Implications of artificial intelligence for coordination under incomplete information (Early stage work)*

Douglas K. G. Araujo[1]

[1]Bank for International Settlements, douglas.araujo@bis.org

Abstract

In a global game with endogenous information acquisition, investors choose how much to invest in the adoption of advanced artificial intelligence (AI), such as large language models, to better extract signal about fundamentals from a common data space. Unlike other information acquisition technologies that increase precision of the private signal, key AI-specific features are included in the model: hallucinations that introduce bias for exponentially high levels of AI adoption and the possibility of technological breakthroughs that enhance models prior to usage. This monotonically increasing technology evolves stochastically in a way that increaseds with the amount of AI resources that investors leave for technologists. Investors then decide to invest on a risky asset with strategic complementarities on information and on the level of technology adoption by peers. The model is simple but allows inference on important topics related to the adoption of AI in finance, including the effect of hallucinations on the coordination of investors, the role of a closed versus open AI model, and the potentially outsized effects of cyber security or other operational risk incidents. JEL codes: D82, G14.

1 Introduction

The increasing capabilities of large language models (LLM) and generative artificial intelligence (gen AI) systems more generally unlock useful new ways to transform data into actionable information, expanding the "data envelope" (Hirshlelfer ((1971)), Goldfarb and Tucker ((2019))). These sophisticated models can be useful even for forecasting and predicting stock returns (eg, Lopez-Lira and Tang ((2024)), Kim, Muhn, and Nikolaev ((2024))), fueling their use in settings of economic importance. More generally, gen AI has the potential to process economically-relevant signals with greater precision. To the extent that many of these settings entail strategic complementarities, as the stock investment example, a baseline approach to understand the effects of gen AI on coordination under incomplete information (Harsanyi ((1995))) is to model it as a standard global game (Carlsson and Van Damme ((1993)), Morris and Shin ((2003))). Under this prism, and with standard assumptions on the common knowledge of the signal generating process, a unique risk-dominant equilibrium obtains where agents choose to invest or not based on a threshold rule. And since this threshold is independent of the signal precision, this baseline suggests gen AI would not affect coordination incentives regardless of its technical prowess.

But specificities of gen AI as an information processing technology require adaptations to the model to enable policy analysis. The breakneck speed of gen AI investment strongly suggests that a key ingredient is endogenous information acquisition (eg, Hellwig and Veldkamp ((2009)), Ming Yang ((2015)), Szkup and Trevino ((2015a))). Other novel elements are also relevant to model gen AI, related to its reasoning and the externalities with AI developers. First, accumulated evidence suggests gen AI systems cannot properly reason and thus may introduce biases in the signal. Gen AI makes mistakes that are increasingly hard to

*This work represents my opinion and not necessarily that of the BIS. An early version of the manuscript circulated as 'Information acquisition in financial markets through artificial intelligence'.

spot because erroneous responses are presented with more confidence and eloquence with more advanced models, the so-called "hallucinations" (Alkaissi and McFarlane ((2023)), Ji et al. ((2023))). Second, the AI technology is still evolving at a fast pace (eg, Jingfeng Yang et al. ((2023))) one breakthrough at a time, but both development and use of these systems require broadly the same core resources: the quintessential example are graphical processing units (GPUs), the computer chips that effectively enable gen AI.[1] In other words, investment in resources to deploy gen AI for information processing reduce its speed of development. This impacts the ex ante information acquisition calculation as investors expect that models will, once acquired, have a positive probability of experiencing a breakthrough.

The main contribution of this paper is to set up a simple but flexible model of endogenous information acquisiton under incomplete information, subject to bias in the private signal (from the "AI hallucinations") and to stochastic technological breakthrough that improves the model's reasoning, tapering the hallucinations. Other than these AI-specific parameterisations, the model follows a standard information acquisition model that can be solved as a global game (Carlsson and Van Damme ((1993)), Morris and Shin ((2003))), yielding a symmetric, unique equilibrium outcome. This baseline model is used to show that the equilibrium choices of AI resource take-up by investors facing a risky payoff are consistent with the level of AI that minimises bias in the private signal. In other words, investors invest the "right" amount in AI for their application at hand, even if the investment decision itself is risk- and not payoff-dominant (Harsanyi and Selten ((1988))).

The baseline 2-investor game plays out in two stages. In the *AI investment* stage, each investor i chooses how much of their endowment E will be allocated to acquiring AI-related resources (eg, skilled practitioners, graphics processing unit chips - GPU) from a limited pool R. The remainder after the investors decide their AI investment is available to technologists, and will influence technological development: $R = R_{AI} + \sum_i R_i$. Next, the *financial investment* stage entails a binary investment decision on the remainder of the endowment, $I_i = E - R_i$. The payoff depends on an unobserved state θ and, for intermediate levels of the fundamental outcome, on coordination between investors. These decisions are taken after each investor observes a private signal that depends non-linearly on the level of AI investment: higher AI amounts increase precision in the private signal, but expose the investor to AI hallucinations. These latter elements are additive biases in the private signal that behave exponentially, like a "hockey stick" line, kicking off strongly at high levels of AI adoption. A final unknown component in the model is the degree of technological advance in between stages. Technologists take the remainer of the resources, R_{AI}, and use it to implement research ideas that are compared with obstacles. Whenever these are net positive, the hallucinations taper off. This represents for example advances in reasoning ability (Araujo ((2024))) or in fine-tuning to make models.

