

Hedging the AI Singularity

Andrew Y. Chen

Federal Reserve Board

April 2025*

Abstract

This paper examines how AI stocks might serve as hedges against a negative AI singularity—a scenario where explosive AI development devastates the representative investor. We develop a parsimonious consumption-based asset pricing model where AI assets can provide insurance against consumption disasters caused by runaway AI progress. Our analysis shows how this hedging value could contribute to high AI stock valuations, even with modest disaster probabilities. This offers a complementary perspective to the common view that AI valuations reflect only future earnings growth. 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 capabilities are advancing at a breathtaking pace. Recent AI models like DeepSeek’s R1 can solve complex mathematical problems that were once the exclusive domain of human experts (DeepSeek-AI et al., 2025). Waymo’s autonomous vehicles have logged millions of driverless miles in increasingly complex environments. These rapid developments have many investors concerned about the future of human labor as AI begins to displace even high-skilled workers (Acemoglu and Restrepo, 2020; Zhang, 2019; Knesl, 2023).

While technological change has occurred throughout history, AI represents something fundamentally different. Previous revolutions like the internet transformed specific sectors, but AI is distinct because there is, in principle, no product or service that AI could not eventually create. For example, the current paper is entirely written by AI, using a few short prompts from a human.¹ You can examine the prompts at <https://github.com/chenandrewy/Prompts-to-Paper/>. Unlike the internet revolution, which augmented human capabilities, AI has the potential to substitute for human labor across virtually all domains. Moreover, AI progress may accelerate suddenly and dramatically—the hypothesized “singularity” (Vinge, 1993; Kurzweil, 2005).

This paper studies how AI stocks are priced given the risk that advanced AI might dramatically reduce human consumption. We develop a simple model where AI assets provide a hedge against a negative AI singularity—a scenario where explosive AI development devastates the representative investor’s consumption. The model shows that even with modest probabilities of disaster, this hedging property can substantially increase AI asset valuations, potentially explaining some of the premium observed in AI stocks.

We should clarify that we are not claiming a negative singularity will happen. Experts remain divided on whether advanced AI systems will benefit humanity or pose existential risks (Bengio et al., 2024). Nevertheless, it is important to consider the financial implications of such tail scenarios, just as we analyze other low-probability, high-impact events in financial markets. Similarly, we are not asserting that this hedging value is already priced into AI stocks. Our model simply illustrates a possible mechanism through which singularity risk could affect asset prices.

Our paper relates to two distinct strands of literature. First, the disaster risk asset pricing literature pioneered by Rietz (1988) and extended by Barro (2006), Gabaix (2012), and Wachter (2013) has established that rare, catastrophic events can explain various asset pricing puzzles. In parallel, a growing literature examines how technological risks affect asset

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

prices. Zhang (2019) shows that firms with routine-task labor maintain a replacement option that hedges their value against macroeconomic shocks, while Knesl (2023) finds that firms with high levels of displaceable labor earn a return premium due to their negative exposure to technology shocks.

Second, recent work on AI risks by Bostrom (2014), Russell (2019), and Bengio et al. (2024) has highlighted potential challenges in ensuring advanced AI systems remain beneficial to humanity. On the economic front, Jones (2024) explores the trade-off between AI-driven growth and existential risk, while Korinek and Suh (2024) analyzes how different AI development scenarios might affect output and wages. However, the financial implications of AI singularity risk remain underexplored.

2 Model

We propose a simple model to capture the essence of our argument about AI asset pricing. The model features 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 preferences determine equilibrium asset prices.

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

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

where γ is the coefficient of relative risk aversion.

We model the representative household's gross consumption growth as a binary process. In normal times, consumption growth is 1 (no growth), while in disaster states, consumption drops by a factor of e^{-b} where $b > 0$. Formally:

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

These disasters represent revolutionary improvements in AI technology that are devastating for the representative household. While such improvements benefit AI owners, they harm the representative household through lost labor income, disruption to their way of life, and diminished sense of meaning. At time $t = 0$, we assume no disasters have occurred yet (the singularity has not happened). Multiple disasters may occur over time, capturing ongoing uncertainty even after an initial singularity event.

The model includes a publicly traded AI asset with dividends that are initially a small fraction of consumption. Let D_t denote the dividend at time t , and n_t denote the number of disasters that have occurred by time t . We assume:

$$\frac{D_t}{C_t} = \delta e^{h \cdot n_t}$$

where $\delta > 0$ is the initial dividend-consumption ratio and h is a parameter that governs how the dividend share changes when disasters occur. Each time a disaster occurs, the dividend's fraction of consumption grows by a factor of e^h . This formulation allows the dividend to grow as a share of consumption during disasters, reflecting the increasing economic importance of AI during technological disruptions.

