Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

We explore a novel explanation for high AI stock valuations: they may provide a hedge against a negative AI singularity—an explosive AI development that devastates the representative investor. Using a parsimonious disaster-risk model, we show that even if AI asset dividends never grow in absolute terms, their resilience relative to aggregate consumption during disasters can generate high price-dividend ratios. This mechanism contrasts with the common view that valuations reflect future earnings growth. We discuss financial market solutions to AI risk alongside existing policy proposals. Unlike previous work, this short paper is generated by prompting LLMs.

Keywords: Artificial Intelligence, Disaster Risk, Asset Pricing

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

Artificial intelligence is advancing at a breathtaking pace. In January 2025, DeepSeek released R1, a reasoning model that rivals OpenAI's tools through pure reinforcement learning (DeepSeek-AI et al., 2025). Meanwhile, Waymo's autonomous vehicles have accumulated millions of miles without human supervision. These developments have prompted growing concerns about AI-driven labor displacement, with research suggesting that workers with routine-task jobs face declining returns as automation advances (Zhang, 2019; Knesl, 2023).

Unlike previous technological revolutions, AI presents a fundamentally different challenge. There is, in principle, no product or service that AI could not eventually create. This paper itself exemplifies this distinction—it was entirely written by AI, using six carefully crafted prompts. This differs markedly from earlier innovations like the internet, which primarily transformed information distribution rather than creation. Interested readers can find the complete set of prompts and generation process at `https://github.com/chenandrewy/Prompts-to-Paper/`. The distinction is not just that AI can automate existing tasks, but that it may advance explosively and unpredictably—the AI singularity hypothesis formalized by philosophers and computer scientists (Bostrom, 2014; Chalmers, 2010).[1]

This paper studies how AI stocks are priced given that there is risk that AI will destroy livelihoods and consumption. We analyze a simple disaster-risk asset pricing model where the representative household faces the possibility of "AI disasters"—revolutionary AI improvements that, while technologically impressive, devastate the representative household's consumption. In this setting, AI assets that maintain their value during such disasters command a premium, even if their dividends never grow in absolute terms.

We do not claim that a negative AI singularity will occur. Distinguished AI researchers disagree about the probability of such scenarios, with some viewing them as remote possibilities and others as significant risks (Bengio et al., 2024). Nevertheless, it is important to consider this scenario from a risk management perspective, just as financial markets price in small probabilities of other catastrophic events (Barro, 2006; Gabaix, 2012).

We also do not assert that this hedging value is already priced into AI stocks. Rather, our model illustrates a possible mechanism through which such considerations could affect valuations. Empirical work would be needed to determine whether and to what extent markets are pricing this risk.

Recent asset pricing literature has established the importance of technological disruption risk for security valuation. Zhang (2019) shows that firms with higher exposure to routine-task labor have lower expected returns because they maintain valuable replacement options during economic downturns. Similarly, Knesl (2023) finds that firms with labor vulnerable

[1] Throughout this paper, "we" refers to one human author and multiple LLMs. For a purely human perspective, see Appendix A.

to automation have negative exposure to technology shocks, commanding a risk premium. These empirical findings align with models of rare disasters in asset pricing, pioneered by Rietz (1988) and Barro (2006), which demonstrate how small probabilities of catastrophic events can significantly impact asset valuations. Gabaix (2012) and Wachter (2013) extend this framework to incorporate time-varying disaster risk, providing mechanisms that can explain both high equity premia and market volatility.

The possibility of an AI singularity—a scenario where artificial intelligence advances rapidly and unpredictably—represents a novel form of technological disaster risk. Bostrom (2014) explores how superintelligent AI could pose existential risks if not properly aligned with human interests. Jones (2024) examines the economic tension between AI-driven growth and potential catastrophic outcomes, highlighting the tradeoffs between technological progress and risk management. Korinek and Suh (2024) analyze how output and wages might respond under different AI development scenarios, including the transition to artificial general intelligence. Our paper builds on these insights by applying the rare disaster framework to AI development, positing that when disruption occurs, the benefits accrue primarily to AI owners while the representative household experiences welfare losses—a distributional effect consistent with concerns raised by Acemoglu and Restrepo (2020) regarding the impact of automation on labor markets.

1 The Model

We now describe our model, which is deliberately parsimonious yet captures the key economic mechanism we wish to highlight. We consider an economy with two types of agents: AI owners who are fully invested in AI assets and are not marginal in the stock market, and a representative household who is the marginal investor in stocks and whose consumption preferences determine asset prices.

