(Generative) AI in Financial Economics*

Hongwei Mo   Shumiao Ouyang

Saïd Business School, University of Oxford

February 2026

ABSTRACT

This review article synthesizes the burgeoning literature on the intersection of (generative) artificial intelligence (AI) and finance. We organize our review around six key areas: (1) the emergent role of generative AI, especially large language models (LLMs), as analytic tools, external shocks to the economy, and autonomous economic agents; (2) corporate finance, focusing on how firms respond to and benefit from AI; (3) asset pricing, examining how AI brings novel methodologies for return predictability, stochastic discount factor estimation, and investment; (4) household finance, investigating how AI promotes financial inclusion and improves financial services; (5) labor economics, analyzing AI's impact on labor market dynamics; and (6) the risks and challenges associated with AI in financial markets. We conclude by identifying unanswered questions and discussing promising avenues for future research.

Keywords: Artificial Intelligence, Large Language Models, Corporate Finance, Asset Pricing, Household Finance, Labor Economics

*Hongwei Mo can be reached at Hongwei.Mo@sbs.ox.ac.uk, and Shumiao Ouyang can be reached at Shumiao.Ouyang@sbs.ox.ac.uk. We are grateful to Alexander Dyck, Jiayin Hu, Manish Jha, Anton Korinek, Leonardo Gambacorta, Joel Shapiro, and Yucheng Yang for their valuable comments and suggestions. We also thank Andreas Charisiadis, Shang Wang, ChatGPT, and Claude for helpful discussions and research assistance. Any remaining errors or omissions are solely ours. The October 2025 version has been published as an invited article in JCEBS, licensed under CC BY 4.0 (Mo and Ouyang, 2025), and this working paper will be continuously updated. The review is also available at AIFinEcon.com.

1 Introduction

Artificial intelligence (AI)[1] has emerged as a transformative force in the global economy—one that spurs investment, fosters innovation, and boosts productivity across sectors (Cockburn et al., 2018; Furman and Seamans, 2019). A recent report by Goldman Sachs (2023) projects that AI could raise global GDP by 7%, amounting to approximately $7 trillion, and increase US productivity growth by 1.5 percentage points annually over the next decade. McKinsey Global Institute (2023, pg. 10) offers even higher projections that AI could contribute between $17.1 and $25.6 trillion to the global economy.[2] Like previous general purpose technologies (GPTs) such as the steam engine and electricity, AI is characterized by fast improvement, wide applicability, and the ability to catalyze complementary innovations (Cockburn et al., 2018; Goldfarb et al., 2023). The rapid advancement of AI, especially the recent breakthroughs in generative AI (GenAI), has begun to reshape the financial system by significantly enhancing its information processing capabilities—the core function through which it allocates resources, manages risk, and supports economic coordination (Aldasoro et al., 2024).

These technological shifts have sparked a surge of interest in AI-related research in financial economics. As shown in Figure 1, both the number of AI-related papers and the number of top journal publications in financial economics increased more than sixfold from 2018 (immediately following Vaswani et al. (2017)) to 2024.[3] Looking forward, leading researchers and AI developers have forecasted that transformative AI capable of performing most cognitive tasks at or above human level could emerge within the next decade.[4] While these predictions remain speculative,

[1]Artificial intelligence is a broad term used to describe a range of advanced technologies that exhibit human-like intelligence (Seamans and Raj, 2018). According to the seminal work by Nilsson (2014), "The fundamental AI ideas underlie applications such as natural language processing (NLP), automatic programming, robotics, machine vision, automatic theorem proving, and intelligent data retrieval, rather than focusing on the subject mattear of the applications." The term "artificial intelligence" was first coined by John McCarthy in a proposal in 1955 (McCarthy et al., 1955). An overview of the history of AI research can be found at https://spectrum.ieee.org/history-of-ai.

[2]Nevertheless, Acemoglu (2025) presents a more conservative outlook, estimating that total AI-driven productivity in the next decade will only experience a modest increase of 0.7%.

[3]Table 1 reports that the number of AI-related academic papers in financial economics rose from 12 in 2018 to 70 in 2024, with 2 (2018) and 14 (2024) appearing in top journals, respectively. The number of GenAI-related papers grew from 0 in 2018 to 30 in 2024, based on a manually curated sample of 234 papers.

[4]Korinek and Suh (2024) project a "baseline" scenario in which GDP doubles over the next decade, while also considering a more optimistic "aggressive" artificial general intelligence scenario where GDP could potentially quadru-

their implications are far-reaching. This prospect, along with the recent wave of AI-related research in financial economics, points to the need to revisit foundational questions in this field, such as how markets function, how firms make decisions, how risks are priced and managed, and how information and technology jointly shape economic behavior and outcomes.

While several prior surveys have made initial efforts to review AI in finance, they tend to focus on selected themes or specific subfields. For example, Eisfeldt and Schubert (2025) specifically discuss GenAI as both a technology shock to firms and a methodological innovation to conduct finance research, Hoberg and Manela (2025) center on the application of NLP in finance research, and Cao et al. (2024c) explore how AI is used to analyze alternative data.[5] To the best of our knowledge, there is no comprehensive survey that systematically examines how AI introduces methodological advances and novel economic implications across the core domains of financial economics. This article fills this gap by offering a structured framework to review and synthesize the evolving literature. By tracing key research themes in financial economics, we aim to provide a comprehensive overview of how AI reshapes our understanding of financial systems and where AI is beginning to play a distinctive role in the economy.[6] We hope to offer a useful roadmap for ongoing and future research, particularly for scholars just starting to explore this field.

We begin in Section 2 by examining current research related to GenAI, focusing on its applications in information analysis, its broader economic impact, and its emerging use cases as economic agents. Section 3 investigates how AI influences firm performance, organizational structure, and corporate decision-making. Section 4 turns to asset pricing, where AI technologies, such as machine learning and deep learning, have improved return predictability, model selection, and investment management. Section 5 explores AI's role in household finance in areas like financial inclusion and financial services. Section 6 examines the labor market impacts

ple. Nobel laureate Geoffrey Hinton, OpenAI CEO Sam Altman, and Anthropic CEO Dario Amodei have all publicly stated that superintelligent systems may arrive within just a few years.

[5]In the area of machine learning, a subfield of AI, Athey and Imbens (2019) discuss its applications in econometrics, and Giglio et al. (2022) review its use in asset pricing. Further, Cao and Zhai (2022) focus on how FinTech improves financial services. Huang et al. (2024b) draw on several recent academic papers and discuss the role of GenAI in financial markets.

[6]Despite AI's relative novelty, this line of research has proliferated, making an exhaustive survey infeasible. We therefore adopt a selective yet representative approach to our review.

of AI, treating it as a technological shock that affects labor demand, occupational exposure, and productivity. Section 7 addresses emerging risks and challenges, including methodological limitations, AI governance, and the policy implications. Finally, Section 8 concludes by discussing promising avenues for future research.

2 Generative AI

GenAI refers to a class of machine learning models typically trained using self-supervised learning and capable of producing outputs that closely resemble human-generated content (Bommasani et al., 2021; Brown et al., 2020).[7] Among the most widely used forms of GenAI are large language models (LLMs), which are designed to generate coherent and contextually appropriate text in response to natural language prompts.[8]. Table 2 shows a range of major LLMs as of January 2026, together with some of their key properties.

Since the public release of ChatGPT in November 2022, interest in GenAI has surged. As illustrated in Figure 2, Google search activity for "ChatGPT" increased dramatically following its launch and rose further after November 2023, when OpenAI introduced GPT-4 into the ChatGPT product. The release of the ChatGPT mobile apps in May 2023 provided an additional boost in mainstream awareness and usage. Search interest continued to rise until it reached its peak in March 2025. The scale of the industry has shifted just as rapidly: the number of available LLMs more than doubled in 2025 alone (253 to over 651), while the cost of using these models has dropped nearly 1,000 times from 2023 to 2025 (Demirer et al., 2025). At the same time, the surge in attention has been accompanied by a growing body of research in financial economics

[7] Table 4 provides a comparison between conventional ML methods and GenAI in financial economics.