3 Equilibrium Asset Prices

We now derive the equilibrium price-dividend ratio for the AI asset at time 0. Under the standard asset pricing framework, the price of an asset equals the expected discounted value of its future payoffs. With CRRA preferences, the representative household's stochastic discount factor (SDF) from time 0 to time t is:

$$M_{0,t} = \beta^t \left(\frac{C_t}{C_0} \right)^{-\gamma}$$

where β is the time discount factor. The time-0 price of the AI asset is therefore:

$$P_0 = E_0 \left[\sum_{t=1}^{\infty} M_{0,t} D_t \right]$$

Given our consumption and dividend processes, we can express consumption at time t as $C_t = C_0 e^{-bn_t}$, where n_t is the number of disasters that have occurred by time t . Similarly, the dividend at time t is $D_t = C_0 \delta e^{(h-b)n_t}$. Substituting these expressions into the price equation:

$$P_0 = C_0 \delta \sum_{t=1}^{\infty} \beta^t E_0 \left[e^{(h-b+\gamma b)n_t} \right]$$

Since disasters follow a binomial process, n_t has a binomial distribution with parameters (t, p) . Using the moment generating function of the binomial distribution, we have:

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

Let $\alpha = h - b + \gamma b$ for notational simplicity. The price equation becomes:

$$P_0 = C_0 \delta \sum_{t=1}^{\infty} \beta^t (1 - p + pe^\alpha)^t = C_0 \delta \sum_{t=1}^{\infty} [\beta(1 - p + pe^\alpha)]^t$$

This is a geometric series that converges when $\beta(1 - p + pe^\alpha) < 1$. The sum of this convergent series is:

$$P_0 = C_0 \delta \frac{\beta(1 - p + pe^\alpha)}{1 - \beta(1 - p + pe^\alpha)}$$

Since $D_0 = C_0 \delta$, the price-dividend ratio at time 0 is:

$$\frac{P_0}{D_0} = \frac{\beta(1 - p + pe^{h-b+\gamma b})}{1 - \beta(1 - p + pe^{h-b+\gamma b})}$$

This formula reveals how the price-dividend ratio depends on the probability of disaster p , the time discount factor β , risk aversion γ , disaster severity b , and the dividend growth parameter h .

To illustrate the quantitative implications of our model, we set $\beta = 0.96$ and $\gamma = 2$. We also assume $h = 0.20 + b$, meaning the dividend's share of consumption grows by a factor of $e^{0.20+b}$ with each disaster. This specification implies that the exponent in our price-dividend formula simplifies to $h - b + \gamma b = 0.20 + b$.

Table 1 shows the price-dividend ratios for various combinations of disaster probability p and severity b . The rows correspond to different values of b ranging from 0.40 to 0.95, while the columns represent different disaster probabilities from 0.0001 to 0.02.

Table 1: Price-Dividend Ratios for Various Parameter Combinations

	$p = 0.0001$	$p = 0.005$	$p = 0.010$	$p = 0.015$	$p = 0.020$
$b = 0.40$	24	27	30	33	39
$b = 0.55$	24	28	32	41	55
$b = 0.70$	24	29	37	55	76
$b = 0.85$	24	31	42	76	249
$b = 0.95$	24	32	52	124	Inf

Several patterns emerge from this table. First, when the disaster probability is very low ($p = 0.0001$), the price-dividend ratio remains around 24 regardless of the disaster severity. This reflects the standard Gordon growth formula result when disaster risk is negligible. Second, as the disaster probability increases, the price-dividend ratio rises, with the effect becoming more pronounced for higher values of b . For instance, when $b = 0.95$ and $p = 0.02$, the price-dividend ratio becomes infinite because $\beta(1 - p + pe^{1.15})$ equals 1, making the denominator in our formula zero.

These results suggest that even modest probabilities of severe AI-driven disasters can substantially increase the valuations of AI assets that hedge against such events. The effect is highly nonlinear, with small changes in parameters potentially leading to large changes in valuations when the disaster severity is high.

4 Model Discussion

Our model intentionally abstracts from many complexities to highlight the core mechanism: AI assets can serve as hedges against negative AI singularity events. However, several important features deserve further discussion.

Market incompleteness is implicit but central to our analysis. The disaster magnitude parameter 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 AI enterprises, and not only fully hedge but potentially benefit from the singularity. In reality, most households cannot invest in many cutting-edge AI labs such as OpenAI, Anthropic, xAI, or DeepSeek. This market incompleteness limits the household’s ability to hedge against AI disaster risk, potentially explaining the high valuations of publicly traded AI stocks.

A more elaborate model would explicitly model the AI owners, their incentives, and interactions with the representative household. Such a model might explore how AI owners’ incentives could lead to a negative singularity, perhaps through competitive pressures that reduce safety investments or accelerate deployment timelines. However, this approach would essentially decorate speculations with mathematics. The resulting model would be costly to analyze and read, while the core economics would remain the same: AI assets can serve as hedges against consumption disasters induced by AI itself.

Our model also abstracts from many aspects of disaster risk modeling developed in the literature. For instance, we could introduce time-varying disaster probabilities (Wachter, 2013), partial recovery from disasters (Nakamura et al., 2013), or multiple disaster magnitudes. While these extensions would add realism, they would complicate the analysis without changing the fundamental insight that AI stocks can serve as a hedge against AI-induced consumption disasters.

