Strategic Wealth Accumulation Under Transformative AI Expectations

Caleb Maresca*
New York University

February 7, 2025

Abstract

This paper analyzes how expectations of Transformative AI (TAI) affect current economic behavior by introducing a novel mechanism where post-TAI labor allocation depends on wealth at the time of invention. Using a modified neoclassical growth model calibrated to contemporary AI timeline forecasts, I find that even moderate assumptions about wealth-based allocation of AI labor generate substantial increases in pre-TAI interest rates. Under baseline scenarios with proportional wealth-based allocation, one-year interest rates rise to 12-16% compared to approximately 6% without strategic competition. The model reveals a notable divergence between interest rates and capital rental rates, as households accept lower productive returns in exchange for the strategic value of wealth accumulation. These findings suggest that evolving beliefs about TAI could create significant upward pressure on interest rates well before any technological breakthrough occurs, with important implications for monetary policy and financial stability.

Keywords: Artificial Intelligence, Economic Growth, Interest Rates, Technological Change, Wealth Distribution

JEL Codes: E43, O33, O40, D31, E21

*Email: ccm7752@nyu.edu. Department of Economics, New York University. Thanks to Philip Trammell, Basil Halperin, J. Zachary Mazlish, and Arden Berg for excellent comments and suggestions.

1 Introduction

The accelerating pace of artificial intelligence (AI) development raises critical questions about its potential to reshape the global economy through two distinct yet interrelated mechanisms. First, AI systems capable of augmenting or replacing human researchers could dramatically accelerate scientific progress and economic growth, enabling parallel deployment of AI agents that rival human capabilities. Second, advanced AI—particularly artificial general intelligence (AGI)—could automate vast swaths of human labor, potentially concentrating economic benefits among capital owners while displacing workers. I term AI systems with these dual disruptive capacities Transformative AI (TAI), focusing on their specific economic implications.[1] This paper develops a theoretical framework to analyze how forward-looking economic agents adjust current decisions in anticipation of TAI's uncertain arrival.

The prospect of AI accelerating scientific advancement is particularly compelling given that the number of scientific researchers appears to be a crucial factor driving economic growth (Jones, 2005). If human-level AI is invented, many instances could be run in parallel, effectively multiplying the researcher population (Jones, 2022). Even without achieving human-level capabilities, AI systems could significantly enhance human researchers' productivity. Moreover, AI's ability to process and synthesize vast amounts of scientific literature could uncover connections that have eluded human scientists, who are necessarily limited in their capacity to absorb information(Agrawal et al., 2018).

While AI's potential to accelerate scientific progress offers cause for optimism, its capacity for widespread automation raises important distributional concerns. Multiple leading AI developers explicitly pursue AGI systems that are "generally smarter than humans,"[2] which could render human labor economically obsolete across most domains. Unlike past

[1]This operationalizes Gruetzemacher and Whittlestone (2022)'s definition of Transformative AI as "Any AI technology or application with potential to lead to practically irreversible change that is broad enough to impact most important aspects of life and society. One key indicator of this level of transformative change would be a pervasive increase in economic productivity."

[2]OpenAI: `https://web.archive.org/web/20250104180629/https://openai.com/about/` Google Deepmind: `https://web.archive.org/web/20250106123809/https://deepmind.google/about/`

technological disruptions that often created as many jobs as they eliminated, AGI could offer superior productivity across virtually all domains, potentially limiting the economic relevance of human labor. Unlike historical automation that created new roles, AGI might enable comprehensive substitution while concentrating returns: AI "laborers" would generate wealth, but ownership of these systems would likely remain with existing capital holders. This represents not merely job displacement but a structural shift in labor's role—from human activity to AI-mediated capital service, benefiting those with more capital at the expense of others.

Crucially, even uncertain TAI prospects could reshape present-day economic behavior. Households anticipating TAI may alter consumption, savings, and investment patterns years before it materializes. These forward-looking adjustments imply that expectations alone—independent of realized technological change—could generate significant macroeconomic effects today. Understanding this anticipatory channel is essential for policymakers and economists navigating AI's economic implications.

This work extends Chow et al. (2024), who model TAI expectations as either explosive growth or existential catastrophe, finding that short-term TAI forecasts elevate long-term interest rates via Euler equation dynamics. While retaining their focus on growth scenarios, I introduce two critical and interrelated innovations: (1) explicit modeling of labor reallocation from human workers to AI systems disproportionately owned by wealthy households, and (2) strategic interactions in savings behavior as households compete for future control over AI labor.

The redistribution mechanism creates novel economic dynamics. Households' post-TAI labor supply depends on accumulated capital, incentivizing strategic savings to secure larger shares of AI-mediated production. Savings thus become both wealth-building tools and claims on future AI labor—a zero-sum competition absent in standard growth models. This disrupts traditional capital-pricing relationships: interest rates must now compensate not just for capital's rental rate but for the expected value of AI labor control rights.