[8] LLMs are part of the broader family of foundation models, which are trained on diverse and extensive datasets and can be adapted to a wide range of downstream tasks. It is useful to distinguish between pretrained language models such as BERT (Devlin et al., 2019), which are encoder models trained with masked language modeling objectives, and large generative language models such as GPT-3 (Brown et al., 2020), GPT-4, Claude, or Gemini, which are decoder-based, typically contain tens of billions of parameters, and are trained with autoregressive next-token prediction objectives for open-ended generation. Other foundation models, such as CLIP (Radford et al., 2021), extend the transformer architecture beyond text to multimodal tasks. The backbone for virtually all of these models is the transformer proposed by Vaswani et al. (2017), which relies on self-attention mechanisms rather than recurrent or convolutional layers. See Korinek (2023) for further discussion of the use cases of modern generative AI in economic research.

examining the implications of GenAI for financial market participants, academic researchers, and policymakers.

In this section, we organize this emerging literature into three main strands. First, we review research that employs GenAI as an analytical tool to assist or automate tasks traditionally performed by humans, such as information extraction, forecasting, and conducting academic research. Second, we summarize studies that treat GenAI as an external shock to financial markets, focusing on how its release and adoption affect market efficiency, investor behavior, and regulatory responses. Third, we highlight a growing body of work that explores GenAI as an economic agent capable of simulating reasoning and decision-making in financial settings or autonomously performing tasks such as forecasting, advising, or trading.

2.1 Generative AI as an analytical tool

Figure 3 illustrates that recent papers employing GenAI as an analytical tool draw on a wide range of information sources, spanning from conventional inputs such as SEC filings to more unconventional ones such as social media posts and images. These studies also address a diverse array of research topics.[9] In the subsections below, we discuss four broad use cases of GenAI in financial economics: (1) prediction, (2) information extraction and semantic analysis, (3) task automation, and (4) data generation.[10]

2.1.1 Prediction

Empirical evidence is mixed regarding the predictive capabilities of LLMs. Several studies confirm LLMs' ability to forecast stock returns. Lopez-Lira and Tang (2023) is the first to study the predictive power of ChatGPT-4 on stock returns by analyzing news headlines. A strategy informed by ChatGPT-4's sentiment predictions yields an average daily return of 38 basis

[9] A brief summary of the key inputs and main findings from these papers is presented in Table 5.

[10] Prediction and information extraction are sometimes performed not by generative models themselves, but by discriminative models (e.g., BERT, RoBERTa), which are often fine-tuned LLMs (Devlin et al., 2019). Discriminative models are trained to "discriminate" between different classes, labels, or outputs given an input. Their goal is to model the decision boundary between classes rather than modeling how the data was generated.

points, with stronger predictability among smaller stocks, negative news, and more complex texts. This predictive ability notably increases with model complexity, and predictability persists for roughly two trading days post-news. However, as the adoption of these sophisticated models becomes widespread, their forecasting advantage diminishes, suggesting improved market efficiency. Chen et al. (2022b) comprehensively examine various LLMs using financial news articles across 16 global equity markets in 13 languages. They find that LLM-based models significantly outperform traditional text models in predicting stock returns. At the aggregate stock market level, Chen et al. (2025b) find that the ratio of good news identified by ChatGPT from Wall Street Journal headlines positively forecasts future market returns with significant economic value and predictive accuracy, especially during downturns or periods of heightened uncertainty, while negative news has no predictive power due to rapid investor assimilation. The authors also show that simpler models like DeepSeek and BERT fail to deliver comparable predictive performance, likely due to their limitations in capturing contextual nuances. Beyond news text, Jha et al. (2024a) use ChatGPT to extract firms' expected investment policies from earnings call transcripts and build a "ChatGPT Investment Score" that predicts both future firm-level capital expenditures and stock returns. Similarly, Gao et al. (2025) extract professional fund managers' structured beliefs from mandated disclosures using ChatGPT-4, finding these beliefs significantly forecast fund trading behavior and subsequent market returns.

Beyond stock return prediction, LLMs have also shown promise in other predictive tasks. Athey et al. (2024) successfully apply LLMs to predict workers' subsequent jobs based on their career histories. Cong et al. (2025b) propose a new LLM-based framework that delivers superior performance in forecasting macroeconomic variables. Likewise, Audrino et al. (2024) develop LLM-based uncertainty indices that exhibit stronger predictive power for macroeconomic indicators, asset returns, and fund flows.

Nevertheless, the literature also suggests notable limitations in LLMs' predictive abilities. Chen et al. (2024b) emphasize both the promise and the caution required when deploying LLMs in financial forecasting. They show that while LLMs are better at gauging risks than humans, they exhibit behavioral biases in predictions, such as over-extrapolating past returns. Along similar lines, Ouyang et al. (2024) show that the ethical alignment of LLMs significantly alters their risk preferences and can systematically bias investment forecasts. Furthermore, Li et al. (2024a) document that GPT-4 underperforms human analysts in forecasting GAAP earnings and in predicting directional changes. Despite its strength in general-purpose textual analysis, its quantitative processing capabilities lack consistency, particularly in low-data environments. In addition, GPT-4's forecast accuracy diminishes beyond its knowledge cutoff.

2.1.2 Information extraction and analysis

Perhaps the most important function of LLMs is information extraction and semantic analysis. Traditional textual analysis has relied heavily on simple representations like bag-of-words (BoW) models or dictionary-based sentiment scores, which ignore word order, context, and deeper semantic relationships, leading to significant information loss and inefficiencies.[11] These BoW-based models generate extremely high-dimensional data that require ad hoc dimensionality reductions. In contrast, LLMs can produce much more sophisticated vector embeddings since they are trained on massive, diverse corpora. Accordingly, LLMs enable financial economists to extract deeper and cleaner information from complex, unstructured text.[12] We provide a more comprehensive comparison between traditional textual analysis and LLM-based analysis in Table 3.

In applications involving financial text, Shaffer and Wang (2024) demonstrate that GPT-4 can automate core earnings estimation by extracting and analyzing unstructured data from 10-K filings. Also using 10-K filings, Serafeim (2024) shows that LLMs are effective at extracting nuanced environmental information from corporate disclosures. The promise of using LLMs to decipher

[11] Chen et al. (2022b) argue that existing finance research has only tapped a small part of the textual data landscape and often uses basic representations that lose important context. In addition, Frankel et al. (2022) find that machine-learning methods outperform dictionary-based methods in capturing disclosure sentiment in 10-K filings and conference-call transcripts.

[12] One caveat is that LLMs can underperform custom-trained classifiers for classification purposes. Dell (2025) conduct experiments involving topic classification on historical newspaper articles. The author shows that GPT-4 and GPT-4o outperform older models but generally do not systematically improve over time. GPT models perform well on straightforward topics (e.g., horoscopes, obituaries) but worse on complex or domain-shifted topics (e.g., politics, World War I).

financial texts is also evident in the analysis of social media data.[13] Using 1.9 million posts and 55,732 images from Reddit's WallStreetBets forum, Huang et al. (2024a) show that memes are often posted in response to negative earnings news, likely serving as a coping mechanism for losses. Meme usage spurs higher engagement and encourages retail investors to temporarily hold or double down on losing stocks. Chen et al. (2024c) investigate how retail investors form beliefs and choose trading strategies by analyzing 96 million social media posts from StockTwits using LLMs. They find that retail investors dynamically adopt stock analysis strategies based on information environments.