A unique value of AI investment minimises the bias in the private signal, ie optimally balances the hallucination with the precision-increasing effects of greater investment in AI. Interestingly, even as the equilibrium definition is not conditional on an optimal first-stage allocation, this happens to be the value that is consistent with the second-stage equilibrium. The model is simple enough to represent the main dynamics of investors' indirect acquisition of information through their investment in AI. At the same time, it contains rich enough representations that allow the study of questions of interest. In this paper I explore three of them. First, the baseline specification speaks to AI's ability to reason. One of the extensions, studying the effect of an eventual much-hyped artificial general intelligence (AGI), simply involves turning off the hallucination. The third extension addresses current policy and academic discussions around the wisedom of openly available versus closed models. Secondly, the model can be slightly adapted to study the potentially outsized effects of cyber security or other operational risk incidents, including issues related to AI vendor concentration.

[1] https://www.nytimes.com/2023/08/16/technology/ai-gpu-chips-shortage.html. Some research has started to consider training processes that circumvent the GPU shortage (Strati et al. ((2024))).

1.1 Literature

This work relates to the literatures on information acquisition in incomplete information games, and on AI and data in finance. The class of global games models address incomplete information settings Carlsson and Van Damme ((1993)). Morris and Shin ((2002)), Morris and Shin ((2003)), Morris and Shin ((2004)). A second stream of papers related to this work discuss endogenous information acquisition, similar to the current paper. **Coordination with information acquisition**. Angeletos and Lian ((2016)), Szkup and Trevino ((2015a)), Szkup and Trevino ((2015b)), Szkup and Trevino ((2021)). Reshidi et al. ((2021)) study the individual and collective information acquisition. Technology adoption canonical model Frankel and Pauzner ((2000)). Private aquisition of information (processing), Hellwig and Veldkamp ((2009)) and Colombo, Femminis, and Pavan ((2014)). Colombo, Femminis, and Pavan ((2014)) highlights the difference between effiencies in information acquisition and its usage. In contrast with that literature, here the AI technology frontier also develops endogenously and responds (negatively) to the resource take-up from technology adoption. (Denti ((2023)) study information acquisition about other players' information)

Another stream of papers discusses **data in finance**: Farboodi et al. ((2022)) models the extraction by investors of *information* on assets from *data*, and how the value of data is related also to characteristics of the asset itself. Begenau, Farboodi, and Veldkamp ((2018)) argues that the existence of more data to larger firms favours them over others. And a more specific line of works examines **AI in finance, including associated risks.** AI can read information better (Araujo et al. ((2023)), Araujo et al. ((2024))), and especially the more sophisticated type of models - large language models (LLMs) can further increase the ability to use data (Korinek ((2023))). Lopez-Lira and Tang ((2024)) show evidence that ChatGPT, the flagship LLM, can successfully pick stocks. Danielsson, Macrae, and Uthemann ((2022)) discuss risks. Increased availability of data (Veldkamp and Chung ((2019))) and lower cost across data pipeline (Goldfarb and Tucker ((2019))). This facilitates the use of new data, or data in a compound way, due to the "discovery" of new data by newer technologies Hirshlelfer ((1971)). An early reference is Ranco et al. ((2015)). New techniques not only process more data more effectively, but they also expand the envelope of data that can be analysed to look for signal.

Finally, the interplay between dynamic aggregate variables and strategic complementarities, which features in my paper from the public signal of the GPU prices, is developed in more detail in other papers. Alvarez, Lippi, and Souganidis ((2023)) analyse how complementarities affect price setting behaviour, and therefore monetary policy effectiveness.

2 Background on AI

- AI is a technology that helps extract signal from more information:
  - these models are not usually restricted by curse of dimensionality
  - when appropriately designed, AI algorithms can regularise well and thus help identify signal-rich parts of the data space
- All of this put together indicates one way to model how AI contributes to information acquisition is to simply use it as a way to improve the precision of the private signal.
  - This also has the advantage that the model is directly comparable with most of the information acquisition literature.

But AI has particular characteristics that are not found in other information acquisition technologies, and in particular are exacerbated in the most recent generation of sophisticated AI models, including LLMs. Those are described next.

2.1 Reasoning and hallucinations

First, AI models hallucinate, meaning that their result seems correct but in fact is made up in an attempt by the model to appease its users. Huang et al. ((2023)) classify hallucinations into two types. The first one are related to facts: these factuality hallucinations are further sub-divided into factual inconsistency between statements and factual fabrication, or the creation of a completely alternate reality. The second type of hallucionations is related to how faithful a source text is reproduced. These are sub-divided into inconsistent ability to follow the instructions, to use the appropriate context, or to complete a logical statement.

