

# Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025\*

## Abstract

This paper proposes that high valuations of AI companies may partly reflect their hedging value against a negative AI singularity. We develop a simple disaster risk model where AI advancement creates potential consumption disasters for the representative household. In this setting, AI assets can attain high valuations even without generating extraordinary growth, because they provide insurance against technological catastrophes that harm most economic participants. The model generates price-dividend ratios consistent with observed market valuations under plausible parameter values. 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. In early 2025, DeepSeek released its R1 model, approaching the reasoning capabilities of OpenAI’s offerings despite having significantly fewer resources and facing export restrictions (DeepSeek-AI et al., 2025). Meanwhile, Waymo’s autonomous vehicles have logged millions of miles in complex urban environments with minimal human intervention. These developments are accompanied by growing concerns about AI’s potential to displace human labor across sectors once thought immune to automation (Zhang, 2019; Acemoglu and Restrepo, 2020; Karabarbounis and Neiman, 2014). Even knowledge workers with advanced degrees now face meaningful substitution risk from increasingly capable AI systems.

Unlike previous technological revolutions, AI represents something fundamentally different. There is no product or service that AI could not, in principle, create or deliver. This paper itself demonstrates this point—it was entirely written by AI, using six carefully crafted prompts.<sup>1</sup> You can examine the exact prompts and generation process at <https://github.com/chenandrewy/Prompts-to-Paper/>. This differs markedly from previous technological shifts like the internet revolution, which primarily transformed information flow and communication but still required human expertise across most domains. Moreover, AI progress may manifest in sudden, discontinuous jumps—a phenomenon some have termed the “technological singularity” (Vinge, 1993; Kurzweil, 2005).

This paper studies how AI stocks might be priced given the possibility that advanced AI could devastate livelihoods and consumption for the representative household. We show that even if AI companies never deliver extraordinary dividend growth, their stock prices could command high valuations simply because they provide a hedge against negative AI singularity events. In a world where AI advancements might simultaneously boost AI company profits while reducing aggregate welfare, owning AI stocks serves as financial insurance.

We are not claiming that a negative AI singularity will happen. Experts disagree substantially about this possibility, and we take no position on its likelihood. Nevertheless, exploring this scenario is important for understanding current asset valuations and potential financial market responses to AI risk. Similarly, we are not asserting that this hedging value is already priced into current AI stock valuations—our model merely illustrates a possible mechanism that could contribute to their high prices.

Our approach builds on the disaster risk literature pioneered by Rietz (1988) and expanded by Barro (2006), who demonstrated that rare catastrophic events can explain asset pricing puzzles. Gabaix (2012) and Wachter (2013) extended this framework by allowing

---

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

disaster probability to vary over time. We apply these insights to technological disasters, specifically the possibility of a negative AI singularity—a concept examined by Vinge (1993) and discussed extensively by Kurzweil (2005).

Recent empirical work has established connections between artificial intelligence and economic risks. Zhang (2019) and Knesl (2023) show that firms with automation potential have distinct risk profiles, with firms exposed to routine-task labor having lower expected returns due to their embedded option to replace workers with technology. Acemoglu and Restrepo (2020) provide evidence that robotics reduces employment and wages in local labor markets. On potential extreme risks, Bengio et al. (2024) articulate how advanced AI systems might pose catastrophic risks through loss of human control, while Bostrom (2014) explores how superintelligent AI could create existential challenges if misaligned with human values.

Our contribution is to consider AI assets as potential hedges against technological disasters, rather than merely valuing them based on expected growth. This perspective offers a novel explanation for their high valuations. Our approach complements recent economic analyses of AI transition scenarios by Korinek and Suh (2024) and the growth-risk tradeoff highlighted by Jones (2024).

## 2 Model

We propose a simple model to formalize our main argument. The model captures the essential tension between AI innovation benefits and their potential negative consequences for the representative household.