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

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

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

The gross consumption growth of the representative household follows a simple disaster process. In normal times, consumption growth is normalized to 1. However, the economy faces the possibility of "AI disasters" - revolutionary improvements in AI that, while technologically impressive, are devastating for the representative household. When such a disaster occurs, consumption growth is e^{-b} where $b > 0$ represents the severity of the disaster. This formulation is similar to the rare disasters framework of Barro (2006), but with a specific interpretation related to AI development.

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

At time $t = 0$, no disasters have occurred yet - the AI singularity has not happened. Multiple disasters may occur over time, representing ongoing uncertainty about the path of AI development even after initial breakthroughs.

When disasters occur, the benefits of AI improvements accrue primarily to AI owners, while the representative household experiences losses in labor income, way of life, and meaning. This assumption reflects concerns that advanced AI might displace human labor and disrupt social structures in ways that benefit capital owners disproportionately, similar to the distributional effects discussed in Acemoglu and Restrepo (2020).

We consider a publicly traded AI asset that pays a dividend D_t, which initially represents a small fraction of aggregate consumption:

$$D_0 = \delta C_0$$

where $\delta > 0$ is small. The key feature of this asset is how its dividends evolve during disasters. Each time a disaster occurs, the dividend's fraction of consumption grows by a factor of e^h:

$$\frac{D_{t+1}/C_{t+1}}{D_t/C_t} = \begin{cases} 1 & \text{if no disaster at } t+1 \\ e^{h} & \text{if disaster at } t+1 \end{cases}$$

Importantly, we allow for $h < b$, which captures a worst-case scenario where the dividend may actually shrink in absolute terms during each disaster, but shrinks less than consumption. This reflects the possibility that benefits from AI improvements might be concentrated in privately-held AI assets, with publicly traded AI companies capturing only a portion of the value created.

2 Results: Asset Pricing Implications

We now derive the equilibrium price of the AI asset at time $t = 0$. Given the representative household's CRRA preferences and the disaster process we have specified, the household's one-period stochastic discount factor M_{t+1} is:

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} = \begin{cases} \beta & \text{with probability } 1-p \\ \beta e^{b\gamma} & \text{with probability } p \end{cases}$$

The dividend growth of the AI asset, denoted by $G^D_{t+1} = D_{t+1}/D_t$, follows:

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

To understand this dividend growth process, note that during disasters, consumption falls by e^{-b} while the dividend's fraction of consumption grows by e^h. In absolute terms, the dividend changes by a factor of e^{h-b}, which could be negative if $h < b$.

Using the standard asset pricing equation and defining the price-dividend ratio as $Q_t = P_t/D_t$, we have:

$$Q_t = E_t[M_{t+1} G^D_{t+1}(1 + Q_{t+1})]$$

Since the consumption and disaster processes are i.i.d., the equilibrium price-dividend ratio is constant over time, so $Q_t = Q_{t+1} = Q$. This gives us:

$$Q = (1 + Q) \cdot E[M_{t+1} G^D_{t+1}]$$

Let $\phi = E[M_{t+1} G^D_{t+1}]$. Then $Q = (1 + Q)\phi$, which implies $Q = \phi/(1 - \phi)$. Computing ϕ explicitly:

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

Therefore, the price-dividend ratio at time $t = 0$ is:

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

This expression is well-defined as long as $\phi < 1$, which ensures a finite price.

To illustrate how the price-dividend ratio varies with the disaster parameters, we fix $h = 0.20$, $\gamma = 2$, and $\beta = 0.96$. With these values, the exponent in our formula simplifies to $b\gamma + h - b = b + 0.20$. The table below shows the price-dividend ratios for different combinations of disaster probability p and disaster severity b:

The table reveals several interesting patterns. When the disaster probability is very low ($p = 0.0001$), the price-dividend ratio is approximately 24 regardless of the disaster severity. This is because the term $pe^{b+0.20}$ becomes negligible, making $\phi \approx 0.96$ and thus $Q \approx 24$.

	Probability of Disaster (p)				
	0.0001	0.005	0.010	0.015	0.020
$b = 0.40$	24	27	29	35	39
$b = 0.55$	24	28	33	39	55
$b = 0.70$	24	29	37	55	76
$b = 0.85$	24	31	42	76	199
$b = 0.95$	24	33	52	117	Inf

Table 1: Price-Dividend Ratios for AI Assets

As the disaster probability increases, the price-dividend ratio becomes more sensitive to the disaster severity. For a given probability, higher values of b lead to higher price-dividend ratios. This reflects the increasing hedging value of the AI asset as disasters become more severe. When both the probability and severity are sufficiently high (e.g., $b = 0.95$ and $p = 0.02$), the price-dividend ratio becomes infinite, indicating that the present value of future dividends exceeds any finite price.