My findings reveal that expectations of TAI can substantially affect current economic conditions, even before any technological breakthrough occurs. Under baseline scenarios with proportional wealth-based allocation of AI labor, I find one-year interest rates rising to 12-16% compared to approximately 6% without strategic competition, highlighting how anticipation of TAI can incentivize aggressive wealth accumulation. The effects strengthen as wealth becomes more important in determining future AI labor allocation, though with diminishing returns. Notably, interest rates diverge markedly from capital rental rates during the transition period - while increased savings drive down the marginal product of capital, interest rates remain elevated due to competition for future AI labor control. This wedge between productive returns and interest rates represents a novel channel through which technological expectations can influence financial markets. The magnitude of these effects varies with the assumed probability distribution of TAI arrival, with more concentrated near-term probabilities generating sharper initial increases in interest rates.

This paper proceeds as follows. Section 2 details the model setup, with a focus on the novel mechanism for AI labor allocation. Section 3 characterizes equilibrium conditions and describes the solution method for computing transition paths from a pre-TAI economy to a post TAI-economy. Section 4 presents quantitative results, analyzing how different assumptions about TAI arrival probabilities and wealth-sensitivity parameters affect interest rates and capital accumulation. Section 5 concludes with policy implications and directions for future research.

2 Model

This framework extends a neoclassical growth model to incorporate two fundamental effects of Transformative AI (TAI): (1) an acceleration of total factor productivity (TFP) growth, and (2) the displacement of human labor by AI labor. Households and firms operate under perfect foresight except for uncertainty about when TAI will be invented, which follows an exogenous arrival process.

2.1 Households and Firms

A continuum of measure 1 households maximizes expected lifetime utility with constant relative risk aversion (CRRA) preferences:

$$\begin{aligned}
&\max_{\{c_t\}_0^\infty} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\eta} - 1}{1-\eta} \\
\text{s.t.} \quad & a_{i,t+1} + c_{i,t} \leq w_t l_i + (1 + r_t) a_{i,t} \\
& c_{i,t} \geq 0 \\
& a_{i,t} \geq 0,
\end{aligned}$$

where $c_{i,t} \geq 0$, $a_{i,t} \geq 0$, and l_i are household i's time t consumption, time t wealth, and labor supply, respectively, w_t, and r_t are the time t wage rate and net interest rate, respectively, β is the discount rate, and η is the coefficient of relative risk aversion. The expectation operator $\mathbb{E}$ accounts for uncertainty over TAI's invention timeline.

A representative firm produces output via Cobb-Douglas technology:

$$Y_t = K_t^{\alpha} (B_t N)^{1-\alpha},$$

where α is the capital share, K_t is aggregate capital, N is aggregate labor, and B_t is total factor productivity (TFP). TFP is assumed to grow at an exogenous rate of g_{SQ} (status-quo) before TAI and g_{TAI} thereafter.

2.2 Invention of TAI and its effects

Households share homogeneous beliefs about TAI's invention timeline, represented by an exogenous probability distribution over arrival dates. This distribution may sum to less than 1, allowing for the possibility that TAI never materializes.[3]

Upon TAI's invention, two immediate regime shifts occur. First, productivity growth rises permanently from g_{SQ} to a g_{TAI}. Second, human labor is fully replaced by AI labor. Aggregate labor supply N remains constant at its pre-TAI level, but its composition shifts from human workers to AI labor units. Crucially, these units are reallocated across households based on wealth at the time of TAI arrival. This design choice serves to isolate redistribution effects: by holding aggregate labor constant, I neutralize scale effects and ensure that all economic growth post-TAI is attributable to the increase in the growth rate of TFP.

In the competition over the control of AI labor, it is assumed that agents who are richer at the time TAI is invented will capture a larger fraction of the wealth created by TAI. This could occur directly through market mechanisms or through political processes. I abstract from the exact mechanism through which this conflict and reallocation occur and instead provide a simple formula by which AI labor is allocated asymmetrically toward the rich. Specifically, while l_i was constant across households prior to TAI, after TAI l_i is reallocated according to the following formula:

$$l_i(a_{i,t_{TAI}}, A_{t_{TAI}}) = \frac{\left(\frac{a_{i,t_{TAI}}}{A_{t_{TAI}}}\right)^{\lambda}}{\int_0^1 \left(\frac{a_{i,t_{TAI}}}{A_{t_{TAI}}}\right)^{\lambda} di} \text{ for } \lambda \in \mathbb{R} \tag{1}$$

Where $a_{i,t_{TAI}}$ is the assets of the individual at the time TAI is invented, $A_{t_{TAI}}$ is the average assets of the population at that time, and λ is a parameter that determines how sensitive the allocation of AI labor is to the wealth of the individual.

Intuitively, this equation allocates AI labor to each individual based on their relative wealth at the time of TAI invention, with the parameter λ determining how strongly wealth differences translate into differences in AI labor allocation. When $\lambda = 0$, AI labor is distributed equally across the population regardless of wealth, implying no reallocation of la-

[3] This does not necessarily correspond with the event that AGI is never invented. Perhaps AGI is invented, but it does not radically alter the economy.

bor. Conceptually, one can imagine that all human labor is automated, but each household is given an AI laborer (by the government or some other redistributing agency) exactly replacing their labor and leaving their economic situation unchanged. Equivalently, one can imagine that this corresponds to the outcome where TAI increases the growth rate but does not automate human labor. Therefore, this case is equivalent to Chow et al. (2024)[4] and will be useful for comparison.

