

We apply these frameworks to the specific context of AI-related risks. Bengio et al. (2024) identify catastrophic risks from advanced AI systems, including potential large-scale social harms and loss of human control. This connects with emerging evidence from Babina et al. (2024) that firms' investments in AI technologies affect their systematic risk profiles. Zhang (2019) demonstrates that firms with automation opportunities maintain a replacement option that hedges their value against unfavorable macroeconomic shocks, a finding consistent with our model of AI stocks providing relative protection during a negative AI singularity. While Babina et al. (2024) show that AI-investing firms experience higher growth and market valuations, our model suggests an additional explanation: these valuations may partly reflect AI stocks' potential to serve as hedges against scenarios where AI advancement proves devastating for the representative household. Other notable works include Jones (2024), who explores the economic tradeoff between AI-driven growth and existential risks, and Korinek and Suh (2024), who analyze how output and wages respond to different AI development scenarios.

## 2 A Simple Model

We present a simple model that captures the essence of our argument while remaining tractable. Consider an economy with two types of agents: AI owners and a representative household. The AI owners are fully invested in private AI assets and are not marginal investors in public markets. The representative household, on the other hand, is the marginal investor in public markets and has constant relative risk aversion (CRRA) preferences over consumption. Since the representative household is the marginal investor, only their consumption matters for asset pricing.

The representative household's gross consumption growth follows a simple process. In normal times, consumption growth is 1 (no growth for simplicity). However, with some probability  $p$ , a "disaster" can occur, in which case consumption growth becomes  $e^{-b}$  where

$b > 0$ . This disaster represents a revolutionary improvement in AI that is devastating for the representative household. When such an improvement occurs, the benefits are captured primarily by the AI owners, while the representative household experiences significant losses in labor income, way of life, and meaning. At time  $t = 0$ , we assume that no disasters have yet occurred (the singularity has not yet happened). Multiple disasters may occur over time, capturing the ongoing uncertainty even after an initial singularity event.

There exists a publicly traded AI asset in this economy. Before any disaster occurs, this asset pays a dividend that represents a small fraction of aggregate consumption, denoted by  $D_t = \delta C_t$  where  $\delta$  is small. Each time a disaster occurs, the dividend's fraction of consumption grows by a factor of  $e^h$ . That is, after  $n$  disasters, the dividend is:

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

This formulation is meant to capture a scenario where publicly traded AI companies might benefit relative to the rest of the economy during an AI singularity, even in a worst-case scenario. Importantly, we allow for  $h < b$ , which means that the absolute dividend may actually shrink during each disaster, consistent with the idea that the most significant AI improvements might be concentrated in privately-held AI assets rather than publicly traded ones.

### 3 Asset Pricing Solution

We now solve for the price of the AI asset in this economy. Since the representative household is the marginal investor, we use their stochastic discount factor to price the asset. With CRRA preferences, the household's one-period stochastic discount factor from time  $t$  to  $t+1$  is:

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

where  $\beta \in (0, 1)$  is the subjective discount factor and  $\gamma > 0$  is the coefficient of relative risk aversion.

Given the i.i.d. nature of disasters in our model, the price-dividend ratio of the AI asset will be constant. Using the standard asset pricing approach, this ratio satisfies:

$$v = \frac{P_t}{D_t} = \frac{E(M_{t+1} g_{t+1}^d)}{1 - E(M_{t+1} g_{t+1}^d)}$$

where  $g_{t+1}^d = D_{t+1}/D_t$  is the gross growth rate of dividends.

To compute  $E(M_{t+1}g_{t+1}^d)$ , we consider the two possible states of the world. In the no-disaster state, which occurs with probability  $(1 - p)$ , consumption remains unchanged, so  $M_{t+1} = \beta$  and  $g_{t+1}^d = 1$ . In the disaster state, which occurs with probability  $p$ , consumption falls by a factor of  $e^{-b}$ , implying  $M_{t+1} = \beta e^{\gamma b}$ . Simultaneously, the dividend growth becomes  $g_{t+1}^d = e^{h-b}$ , reflecting both the economy-wide contraction and the relative gain in the AI asset's share of consumption.

Combining these outcomes, we get:

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

Let  $x = (1 - p) + pe^{(\gamma-1)b+h}$ . Then the price-dividend ratio is:

$$v = \frac{\beta x}{1 - \beta x}$$

This expression reveals how the AI asset's valuation depends on disaster risk. When  $\beta x < 1$ , the price-dividend ratio is finite and well-defined.

To illustrate the quantitative implications of our model, we set  $\beta = 0.96$ ,  $\gamma = 2$ , and  $h = 0.20$ . The latter parameter means that each time a disaster occurs, the AI asset's share of aggregate consumption increases by about 22%. Table 1 shows the price-dividend ratios for various combinations of disaster probability  $p$  and disaster magnitude  $b$ .