The economic value of LLM-derived text analytics can be comparable to that of traditional information sources or even quantitative content. Lv (2024) examines analyst reports with LLMs and show that the text component explains more return variation than quantitative forecasts. Li et al. (2026) use LLMs to extract cause-effect links between corporate culture and firm outcomes from diverse information sources such as analyst reports, earnings calls, and employee reviews.[14] Comparing perspectives across analysts, management, and employees, the authors document systematic differences in how each group perceives the causes and effects of corporate culture, shaped by their roles and incentives. Moreover, Bastianello et al. (2024) use LLMs to derive analysts' reasoning from 1.6 million reports. They show that this LLM-derived qualitative reasoning can explain forecast errors and pricing anomalies. Chen and Wang (2024) combine LLM-based text analysis of over 140,000 US AI patents with detailed microdata on worker flows from Revelio Labs, categorizing AI innovations into seven functional areas. They find that different types of AI have heterogeneous effects on occupational employment: creativity-, engagement-, and learning-related AI augment labor demand, while perception-related AI displaces workers. Furthermore, Yoon (2025) shows that incorporating a fine-tuned LLM into iBuyer pricing models

[13] Social media has had an increasing impact on financial markets, driven by the rapid growth of finance-specific platforms such as StockTwits, FinTwit, Seeking Alpha, and Reddit's WallStreetBets. For a comprehensive overview, we refer interested readers to the recent survey on social media and finance by Cookson et al. (2024).

[14] Li et al. (2026) emphasize key design strategies for effectively using GenAI in financial text analysis, including employing step-by-step, chain-of-thought prompting for complex reasoning tasks (Wei et al., 2022), addressing LLMs' limitations with long documents by segmenting texts (Liu et al., 2024), and dynamically augmenting inputs with relevant information from full reports.

substantially mitigates adverse selection by extracting latent home quality signals from unstructured listing text. The resulting "text score" reduces expected iBuyer losses by about 44% and raises the profitability threshold for upfront-payment contracts from roughly 70%.

LLMs have also proven powerful for risk management. Kim et al. (2023b) show that GPT-based risk measures outperform traditional bigram-based methods (e.g., Hassan et al., 2019) by capturing implicit and contextual risk information and offering explanatory narratives. GPT-derived proxies better predict future stock volatility, especially for political and climate risks. Krockenberger et al. (2024) introduce CovenantAI, trained on 580,000 SEC filings, to detect covenant violations. They find that CovenantAI achieves higher accuracy and more reliable violation identification than traditional methods. Beckmann et al. (2024) design a three-step prompting strategy using ChatGPT to identify "unusual" financial communication in earnings calls, such as a CEO displaying unpreparedness, addressing non-traditional topics, or responding aggressively. These types of communication introduce ambiguity and informational frictions and can serve as red flags for investors. Further, Wu et al. (2025b) apply ChatGPT on the loan assessments reports of 2,460 micro and small business loans. Human-written texts differ from ChatGPT-refined texts in terms of text length, semantic similarity, and linguistic representations. The authors show that incorporating unstructured, ChatGPT-refined text alongside conventional structured data significantly improves credit default predictions.

Hansen and Kazinnik (2023) apply LLMs in the macro-policy domain. They evaluate GPT-3.5 and GPT-4 models for interpreting Federal Reserve communications, showing that LLMs outperform traditional NLP methods in classifying policy stance. The study also finds that market reactions to FOMC announcements intensified after the release of GPT-4, suggesting that LLMs can improve public understanding of central bank communications. Similarly focused on policy, Fang et al. (2024) conduct a large-scale analysis of China's industrial policy using over 3 million government documents from 2000 to 2022 with Gemini-1.5. The study identifies 770,000 industrial policies and documents rich variation across regions, industries, and over time, highlighting how local governments select sectors and experiment with policy tools that evolve with industry development. They show that industrial policies boost firm entry and sometimes productivity, though effects depend heavily on implementation quality and tool choice.

Extending the use of LLMs to historical and cultural contexts, Jha et al. (2025) apply BERT-based embeddings to millions of books and construct a financial sentiment index covering eight large economies from 1870 to 2009. They find that declines in public sentiment toward finance predict future banking crises, even after controlling for traditional indicators like credit growth. This suggests that public trust in finance is an underappreciated early warning signal for financial instability. Their study demonstrates the power of LLMs to extract nuanced, context-aware sentiment from unstructured historical texts at scale, something previously infeasible using traditional sentiment analysis methods.

Furthermore, Gabaix et al. (2024) propose a natural and promising extension of embedding methods—vector representations learned from portfolio holdings. They treat assets in an investor's portfolio like words in a sentence. They train models (e.g., BERT) to predict missing assets based on the rest of the portfolio, analogous to predicting a missing word in a sentence. The authors show that asset embeddings explain relative valuations, stock return comovement, and institutional portfolio choices better than traditional firm characteristics or text-based embeddings. Kakhbod et al. (2024) use LLMs to embed textual descriptions of 10-K filings and patent abstracts from the USPTO and compute their semantic similarity. They develop a firm-level measure of Innovation Displacement Exposure (IDE) to quantify how innovations by other firms can disrupt a given firm's future growth through technological obsolescence. They show that higher IDE is strongly and persistently associated with lower future profit growth, particularly over long horizons. Additionally, IDE also predicts declines in employment, output, and intangible capital, supporting the interpretation of IDE as a form of obsolescence risk.

2.1.3 Automating tasks

LLMs are increasingly used to automate tasks traditionally requiring significant human expertise. Kim et al. (2023a) show that GPT-3.5 can automate the task of processing complex financial disclosures by generating concise summaries that better capture decision-relevant information than full documents. By comparing summary sentiment and LLM-generated textual embeddings to those of original documents, they find that summaries are more predictive of both concurrent and future returns. By filtering out irrelevant "bloat" disclosures, LLMs help investors quickly access key insights, reducing information processing costs traditionally handled by human analysts. Huang et al. (2023a) use the BERT algorithm to build a finance-specific LLM, FinBERT. They train FinBERT on financial texts and fine-tune it for tasks like sentiment and ESG classification. The authors find that FinBERT outperforms traditional dictionaries, classical ML models, deep learning algorithms, and even standard BERT in identifying negative sentiment and capturing market-relevant information in earnings calls. From a practitioner's perspective,

The rapid advancement of LLMs, particularly in analytics and reasoning abilities, has led to a range of new scientific applications, including solving complex mathematical problems (Trinh et al., 2024), proof writing (Collins et al., 2024), retrieving related literature (Ajith et al., 2024; Press et al., 2024), and generating code for analytical and computational tasks (Huang et al., 2023b; Tian et al., 2024). Although these developments have the potential to significantly boost research productivity, it remains uncertain whether LLMs can tackle the more creative and intellectually demanding aspects of the research process: research ideation. Si et al. (2024) are among the attempts to answer this question by conducting a large-scale, controlled experiment comparing LLM-generated ideas within the field of NLP to those produced by over 100 expert NLP researchers. They find that LLM ideas are judged significantly more novel than human ideas, though slightly less feasible. Similarly, Meincke et al. (2024) find that GPT-4 can produce new product ideas with higher purchase intent than those generated by top MBA students. Furthermore, LLMs can facilitate rigorous academic research by systematically generating and justifying candidate instrumental variables through carefully designed, multi-step, role-playing prompts, thus aiding in the establishment of causal inference (Han, 2024).

More relevant to the field of financial economics, Korinek (2023) identifies six major use cases for LLMs in economic research: ideation and feedback, writing, background research, coding, data analysis, and mathematical derivations. They provide instructions and a systematic rating of LLMs' usefulness in each area. While LLMs are already highly effective for automating tasks in research, the authors stress the need for critical human oversight due to LLMs' tendency to hallucinate and produce authoritative but inaccurate content. In a specific research topic, Novy-Marx and Velikov (2025) demonstrate that LLMs can mine accounting data to identify over 30,000 potential return predictors and find 96 significant signals. For each predictor, the authors use GPT-3.5-turbo and Claude 3.5-Sonnet to automatically generate full academic papers, including hypotheses, theoretical explanations, data analysis, results, and conclusions. Nevertheless, they raise serious concerns that AI's ability to mass-produce post-hoc explanations could overwhelm peer review processes, artificially inflate citation metrics, and undermine the credibility of the work.