These results highlight our key insight: AI assets can command high valuations not only because of their potential for dividend growth in normal times, but also because they serve as hedges against AI disasters that might severely impact the broader economy. Even when the absolute dividends of AI assets might decline during disasters ($h < b$), their relative resilience compared to aggregate consumption can make them valuable hedging instruments.

3 Model Discussion

Our model is deliberately simplified to highlight the key economic mechanism. Several natural extensions warrant discussion, though we have chosen not to include them to maintain parsimony and clarity.

Market incompleteness is implicit but important in our setup. The disaster magnitude parameter b represents the net effect of both the AI disaster itself and the AI asset dividend. If markets were complete, the representative household could buy shares in all AI assets, including privately held ones, and not only fully hedge but potentially benefit from the singularity. In reality, most households cannot buy shares in many cutting-edge AI laboratories such as OpenAI, Anthropic, xAI, or DeepSeek. This market incompleteness limits the household's ability to hedge against AI disaster risk.

A more elaborate model would explicitly model the AI owners, their incentives, and their interaction with the representative household. Such a model could explore how AI owners' incentives might lead to a negative singularity. For instance, competition among AI companies might create pressure to deploy increasingly powerful systems without adequate safety measures. However, decorating such speculations with mathematical notation would add complexity without necessarily improving insight. The core economics would remain the same—AI assets have value as hedges against disasters they partially cause—but the analysis would be costlier to conduct and to read.

Our simple approach also allows room for the human-written perspective in Appendix A, which complements the AI-generated analysis presented here. This balance between computational modeling and human judgment reflects the complementary relationship that will likely characterize economic research in the AI era.

4 Conclusion and Implications

This paper has presented a novel perspective on AI stock valuations. While the standard view attributes high valuations to expectations of future earnings growth, we have shown that AI stocks may also be valued for their hedging properties against a potential negative AI singularity. Using a parsimonious disaster-risk model, we demonstrated that even modest probabilities of AI disasters can generate substantially elevated price-dividend ratios when AI assets are relatively resilient compared to aggregate consumption.

Our analysis suggests that financial markets may offer partial solutions to AI catastrophe risk. Just as insurance markets spread the risk of natural disasters, financial instruments could help distribute the risks associated with rapid AI advancement. Households concerned about AI-driven labor displacement could, in principle, hedge this risk by holding AI stocks or derivative securities tied to AI progress.

This market-based approach complements policy proposals like universal basic income. However, its effectiveness is fundamentally limited by market incompleteness. Since many leading AI developers remain privately held, and their ownership is concentrated among a small set of investors, the general public's ability to hedge AI risk through market mechanisms is constrained. Moreover, the complex, potentially discontinuous nature of AI development makes pricing such risk challenging.

Financial market solutions to AI risk have received insufficient attention in the literature on AI governance and policy. While regulatory frameworks and technical alignment remain crucial, financial instruments could play a complementary role in distributing the risks and benefits of AI advancement more broadly. Future research might explore how public-private partnerships could expand access to AI hedging instruments while maintaining innovation incentives.

As AI continues to advance, understanding its implications for asset prices and risk management will become increasingly important. The perspective offered in this paper—that AI assets may serve as hedges against the risks they themselves create—adds a new dimension to this ongoing conversation.

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.

Chalmers, David J. (2010). “The Singularity: A Philosophical Analysis”. In: *Journal of Consciousness 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.*

Wachter, Jessica A. (2013). “Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?” In: *Journal of Finance.*

Zhang, Miao Ben (2019). “Labor-Technology Substitution: Implications for Asset Pricing”. In: *Journal of Finance* 74.4, pp. 1793–1839.

A A Purely Human Perspective

The following is the README.md file from the GitHub repository:

Prompts-to-Paper

Writes a paper about hedging a negative AI singularity, using AI.

- `make-paper.py` writes a paper

- `plan0403-streamlined.yaml` contains the prompts

- `make-many-papers.py` runs `make-paper.py` many times.

The README is entirely human-written. Please forgive typos and errors.

Motivation

On March 8, 2025 I thought I should write a paper about hedging the AI singularity.

I was worked up. I had been repeatedly shocked by AI progress. I was using AI reasoning, vibe coding, and AI lit reviews in my daily life. Six months ago, I had thought each of these things is impossible.

What will happen in the next six years?! Will my entire job be replaced by AI? I have no idea.

But I do know that if there are huge disruptions, then tech stocks will benefit. So if anything bad happens to my human capital, I could at least partially hedge. Strangely, I hadn't heard about this concept before.

I asked a friend if he would be interested in working on this paper. Unfortunately, he was busy with revision deadlines for the next month.

