# Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

**Abstract**

We present a simple asset pricing framework where AI stocks may command high valuations partly because they hedge against a negative AI singularity—a scenario where explosive AI development devastates the representative investor's consumption. Using a consumption-based model with rare disasters, we demonstrate how AI assets that increase their share of consumption during disasters can have high price-dividend ratios even without dividend growth. This mechanism offers financial market solutions to AI-driven economic disruption, complementing policy proposals like universal basic income. 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, achieving reasoning capabilities comparable to OpenAI's o1 through pure reinforcement learning (DeepSeek-AI et al., 2025). Meanwhile, Waymo's autonomous vehicles have logged millions of driverless miles across multiple cities. These achievements represent merely the visible edge of a technological revolution that many believe will fundamentally reshape labor markets and economic structures. As AI capabilities expand, investors may naturally worry about their wages being displaced, creating substantial labor income risk (Zhang, 2019; Knesl, 2023; Betermier et al., 2012).

While technological change has transformed economies throughout history, AI differs fundamentally from previous revolutions. Unlike the printing press, steam engine, or even the internet, there is virtually no product or service that AI could not, in principle, create. This paper itself exemplifies this distinction—it is entirely written by AI, using a few short prompts from a human researcher.[1] The complete prompts and generation process are available at `https://github.com/chenandrewy/Prompts-to-Paper/`. Unlike the internet revolution, which primarily transformed information access and communication, AI progress may eventually touch every sector of the economy simultaneously. Moreover, this progress could arrive with shocking suddenness—what researchers call the "AI singularity," a potential explosion in capabilities that fundamentally alters our economic and social landscape.

This paper studies how AI stocks might be priced given the risk that artificial intelligence will significantly disrupt livelihoods and consumption. We develop a simple consumption-based asset pricing model with disaster risk, where disasters represent negative AI singularity events. In our framework, AI assets increase their share of consumption during these disasters, making them valuable hedges against catastrophic AI outcomes. This hedging property can generate high valuations for AI stocks, even when their dividends never grow in absolute terms—a perspective that contrasts with the common view that AI valuations are driven primarily by expectations of future earnings growth.

We are not claiming that a negative singularity will definitely occur. Many AI researchers and ethicists hold divergent views on the probability and nature of such scenarios (Bengio et al., 2024; Russell, 2019). Nevertheless, it is important to consider how financial markets might price this tail risk, especially as AI capabilities continue their rapid advancement. Even a small probability of a significant consumption disaster can have outsized effects on asset prices, as demonstrated by the rare disaster literature (Barro, 2006; Gabaix, 2012; Wachter, 2013).

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

We are also not claiming that this hedging value is already fully priced into current AI stock valuations. Our model simply illustrates a possible mechanism that could contribute to high valuations alongside growth expectations. The goal is to highlight how conventional asset pricing frameworks can accommodate the unique risks posed by transformative AI.

Our work connects several strands of literature. Recent papers establish important links between technological innovation and asset prices, with Zhang (2019) and Knesl (2023) showing that firms with automation potential can serve as hedges against labor displacement risk. Meanwhile, the rare disaster framework pioneered by Rietz (1988) and extended by Barro (2006) and Gabaix (2012) has successfully explained many asset pricing puzzles by incorporating low-probability catastrophic events. We bridge these approaches to model AI stocks as potential hedges against a negative singularity—a concept discussed by Bostrom (2014) and explored in economic terms by Jones (2024) and Korinek and Suh (2024).

Our paper extends the multiple-tree framework of Cochrane, Longstaff, and Santa-Clara (2008) to include a specific disaster process where AI improvements may benefit AI stockholders while harming the representative household. This approach is conceptually related to Santos and Veronesi (2006), who model an economy with both financial income and labor income, showing how the ratio between them affects asset prices. Our contribution is to explicitly model how AI disasters could shift consumption shares toward AI assets, potentially explaining their high valuations even with modest disaster probabilities.

# 2 Model

We now describe a simple model that captures the essence of our argument. Consider an economy with two types of agents: AI owners and a representative household. The AI owners are fully invested in AI companies and are not marginal investors in the stock market. The representative household, on the other hand, is the marginal investor whose consumption-based pricing kernel determines asset prices.