These hallucinations have several causes, including some structural ones that are difficult to adress with the current technology (Huang et al. ((2023))). The dataset used to train AI models can include low-quality, factually wrong or biased data. In addition, it could be outdated. But even perfect data does not guarantee a hallucination-free model: the transformer architecture is known to suffer from defficiencies that reflect in a non-negligible chance of hallucination. When the models are trained to be "generative", ie create text, the task of predicting the following tokens lowers the ability to reflect more nuanced contextual dependencies. Another problem occurs when the attention is diluted across a large swath of tokens in a context, making the answers too unstable. The training strategies, including the practice of adjusting trained models to better adjust to human behaviour (the so-called "reinforcement learning from human feedback"), can further deteriorate performance. Hallucination can also result from the inference process, such as when the context is insufficient (especially in the absence of sotred commonsense knowledge) to help models "reason".

Even in applications in finance, which typically use more scalable applications than chatbots, suffer from the fact that these models do not properly reason (even as they appear to gain capabilities with their growing size). All of this is compounded by the fact that these models are increasing unscrutinable.

LLMs might provide human-like, confident but incorrect answers to trivial questions (Perez-Cruz and Shin ((2024))), flip initially correct answers when questioned by the user (Laban et al. ((2023))) or double-down on previous mistakes during a chat(Zhang et al. ((2023))).

2.2 AI development

A second characteristic of AI is that the technology is still undergoing very fast levels of development (eg, Jingfeng Yang et al. ((2023))). This matters because of the influence it can have on decisions related to information acquisition and ultimately on the financial payoff, since agents expect some possibility that these breakthroughs will happen.

The development for these models require considerable computing power and technical staff resources.

2.3 Policy questions related to AI in finance

The use of AI in finance matters not only for welfare and potentially distributional reasons, but also for financial stability. Regulators in particular have ramped up work on the oversight of AI use by banks and other financial firms. For example, the Basel Committee on Banking Supervision announced work on "the potential implications of broader usage of AI/ML models for the resilience of individual banks and more broadly, for financial stability".[2] National regulators such as the Bank of England and Prudential Regulation Authority are actively engaging with the financial industry, as are others.[3]

Important policy questions include: * ability of financial market participants (broadly referred to in this paper as "investors") to secure resources - especially adequately skilled human resources - to deploy AI

[2] https://www.bis.org/publ/bcbs_nl27.htm

[3] https://www.bankofengland.co.uk/prudential-regulation/publication/2023/october/artificial-intelligence-and-machine-learning

* potential for hallucinations, biases and other "silent mistakes" * vendor concentration as a third-party service provider risk

And more recently, another potential safety concern appears in the debate between closed vs open models. Open models lower the adoption barrier, democratising AI and favouring "public good" effects similar to other open source software (Lerner and Tirole ((2005))). But on the other hand they also enable malicious users such as impersonators or cyber attackers.

3 Model

The model draws heavily from the setup in Szkup and Trevino ((2021)), with important additions that represent key features of AI as an information technology. Second, the technology itself evolves stochastically over time, including in response to factor prices as well. And third, the process by which AI improves the signal is laid out in more detail to highlight the cases where the technology can transform more data text into information but cannot yet *reason* about it.

3.1 Setup

The simplest version of the model is a two-investor setup as follows.[4]

The state of economic fundamentals is a random variable with normal distribution $\theta \sim N(\mu_\theta, \sigma_\theta^2)$. θ is only observed indirectly by each investor i as a noisy signal, $x_i = \theta + \sigma_i \epsilon_i$. An AI technology $\alpha(R) > 0$ uses finite specialised resources R (eg, AI scientists, data engineers, graphics processing units chips - GPUs, etc) to improve the precision of the signal of the existing data.[5] The total amount of these resources is divided into AI developers and the investors: $R = R_{\text{AI}} + \sum_i R_i$ for $R_{\text{AI}} > 0$ and $\forall R_i \geq 0$. The technology has decreasing returns to scale with $\alpha'_R > 0, \alpha''_R < 0$, and each individual precision defined as $\sigma_i = \sigma/\alpha(R_i)$. α itself is common amongst players, reflecting the current relevance of open source and open weight models in the high end AI market. One application later in the text relaxes this definition so that investors can also purchase a unique α_i technology.