2.1.4 Deriving new data

Early studies in finance and economics applied textual analysis to generate new structured data from unstructured sources. Notable examples include, among others, text-based network industry classifications (Hoberg and Phillips, 2016), firm-level political risk measures (Hassan et al., 2019), market-implied volatility based on news articles (Manela and Moreira, 2017), and firm-level measures of financial constraints (Buehlmaier and Whited, 2018).[15]

Building on these foundations, recent studies have begun to integrate GenAI into creating new economic measurements. For example, Jha et al. (2024b) use ChatGPT to construct forward-looking measures based on earnings call transcripts. These AI-based expectation indices can predict future GDP growth, industrial production, employment, and firm performance. Cao et al. (2025) use Google's Bard to identify product market peers for publicly listed US firms from 2003 to 2022. They find that AI-identified peers overlap significantly with those selected by human experts and outperform traditional methods like text-based similarity (TNIC) in capturing stock return and accounting fundamental correlations. Bartik et al. (2024) create a comprehensive clas-

[15]See Hoberg and Manela (2025) for a comprehensive review.

sification of US municipal housing regulations with LLMs. Their method achieves near-human accuracy: 96% on binary classification tasks and a correlation of 0.87 on continuous regulatory measures.

At a global level, Breitung and Müller (2025) construct global, time-varying business networks by applying GPT-3 to historical business descriptions from over 63,000 firms across 67 countries. Their similarity-based networks outperform traditional word-based methods (e.g., Hoberg and Phillips, 2016) in predicting lead–lag effects in stock returns and identifying M&A targets. On a similar scale, Fetzer et al. (2024) introduce an AI-generated global production network mapping input-output relationships across more than 5,000 product categories. They employ a novel "build-prune" methodology: first, they use LLMs to generate a large distribution of potential links between thousands of product categories, and second, the candidate connections generated during the build phase are rigorously evaluated and filtered to remove irrelevant or incorrect links. They use this production network to show that recent global trade has shifted toward more upstream, central products such as semiconductors and critical minerals. Furthermore, using a global sample of 3,769 firms across 35 countries, Dyck et al. (2025) employ a retrieval-augmented generation (RAG) LLM to measure family owners' environmental preferences based on public information such as philanthropy, advocacy, and green investments. Interestingly, they find that these closely controlled firms are not cleaner than widely held firms on average. Only when controlling families exhibit strong environmental preferences and the marginal cost of reducing emissions is low do these firms significantly lower their carbon output.

2.2 Generative AI as an external shock

The emergence of GenAI has acted as a profound external shock to financial markets, reshaping investor behaviors, redistributing informational advantages, and altering labor market dynamics. Several recent studies document how GenAI's sudden availability, or its temporary absence, affects economic activity.

At the investor level, Cheng et al. (2025) show that both retail and institutional investors rely on GenAI for trading-related tasks. They document a significant decline in stock trading volume during eight major ChatGPT outages. Chang et al. (2023) find that the release of ChatGPT narrowed the information gap between retail investors and institutional short sellers. Retail investors' trading decisions became significantly more aligned with AI-generated sentiment, suggesting that GenAI can democratize access to sophisticated information processing. Similarly, Lu et al. (2023) demonstrate that ChatGPT can act as a retail investor's robo-advisor that interprets complex information and provides stock recommendations based on news articles and policy announcements. They also find that ChatGPT outperforms traditional textual analysis methods, such as cosine similarity, in both predictive power and recommendation quality. Extending this line of research, Lu (2025) conducts a large-scale randomized controlled trial with over 28,000 investors at a major Chinese brokerage firm. They find that GenAI-powered robo-advisors significantly improve financial literacy and shift investor behavior toward more diversified, cost-efficient, and risk-aware investment choices.

At a more granular scale, using 1.7 million query answer pairs from a major Chinese GenAI platform in early 2024, Ecker et al. (2026) study how retail investors use GenAI for stock research. They show that investors rely on GenAI mainly to process and interpret firm specific information rather than to directly predict stock prices. GenAI supports a transition from basic awareness to deeper analytical tasks, and users learn from earlier interactions by adjusting subsequent queries in response to prior answers. Higher GenAI usage is associated with more active and more information driven trading, although these relationships are correlational rather than causal.

However, there is large heterogeneity in the usage of GenAI. Ecker et al. (2026) find that more engaged and financially sophisticated users rely on GenAI for integration tasks, while casual users mostly seek basic awareness. Consistent with this evidence, Blankespoor et al. (2024) find that sophisticated investors are far more likely to integrate GenAI tools than novice investors, implying that while GenAI reduces some barriers, it may also exacerbate inequality depending on access and skills.

Distributional effects are also observed in the study of Guo et al. (2022). The authors find that while GenAI-based investment consultants improve individual investors' investment returns, novice and risk-averse investors gain less, suggesting that GenAI may unintentionally widen performance gaps among investors. Hirshleifer et al. (2025) document stark heterogeneity across social media platforms: GenAI adoption reduces communication frictions and improves price discovery on Seeking Alpha, a well governed and expert oriented platform, but is instead associated with speculative episodes and noise trading on WallStreetBets, where users are less sophisticated and governance is weaker.

GenAI also provides incremental benefits to more sophisticated market participants. Sheng et al. (2024) measure hedge funds' GenAI adoption using a novel Reliance on AI Information (RAI) index. They find that AI adoption surged in 2022, especially among large, active, and high-performing funds. Higher RAI is associated with significantly better fund performance, suggesting GenAI enhances decision-making. However, the benefits are unevenly distributed, with smaller and passive funds seeing little improvement, implying that AI may widen performance disparities. Christ et al. (2024) provide survey-based evidence on the adoption and impact of AI tools in sell-side equity research. Based on responses from 190 equity analysts and follow-up interviews, the authors find that 58% of analysts use AI, primarily to streamline existing processes such as summarizing text and collecting data, rather than for novel forecasting or independent stock recommendations. Analysts who use AI more frequently produce more timely and accurate forecasts. Complementing these findings, Bertomeu et al. (2025) use the temporary ban of ChatGPT in Italy as an exogenous shock to show that ChatGPT affects the information processing capabilities of financial analysts. They find that domestic analysts, especially those with technical backgrounds, reduce their use of AI-generated content during the ban, leading to fewer and less accurate forecasts. The resulting informational efficiency of capital markets also declines. In a related vein, Da et al. (2025) show that the 2023 rollout of Chinese LLMs reduced the long-standing A-share premium over identical H-shares by about 5%, indicating that Chinese LLMs mitigated information asymmetry by exposing domestic investors to less censored, more comprehensive information.

In the broader labor market, GenAI also represents an unexpected shock. Eloundou et al. (2024) find that about 80% of US workers could see at least 10% of their tasks affected by LLMs, and 19% could see over 50% affected. Interestingly, higher-wage jobs exhibit greater exposure, as a larger share of their tasks can potentially be performed or augmented by GenAI. While impacts may vary across occupations and industries, GenAI is rapidly integrating into the workplace and meaningfully influences productivity. A national survey by Bick et al. (2024) on GenAI adoption in the US in 2024 finds that approximately 26% of employed respondents use GenAI at work, a rate comparable to early personal computer adoption but faster than that of the internet. Occupations with higher predicted exposure to GenAI tasks show correspondingly higher actual adoption rates. Users report average time savings of 5.4%, suggesting a potential aggregate productivity gain of approximately 1.1% at current adoption levels. Further, Brynjolfsson et al. (2025) provide early evidence of task-level effects of GenAI deployed at scale in the workplace: GPT-based chat assistants improved the productivity of customer service agents by 15%, particularly among lower-skilled workers. Nevertheless, Humlum and Vestergaard (2025) suggest that GenAI has small labor market effects using two large-scale surveys in Denmark of 25,000 workers covering 11 exposed occupations. While AI chatbots have been rapidly adopted, workers report only modest time savings and no meaningful effects on earnings, hours, or employment. Their findings suggest that the transformative potential of GenAI may be overstated in the short run and highlight the importance of organizational context in realizing productivity and economic benefits.