The representative household has constant relative risk aversion (CRRA) preferences over consumption, with utility function

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

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

The representative household's gross consumption growth follows a simple disaster process. In normal times, consumption grows at a rate of 1 (no growth for simplicity). However, the economy faces the possibility of disasters, which we interpret as revolutionary improvements in AI that are devastating for the household. When a disaster occurs, consumption drops by a factor of $e^{-b}$ where $b > 0$. Formally, gross consumption growth is given by

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

These disasters represent a singularity scenario where AI improvements primarily benefit AI owners while harming the representative household through displacement of labor, disruption of way of life, or loss of meaning. We assume that at $t = 0$, no disasters have yet occurred. Multiple disasters may happen over time, capturing ongoing uncertainty about the evolution of AI capabilities.

We consider a publicly traded AI asset with dividend $D_t$. Prior to any disaster, the dividend represents a small fraction of aggregate consumption. Each time a disaster occurs, the dividend's share of consumption grows by a factor of $e^h$. Specifically, if we denote by $n_t$ the number of disasters that have occurred by time $t$, then

$$\frac{D_t}{C_t} = d_0 e^{hn_t}$$

where $d_0 > 0$ is the initial dividend-consumption ratio and $h$ is a parameter that governs how the dividend share changes during disasters.

This specification is meant to capture a worst-case scenario for the representative household. Even if $h > 0$, the dividend may actually shrink in absolute terms during disasters if $h < b$. This reflects the possibility that benefits from AI improvements are concentrated in privately-held AI assets, with publicly traded AI companies capturing only a portion of the value created by advances in artificial intelligence, similar to the framework in Barro (2006) and Gabaix (2012).

# 3 Asset Pricing Implications

We now derive the price-dividend ratio for the AI asset in our model. Under the standard consumption-based asset pricing approach, the price of the asset at time 0 is the expected present value of all future dividends:

$$P_0 = E_0 \left[ \sum_{k=1}^{\infty} M_1 M_2 \cdots M_k D_k \right]$$

where $M_t$ is the stochastic discount factor from period $t-1$ to $t$. With CRRA preferences and no separate time discounting (for simplicity, we set $\beta = 1$), the stochastic discount factor is given by $M_{t+1} = (C_{t+1}/C_t)^{-\gamma}$.

In our model, the stochastic discount factor takes on two possible values. If no disaster occurs between periods $t$ and $t+1$, then $C_{t+1}/C_t = 1$ and $M_{t+1} = 1$. If a disaster occurs, then $C_{t+1}/C_t = e^{-b}$ and $M_{t+1} = e^{b\gamma}$.

To compute the price-dividend ratio, we first express the cumulative discount factor and dividend growth in terms of the number of disasters. If $n_k$ disasters have occurred by time $k$, then:

$$M_1 M_2 \cdots M_k = e^{b\gamma n_k}$$

and

$$\frac{D_k}{D_0} = e^{n_k(h-b)}$$

Therefore, the discounted dividend at time $k$ is:

$$M_1 M_2 \cdots M_k \frac{D_k}{D_0} = e^{n_k[h-b(1-\gamma)]}$$

Since disasters occur independently with probability $p$ in each period, the number of disasters $n_k$ follows a binomial distribution with parameters $k$ and $p$. Using the moment generating function of the binomial distribution, we have:

$$E_0\left[e^{n_k[h-b(1-\gamma)]}\right] = (1-p+pe^{h-b(1-\gamma)})^k$$

The price-dividend ratio at time 0 is thus:

$$X(0) \equiv \frac{P_0}{D_0} = \sum_{k=1}^{\infty}(1-p+pe^{h-b(1-\gamma)})^k$$

This is a geometric series that converges if $1-p+pe^{h-b(1-\gamma)} < 1$, which simplifies to $h < b(1-\gamma)$. This condition has a clear economic interpretation: each disaster must reduce marginal utility (due to the consumption drop) sufficiently more than it increases the AI dividend share for the household's valuation to remain finite.

When the convergence condition is satisfied, the price-dividend ratio has the closed-form solution:

$$X(0) = \frac{1-p+pe^{h-b(1-\gamma)}}{p[1-e^{h-b(1-\gamma)}]}$$

To provide more realistic numerical results, we now incorporate a time discount factor $\beta = 0.96$ and set the risk aversion parameter to $\gamma = 2$. The one-period pricing kernel becomes:

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

With this modification, the price-dividend ratio is:

$$X(0) = \frac{\beta[(1-p) + p \cdot e^{h-b(1-\gamma)}]}{1 - \beta[(1-p) + p \cdot e^{h-b(1-\gamma)}]}$$