Table 1: Price-Dividend Ratios for the AI Asset

	Disaster probability ( $p$ )				
	0.0001	0.005	0.01	0.015	0.02
$b = 0.40$	24	27	30	35	39
$b = 0.55$	24	28	32	41	55
$b = 0.70$	24	29	37	55	76
$b = 0.85$	24	31	42	76	199
$b = 0.95$	24	32	52	124	Inf

The table demonstrates several key insights. First, even with a very low disaster probability of 0.0001, the AI asset commands a substantial price-dividend ratio of 24. Second, as either the disaster probability or the disaster magnitude increases, the price-dividend ratio rises significantly. For instance, when  $b = 0.85$  and  $p = 0.02$ , the ratio reaches 199, more than eight times the baseline value. Finally, for sufficiently severe disasters and probabilities (e.g.,  $b = 0.95$  and  $p = 0.02$ ), the price becomes infinite as  $\beta x$  approaches or exceeds 1.

These results suggest that even a small probability of a negative AI singularity can substantially boost AI asset valuations if these assets are expected to perform relatively

better than the overall economy during such an event. This hedging property could explain part of the high valuations we observe in AI stocks today, complementing the traditional growth-based explanation.

## 4 Model Discussion

Our model intentionally simplifies many aspects of the economy to highlight the key mechanism through which AI stocks might serve as hedges against a negative AI singularity. Several natural extensions deserve discussion, though we leave their formal analysis to future work.

An important implicit feature of our model is market incompleteness. This is embodied in the disaster magnitude parameter  $b$ , which represents the net effect of both (1) the AI disaster itself and (2) the AI asset dividend. If markets were complete, the representative household could buy shares in all AI assets, including privately held ones like OpenAI, Anthropic, xAI, and DeepSeek. In such a scenario, households could not only fully hedge against the negative effects of an AI singularity but potentially benefit from it. In reality, most households cannot invest in cutting-edge AI labs, which creates the wedge that drives our results.

A more elaborate model would explicitly represent the AI owners, their incentives, and their interaction with the representative household. Such a model might explore how AI owners' profit motives could lead to a negative singularity, perhaps through competition that sacrifices safety for capability gains. However, this additional complexity would largely decorate speculations with mathematics rather than deliver fundamental new insights. The resulting model would be costlier to analyze and to read, while the core economics would remain essentially unchanged. Moreover, our streamlined approach leaves room for the human-written perspective in Appendix A.

Another potential extension would be to incorporate time-varying disaster probabilities, as in Wachter (2013). This would generate additional asset price volatility and possibly explain some of the boom-bust cycles we observe in AI stocks. Similarly, one could extend the model to include multiple assets with different exposures to AI disaster risk, leading to cross-sectional predictions about expected returns.

Despite these possible extensions, our simple model captures the core insight: AI stocks may command high valuations partly because they serve as hedges against scenarios where AI advancement severely disrupts the broader economy.

## 5 Policy Implications and Conclusion

Our analysis suggests an intriguing policy implication: financial markets might offer partial solutions to AI disaster risk that complement more commonly discussed approaches like universal basic income. Just as insurance markets allow individuals to hedge against personal disasters, broader financial market innovations could help society hedge against technological disruption.

The key economic insight is that the effectiveness of this hedge is limited by market incompleteness. When significant portions of AI development occur in private companies with restricted ownership, the representative household cannot fully diversify away the risk of income and consumption loss from AI advances. This suggests potential benefits from policy initiatives that broaden ownership of AI companies or create derivative securities tied to AI development milestones.

Financial solutions to AI risk have received relatively little attention compared to regulatory approaches (Bengio et al., 2024; Russell, 2019). Yet the history of capitalism shows that when property rights are well-defined and markets function properly, they can be powerful tools for distributing risk and aligning incentives. A more deliberate approach to structuring financial claims on AI development could help ensure that the benefits of this technology are widely shared, even if its development leads to substantial economic disruption.

In conclusion, our paper offers a new perspective on the high valuations of AI stocks. Beyond the standard growth narrative, these valuations may reflect the assets' role as hedges against potential negative AI singularity events. While simple, our model demonstrates that even small probabilities of disaster can generate substantial pricing effects when the assets provide relative protection during these events. This mechanism could be part of the explanation for the remarkable valuations we observe in AI-related companies, even when their current profits seem disproportionately small relative to their market capitalization.

The rapid advancement of artificial intelligence presents both tremendous opportunities and significant risks. By better understanding how these risks may already be priced into financial markets, we gain insights into both market behavior and potential policy responses. Indeed, the financial hedging mechanisms highlighted in this paper may prove to be important complements to regulatory approaches in managing the transition to an AI-influenced economy.

## References

- Acemoglu, Daron and Pascual Restrepo (2020). “Robots and Jobs: Evidence from US Labor Markets”. In: *Journal of Political Economy*.
- Babina, Tania, Anastassia Fedyk, Alex 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>.
- Benzoni, Luca, Pierre Collin-Dufresne, and Robert S. Goldstein (2007). “Portfolio Choice over the Life-Cycle when the Stock and Labor Markets Are Cointegrated”. In: *Journal of Finance* 62.5, pp. 2123–2167.
- 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*.
- 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.
- 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 many 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  $e^{\lambda h}$
  - 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.