The simplicity of our approach allows us to focus on the core mechanism while leaving room for the human-written Appendix A, which provides additional context and perspective on the paper’s origins and motivation.

5 Policy Implications and Conclusion

Our analysis suggests that financial markets can provide partial solutions to AI disaster risk through the natural hedging properties of AI stocks. This perspective complements discussions of universal basic income (UBI) and other policy responses to widespread technological unemployment. While UBI might address income loss, financial market solutions directly link households' wealth to the success of AI technologies, potentially creating better incentives and more efficient risk-sharing.

However, the effectiveness of this hedging mechanism is limited by market incompleteness. As noted earlier, most households cannot invest in private AI companies that may capture the largest share of value from advanced AI. This suggests potential benefits from policies that broaden access to AI investments, such as encouraging more AI companies to go public or creating financial instruments that track the value of private AI ventures.

The rapid development of AI technologies presents both opportunities and challenges for society. Our model highlights a mechanism through which financial markets might naturally address some of these challenges through asset prices. While we take no position on the likelihood of a negative AI singularity, our analysis shows that even small probabilities of such events can have significant pricing implications.

Future research might explore more complex models that explicitly incorporate AI owners' incentives, multiple sectors differentially affected by AI, or international dimensions of AI risk hedging. Empirical work could test whether AI stocks already exhibit pricing patterns consistent with disaster-risk hedging, perhaps by examining their performance during events that increase perceived AI risk.

In conclusion, our simple model offers a novel perspective on AI asset pricing by focusing on the hedging value rather than just the growth potential of AI stocks. This perspective may help explain high valuations in the AI sector and points to financial markets as one mechanism for addressing the risks of transformative AI.

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>.
- 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.
- Nakamura, Emi, Jón Steinsson, Robert Barro, and José Ursúa (2013). “Crises and Recoveries in an Empirical Model of Consumption Disasters”. In: *American Economic Journal: Macroeconomics* 5.3, pp. 35–74. DOI: 10.1257/mac.5.3.35.
- Rietz, Thomas (1988). “The Equity Risk Premium: A Solution?” In: *Journal of Monetary Economics*.
- Russell, Stuart (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking Press.
- 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 quality writing varies across each run of the code. There is both a good tail and a bad tail. Some drafts, I found quite insightful! Others, had flagrant errors in the economics.

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

Lessons about Research

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

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

1. Readers want to learn something interesting and true.
2. Readers don't want to check all the math.
3. A system of author reputations makes 1 and 2 possible.

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

Limitations of the Current AI (April 7, 2025)

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

But right now, AI is like a junior co-author with a talent for mathematics and elegant writing, but sub-par economics reasoning. Put another way, the writing can fail to portray the mathematics accurately.

For example, 3.7 Sonnet sometimes fails to recognize that the economic model does not capture an important channel. This is a common scenario in economics writing (no model can capture everything). The standard practice is to dance gingerly around the channel in the writing. A decent PhD student can recognize this. But Sonnet cannot. Instead, 3.7 Sonnet will write beautiful prose about the channel anyway, even though it's not really being studied properly.

AI also cannot generate satisfying mathematics on its own (at least not satisfying to me). I tried asking o1 and Sonnet to generate a model to illustrate the point I'm trying to make. The resulting models were either too simplistic or did not lead to a clean analysis. They often introduced complications that I found unnecessary.

There could be models with capabilities that I missed. But my sense is that ChatGPT-o1 and Claude 3.7 Sonnet are close to the best for producing economic research.

But more importantly, how long will these limitations last?

The Future of AI and Economics Research

At some point, 2024-style economic analysis will be "on tap." You'll be able to go to a chatbot and ask "write me a paper about hedging AI disaster risk," and it will return you something like this paper (or perhaps something better).

"Economics on tap" could be a disaster for the economics labor market. It would certainly mean that AI is an extremely cheap substitute for at least some economists' labor. It's possible that this would result in a strong substitution away from labor.

The optimistic argument is that AI also complements economists' labor. Perhaps, the number of economists will remain the same, but research output increases in terms of both quantity and quality.

But I think there are reasons why total research output is limited. Two key factors in academic publishing are attention and reputation (Klamer and van Dalen 2001, *J of Economic Methodology*). Readers can only pay attention to so many scholars. These scholars, in turn, can only pay attention to so many projects.

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

B Prompts Used to Generate This Paper

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

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

The system prompt and instructions are listed below.

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

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

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

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

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

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

Be modest. Do not overclaim.

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

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

Draft the model description. The model is purposefully simple and captures the essence of the main argument. Only describe the assumptions. No results or insights.

- Two agents
 - AI owners: Fully invested in AI, not marginal investors in stocks
 - Representative household: Marginal investor, only their consumption matters, CRRA
- Representative household's gross consumption growth
 - is either 1 or e^{-b} (disaster)

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

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

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

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

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

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

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

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

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

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

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

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

Title page:

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

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

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

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