However, unless the AI model can actually reason, very precise signals (low σ_i) also increase the risk of belieavable wrong answers (such as "hallucinations") or other forms of "silent mistakes", which bias the investor's perception about fundamentals. Such a problem is of course compounded by the high confidence the investor has in the signal given its acquired low σ_i. This is modelled through a *reasoning filter* ϕ . This function is the identity function if the AI cannot reason and 0 if it can reason. Having $\eta \sim N(0, \sigma_\eta)$ as the baseline level of noise[6] for all private signals, then each investor will observe $\epsilon_i = \eta + \phi e^{\lambda \alpha_i(R)}$, in which λ is an inconsequential positive constant for scaling only. Putting all of this together, the private signal about the fundamentals is:

$$x_i = \theta + (\eta + \phi e^{\lambda \alpha(R_i)}) \sigma / \alpha(R_i) \tag{1}$$

Given this scenario for technology investment, two ex ante identical investors $i \in \{1, 2\}$ choose how much R_i to acquire. Because resources are finite and technologists pick up the residual resources not acquired

[4]More advanced or more generalised perturbations, where the information structure is more richer than a linearly additive error (eg, Morris, Shin, and Yildiz ((2016)) with rank beliefs or Morris and Yang ((2022)) with a stochastic continuous choice) could in principle be explored. But since the goal of this work is to add a richer structure to a model, a canonical global games model with linear information (Morris and Shin ((2003))) is used as a well-studied benchmark.

[5]$\alpha > 0$ even when $R_i = 0$ as investors can always implement simplistic tools to get *some* information from data. This structure, which can be seen as $\alpha = g(R_i) + \xi$ for an infinitesimal but positive ξ, also helps make the main model simpler by introducing α directly in the private signal equation.

[6]Reflecting, for example, technology frictions in the production and dissemination of data, or more fundamentally even the sparsity in the actual signal from the manifold hypothesis.

by investors, the investors face a supply curve and the technologists are assumed for simplicity to be price takers. Given this structure, prices are normalised as the proportion of resources taken by investors, $\rho = (1/R) \sum_i R_i$.[7]. The prices are public information and the market for AI resources clear.

The two investors decide in the first stage how much of their endowment to invest in AI usage, with the remainder available for the next stage where they decide whether or not to invest, $a_i \in \{0, 1\}$. This equality is represented as $E_i = R_i + I_i$. In the second period, the investors decide how to allocate I_i, in a safe or risky asset (the allocation is binary). The safe asset does not have a cost, and yields zero regardless of θ or the number of investors who choose it. Conversely, the risk asset's payoff can be successful in either of the following situations: (a) if $\theta \geq \bar{\theta}$ or (b) if $\theta \geq \underline{\theta}$ and $A_i = A_j = 1$. While this payoff structure follows Szkup and Trevino ((2021)) closely, the current model differs from that one because investors only allocates I_i. Because investing in the risky asset entails a cost T, the risky asset yields $\theta I_i - T$ in case of success or alternatively, $-T$.

Each investor's choice a_i depends on the observed signal x_i and the level of use of AI chosen in the preceding step, R_i. As in Szkup and Trevino ((2021)), the level of precision (from the investment in AI resources) is common knowledge in this simple game, but here it is only incidentally so: this settings comports only two investors and a common price that reflects their joint AI investments.[8] Each investor's utility is a mapping of the form $u : \{0, 1\} \times \{0, 1\} \times \mathbf{R} \times [0, 1] \to \mathbf{R}$, with $u(A_i, A_j, x_i, R_i)$ representing investor i's pay-off as a function of their own action, the other investor's action, the signal observed by i, and its investment in the AI technology.

Innovations in AI reasoning happen in between periods: after investors have decided ρ, the remainder R_{AI} determines the probability of a major breakthrough in reasoning ability. This is modelled as follows.[9] Two independent random draws from U(0, 1), $\pi_{\text{idea}}, \pi_{\text{obstacle}}$, correspond respectively to innovative ideas and to practical obstacles to innovation related to those ideas. The idea requires resources for implementation, and thus a technological advance only happens if the idea, once actually implemented, overcomes the barrier to innovation. Formally:

$$\phi = 1 - \max(0, \underbrace{\pi_{\text{idea}} * (1 - \rho)}_{\text{Implemented idea}} - \pi_{\text{obstacle}}). \tag{2}$$

ϕ tapers off the noisiness to help the private signal get closer to $\theta + \eta$. Note that the only chance of a full shutdown of the added noise by the AI model - for example, through the concept of *artificial generalised intelligence*, is ruled out and would require all of the AI-related resources to be available to the technologists only.

> If a breakthrough occurs, productivity increases (or at least is expected to) and asset prices increase more easily, ie the threshold from which investment is considered a success is now lower.

3.2 Bias in private signal

With a nonzero level of AI adoption and with an AI that cannot robustly reason (and thus hallucinates), the private signal will always on expectation have a positive bias. However, there is exactly one value of $\alpha(R_i)$ that results in an minimally biased private signal.

[7] The actual prices could be proportional to this ratio, but the added clutter to notation does not justify it.

[8] A richer setting would see not only more investors, including atomic ones, but also have only the global AI resource expense be public, not its distribution to investors. The type of challenge it would bring to the current model includes for example a non-trivial correlation between the use of AI and the perceived signal. This interesting case demands a dedicated exposition and not further dealt with in this paper.