The table below shows the price-dividend ratios for various values of the disaster magnitude $b$ and probability $p$, with $h = 0.20$:

| | $p = 0.0001$ | $p = 0.005$ | $p = 0.01$ | $p = 0.015$ | $p = 0.02$ |
|---|---|---|---|---|---|
| $b = 0.40$ | 24 | 27 | 30 | 35 | 39 |
| $b = 0.55$ | 24 | 28 | 33 | 41 | 55 |
| $b = 0.70$ | 24 | 29 | 37 | 52 | 76 |
| $b = 0.85$ | 24 | 30 | 42 | 76 | 199 |
| $b = 0.95$ | 24 | 32 | 52 | 124 | Inf |

Table 1: Price-dividend ratios for AI assets with $h = 0.20$, $\beta = 0.96$, and $\gamma = 2$

Several patterns emerge from these results. First, when the disaster probability is very small ($p = 0.0001$), the price-dividend ratio is approximately 24 regardless of the disaster magnitude $b$. This is because the term $p \cdot e^{h-b(1-\gamma)}$ becomes negligible, making the discount-growth factor nearly equal to $\beta = 0.96$ in all cases.

Second, as the disaster probability increases, the price-dividend ratio rises, particularly for larger values of $b$. This reflects the increasing value of the AI asset as a hedge against disasters. For instance, when $b = 0.85$, the price-dividend ratio increases from 24 to 199 as $p$ rises from 0.0001 to 0.02.

Finally, for sufficiently high values of $b$ and $p$, the price-dividend ratio can become infinite. This occurs when $\beta[(1-p) + p \cdot e^{h-b(1-\gamma)}] \geq 1$, as seen in the case of $b = 0.95$ and $p = 0.02$. Economically, this represents a situation where the AI asset becomes so valuable as a hedge that investors are willing to pay any price for it.

These results highlight how the possibility of AI-driven disasters can significantly impact the valuation of AI assets, even when the probability of such disasters is relatively small. The hedging value of these assets increases with both the probability and magnitude of potential disasters, potentially explaining part of the high valuations observed in AI stocks today.

# 4 Model Discussion

While our model is deliberately simple, it captures a key insight: AI assets can derive substantial value from their role as hedges against negative AI singularity scenarios, even if they never deliver extraordinary dividend growth. Several extensions and refinements could be made to the model, though we argue that they would mainly elaborate rather than fundamentally alter this central insight.

Market incompleteness is implicit but important in our analysis. The disaster magnitude $b$ represents the net effect of (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—and not only fully hedge but potentially benefit from the singularity. In reality, most households cannot invest in many cutting-edge AI labs, creating a fundamental market incompleteness that limits hedging opportunities.

The model could be extended to explicitly model AI owners, their incentives, and their interaction with the representative household. One might ask: How might AI owners' incentives lead to a negative singularity? Wouldn't competitive forces drive AI developers to create beneficial rather than harmful AI? These are important questions, but addressing them formally would require speculative assumptions about AI development trajectories, competitive dynamics, and potential coordination failures. Such an extension would decorate speculations with mathematics without necessarily improving our understanding of the core economic mechanism.

Similarly, we could introduce multiple assets with different exposures to AI risk, endogenize the probability of disasters, or incorporate learning dynamics. While these extensions might add realism, they would come at the cost of analytical tractability and clarity. The simple model presented here achieves its purpose: it illustrates how AI assets can derive significant value from their hedging properties, even absent extraordinary growth prospects.

A more elaborate model would also leave less room for the human-written Appendix A, which provides valuable perspective on the experiment of creating this paper through AI. Given our focus on the financial market implications of AI risk, we believe the current level of modeling detail strikes the right balance between rigor and accessibility.

# 5 Policy Implications and Conclusion

Our analysis highlights an important but underappreciated dimension of the AI revolution: financial markets may naturally develop hedging instruments against AI-driven economic disruption. As AI continues to advance, creating targeted financial products that provide explicit insurance against AI-driven job displacement could complement policy proposals like universal basic income (UBI).

These market-based solutions have three key advantages. First, they can be implemented without the extensive government infrastructure required for UBI. Second, they allow for more nuanced pricing of AI risk across different sectors and worker groups. Third, they create financial incentives for AI developers to minimize harm to human workers, as doing so would reduce the cost of hedging liabilities.

However, the effectiveness of such financial market solutions is fundamentally limited by the market incompleteness highlighted in our model. If households cannot invest in the full spectrum of AI assets—particularly the private companies at the cutting edge of AI development—then their hedging opportunities remain constrained. This limitation suggests a potential role for policy in broadening access to AI investment opportunities, perhaps through public-private partnerships or regulated investment vehicles that provide retail investors with exposure to private AI firms.