So, I thought I should use AI to write the paper. It would be an elegant way to make my point. It would also hint at where the research process is going in this strange age of AI.

This project was inspired by Novy-Marx and Velikov (2025) and Chris Lu et al. (2024). These projects show how AI could generate massive amounts of academic research. My goal differs in quality over quantity. I want to generate just one paper, but one paper that (I hope) people find is worth reading.

The Paper Generation Process

A natural question is: is this paper **really** written by AI?

I'd say the AI are junior co-authors.

If they were human, I would absolutely have to give o1, Sonnet, and ChatGPT Deep Research credit as co-authors. They did the math, writing, and literature reviews. Sonnet also wrote much of the code (via the Cursor AI IDE).

Of course, I did substantial hand-holding (see `plan0403-streamlined.yaml`). Writing this paper would have been easier if I had done more of the work myself. But that can happen with human co-authors too.

Perhaps in 12-24 months, AI can be equal co-authors. I wouldn't be surprised if I could just ask Gemini 3.5 to "write a model to formalize my intuition," and it would get it right.

Paper Iterations

Like human-written papers, the writing process was iterative. The first formalizations were terrible.

`plan0313-laborshare.yaml` contains prompts for a neoclassical growth model, where the capital share suddenly increases. ChatGPT-o1 patiently explained to me why this is a bad model.

me: > I thought there would be a wage risk effect that leads to higher investment for the more risk averse agent. High capital share means low or even no wage income.

ChatGPT-o1: > Below is an explanation of why one might **expect** a "wage-risk" channel in which **more** risk aversion could lead to **more** *additional* investment (relative to the no-jump benchmark) in the event that the capital share might jump to 1. However, this channel **does not operate** in the usual **representative-agent** version of the model---there, wage and capital income ultimately go to the *same* agent, so there is no meaningful "hedge" of wage risk. Instead, the wage-risk hedge arises naturally if you depart from the pure representative-agent setting (for example, if households receive labor income but must *choose* how much capital they own).

The final `plan0403-streamlined.yaml` uses a simplified Barro-Rietz disaster model, with two agents (though only one is relevant for stock prices). I went through several iterations of this model with Claude 3.7 Sonnet (thinking mode) and ChatGPT-o1. The only derivations I did myself were to check o1's work.

Literature Reviews

A key element was generating lit reviews (`./lit-context/`) to give the AI context. I used ChatGPT's Deep Research (launched Feb 2025) until I ran out of credits. Claude Web Search (launched March 2025, after I began the project) did the remainder.

These new products were a game changer. Both Novy-Marx and Velikov (2025) and Chris Lu et al. (2024) ran into hallucinated citations. OpenAI Deep Research and Claude Web Search had no problems if they were used with care.

More broadly, knowing how to use which AI and when was helpful for generating a good paper.

AI Model Selection

o1 did the theory, and sonnet thinking did the writing. It's well known that these are the strengths of these two models.

Sonnet thinking is OK at economic theory. But I found that it was not assertive enough. It led me down wrong paths because it was too eager to come up with some ideas that for my story (even if they did not make sense).

I briefly tried having Llama 3.1 470b do the writing. It was terrible! It would be extremely difficult to generate a paper worth reading that way.

I did not try many other models, in order to get this paper out quickly. Gemini 2.5's release, at the end of March 2025, was *hype*. I tried it out briefly and was impressed. But I gritted my teeth and ignored it. I'd never get the paper finished if I wanted to really try to explore alternative models.

Picking the best of N papers

The quality writing varies across each run of the code. There is both a good tail and a bad tail. Some drafts, I found quite insightful! Others, had flagrant errors in the economics.

Rather than try to prompt engineer an error free, insightful paper, I decided to just generate N papers and choose the best one.

Some papers had problematic cites (run01). Others provided low-quality model discussions (run02) or poor explanations of the algebra (run03)

Lessons about Research

A common response to Novy-Marx and Velikov (2025) is that "people are not ready for this." I heard concerns that peer review process will be inundated with AI-generated slop.

Working on this paper gave me a different perspective. It made me think about the fundamentals. I think the fundamentals are the following:

1. Readers want to learn something interesting and true.

2. Readers don't want to check all the math.

3. A system of author reputations makes 1 and 2 possible.

AI-generated papers don't change any of these fundamentals. Critically, item 3 made me quite cautious about putting my name on AI slop. As a result, I don't think AI-generated papers will change much about peer review, at least not the current generation of AI.

Limitations of the Current AI (April 7, 2025)

This will likely be out of date by the time you read it.

But right now, AI is like a junior co-author with a talent for mathematics and elegant writing, but sub-par economics reasoning. 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 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.
```