[9] Recall that ϕ determines the reasoning ability in a way that is orthogonal to model performance.

Proposition 3.1 (AI use with minimal bias in private signal)**.** *There is only one specific value* $\alpha^* = \alpha(R^*)$ *for which the private signal is the closest to* θ *in expectation. The subscript is irrelevant because all investors are ex ante similar.*

Proof. The first order condition only holds for one value of $\alpha > 0$. Starting with the first derivative,

$$\frac{d}{d\alpha^*}(\eta + \phi e^{\lambda\alpha^*})\sigma/\alpha^* = 0$$

the expression can be manipulated to facilitate isolating α^* in the numerator to the left side:

$$\frac{\phi\lambda e^{\lambda\alpha^*}\sigma}{\alpha^*} - \frac{\phi e^{\lambda\alpha^*}\sigma}{(\alpha^*)^2} = \frac{\eta\sigma}{(\alpha^*)^2}.$$

Multiplying both sides by $(\alpha^*)^2$ obtains

$$\alpha^*\lambda\phi e^{\lambda\alpha^*}\sigma - \phi e^{\lambda\alpha^*}\sigma = \eta\sigma,$$

which is the same as

$$\phi e^{\lambda\alpha^*}(\lambda\alpha^* - 1) = \eta,$$

and thus if $\phi > 0$, the only possible solution with a positive value is $\alpha^* = 1/\lambda$.

The second derivative obtained by substituting this equality above is positive, confirming that α^* minimises the bias on the signal.

$$\frac{d^2}{d(1/\lambda)^2}(\eta + \phi e)\sigma\lambda = (\sigma\lambda^2\phi e + \sigma\lambda\phi e)\lambda^3$$

□

Proposition 3.1 highlights the problem with AI investment as a technology to process information: optimising on precision alone (lowering σ_i) exposes the investors to more biases in the signal at very high levels of technology adoption. Higher levels of R_i monotonically increase precision but there is only one level of R that minimises bias in information.

Note also that it seems to be a repetition of the classical bias-variance trade-off (BVTO) but it is in fact a different phenomenon. BVTO implies that more flexible and complex models would lead to lower bias and high out-of-sample variance, as more sophisticated models obtain a better fit to existing data at the risk of not generalising too well. The current setting shines a light on a different problem that incorporates economic and information- and game-theoretic elements: even if the model obtains lower variance by usefully increasing out-of-sample precision (real life examples include the ability to digest unstructured data such as text to improve signal pickup), the fact it *can* hallucinate and at the same time is increasingly trusted both by its practicality and precision, introduces uncertainty in the model.

4 Equilibrium

The equilibrium analyses draw from Szkup and Trevino ((2021)), who extend the monotone supermodular games result from Van Zandt and Vives ((2007)) in games with unbounded utility functions. In particular, Szkup and Trevino ((2021)) show that there exist both a least and a greatest bounds in Bayesian Nash equilibra, and that these thresholds correspond to a univalent mapping to a unique outcome.

The equilibrium is found by backward induction, starting with the financial investment stage. Collect the allocations of R in $r = \{R_i, R_j, R_{AI}\}$. Investor strategies $A : \mathbb{R} \times [0, 1]^3 \times (0, 1] \rightarrow \{0, 1\}$ map the private signal, the allocation r and the AI reasoning ability ϕ to a binary investment decision, where 1 is associated with the investing choice. Using the global games' threshold strategies and the uniqess of equilibrium outcomes due to Szkup and Trevino ((2021)) extending the results from Van Zandt and Vives ((2007)), the investor decides to invest if the private signal is higher than a specific threshold, as in

$$A(x_i; r; \phi) = \mathbf{1}[x_i \geq x_i^*(r, \phi)]. \tag{3}$$

The value x_i^* in Equation 3 that makes the investor indifferent to investing or not investing is the optimal threshold. The following specification considers both the scenario in which fundamentals are intermediate and the investment requires coordination, and the scenario in which fundamentals are good enough that success does not necessitate coordination. The equation

$$E[\theta I_i \Pr(A_j(\theta) = 1)|x_i = x_i^*, \theta \in [\underline{\theta}, \bar{\theta})] + E[\theta I_i | x_i = x_i^*, \theta \geq \bar{\theta}] = T \tag{4}$$

represents the situation where the expected payoff of an I_i investment is set to zero. Now moving backwards to the AI investment stage, consider $\nu_i(\cdot)$ as the expected investment payoff to i afer observing their own private signal x_i and believing that investor j will also optimise. Consider also a belief function $\mu_i : [0, 1] \rightarrow [0, 1]$ as i's belief on the probability that j chose a specific value of R_j. In general terms, the expected utility of each investor is:

$$U_i(r) = E[\mathbf{1}[x_i \geq x_i^*(e, \phi)]\nu_i(x_i, xj^*(r, \phi); r)]f(x_i; R_i, \phi)dx_i. \tag{5}$$

All elements to define the equilibrium are now in place.