2.3 Generative AI as economic agents

GenAI, particularly LLMs, can be viewed as implicit computational models of human behavior or economic agents due to the way they are trained and structured (Horton, 2023). In this framing, LLMs function as decision-making entities with embedded preferences and beliefs, capable of participating in economic environments much like individuals or firms. Treating an LLM as an economic agent involves assigning it an objective and analyzing its responses to economic tasks, such as making consumption choices under budget constraints, negotiating prices, forecasting financial trends, or responding to incentives, as one would analyze the behavior of a human subject. This perspective has sparked a growing body of research investigating whether LLMs behave in economically rational ways, whether they display human-like cognitive biases, and whether they can simulate or predict real-world economic outcomes. In essence, the possibility of GenAI being a stand-in for human agents allows economists to test theories and conduct virtual experiments in ways that were not previously possible.

Current studies in this emerging field reveal a mix of human-like and unique traits in LLM behavior. Ouyang et al. (2024) explicitly frame LLMs as AI decision-makers, demonstrating that these models exhibit stable but diverse risk preferences and that ethical alignment systematically shifts their willingness to take economic risks. This reinforces the notion that LLMs can be treated as autonomous agents in economic analysis. Ross et al. (2024) investigate whether LLMs conform to the classical "homo economicus" ideal (i.e., perfectly rational economic agents) or display systematic biases similar to humans. They find that, compared to humans, LLMs generally exhibit weaker loss aversion, similar risk aversion, and stronger time discounting. However, LLM behavior often lacks consistency across different contexts, and prompting techniques can unpredictably influence their biases. Similar findings are reported by Chen et al. (2023b), who examine GPT's economic decision-making across various budget allocation tasks, comparing its performance to that of a representative sample of 347 humans. They show that GPT achieves higher rationality scores than humans across all domains, although GPT's responses remain sensitive to question framing—a trait reminiscent of human decision-making. Extending this analysis, Bini et al. (2025) test major LLM families (ChatGPT, Claude, Gemini, and Llama) across multiple versions and scales. They find that more advanced and larger-scale LLMs become more human-like (and less rational by expected utility standards) in preference-based decisions but more rational in belief-based decisions, while small-scale LLMs tend to behave more irrationally.

A prominent question is whether GenAI can forecast economic outcomes or interpret information as a human expert would. LLMs have been found to demonstrate notable predictive capacities in financial contexts, though the performance varies. Chen et al. (2024b) examine ChatGPT's ability to forecast stock returns from historical price data. They find that, similar to human biases documented in prior studies, the AI's forecasts display an extrapolation bias and overweight recent data. Moreover, ChatGPT's predictions were overly optimistic on average; however, it demonstrates greater calibration accuracy than human forecasts regarding risk and return intervals. Bybee (2023) introduces an approach to "generate beliefs" from news: the author feeds historical news articles into GPT-3.5 and queries it for economic expectations (such as inflation or growth forecasts). Strikingly, the expectations the AI produces closely track actual human expectations from contemporary survey data. Hansen et al. (2024) take a further step forward to mimic human experts. They create synthetic forecaster personas based on detailed participant characteristics and provide them with real-time macroeconomic data to generate simulated responses. They find that LLM-generated forecasts are generally similar to human forecasts but often more accurate, especially at medium- and long-term horizons. This superior performance stems from LLMs' ability to extract latent information from past human forecasts while avoiding systematic biases and noise.

A more complex application of GenAI is to simulate multiple agents and complex interactions. For example, Fish et al. (2024) integrate GPT-4 into an algorithmic pricing task. They let two LLM-based agents set prices in a simplified market environment and found that the AI agents quickly learned to avoid undercutting each other, effectively arriving at tacit collusion that kept prices high. Nevertheless, minor variations in the prompts significantly influenced pricing behavior. Horton (2023) conducts a series of classic game-theoretic experiments with GPT-based agents (dubbed "Homo silicus") in roles such as bidders, bargainers, or players in a Prisoner's Dilemma. He finds that LLM agents can indeed engage in strategic behavior: they often learn to cooperate in repeated Prisoner's Dilemma rounds and make decisions in dictator and ultimatum games that fall within the range of human behavior. Extending this idea, Manning et al. (2024) present an automated social science framework where LLMs are used both as subjects and as the experimenter. They formalize scenarios (like an auction or a job interview) using a structured causal model and let LLM agents interact according to this setup. Impressively, the outcomes of these LLM-driven simulations reflected known economic principles.

Within the field of financial economics specifically, **kazinnik2023bank** simulates a bank run by assigning GPT-based agents and having them decide whether to withdraw funds under rumors of bank distress. The aggregate withdrawal patterns from this simulated population showed remarkable alignment with real-world behavior observed in past bank runs. Zarifhonarvar (2024) uses GPT models as respondents in a macroeconomic survey to state their inflation expectations. They find that these AI agents produce expectation distributions and reaction patterns that mirror those of human surveys.

3 AI and Corporate Finance

This section examines how AI influences firm performance, organizational change, risk and disclosure behavior, and how corporate information is produced, interpreted, and strategically managed.

3.1 AI investment and firm performance

Industrial firms have now entered a new era empowered by AI-related technologies, and businesses across the economy are going through technological transformation (Brynjolfsson, 2014).[16] Nevertheless, the benefits of AI adoption for firms are not obvious and have been an open question. On the one hand, firms can improve efficiency using AI to automate high-skilled tasks (Webb, 2019). On the other hand, AI might still be in its infancy to have a meaningful impact on firm performance (Brynjolfsson et al., 2019). There are two broad mechanisms through which AI can potentially benefit firms. First, AI can facilitate product innovation, a key driver of business growth (e.g., Klette and Kortum, 2004). AI has the capability to foster knowledge accumulation by

[16] In finance, FinTech represents the first wave of applying AI-related technologies to drive substantial innovation in financial services (e.g., Chen et al., 2019; Berg et al., 2022; Hau et al., 2024; Chen et al., 2025a) and promote the growth of the real economy (e.g., Agarwal et al., 2025).

streamlining experimentation and lowering the costs associated with product innovation (Bustamante et al., 2019). Executive surveys have indicated that the most common application of AI to date has been to improve existing products and services, as well as to develop entirely new product offerings.[17] The second mechanism is through reducing the costs of process innovation. The cost reduction can be achieved in at least two ways: cutting per-unit labor costs and increasing operational and production efficiency through its forecasting ability.

There are two central challenges in estimating the firm-level impact of AI: the paucity of granular data on investments in AI and the lack of causal evidence (Seamans and Raj, 2018; Frank et al., 2019). Babina et al. (2024) is among the first to provide both firm-level data and evidence of causality. They measure the stock of and demand for AI-related workers of each firm using detailed resume and job posting data, respectively. They find that AI investment and firm size form a positive feedback loop: larger firms invest more in AI, which subsequently increases their sales, employment, and market share. This AI-driven growth stems from product innovation and increased product offerings rather than cost cutting. These findings are important as they differentiate the effects of AI from the productivity shock of information technology (IT) in the 1980s and 1990s (Dedrick et al., 2003). For causality, the authors employ an instrumental variable based on firms' ex-ante exposure to future AI talent supply, measured by the hiring networks of universities with a strong AI research history.

In a related study, Fedyk et al. (2022) analyze over 310,000 employee resumes from 36 major US auditing firms to show that AI talent improves product quality by reducing financial restatements and fees while boosting productivity. However, the benefits are unevenly distributed: senior partners gain, whereas junior employees face increased displacement risk. Adams et al. (2025) use US online job postings since 2010 and show that AI pricing adoption has grown rapidly and is concentrated in large productive and R&D intensive firms. Adopting firms grow faster earn higher markups and have stock returns that are more sensitive to monetary policy shocks. Zhang (2024) measures AI adoption of mutual funds based on job postings and shows that mutual funds

[17] The survey by Deloitte can be accessed at https://www2.deloitte.com/content/dam/insights/us/articles/4780_State-of-AI-in-the-enterprise/DI_State-of-AI-in-the-enterprise-2nd-ed.pdf

with higher AI adoption outperform their peers. A long-short strategy based on this AI ratio yields an annualized excess return of 2.89%. The performance gains are attributed to enhanced stock-picking ability, especially for stocks with high volumes of public information, while AI is less effective for opaque or underreported stocks. Wu et al. (2025a) study companies' AI analytics by combining firm-level job postings, hiring patterns, and skill requirements. They reveal a mitigating role of AI in the decline in innovation post-IPO: acquiring AI talent helps ameliorate short-term financial and disclosure pressures, allowing firms to sustain innovation after going public. In the context of credit markets, Gambacorta et al. (2025) use unique data on Italian banks to demonstrate that AI-based credit scoring improves firms' borrower evaluation accuracy by relying on hard, verifiable data, hence reducing reliance on soft information. Zheng (2025) uncovers an indirect channel through which AI affects firm performance. He finds that ML-assisted patent screening increases patent quality by 17–26%. Such improvements translate into higher public firms' ROA and raise private firms' likelihood of IPO or M&A.