When $\lambda > 0$, wealthier individuals receive disproportionately more AI labor, with higher values of λ leading to a concentration of AI labor among the wealthy. $\lambda = 1$ corresponds to the case where AI labor is allocated proportionally to wealth, a household that is 10% richer when TAI is invented will be allocated 10% more AI labor. There is good reason to believe that λ may be significantly greater than one. Namely, race dynamics may cause differences at the top of the wealth distribution to be pivotal, whereas everyone else gets close to nothing.

One can also consider $\lambda < 0$, whereby AI labor is distributed disproportionately toward poorer individuals. However, this case is less economically relevant given the assumption that wealth provides advantages in securing AI resources.

2.3 TAI Timelines

Household beliefs regarding the year in which TAI will be invented are treated as exogenous. While this assumption simplifies the analysis, it abstracts from plausible feedback mechanisms. For instance, the pace of AI development likely depends on endogenous factors such as private R&D investment and policy interventions, which could correlate with household expectations—especially under high values of λ, where wealthier agents anticipate disproportionate gains from automation. Nevertheless, the exogenous timeline assumption provides a tractable foundation for isolating the economic effects of belief-driven behavior.

TAI invention is modeled as a stochastic process requiring n technical breakthroughs, each

[4]Ignoring the possibility of existential catastrophe.

with heterogeneous difficulty. This motivates using a negative beta binomial distribution (NBN), which generalizes the negative binomial by allowing success probabilities to follow a beta distribution.[5] Each trial represents a potential breakthrough in a given month. These monthly probabilities are then aggregated to obtain annual probabilities.

To capture uncertainty about the number of breakthroughs required for TAI, I treat n as a discrete random variable drawn from a bounded support $\{n_{\min}, ..., n_{\max}\}$. The resulting compound distribution integrates uncertainty over both breakthrough requirements and success probabilities. This distribution is truncated at 60 years for computational feasibility, with residual probability mass reallocated to the event that TAI is never invented.

Households update timeline probabilities annually via Bayesian filtering. For tractability, current results focus on passive learning, wherein probability mass shifts from elapsed years to remaining possibilities as time progresses without TAI being invented. A possible extension could involve active learning, wherein households would update their posterior beliefs in accordance with the number of observed breakthroughs each year.

2.4 Calibration

The yearly TAI probabilities are calibrated using two primary sources of AI timeline estimates: forecasts by Ajeya Cotra,[6] a Senior Advisor at Open Philanthropy who has conducted extensive research on AI development timelines, and aggregate predictions from Metaculus,[7] a reputation-weighted forecasting platform. These sources offer well-reasoned probability distributions for TAI development, though it's important to emphasize that such distributions are inherently speculative and represent different possibilities rather than objective truths. The source probabilities and the fitted distributions are displayed in Figure 1.

It's worth noting that these probability distributions come from individuals and commu-

[5] Following the convention where the negative binomial counts total trials until n successes, not failures.

[6] I selected yearly probabilities to roughly align with `https://www.alignmentforum.org/posts/AfH2oPHCApdKicM4m/two-year-update-on-my-personal-ai-timelines` and `https://www.alignmentforum.org/posts/K2D45BNxnZjdpSX2j/ai-timelines`

[7] `https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/`

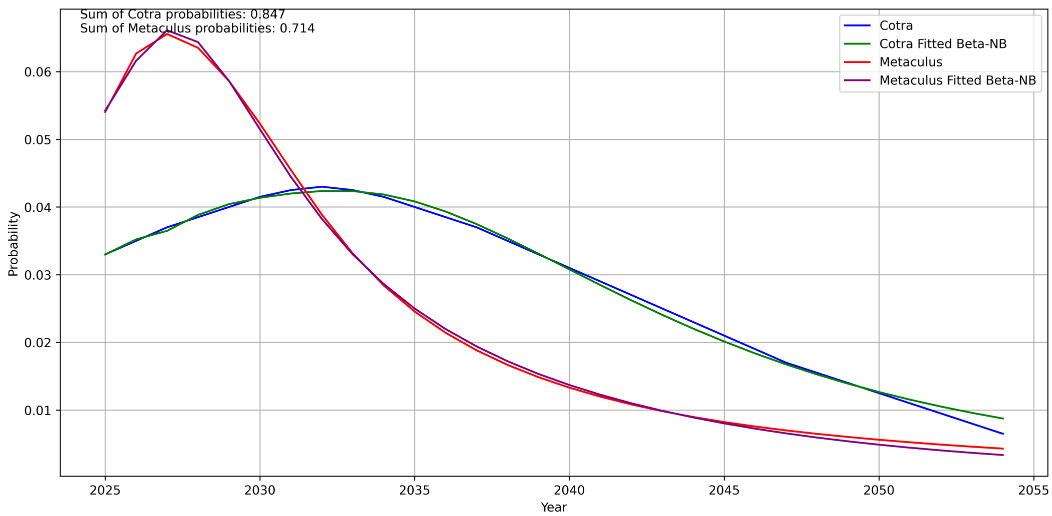

Figure 1: Predicted Probabilities of TAI Arrival. The figure shows fitted negative beta-binomial distributions based on forecasts from Cotra (2023) and Metaculus predictions.