As Jones (2024) and Korinek and Suh (2024) have argued, there is a fundamental tension between promoting AI innovation and ensuring its benefits are widely shared. Financial market solutions offer one pathway to managing this tension, complementing rather than replacing other policy approaches. By allowing investors to hedge against AI risk, markets can provide valuable signals about where disruption is most likely to occur and how it is expected to unfold.

In conclusion, our simple asset pricing model illustrates how AI stocks might command high valuations partly because they hedge against a negative AI singularity. This perspective complements the common view that AI valuations reflect growth expectations, suggesting that investors may be willing to pay a premium for AI assets even if their future cash flows are uncertain. By understanding this hedging dimension, investors, policymakers, and AI developers can better navigate the complex landscape of risks and opportunities presented by advanced artificial intelligence.

# 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`.

Betermier, Sebastien, Thomas Jansson, Christine Parlour, and Johan Walden (2012). “Hedging Labor Income Risk”. In: *Journal of Financial Economics* 105.3, pp. 622–639.

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

Cochrane, John H., Francis A. Longstaff, and Pedro Santa-Clara (2008). “Two Trees”. In: *Review of Financial Studies*.

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

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

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

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

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

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

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 may
projects.

I'm not saying that I *expect* a disaster for the economics labor market.  But
it's definitely a scenario that economists should think about.
```

# B Prompts Used to Generate This Paper

Each prompt consists of context and instructions. The context consists of the responses to the previous prompts, and may include literature reviews (all AI generated). For writing tasks (using Claude 3.7 Sonnet), a system prompt is also included.

For further details, see `https://github.com/chenandrewy/Prompts-to-Paper/`.

The system prompt and instructions are listed below.

## System Prompt (model: claude-3-7-sonnet-20250219)

```
You are an asset pricing theorist who publishes in the top journals
   (Journal of Finance, Journal of Financial Economics, Review of
```

```
   Financial Studies). You think carefully with mathematics and
   check your work, step by step.

Your team is writing a paper with the following main argument: the
   high valuations of AI stocks could be in part because they hedge
   against a negative AI singularity (an explosion of AI development
    that is devastating for the representative investor). This
   contrasts with the common view that AI valuations are high due to
    future earnings growth. Since the AI singularity is inherently
   unpredictable, the paper is more qualitative than quantitative.
   The goal is to just make this point elegantly.

Write in prose. No headings and no bullet points. But do use display
    math to highlight key assumptions. Cite papers using Author (
   Year) format.

Be conversational yet rigorous. Favor plain english. Be direct and
   concise. Remove text that does not add value. Use topic sentences
   . The first sentence of each paragraph should convey the point of
    the paragraph.

Be modest. Do not overclaim.

Format the math nicely. Use we / our / us to refer to the writing
   team.
```

### Instruction: 01-model-prose (model: claude-3-7-sonnet-20250219)

```
Draft the model description. The model is purposefully simple and
   captures the essence of the main argument. Only describe the
   assumptions. No results or insights.
  - Two agents
    - AI owners: Fully invested in AI, not marginal investors in
       stocks
    - Representative household: Marginal investor, only their
       consumption matters, CRRA
  - Representative household's gross consumption growth
    - is either 1 or e\\^(-b) (disaster)
```

```
      - A disaster is a revolutionary improvement in AI that is
         devastating for the household
      - Benefits of AI improvement are captured by the AI owners
      - For the household, labor income, way of life, meaning is
         lost
      - At t=0, no disasters have happened (singularity has not
         occurred)
      - Multiple disasters may happen, capturing ongoing uncertainty
          if a singularity occurs
  - A publicly traded AI asset
    - Dividend is a small fraction of consumption before the
       singularity
    - Each time a disaster occurs, the dividend's fraction of
       consumption grows by a factor of e\\^h
    - Meant to capture a worst case scenario, where the dividend may
        actually shrink in each disaster
      - i.e. AI improvements are concentrated in privately-held AI
         assets
```

### Instruction: 02-result-notes (model: o1)

```
Find the price/dividend ratio of the AI asset at t = 0. Show the
   derivation, step by step.
```

### Instruction: 03-table-notes (model: o3-mini)

```
Make a table of the price/dividend for b from 0.40 to 0.95 and prob
   of disaster from 0.0001 to 0.02. Here, fix h = 0.20, CRRA = 2,
   time preference = 0.96. If the price is infinite, use "Inf".
   Round to the nearest whole number.
```

### Instruction: 04-resultandtable-prose (model: claude-3-7-sonnet-20250219)

```
Convert the notes in '02-result-notes' and '03-table-notes' into
   prose. The prose is intended to immediately follow '01-model-
   prose' and should flow naturally. Include the table.
```

### Instruction: 05-litreview-prose (model: claude-3-7-sonnet-20250219)

```
Write a short two paragraph lit review based on the "prose-response"
    and "lit-" context.