Definition 4.1 (Pure strategy Bayesian Nash equilibrium). A vector of AI investment allocations $r^* = \{R_i^*, R_j^*, R_{AI}^*\}$, belief function μ_i, and optimal financial investment decision $A_i(x_i; r; \phi)$ is a pure streategy Bayesian Nash equilibrium if, for each $i \in \{1, 2\}$, it complies with the conditions below:

C1 - no incentives to deviate. $U_i(r^*) \geq U_i(r') \forall r' \neq r^*$,

C2 - correct strategic anticipation. $\mu_i(R_j^*) = 1, \mu_i(R_j^{'}) = 0, r' \neq r^*$; and

C3 - optimal financial investment even if sub-optimal AI investment.

$$\begin{gathered} A(x_i; r; \phi) = \mathbf{1}[x_i \geq x_i(\{R_i, R_j^*, R_{AI}\}, \phi)], \text{s.t.} \\ x_i^* \in \{x_i : \nu(x_i(\{R_i, R_j^*, R_{AI}\}, \phi), x_j^*(r^*, \phi); r) = 0\} \end{gathered}$$

The first condition, C1, lays out that no investor has an incentive to deviate from the equilibrium because they will not extract a higher utility. C2 is necessary to ensure that the equiibrium obtains from a situation that each investor assigns positive probability to the amounts of AI investment that the other investor would choose. And the third condition pins down the idea that the signal threshold associated with the investment action can also be associated with of each investors' own AI investment that is not necessarily the optimal.

Proposition 4.1 (Equilibrium)**.** *The equilibrium as defined in Definition 4.1 exists and is unique.*

4.1 A discussion on the uniqueness of the equilibrium

Equilibrium selection from perturbations of complete information games is widely studied (Carlsson and Van Damme ((1993)), Morris and Shin ((2003))). More recently, this result has been increasingly generalised: Morris, Shin, and Yildiz ((2016)) shows that the equilibrium selection comes from the common uniform rank beliefs and Szkup and Trevino ((2015a)), Ming Yang ((2015)) and Morris and Yang ((2022)) generalise this further to the infeasibility of acquiring an information precision that perfectly discriminates between states. In fact, while this paper models the information acquisition and investment stages as separate steps, they can be understood as a single step, two-signal game where one of the signals is related to the continuous stochastic choice (Morris and Yang ((2022))): to the extent that players are not able to sharply distinguish between states when they get close to a state threshold, then uniqueness of outcome results.

4.2 Equilibrium efficiency

Is the equilibrium under information acquisiton through AI with hallucination and technology externalities efficient?

- Szkup and Trevino ((2015a)) study equilibrium efficiency under endogenous information acquisition.

4.3 Information-efficient use of AI

In the base scenario that AI models do not reason robustly, they always let at least *some* hallucination pass through ($\phi > 0$). A natural question is whether the value of AI investment that minimises bias in the private signal (from Proposition 3.1) is consistent with the equilibrium outcomes. The following exercise focuses on the case of intermediate fundamental values, as they require coordination for success.

Define $r^{\S} = \{R_i^{\S}, R_j^{*}, R_{AI}^{\S}\}$ as a the AI investment by investor i driven only by their need to optimise the private signal, while j continues to allocate their investment according to the full-game perspective. Assume C2 and C3 of Definition 4.1 hold. Then it suffices to check if C1 still holds, but with $r^{\S}$. In particular, whether:

$$E[\theta I_i^{\S}\Pr(A_j(\theta) = 1)|x_i = x_i*, \theta \in [\underline{\theta}, \bar{\theta})]] = E[\theta I_j^{*}\Pr(A_i(\theta) = 1)|x_j = x_j*, \theta \in [\underline{\theta}, \bar{\theta})]].$$

Factoring out the common expectation components related to θ, we have:

$$I_i^{\S}\Pr(A_j(\theta) = 1) = I_j^{*}\Pr(A_i(\theta) = 1),$$

which are the same since the equilibrium values are robust to AI investment misspecification (C3), and therefore the right hand side stays the same. Since $I_i^{\S} > 0$ the equation still holds, and from the threshold strategies, the two probabilities are equal, thus also $I_i^{\S} = I_j^{*}\S*$, confirming the value of r that is consistent with the equilibrium outcome is $r^{*} = r^{\S}$.

5 Open vs closed models

What is the influence, if any, of the possibility of investors to use exclusive, closed models to the equilibrium outcomes laid out above? This section relaxes the definition of α to allow for the possibility that investors spend some money in the first period to acquire or develop such models.[10]

[10]This section is still very premature.

Assume that the development of a closed-model entails a cost $C > 0$ whereas the use of a public-knowledge open model is free (consistent with real life). In this case, then investors would choose the closed-model if $\alpha_i(R_i) > \alpha(R_i)$ and the final net payoff would still be positive even with the safe. And, for conservativeness, to reflect concerns that open source gen AI can facilitate cyber crime and other deleterious activities such as bomb building (Seger et al. ((2023))), suppose also that the realised technology at the second stage can serve with a probability δ to lead to the same outcome as the non-invest for both players.