nities that are particularly knowledgeable about and engaged with AI development. They likely differ substantially from the beliefs of the average household, many of whom may not be attentive to changes in the AI landscape and who expect current economic conditions to persist. This suggests an interesting extension of the model to incorporate heterogeneous beliefs, where some fraction of households maintain status quo expectations while others anticipate transformative change, possibly with varying timeline distributions. Such an extension could provide insights into how differential expectations about technological change affect wealth accumulation and economic inequality.

Using a functional form instead of directly applying the predictions from Cotra and Metaculus help to smoothen the probabilities. In addition, the negative beta binomial framework offers an intuitive way for individuals to construct their own probability distributions and see how they affect the model. By considering their beliefs about the number of necessary breakthroughs and the difficulty of achieving them, people can naturally translate their technological expectations into the model's distributional parameters.

The TFP growth rate after TAI (g_{TAI}) is assumed to be 30% as per Davidson (2021). This is of course much higher than the 2% historical average, though an increase of more than an order of magnitude is not without precedent. Prior to the industrial revolution, the world economy growth rate was near zero for most of human history (Roser et al., 2023). Whether or not another increase in growth rates of such magnitude is possible is left to the reader.

For λ, I consider a baseline case with $\lambda = 1$ wherein AI labor is allocated proportionally to wealth. I compare with a no strategic competition case of $\lambda = 0$, and cases of more extreme competition with $\lambda = 2$ and $\lambda = 4$. The rest of the parameters are standard: $\eta = 1$, $\beta = 0.96$, $\alpha = 0.36$, $\delta = 0.025$, and $g_{SQ} = 0.018$.

3 Equilibria and Transition Paths

3.1 Stationary Equilibria

This model has two important stationary equilibria: the status-quo equilibrium before households develop beliefs regarding future TAI and the post-TAI equilibrium. Each is the standard stationary equilibrium of the neo-classical growth model with respect to their parameters, wherein the interest rate is equal to

$$r_j = \frac{(1 + g_j)^\eta}{\beta} \tag{2}$$

for $j = SQ$ and $j = TAI$ respectively.

There are two important notes to make here. First, the economy will not remain in the status-quo equilibrium until the invention of TAI, but rather only until households develop beliefs regarding the invention of TAI. As soon as households realize that TAI is possible, they will begin to prepare for this uncertain future, causing the economy to exit the status-quo stationary equilibrium. The economy will then follow a transition path that will settle into a new stationary equilibrium only after TAI is invented. Alternatively, if TAI is never invented the economy will eventually settle back into the status-quo stationary equilibrium after the households believe such a transition is no longer likely.

Second, it is clear from Equation 2 that the higher rate of post-TAI TFP growth implies a higher stationary equilibrium interest rate. While this higher post-TAI equilibrium interest rate follows mechanically from the standard neoclassical growth model's properties, the more economically interesting phenomenon is the behavior of interest rates along the transition path, particularly in the periods before TAI arrives.

These two points highlight the importance of understanding the economy's dynamic adjustment process. The transition from the status-quo equilibrium begins as soon as households develop beliefs about future TAI, and the resulting path of interest rates reflects both current economic conditions and expectations about future technological change. In the following section, I examine these transition dynamics in detail.

3.2 Transition Paths

The economy's transition dynamics are characterized by multiple potential paths, reflecting the uncertainty about when TAI will arrive. Once households develop beliefs about future TAI, the economy deviates from its initial equilibrium, with the path forward branching at each potential arrival date. Specifically, in each year where there is a positive probability of TAI being invented, the economy's path splits: one branch represents the scenario where TAI arrives in that year, leading eventually to the post-TAI equilibrium, while the other branch continues to the next period maintaining the possibility of future TAI arrival. This process creates S+1 distinct paths, where S is the number of years in which TAI has a positive probability of being invented: one path for TAI arriving in each possible year, plus one path where TAI never materializes and the economy eventually returns to its initial equilibrium.

Formally, a household's optimization problem before TAI reflects this branching structure. In each period, households must consider both the immediate possibility of TAI arriving and how their wealth at that moment would determine their share of future AI labor if TAI does arrive. Their value function is:

$$
\begin{aligned}
&V_N(a_t) = u(c_t) + \beta\left[p_{t+1}V_{TAI}(a_{t+1}, l_i(a_{t+1}/A_{t+1})) + (1 - p_{t+1})V_N(a_{t+1})\right] \\
&\text{s.t.} \quad c_t + a_{t+1} = (1 + r_t)a_t + w_t
\end{aligned} \tag{3}
$$

where p_{t+1} is the probability of TAI arriving in the next period, and $l_i(a_{t+1}/A_{t+1})$ determines the household's share of AI labor based on their relative wealth at the time of TAI's arrival using Equation 1.