While much of the existing literature focuses on firms' adoption of AI technologies, a complementary line of research has begun to examine the production of AI innovation. An increasing number of studies use patent data as a proxy for AI-related technological activity (e.g., Webb et al., 2018; Cockburn et al., 2018). However, firm-level evidence linking AI innovation to productivity and employment remains limited. Alderucci et al. (2020) identify AI-related patent grants by the United States Patent and Trademark Office (USPTO) and match them with the US Census microdata. They find that AI innovation is associated with increases in revenue, value-added per employee, and within-firm wage inequality. Not surprisingly, these effects are more pronounced in large, R&D-intensive firms. Although the findings are suggestive, the authors caution against drawing causal conclusions. Expanding this line of work, Ahmadi et al. (2023) explore patent data from the USPTO and document that the share of AI innovation in all patenting activity rises from 5% in 1990 to between 15% and 35% today. They show that AI innovation increases labor productivity, reduces physical capital intensity, and improves financial flexibility, while leaving firm size and employment largely unchanged. This suggests AI complements rather than substitutes labor.[18] The effects of AI innovation depend on its type. Chen and Wang (2024) show that augmenting AI (e.g., creativity, learning) increases employment and firm productivity, while displacing AI (e.g., perception) reduces labor costs without raising productivity. Despite these differences, all AI types raise firm valuation (Tobin's Q), with the gains shaped by labor market frictions. These findings are established using instrumental variables based on quasi-random patent examiner (different leniency) assignment to address endogeneity.

AI innovations are increasingly valuable to firms. Using US patent data from 1995 to 2020, Chen et al. (2024d) find that AI patents are approximately 9% more valuable than non-AI patents, a premium driven by higher knowledge spillovers (evidenced by 26% more forward citations) and greater commercialization potential, reflected in improved future profit margins. The value premium of AI innovations has risen steadily over time, particularly in sectors where occupational tasks align with AI capabilities. The study also reveals sectoral specialization: IT firms primarily develop general AI technologies, while non-IT firms focus on application-specific innovations.

Chen et al. (2024d) identify two key policy changes that significantly enhanced the economic value of AI innovations. First, the implementation of the 18-month patent publication rule under the American Inventor Protection Act (AIPA) in 2000 mandated that patent applications be published 18 months after filing, regardless of whether the patent was ultimately granted. This policy accelerated the public disclosure of technical information, including applications that were never granted, thereby expanding the pool of shared knowledge and enabling broader access to cutting-edge innovation. This policy was associated with a 5% increase in the relative value of AI patents in application sectors. Second, Google's decision in 2015 to open-source TensorFlow (a deep learning framework) marked a significant inflection point in the democratization of AI development tools. This unexpected move not only lowered barriers to entry for AI experimentation and deployment across industries, but also encouraged rapid diffusion of technical expertise beyond the IT sector. They find that this event increased the relative value of AI patents in ap-

[18] To establish causality, the authors use an instrumental variable based on the interaction between a firm's tax credit–induced R&D capital (capturing exogenous variation in innovation capacity) and its industry's exposure to AI.

plication sectors by 2%.

Yang (2022) provides international evidence from Taiwan's electronics industry. They find that while AI technology is positively associated with productivity and employment, non-AI patents also generate pro-productivity and pro-employment effects with a magnitude similar to that of AI technology. They further show that AI innovation is not universally productivity-enhancing and may have varied implications depending on firm-specific conditions, task types, and skill levels.

The market appears to price in AI's impact on firm value. Rock (2019) studies the effect of Google's open-sourcing of TensorFlow and find that firms with greater pre-existing AI talent experienced significant valuation gains—up to $3.3 million per 1% increase in AI-skilled employees—driven by revised investor expectations rather than immediate productivity gains.[19] Eisfeldt et al. (2023) use information on the tasks involved in each occupation and evaluate firms' exposure to ChatGPT based on whether each task can be done productively by ChatGPT. They find that ChatGPT represents a shock to corporate valuations. An "Artificial-Minus-Human" (AMH) portfolio that is long high-exposure firms and short low-exposure firms generated daily returns of 0.44% in the two weeks following the introduction of ChatGPT. The valuation effects exhibit large industry variation. Publishing, information, and computing-related industries experience positive returns following the release of ChatGPT, while finance and transportation-related industries have negative returns overall. Nevertheless, the authors do not provide direct evidence on the mechanisms behind this impact. International evidence is also consistent. Lu et al. (2024) find that AI adoption disproportionately benefits large Chinese firms with better access to data and technical talent, leading to reduced exit rates, increased concentration, and higher attractiveness to equity investors.

[19] Lee et al. (2025) also show that firms with a higher proportion of jobs exposed to AI exhibited more significant stock price responses after the introduction of low-cost, open-source AI models.

3.2 AI-driven organizational transformation

AI can be an important driver of skill-based organizational change.[20] Babina et al. (2023a) show that AI investments can potentially reduce the reliance on traditional hierarchical layers and enable lower-level employees to perform more complex tasks independently. Using resume and job posting data, they find that firms increasing their AI investments tend to employ a more highly educated workforce, predominantly with STEM degrees. These firms experience shifts towards flatter organizational hierarchies with an increase in junior roles and a decrease in middle-management and senior positions. Consistent with these findings, Abis and Veldkamp (2024) find that just like industrialization once allowed workers to produce more goods, AI now allows knowledge workers to do more with data. They study job postings, wages, and hiring patterns and show that data become more powerful with machine learning—firms benefit more from each unit of data, and they need fewer people per unit of data.

This skill-based change is further fueled by the increasing importance of big data due to machine learning's capability to reduce diminishing returns to data (Farboodi et al., 2022). While Farboodi et al. (2022) suggest that data could be valued differently by different entities, Jones and Tonetti (2020) argue that data is nonrival—that is, data can be used simultaneously by multiple parties without depletion. Because of this property, data has the potential to generate large social gains if broadly shared.[21] Beraja et al. (2023) confirm the benefit of shared data by showing that access to government data spurs commercial AI innovation. Using China as an experimental ground, where AI firms access rich surveillance data through public security contracts, the authors find that firms receiving data-rich contracts increase their commercial software output. Mihet et al. (2025) argue that market power in AI driven markets depends as much on data access

[20] At the global level, a survey-based study by Bughin (2020) on 3,000 firms across ten countries reveals that AI adoption is far from uniform. While AI adoption is still nascent—fewer than 10% of firms have broadly deployed AI tools—firms express varied expectations: some foresee workforce reductions, while many anticipate employment growth or corporate reorganizations.

[21] From the perspective of the overall welfare of the economy, Jones and Tonetti (2020) show that firm ownership of data can lead to excessive data hoarding, while government restrictions, though protecting privacy, may greatly reduce the economic value of data. In contrast, consumer ownership of data, where individuals balance privacy with incentives to sell their data, yields outcomes closer to the social optimum by enabling widespread use of data and maximizing welfare.

as on AI algorithms themselves. Using shocks from AWS in 2006 and transformer models in 2017, the authors show that cheaper compute favors data intensive firms, whereas better processed data benefits low AI firms. Bian et al. (2025) develop the first large-scale mapping of inter-firm data-sharing networks and show that data functions as a form of productive, non-rival capital that creates strong economic linkages across firms. They find that firms connected via shared data channels experience significant comovement in operational performance and stock returns.