6 Operational risks[11] (eg, cyber disruptions and AI vendor dependence)

Suppose now that α has a small but nonzero chance of being set to zero during the investment stage. Such a scenario would be akin to an operational risk incident, such as a cyber attack. Building on financial supervisors' work on risks from third party service providers, such as AI providers,[12] this section explores AI shutdown effects that are proportional to ρ.

Take Equation 1, but now assume that μ is a constant that always multiplies α. The constant goes to zero with probability ρ, representing an operational incident.

7 AI breakthrough and asset prices[13]

Suppose an AI breakthrough leads to an increase in asset prices, all else constant. This could reflect, for example, expectations about increase in productivity such as Aldasoro et al. ((2024)). Or it could be due to higher (unobserved) expectation of future cash flows. To model this possibility, consider that there is uncertainty about the levels of $\underline{\theta}$ and $\bar{\theta}$. For simplicity, let $\underline{\theta} = \underline{\theta}_0 - \Delta(1-\phi)$ and $\bar{\theta} = \bar{\theta}_0 - \Delta(1-\phi)$ for Δ some positive number such that $\underline{\theta}, \bar{theta}$ remain within bounds.

This situation makes it easier to coordinate while also increasing the share of states of the world for which no coordination is needed for success in the investment game stage.

8 "Data reflexion" problem[14]

One of the concerns about the widespread use of generative AI is that over time, the data available to train new models will be increasingly itself the output of an AI. This can be conceptualised as a similar application to Morris and Shin ((2018)), who studied the reflexion of market prices and central bank guidance. For this reason, I use the same machinery to sketch what the implications of heightened data reflexion could look like.

9 Preliminary considerations[15]

AI is not simply a generic information technology. In addition to resource distribution considerations, introducing the use of AI to process information entails both an increase in precision (in line with existing information acquisition models), but also changes that reflect hallucinations and other silent mistakes and a chance for stochastic technology improvements that agents may expect to take place.

[11] This section is still very premature.
[12] For example, the Basel Committee on Banking Supervision recently ((2022)) exorted banks to address risks related to concentration of third party service providers.
[13] This section is still very premature.
[14] This section is still very premature.
[15] This section will become the conclusion section.

The present model is simple but sufficiently rich to enable studies of important phenomena of interest at the intersection of AI and finance. For example, this model can help estimate the effects of hallucinations, operational risk issues, welfare implications from closed vs open models and others.

The joint determination in equilibrium of AI investment in finance industry and academia, together with the levels and returns of financial investments, bear interesting results. This paper shows that the global game results in AI take-up that is consistent with the minimisation of private noise in expectation. This is an interesting result because the game-level result itself is the risk-dominant strategy, not the most efficient payoff-dominant.

References

Aldasoro, I., S. Doerr, L. Gambacorta, and D. Rees. (2024): The Impact of Artificial Intelligence on Output and Inflation, BIS Working Papers, Bank for International Settlements.

Alkaissi, H., and S. I. McFarlane. (2023): "Artificial hallucinations in ChatGPT: Implications in scientific writing," *Cureus*, 15.

Alvarez, F., F. Lippi, and P. Souganidis. (2023): "Price setting with strategic complementarities as a mean field game," *Econometrica*, 91, 2005–39.

Angeletos, G.-M., and C. Lian. (2016): "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," in *Handbook of macroeconomics*, Elsevier, 1065–1240.

Araujo, D. K. G. (2024): "Benchmarking Economic Reasoning in Artificial Intelligence Models,"

Araujo, D. K. G., G. Bruno, J. Marcucci, R. Schmidt, and B. Tissot. (2023): "Data science in central banking: Applications and tools," *IFC Bulletin*, 59.

Araujo, D. K. G., S. Doerr, L. Gambacorta, and B. Tissot. (2024): "Artificial intelligence in central banking," *BIS Bulletin*, 84.

Basel Committee on Banking Supervision. (2022): "Newsletter on Third- and Fourth-Party Risk Management and Concentration Risk,"Bank for International Settlements.

Begenau, J., M. Farboodi, and L. Veldkamp. (2018): "Big data in finance and the growth of large firms," *Journal of Monetary Economics*, 97, 71–87.

Carlsson, H., and E. Van Damme. (1993): "Global games and equilibrium selection," *Econometrica: Journal of the Econometric Society*, 989–1018.

Colombo, L., G. Femminis, and A. Pavan. (2014): "Information acquisition and welfare," *The Review of Economic Studies*, 81, 1438–83.

Danielsson, J., R. Macrae, and A. Uthemann. (2022): "Artificial intelligence and systemic risk," *Journal of Banking & Finance*, 140, 106290.

Denti, T. (2023): "Unrestricted information acquisition," *Theoretical Economics*, 18, 1101–40.