After TAI arrives, the household's problem simplifies to:

$$
\begin{aligned}
&V_{TAI}(a_t, l_i) = u(c_t) + \beta V_{TAI}(a_{t+1}, l_i) \\
&\text{s.t.} \quad c_t + a_{t+1} = (1 + r_t)a_t + w_t l_i
\end{aligned} \tag{4}
$$

where l_i represents their fixed allocation of AI labor determined by their relative wealth when TAI arrived.

I assume homogeneous initial wealth and beliefs across households, which generates identical savings decisions across households and preserves wealth homogeneity in every period. While this symmetry assumption sacrifices realism, it permits tractable analysis of the core mechanism. Future work could relax these constraints to study wealth inequality dynamics.

To solve for these transition paths, I employ a numerical approach based on solving for the derivatives of the value functions with respect to assets. This gradient-based method is more computationally tractable than directly solving for the value functions, as it avoids having to calculate the value function for every possible asset level. The key derivatives are:

For the pre-TAI value function:

$$
\frac{\partial V_N(a_t)}{\partial a_{t+1}} = -u'(c_t) + \beta\left(p_{t+1}\frac{\partial V_{TAI}(a_{t+1}, f_z(a_{t+1}/A_{t+1}))}{\partial a_{t+1}} + (1 - p_{t+1})\frac{\partial V_N(a_{t+1})}{\partial a_{t+1}}\right) \tag{5}
$$

For the post-TAI value function:

$$\frac{\partial V_{TAI}(a_t, l_i)}{\partial a_{t+1}} = -u'(c_t) + \beta \frac{\partial V_{TAI}(a_{t+1}, l_i)}{\partial a_{t+1}} \tag{6}$$

To find the equilibrium paths, I begin at a distant terminal period T where I assume the economy has reached its final steady state, either the post-TAI equilibrium or the initial equilibrium if TAI never arrives. Working backwards from this terminal condition, I solve for optimal household behavior in each period along each possible path.

These equations highlight a crucial feature of the model: in the pre-TAI period, households must account for how their savings will affect not only their future wealth but also their potential share of AI labor. This additional strategic motivation for saving causes households to engage in what is effectively a wealth-accumulation race. The resulting strategic behavior fundamentally alters the relationship between interest rates and capital returns during the transition period.

In the standard model, the interest rate equals the marginal product of capital. However, in this model, capital ownership provides not only traditional returns but also a claim on future AI labor through the wealth-based allocation mechanism. This strategic component appears as an additional term in Equation 5 compared to the standard model's value function derivative in Equation 6. Through its effect on the Euler equation, this term drives a wedge between interest rates and capital returns: interest rates must compensate households not just for the standard opportunity cost of capital but also for giving up the strategic advantage that capital ownership provides in securing future AI labor.

Even in periods well before TAI's potential arrival, the anticipation of this wealth-based allocation of AI labor influences household saving decisions and, consequently, equilibrium interest rates. The exact magnitude of this effect depends crucially on the parameter λ, which determines how strongly relative wealth differences translate into AI labor allocation differences.

These theoretical insights set the stage for quantitative analysis. In the following section, I examine how different assumptions about the timing of TAI arrival and the strength of the wealth-based allocation mechanism affect key macroeconomic variables, with particular attention to saving rates and interest rates during the transition period.

4 Results

The model generates several key insights about how expectations of Transformative AI (TAI) affect interest rates and capital rental rates. Figure 2 presents baseline results under both Cotra and Metaculus probability distributions, while subsequent figures show comparative results under different values of λ, which governs the wealth-sensitivity of future AI labor allocation. All figures track pre-TAI rates—that is, rates in each year conditional on TAI not having occurred. If TAI does occur, interest rates quickly converge to a new equilibrium of approximately 35%, as implied by Equation 2 given the assumed post-TAI growth rate of 30%.