These organizational changes can be path-dependent. Schubert (2025) present evidence of the existence of an "organizational technology ladder": initial technological adaptations determine subsequent technology adoption trajectories. The author finds that the initial shift in building digital infrastructure necessary for effective remote operations during the COVID-19 pandemic sets the stage for companies' subsequent integration of GenAI. Firms that hire remotely also hire for GenAI skills. This path dependency can exacerbate heterogeneity in corporate productivity.

AI itself is also a useful tool in facilitating corporate decision making. Campello et al. (2023) argue that managerial decision-making is essentially a robust control problem characterized by high dimensionality, nonlinearity, dynamic learning, and evolving complexity. The authors introduce AlphaManager, a model leveraging deep learning and offline reinforcement learning to analyze managerial decisions and optimize firm outcomes based on real-world data. The study shows that AI can generate high-dimensional, dynamic, and counterfactual insights that guide corporate policy more effectively than conventional econometric models. Using Compustat and CRSP data on nearly 20,000 US firms from 1976 to 2023, AlphaManager demonstrates high predictive accuracy in corporate outcomes and achieves strong performance gains: optimizing short- and long-term market capitalization results in quarterly outperformance of 10.1% and 8.7%, respectively, over real managerial decisions.

While AI improves managers' ability to assess the probability of a project's success, Chen and Han (2024) advocate the necessity of designing a corporate governance structure to effectively address the agency problem in AI-augmented decision-making. They argue that AI adoption can unintentionally backfire, leading to suboptimal decisions that reduce shareholder benefits. They show that additional information provided by AI raises the likelihood that the manager's estimate of the project's success probability falls within a range where managerial compensation is hardly linked to decision making, resulting in the pursuit of private benefits over shareholder interests.

In the same spirit as Campello et al. (2023), Erel et al. (2021) demonstrate that AI has the potential to inform better corporate governance. They treat director selection as a prediction problem and apply multiple machine learning algorithms to predict director performance. Based on firm, board, and candidate characteristics from a large dataset of US public firms between 2000 and 2014, they show that the algorithm-selected directors would outperform many real-world selections, since director selection by humans may be influenced by bias or agency issues rather than optimal decision-making. In a related vein, Bubb and Catan (2022) shed light on mutual funds' corporate governance preferences via machine learning methods. Analyzing votes on over 180,000 proposals across nearly 5,800 companies by more than 4,600 funds, the authors apply unsupervised learning to identify key dimensions that explain how mutual funds vote. They show that large passive funds are more likely to challenge management on key governance issues than previously assumed, while smaller passive funds relying on proxy advisors may contribute to their growing influence.

Overall, AI technologies are redefining how firms operate and govern. Yet, realizing their full potential in corporate settings requires thoughtful integration, complementary investments in human capital, and robust governance structures to address emerging agency concerns. Furthermore, the increasing integration of AI into corporate decision-making raises important questions about how humans and AI can best collaborate. Since AI is more likely to augment rather than replace human decision-making, designing effective human-AI interaction frameworks will be central to promoting sustainable and productive organizational transformation.

3.3 Technological shift and firm-level risk

Theoretically, the link between firms' risk and their AI investments is unclear ex ante. As a prediction technology (Agrawal et al., 2019), AI can enhance firms' ability to adapt to evolving market conditions, potentially reducing their exposure to systematic risk. However, two channels suggest that AI adoption could heighten firms' systematic risk. First, AI may introduce greater fragility, exacerbating downside risk during market downturns. Second, AI can increase systematic risk by generating growth options, which is likely to make firms more sensitive to market movements, particularly in upswings (e.g., Carlson et al., 2004; Pástor and Veronesi, 2009), which is consistent with the role of AI in driving product innovation.

Babina et al. (2023b) provide the first empirical evidence showing AI investments lead to higher firms' systematic risks. Specifically, a one-standard-deviation increase in the AI-related labor share is associated with a 0.05 increase in market beta. This increase in systematic risk is concentrated in upside beta and is more pronounced during periods of heightened media coverage and investor attention to AI. Despite the increase in systematic risk, total return volatility remains unchanged due to offsetting declines in idiosyncratic and cash flow volatility, suggesting a reallocation rather than an amplification of risk. Importantly, these effects appear unique to AI. Similar increases in beta are not observed for other forms of technological investment, such as IT, robotics, R&D, or organizational capital. Moreover, the rise in systematic risk is not attributable to changes in investor composition, firm size, or financial leverage, nor is it concentrated within the tech sector. Instead, AI-investing firms become more correlated not only with their own industries but also with the broader market, reinforcing the notion that AI is fundamentally reshaping firms' risk exposure and market comovement.

In contrast, Ahmadi et al. (2023) document that firms increasing AI innovation experience a reduction in both systematic and idiosyncratic volatility. They employ instrumental variables based on R&D tax credits and occupation-level AI exposure to show that a 10% increase in AI patenting leads to a 5–6% reduction in the volatility of net income and profit margins, as well as a roughly 2% decline in stock return volatility. These reductions in cash flow and return volatility suggest that AI innovation enhances operational stability. This evidence underscores the stabilizing effect of successful AI development and distinguishes the risk implications of AI production from those of mere adoption or investment in other technologies.

3.4 Dissecting corporate information in the age of AI

The rise of AI is fundamentally transforming how corporate information is produced, communicated, and interpreted in financial markets. AI-driven data analytics are being increasingly adopted by corporate stakeholders to process financial information. Firm information has been increasingly scrutinized by machines. According to Cao et al. (2023a), the share of machine downloads of SEC filings (Forms 10-K and 10-Q) increased from under 40% in 2003 to over 80% after 2015. These algorithmic syntheses and disseminations of information are generally less biased (Cardinaels et al., 2019) and can increase firms' trading volume and liquidity (Blankespoor et al., 2018). Textual analysis, such as computational linguistics and NLP, has been widely adopted to analyze corporate disclosures in extensive academic research.[22] More recently, academics have started to employ ML techniques to examine corporate disclosure. For example, they use financial statements to predict future earnings changes (Chen et al., 2022a), to detect accounting fraud (Bao et al., 2020), and to conduct financial statement analysis (Amel-Zadeh et al., 2020).

Beyond mandatory filings and voluntary corporate disclosures, data can be generated and extracted with AI tools from many alternative sources. In a survey article by Cao et al. (2024c), they highlight the expanding importance of alternative data, such as earnings calls, ESG disclosures, and social media. They offer a comprehensive overview of how AI is reshaping the information environment, fundamentally changing both how financial information is processed and how firms engage with stakeholders. AI-driven data analytics are also extended to analyze firms' complex firm attributes. Li et al. (2021) apply machine learning methods to measure corporate culture from earnings call transcripts. By analyzing unscripted Q&A sections from over 200,000 calls across 7,500 US firms, they generate firm-level culture scores that are validated against known indicators and shown to correlate with stronger operational efficiency, risk-taking, long-term executive incentives, and higher firm value—especially during downturns. The study also finds that

[22] It is unrealistic to provide an exhaustive review of the textual analysis literature in financial economics. Therefore, we highlight a few survey articles that review major methodological and empirical developments from different perspectives (Li et al., 2010; Loughran and McDonald, 2016; Loughran and McDonald, 2020; Hoberg and Manela, 2025).

cultural alignment influences M&A activity and that acquirers' cultures shift post-merger.

Beyond textual disclosures, researchers have explored nontraditional channels of communication. Mayew and Venkatachalam (2012) investigate how vocal cues from managers during earnings conference calls affect capital markets. Using audio files and Layered Voice Analysis (LVA) software, they measure managers' emotional states during the unscripted Q&A portions of calls, finding that both positive and negative affective cues are significantly associated with contemporaneous stock returns, future earnings surprises, and the tone of future firm disclosures. Baik et al. (2024) develop a deep learning–based measure of managers' vocal delivery quality (VDQ) from earnings call audio and finds that VDQ declines when firms report bad or uncertain news. Lower VDQ reduces real-time trading volume and price reactions—especially among retail investors—and weakens analyst engagement and media response. Hobson et al. (2012) adopt similar methods to measure vocal indicators of psychological discomfort using a large sample of CEO speech. They find that higher vocal dissonance is significantly associated with future financial restatements due to irregularities, even after controlling for traditional financial and linguistic predictors. This suggests that vocal cues offer incremental value in detecting deception.