Farboodi, M., A. Matray, L. Veldkamp, and V. Venkateswaran. (2022): "Where has all the data gone?" *The Review of Financial Studies*, 35, 3101–38.

Frankel, D., and A. Pauzner. (2000): "Resolving Indeterminacy in Dynamic Settings: The Role of Shocks*," *The Quarterly Journal of Economics*, 115, 285–304.

Goldfarb, A., and C. Tucker. (2019): "Digital economics," *Journal of economic literature*, 57, 3–43.

Harsanyi, J. C. (1995): "Games with incomplete information," *The American Economic Review*, 85, 291–303.

Harsanyi, J. C., and R. Selten. (1988): "A general theory of equilibrium selection in games," *MIT Press Books*, 1.

Hellwig, C., and L. Veldkamp. (2009): "Knowing what others know: Coordination motives in information acquisition," *The Review of Economic Studies*, 76, 223–51.

Hirshlelfer, J. (1971): "The private and social value of information and the reward to inventive activity," *The American Economic Review*, 61, 561–74.

Huang, L., W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, et al. (2023): "A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions," *arXiv* *preprint arXiv:2311.05232*,.

Ji, Z., N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung. (2023): "Survey of hallucination in natural language generation," *ACM Computing Surveys*, 55, 1–38.

Kim, A., M. Muhn, and V. V. Nikolaev. (2024): "Financial statement analysis with large language models," *Chicago Booth Research Paper Forthcoming, Fama-Miller Working Paper*,.

Korinek, A. (2023): "Generative AI for economic research: Use cases and implications for economists," *Journal of Economic Literature*, 61, 1281–1317.

Laban, P., L. Murakhovs' ka, C. Xiong, and C.-S. Wu. (2023): "Are you sure? Challenging llms leads to performance drops in the flipflop experiment," *arXiv preprint arXiv:2311.08596*,.

Lerner, J., and J. Tirole. (2005): "The economics of technology sharing: Open source and beyond," *Journal of Economic Perspectives*, 19, 99–120.

Lopez-Lira, A., and Y. Tang. (2024): "Can chatgpt forecast stock price movements? Return predictability and large language models," *SSRN*,.

Morris, S., and H. Shin. (2003): "Global Games: Theory and Applications. Advances in Economics and Econometrics, m Dewatripont, l Hansen, and s Turnovsky,"Cambridge University Press, NY.

Morris, S., and H. S. Shin. (2002): "Social value of public information," *american economic review*, 92, 1521–34.

---. (2004): "Liquidity black holes," *Review of Finance*, 8, 1–18.

---. (2018): "Central Bank Forward Guidance and the Signal Value of Market Prices," *AEA papers and proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 572–77.

Morris, S., H. S. Shin, and M. Yildiz. (2016): "Common belief foundations of global games," *Journal of Economic Theory*, 163, 826–48.

Morris, S., and M. Yang. (2022): "Coordination and continuous stochastic choice," *The Review of Economic Studies*, 89, 2687–2722.

Perez-Cruz, F., and H. S. Shin. (2024): Testing the Cognitive Limits of Large Language Models,Bank for International Settlements.

Ranco, G., D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič. (2015): "The effects of twitter sentiment on stock price returns," *PloS one*, 10, e0138441.

Reshidi, P., A. Lizzeri, L. Yariv, J. H. Chan, and W. Suen. (2021): Individual and Collective Information Acquisition: An Experimental Study,National Bureau of Economic Research.

Seger, E., N. Dreksler, R. Moulange, E. Dardaman, J. Schuett, K. Wei, C. Winter, et al. (2023): "Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives,"

Strati, F., P. Elvinger, T. Kerimoglu, and A. Klimovic. (2024): "ML Training with Cloud GPU Shortages: Is Cross-Region the Answer?" *Proceedings of the 4th workshop on machine learning and systems*,.

Szkup, M., and I. Trevino. (2015a): "Information acquisition in global games of regime change," *Journal of Economic Theory*, 160, 387–428.

---. (2015b): "Information acquisition and transparency in global games," *Journal of Economic Theory*, 160, 387–428.

---. (2021): Information Acquisition and Self-Selection in Coordination Games,mimeo.

Van Zandt, T., and X. Vives. (2007): "Monotone equilibria in bayesian games of strategic complementarities," *Journal of Economic Theory*, 134, 339–60.

Veldkamp, L., and C. Chung. (2019): "Data and the aggregate economy," *Journal of Economic Literature*,.

Yang, J., H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, B. Yin, and X. Hu. (2023): "Harnessing the power of LLMs in practice: A survey on chatgpt and beyond," *arXiv preprint arXiv:2304.13712*,.

Yang, M. (2015): "Coordination with flexible information acquisition," *Journal of Economic Theory*, 158, 721–38.

Zhang, M., O. Press, W. Merrill, A. Liu, and N. A. Smith. (2023): "How Language Model Hallucinations Can Snowball,"