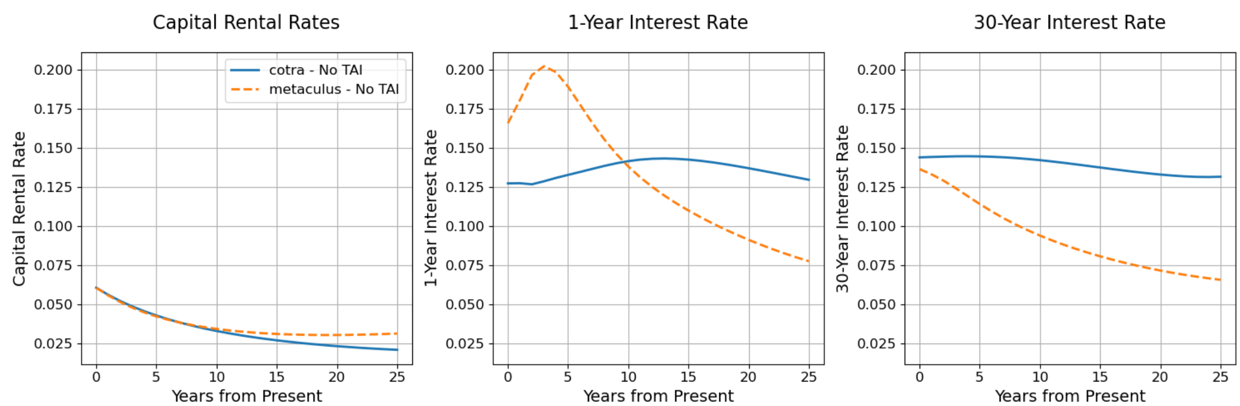

Figure 2: Baseline Economic Outcomes ($\lambda = 1$). The figure shows predicted paths for capital rental rates and interest rates under proportional wealth-based AI labor allocation. Left panel shows declining capital rental rates due to increased savings. Center and right panels show elevated interest rates reflecting both productivity growth expectations and strategic competition for future AI labor. Solid blue lines represent Cotra probabilities; dashed orange lines represent Metaculus probabilities.

4.1 Baseline Case and Strategic Competition

Baseline simulations reveal substantial effects of TAI expectations on interest rates. As shown in Table 1, even with moderate assumptions about wealth-based allocation of AI labor ($\lambda = 1$), one-year interest rates begin at 12.71% under Cotra probabilities and 16.56% under Metaculus probabilities—far above historical averages. Thirty-year rates show similar elevation, at 14.38% and 13.63% respectively.

The time paths of these rates, shown in Figure 2, exhibit distinct patterns between the probability distributions. Under Metaculus probabilities, which assign higher likelihood to near-term TAI arrival, one-year rates spike dramatically in the first five years before declining, while Cotra probabilities produce a more gradual increase followed by a modest decline. This difference reflects the more concentrated near-term probability mass in the Metaculus distribution.

	1y Interest Rate	30y Interest Rate
$\lambda = 0$		
Cotra	6.19%	10.85%
Metaculus	6.24%	9.53%
$\lambda = 1$		
Cotra	12.71%	14.38%
Metaculus	16.56%	13.63%
$\lambda = 2$		
Cotra	14.81%	15.61%
Metaculus	19.57%	15.54%
$\lambda = 4$		
Cotra	16.51%	16.74%
Metaculus	21.94%	17.61%

Table 1: Interest Rates in Initial Year

4.2 Role of Strategic Competition

The importance of strategic competition for future AI labor becomes apparent when comparing the baseline results to the no-competition scenario ($\lambda = 0$) shown in Figure 3. Without wealth-based allocation of AI labor, initial interest rates are dramatically lower: 6.19% and 6.24% for one-year rates under Cotra and Metaculus probabilities respectively. This marked difference from the baseline case—where rates are roughly twice as high—demonstrates how competition for future AI labor significantly amplifies saving incentives.

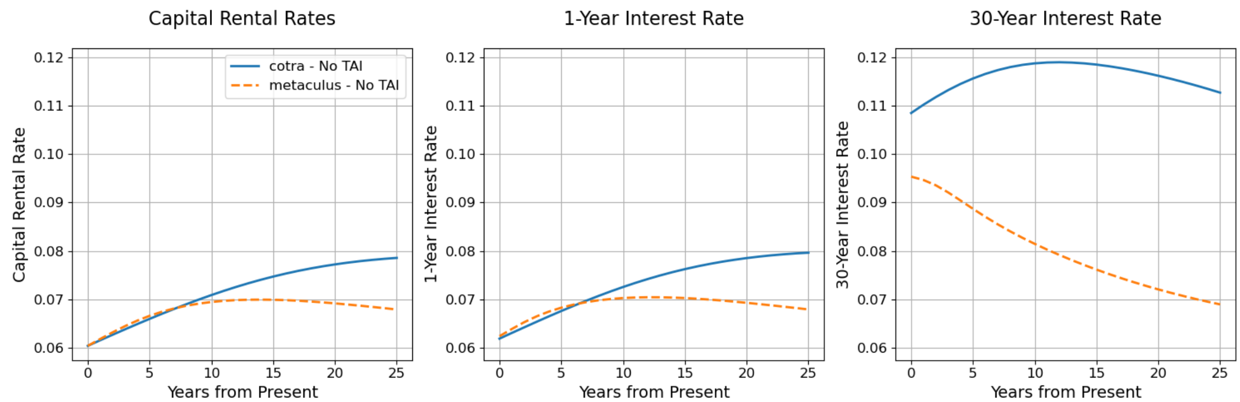

Figure 3: Economic Outcomes without Strategic Competition ($\lambda = 0$).

4.3 Sensitivity to Wealth-Based Allocation

Figures 4 and 5 explore scenarios with stronger wealth sensitivity in AI labor allocation ($\lambda = 2$ and $\lambda = 4$ respectively). As λ increases, both short and long-term interest rates rise monotonically, reflecting intensified competition for future AI labor control. However, this effect exhibits diminishing returns: the increase in rates from $\lambda = 0$ to $\lambda = 1$ is substantially larger than subsequent increases.