Consider an economy with two types of agents: AI owners and a representative household. AI owners are fully invested in AI assets and are not marginal investors in the stock market. The representative household, on the other hand, is the marginal investor whose consumption-based preferences determine asset prices. Following the standard approach in asset pricing, we assume the representative household has constant relative risk aversion (CRRA) preferences:

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

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

The representative household’s gross consumption growth follows a simple disaster process. In normal times, consumption growth is normalized to 1, but with some probability  $p$ , consumption experiences a disaster and becomes  $e^{-b}$ , where  $b > 0$ . These disasters represent revolutionary improvements in AI that, while technologically impressive, are devastating for

the representative household. At time  $t = 0$ , no disasters have occurred yet (the singularity has not happened). Multiple disasters may occur over time, capturing the ongoing uncertainty if a singularity occurs.

Crucially, the benefits of AI improvements are captured predominantly by AI owners, while the representative household experiences negative consequences such as loss of labor income, disruption to their way of life, and diminished sense of meaning or purpose.

We consider a publicly traded AI asset with dividend  $D_t$  that represents a small fraction of consumption before any singularity:

$$D_t = \delta C_t$$

where  $\delta$  is a small positive constant. Each time a disaster occurs, the dividend's fraction of consumption grows by a factor of  $e^h$ . This reflects the increased value captured by AI assets during technological disruptions. The parameter  $h$  can be positive or negative, allowing us to model scenarios where dividends may actually shrink in absolute terms during disasters.

### 3 Price-Dividend Ratio

We now solve for the price-dividend ratio of the AI asset at time 0, before any singularity has occurred. To derive this ratio, we use the representative household's stochastic discount factor under CRRA preferences in a discrete-time, i.i.d. setting.

Let us denote the subjective discount factor as  $\beta \in (0, 1)$ . In each period, a disaster (singularity) occurs with probability  $p$ , independent across time. When a disaster occurs, consumption shrinks by a factor  $e^{-b}$ . The representative household's stochastic discount factor  $M_{t,t+1}$  takes two values:

$$M_{t,t+1} = \begin{cases} \beta & \text{with probability } 1 - p \\ \beta e^{\gamma b} & \text{with probability } p \end{cases}$$

Similarly, the one-period dividend growth ratio  $D_{t+1}/D_t$  is:

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

In a stationary environment, the price-dividend ratio  $\phi \equiv P_t/D_t$  is constant across time. Using the fundamental asset pricing equation and our assumptions, we can derive:

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

This expression reveals how the price-dividend ratio depends on the disaster probability  $p$ , the risk aversion  $\gamma$ , the consumption impact  $b$ , and the dividend growth factor  $h$ .

To illustrate the implications of our model, we set  $h = 0.20$ ,  $\gamma = 2$ , and  $\beta = 0.96$ , while varying both  $b$  and  $p$ . The following table shows the price-dividend ratios for different combinations of disaster magnitude  $b$  and probability  $p$ :

	$b$				
$p$	0.40	0.55	0.70	0.85	0.95
0.0001	24	24	24	24	24
0.005	27	28	29	31	33
0.01	30	32	37	42	52
0.015	35	39	55	76	120
0.02	39	55	82	199	Inf

The table reveals several important patterns. When the disaster probability is very small ( $p = 0.0001$ ), the price-dividend ratio remains stable at around 24 regardless of the disaster magnitude. However, as the probability increases, the ratio becomes increasingly sensitive to the disaster magnitude. For instance, with  $p = 0.02$  and  $b = 0.95$ , the price-dividend ratio becomes infinite, indicating that the asset becomes extremely valuable as a hedge against severe AI-related disasters.

This result illustrates our main argument: AI assets can command high valuations not just because of their growth potential, but also because they serve as a hedge against negative AI singularity events. When the representative household faces a higher probability of experiencing consumption disasters due to AI developments, AI assets become more valuable as they provide a form of insurance against these adverse outcomes.