Be careful to avoid incorrect citations. Make sure the papers cited
   make the claims they are cited for.
```

### Instruction: 06-full-paper (model: claude-3-7-sonnet-20250219)

```
Write a paper titled "Hedging the AI Singularity" based on the "
   prose-response" context.

Title page:
- Title: "Hedging the AI Singularity"
- Abstract (less than 100 words)
  - Goal is to make a simple point
  - Secondary goal: bring attention to financial market solutions to
      AI disaster risk
  - At the end, say: unlike previous work, this short paper is
     generated by prompting LLMs.

The start of the Introduction is important. You need to bring the
   reader in, catch their eye, and establish credibility.

Start with background. Describe how AI progress is happening quickly
    (e.g. Deepseek R1, Waymo), and investors may be concerned about
   their wages being displaced (cite papers).

Then describe how technological change has occurred before, but AI
   is distinct because there is no product or service that AI could
   not, in principle create.  An example is the current paper, which
    is entirely written by AI, using a few short prompts. Provide a
   link to the github site, which is https://github.com/chenandrewy/
   Prompts-to-Paper/. This differs from say, the internet revolution
   . AI progress may also be incredibly sudden (the AI singularity).
    Include a footnote: "we" refers to one human author and multiple
    LLMs. For a purely human perspective see \\hyperref[app:readme
   ]\\{\\textcolor\\{blue\\}\\{Appendix \\ref\\{app:readme\\}\\}\\}.
```

```
Then describe what the paper does. It studies how AI stocks are
   priced, given that there is the risk that AI will destroy
   livelihoods and consumption.

Afterwards, the text should discuss:
- We are not saying a negative singularity will happen
  - But it is nevertheless important to consider this scenario
- We are also not saying that this hedging value is priced in
   already
  - Model illustrates a possible mechanism
- Related lit at end of Introduction
  - Cite papers in `05-litreview-prose`
  - Add Jones (2024) "AI Dilemma" and Korinek and Suh (2024) "
     Scenarios" if they're not already cited
- Model is the simplest possible to make the main argument
- Derivation of the key formulas
- High price/dividend ratios, even though dividends never grow
- A "Model Discussion" section that discusses natural model
   extensions and why they are not included
  - Market incompleteness is implicit but important
    - Implicit in the disaster magnitude 'b'
    - 'b' is the *net* effect of (1) AI disaster and (2) AI asset
       dividend
    - If markets were complete, representative household could buy
       shares in all AI assets (including private AI assets), and
       not only fully hedge but benefit from the singularity
    - In reality, most households cannot buy shares in many cutting
       edge labs (e.g. OpenAI, Anthropic, xAI, DeepSeek)
  - A more elaborate model would explicitly model the AI owners,
     their incentives, and interaction with the representative
     household
    - How might AI owners' incentives lead to a negative singularity
       ?
    - But wouldn't this just decorate speculations with math?
    - This would be costly to analyze, as well as to read
    - The core economics will remain the same
  - A short model analysis allows room for the human-written
     Appendix \\ref\\{app:readme\\}
```

```
- A "Policy Implications and Conclusion" section that discusses
   financial market solutions to AI disaster risk
  - These solutions are an alternative to UBI
    - Key economics: this hedge is limited by market incompleteness
  - These solutions to AI disaster risk are not discussed enough in
     the literature (cite papers)
  - Be very centrist (see below)

Text should avoid
- Being overly academic
- Politically-charged topics: sovereign wealth funds, industrial
   policy, redistribution, extolling free markets
- Overselling the model (it's just a simple illustration)
- Incorrect citations
  - Make sure papers cited make the claims they are cited for

Style Notes:
- Be conversational and direct, yet rigorous
- A touch of wit and wry humor are OK
- No bulleted lists
- No subsections (e.g. Section 1.2) though sections are OK (Section
   1)

Output a complete latex document, including preamble. Cite papers
   using \\cite, \\citep, \\citet. Use ‘template.tex‘ and keep the
   appendix that is already in the template.
```