For instance, under Metaculus probabilities, one-year rates increase by over 10 percentage points when moving from $\lambda = 0$ to $\lambda = 1$, but only by about 3 percentage points when moving from $\lambda = 2$ to $\lambda = 4$. This pattern suggests that while strategic competition for AI labor significantly affects interest rates, extreme sensitivity to wealth differences may not proportionally intensify these effects.

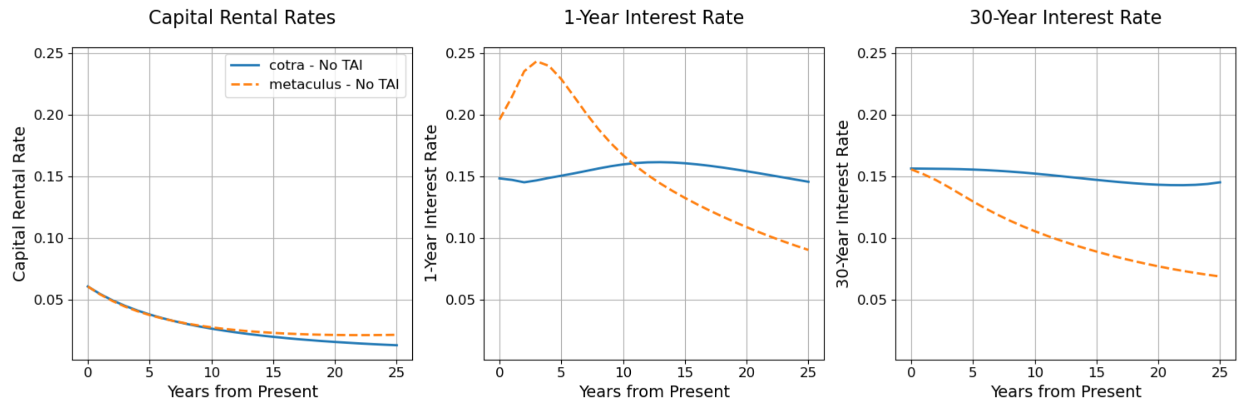

Figure 4: Economic Outcomes with Enhanced Strategic Competition ($\lambda = 2$).

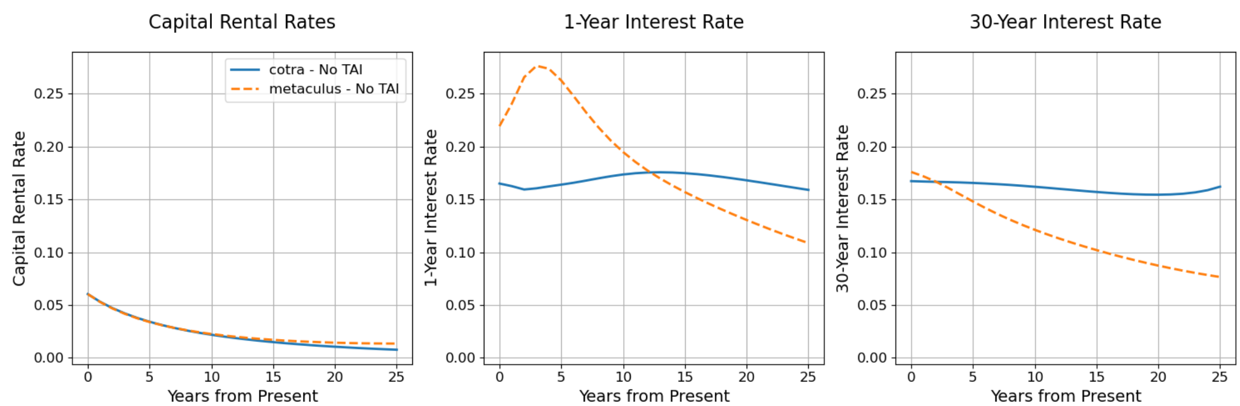

Figure 5: Economic Outcomes with Strong Strategic Competition ($\lambda = 4$).

4.4 Capital Rental Rates

Capital rental rates show less dramatic variation across scenarios than interest rates and follow a generally declining pattern over time, reflecting capital accumulation. This divergence between interest rates and rental rates is particularly noteworthy: despite increasing savings driving down the marginal product of capital (as shown by falling rental rates), interest rates remain elevated due to strategic competition for future AI labor. This demonstrates how the prospect of TAI can break the traditional link between capital returns and interest rates, as households accept lower productive returns in exchange for the strategic value of wealth accumulation.

4.5 Practical Implications

Notably, all scenarios with positive λ produce interest rates substantially higher than current real-world rates. Even in the no-competition case, the model generates rates well above historical averages, suggesting that TAI expectations alone could significantly influence financial markets. This finding is shared with Chow et al., who's representative agent model is equivalent to my model with $\lambda = 0$. Importantly, λ is plausibly greater than or equal to one, in which case TAI expectations could cause interest rates to rise far higher than even the already high rates predicted by Chow et al., as shown in Table 1. These findings has important implications for monetary policy and financial stability, as it suggests that evolving beliefs about TAI could create strong upward pressure on interest rates well before any technological breakthrough occurs.