## 4 Model Discussion

Our model deliberately abstracts from many realistic features to focus on the core mechanism. One key simplification is market incompleteness, which is implicit but important in our framework. The disaster magnitude  $b$  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 private ones), and not only fully hedge but actually benefit from the singularity. In reality, most households cannot buy shares in many

cutting-edge labs such as OpenAI, Anthropic, xAI, or DeepSeek. This market incompleteness limits the household’s ability to hedge against AI risks.

A more elaborate model would explicitly represent AI owners, their incentives, and their interaction with the representative household. Such a model might explore how AI owners’ incentives could lead to a negative singularity. For instance, competition among AI labs might create a race to develop increasingly powerful systems without adequate safety precautions. However, decorating these speculations with mathematical formalism would add complexity without necessarily improving insight. The core economics—that AI assets can serve as hedges against technological disasters—would remain the same.

Our model also simplifies by assuming that AI disasters affect consumption directly. A more realistic approach might model how AI progress affects labor income first, which then impacts consumption. However, this additional layer would complicate the model without changing the fundamental insight about the hedging value of AI assets.

The pricing formula reveals an interesting feature: AI assets can have high price-dividend ratios even if their dividends never grow in absolute terms. If  $h < b$ , dividends actually shrink during disasters, yet the asset’s price can still be high due to its hedging properties. This occurs because the stochastic discount factor increases dramatically during disasters (by  $e^{\gamma b}$ ), making payoffs in those states particularly valuable.

The simplicity of our approach allows room for the human-written Appendix A, which provides additional perspective on the model and its implications.

## 5 Conclusion and Implications

This paper has presented a simple model showing how AI assets can be valued not just for their growth potential but also for their hedging properties against negative AI singularity events. The model generates high price-dividend ratios for AI assets under plausible parameter values, even when the probability of catastrophic AI events is small. This offers a novel perspective on the current high valuations of AI companies, suggesting they may partly reflect the market’s pricing of technological disaster risk.

Financial markets may thus provide a natural mechanism for addressing AI catastrophe risk. By investing in AI assets, households can partially hedge against negative AI outcomes—a sort of technological disaster insurance. This hedging opportunity represents an alternative to policy proposals like universal basic income (UBI) for addressing AI displacement concerns. However, the effectiveness of this hedge is inherently limited by market incompleteness. Most households cannot invest in privately held AI enterprises, which may capture the largest share of benefits from advanced AI.

The idea of financial market solutions to AI risk has received surprisingly little attention in discussions about AI safety and ethics. Most discussions focus on regulatory approaches (Bengio et al., 2024), technical safeguards (Bostrom, 2014), or social policies like UBI. Yet financial markets already provide mechanisms for risk sharing that could be enhanced to address AI transition risks. Creating more accessible investment vehicles that track AI progress could help spread the benefits of AI more broadly while simultaneously creating incentives for safer development.

Our analysis suggests that researchers and policymakers should consider financial market mechanisms alongside other approaches when addressing the challenges of AI transition. By enabling broader participation in AI’s upside, financial innovation could complement other policy responses and help ensure that technological progress benefits society broadly rather than concentrating its gains among a small set of owners. Of course, no single approach—whether through markets, regulation, or social policy—will fully address the complex challenges posed by transformative AI, but a portfolio of responses will likely be needed as we navigate this uncertain technological frontier.

## References

- Acemoglu, Daron and Pascual Restrepo (2020). “Robots and Jobs: Evidence from US Labor Markets”. In: *Journal of Political Economy*.
- 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>.
- Karabarbounis, Loukas and Brent Neiman (2014). “The Global Decline of the Labor Share”. In: *Quarterly Journal of Economics*.
- 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.
- Kurzweil, Ray (2005). *The Singularity Is Near: When Humans Transcend Biology*. Viking Press.
- Rietz, Thomas (1988). “The Equity Risk Premium: A Solution?” In: *Journal of Monetary Economics*.
- 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 writing quality 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 a satisfying economic model on their 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  $e^h$
  - 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.