As for savings, the capital rental rates across all scenarios show a generally declining pattern over time, reflecting increasing capital accumulation as households prepare for potential TAI arrival. This decline is more pronounced in scenarios with higher λ values, indicating that stronger strategic motives for wealth accumulation lead to greater capital deepening in the pre-TAI period.

5 Conclusion

This paper develops a theoretical framework for analyzing how expectations of Transformative AI (TAI) could influence current economic behavior, with particular attention to the effect of wealth-based allocation of automated labor. The model reveals that anticipation of TAI can significantly affect present-day interest rates and capital accumulation patterns through two distinct channels: expectations of higher future growth and strategic competition for future AI labor control.

The results demonstrate that even moderate assumptions about wealth-based allocation of AI labor can generate substantial increases in interest rates well above predictions without strategic competition over labor control. Under baseline scenarios with proportional wealth-based allocation ($\lambda = 1$), one-year interest rates rise to 12.71% and 16.56% under Cotra and Metaculus probability distributions respectively, compared to 6.19% and 6.24% in scenarios without strategic competition ($\lambda = 0$). This dramatic difference highlights how competition for future AI labor can amplify saving incentives and elevate interest rates even before any technological breakthrough occurs.

A key finding is that the strength of wealth-based allocation in determining future AI labor shares (parameterized by λ) monotonically increases both short and long-term interest rates, though with diminishing returns. This suggests that while strategic competition for AI labor significantly affects financial markets, extreme sensitivity to wealth differences may not proportionally intensify these effects. Moreover, the model reveals a notable divergence between interest rates and capital rental rates, as households accept lower productive returns in exchange for the strategic value of wealth accumulation.

Several promising directions for future research emerge from this analysis. First, relaxing the assumption of homogeneous initial wealth could provide insights into how TAI expectations might affect wealth inequality dynamics. Second, incorporating heterogeneous beliefs about TAI arrival probabilities would better reflect real-world variation in technological expectations across different economic agents. Third, extending the model to include active belief updating based on observed technological progress could capture how evolving information about AI development influences economic behavior.

Additional extensions could explore the role of takeoff speed in shaping economic responses to TAI expectations. While the current model assumes an immediate transition to higher productivity growth, a more gradual takeoff might generate different patterns of anticipatory behavior. TAI could even result in superexponential growth (Aghion et al. 2017, Trammell and Korinek 2023), which could boost interest rates even further. Furthermore, building on Chow et al.'s findings, incorporating TFP shocks could illuminate how increased growth volatility might counteract or amplify the interest rate effects identified in this paper.

These findings have important implications for both policy and theory. For policymakers, the model suggests that evolving beliefs about TAI could create significant upward pressure on interest rates well before any technological breakthrough occurs, with potential implications for monetary policy and financial stability. For economic theory, the results demonstrate how anticipation of transformative technological change and zero-sum competition can create novel strategic interactions in saving behavior, distinct from standard growth model dynamics.

In conclusion, this analysis demonstrates that expectations of TAI can substantially influence current economic behavior through both growth expectations and strategic wealth accumulation motives. As AI technology continues to advance, understanding these anticipatory channels becomes increasingly important for economic policy and planning. Future research extending this framework along the directions outlined above will be crucial for developing a more complete understanding of how technological expectations shape economic outcomes.

References

Aghion, P., Jones, B. F., and Jones, C. I. (2017). *Artificial intelligence and economic growth*, volume 23928. National Bureau of Economic Research Cambridge, MA.

Agrawal, A., McHale, J., and Oettl, A. (2018). Finding needles in haystacks: Artificial intelligence and recombinant growth. In *The economics of artificial intelligence: An agenda*, pages 149–174. University of Chicago Press.

Chow, T., Halperin, B., and Mazlish, J. Z. (2024). Transformative ai, existential risk, and real interest rates. Working Paper.

Davidson, T. (2021). Could advanced ai drive explosive economic growth. *Open Philanthropy*, 25.

Gruetzemacher, R. and Whittlestone, J. (2022). The transformative potential of artificial intelligence. *Futures*, 135:102884.

Jones, C. I. (2005). Growth and ideas. In *Handbook of economic growth*, volume 1, pages 1063–1111. Elsevier.

Jones, C. I. (2022). The past and future of economic growth: A semi-endogenous perspective. *Annual Review of Economics*, 14(1):125–152.

Roser, M., Arriagada, P., Hasell, J., Ritchie, H., and Ortiz-Ospina, E. (2023). Data page: Global GDP over the long run. `https://ourworldindata.org/grapher/global-gdp-over-the-long-run`. Part of "Economic Growth". Data adapted from World Bank, Bolt and van Zanden, Angus Maddison.

Trammell, P. and Korinek, A. (2023). Economic growth under transformative ai. Technical report, National Bureau of Economic Research.