Visual content has also gained prominence as a source of market-relevant information. Cao et al. (2024a) analyze over 17,000 corporate executive slide decks and use deep learning to classify images into categories such as forward-looking operations and current summaries. They find that forward-looking visual content, including product roadmaps or strategic blueprints, predicts both short-term stock returns and long-term performance. These effects are predominantly driven by trades from AI-equipped institutional investors, pointing to an emerging "AI divide" in financial markets, where retail investors and traditional institutional investors are put at a disadvantage. Borgschulte et al. (2025) combine machine learning–based facial age estimation with mortality data to assess how industry distress shocks and shifts in corporate governance affect CEO aging and life expectancy. Using neural network algorithms to estimate "apparent" age from 3,002 CEO facial images around the Great Recession, they find that exposure to industry distress accelerates visible aging by about one year.

Hu and Ma (2025) explore another dimension of nontraditional communication: how delivery style in startup pitch videos affects investor behavior. Using machine learning to quantify vocal, visual, and verbal delivery features, they construct a "Pitch Factor" and find that higher scores associated with enthusiasm, warmth, and positivity significantly increase funding success. However, these delivery cues do not correlate with startup performance. The authors find that these persuasion effects arise mainly through belief distortions, as investors are often swayed by style rather than substance.

Given the expanding role of AI in data analytics, firms are beginning to strategically adapt to the presence of AI readers. Cao et al. (2023a) find that companies increasingly tailor their disclosures to optimize for machine readability and sentiment interpretation. Using proxies such as bulk downloads from the EDGAR system, AI-related job postings, and local AI talent supply, they show that firms expecting greater machine readership reduce the use of negatively connoted words, especially following the publication of standard sentiment dictionaries. These adjustments are more pronounced among firms with greater financing needs and moderated by litigation risk. Importantly, such disclosures trigger faster market reactions, particularly from institutions with AI capabilities, raising broader concerns about manipulation in disclosure language.

While AI improves information consumption, it may discourage information production, raising concerns about transparency in the AI era. Bertomeu et al. (2024) find that as investors increasingly rely on AI, firms are more likely to withhold information, reducing their voluntary disclosures (e.g., management forecasts). Using the launch of ChatGPT 3.5 as an exogenous shock, the authors show that firms covered by AI-savvy analysts reduced disclosures by 20% post-launch. While AI improves information processing speed, it does not increase the overall informativeness of stock prices, as the crowding-out of firm disclosures offsets this benefit. Analysts' reactions to management forecasts also diminish, and non-disclosing firms experience higher valuations.

Furthermore, the rise of AI also seems to incentivize firms to engage in misleading disclosure, particularly "AI washing." Using 10-K filings, earnings calls, and conference transcripts, combined with firm-level employment data, Barrios et al. (2024) find that AI disclosures are increasing across firms but often lack corresponding in-house AI talent. Firms with high AI disclosure and substantial AI employment show stronger future productivity, innovation, and long-term stock performance, while firms suspected of "AI washing" (high disclosure but low AI employment) do not experience similar gains and may even underperform. The market reacts similarly to all firms' AI disclosures in the short term, but only rewards those with genuine AI investments over time. The authors warn of potential misrepresentation in the absence of clear, investment-based verification of AI disclosures.

4 AI and Asset Pricing

This section synthesizes recent research at the intersection of AI and asset pricing, organized around three core domains: (1) return predictability, (2) estimating SDF and factor discovery, and (3) portfolio construction and investment strategy. Within each theme, we highlight how state-of-the-art AI techniques introduce new methodologies and how they overcome classical challenges like dimensionality, overfitting, and model misspecification. While this section is titled "AI and Asset Pricing" to reflect the broader technological shift, most existing studies primarily focus on machine learning (ML) techniques—a narrower subset of AI.[23] ML's flexibility and predictive power have positioned it as a promising alternative to traditional approaches. Table 6 summarizes each study reviewed in terms of its methodologies and research questions.

[23]Supervised learning is the most commonly used class of ML techniques in asset pricing, where the goal is to predict an outcome variable based on a set of input features, typically firm or macroeconomic characteristics (Nagel, 2021). They include linear methods like linear regression and logistic regression. These serve as foundational tools and are often extended through regularization techniques such as Lasso, Ridge, and Elastic Net to handle high-dimensional data and prevent overfitting; tree-based models, including decision trees, random forests, and gradient boosting machines (e.g., XGBoost, LightGBM), are popular for their ability to capture nonlinear interactions and variable importance in an interpretable way; neural networks, especially feedforward neural networks, recurrent neural networks (RNNs), LSTMs, and more recently, transformers, are employed to model complex temporal dependencies and extract latent patterns from sequential or textual data.

4.1 Return predictability

A major strand of the literature uses ML to improve stock return predictability. Forecasting returns is inherently difficult due to low signal-to-noise ratios and the risk of overfitting in a high-dimensional predictor space. OLS tends to overfit noise rather than identify the true signals if the right-hand side contains hundreds of characteristics. ML provides powerful tools to address these issues. The seminal work by Gu et al. (2020) compares a wide range of ML methods for predicting US equity risk premia using a large sample of stocks over 60 years with over 900 predictors. They find that ML models significantly outperform traditional linear regressions in out-of-sample stock return prediction, with trees and neural networks being the best-performing techniques. At the aggregate US market level, Dong et al. (2022) use ML shrinkage techniques to uncover a strong link between cross-sectional anomalies and time-series market return predictability. They show that returns from 100 long-short anomaly portfolios significantly predict the market excess return out-of-sample. The predictive power is economically meaningful, especially during periods of heightened limits to arbitrage, and is driven by persistent correction of overpricing in short-leg portfolios.

The strong predictive power of ML models is also observed in international equity markets. For example, Leippold et al. (2022) apply an array of ML methods to forecast stock returns in China's unique market dominated by retail investors and influenced by state-owned firms and trading frictions. They find that neural networks outperform other ML techniques in predicting Chinese stock returns. Notably, return predictability is stronger in China than in the US, and even long-only portfolios yield significant abnormal profits after transaction costs, performing well even through the 2015 crash and early 2020 COVID period.

In fixed income markets, Bianchi et al. (2021) employ randomized trees and deep neural networks to forecast US Treasury bond excess returns. They find that nonlinear models such as extreme trees and deep neural networks outperform linear benchmarks and traditional dimension-reduction methods like PCA, even in low-dimensional settings. Importantly, yield and macroeconomic data significantly enhance model predictive accuracy. Further, the study introduces a novel "group-ensembled" neural network architecture that leverages economic priors by grouping variables by macro categories, finding that nonlinearities within economic groups (rather than across them) are the primary driver of improved prediction. In a corporate bond setting, Bell et al. (2024) apply an Explainable Boosting Machines to predict bond returns using a rich set of bond, firm-level, and macroeconomic variables. The model outperforms linear benchmarks and performs on par with black-box alternatives, achieving an out-of-sample R^2 exceeding 12%. Similar to the findings in Bianchi et al. (2021), the most influential predictors are macroeconomic and financial uncertainty indices and the term structure factor.

In option pricing, Chen et al. (2023a) propose a transfer learning (TL) framework that integrates economic theory from structural models with the flexibility of ML to enhance predictive modeling in financial applications. Instead of treating potentially flawed structural models as hard constraints, the approach uses synthetic data generated from structural models to pre-train neural networks, then fine-tunes them with empirical data. This method helps mitigate overfitting, improves generalizability, and allows insights from theory to inform empirical predictions without rigid adherence. The authors test this method using the Black-Scholes option pricing model as the structural source. The TL model significantly outperforms deep learning and traditional models like Heston (Heston, 1993) in terms of accuracy and stability, especially in volatile markets, for out-of-the-money options, and when empirical data are limited. Unlike other methods such as boosting or constrained optimization, the TL model benefits from informative initialization and fine-tuned flexibility, therefore, it can be applied to other financial problems beyond option pricing.

The return predictive ability of ML methods directly challenges the weak-form efficient market hypothesis (Fama, 1970). Murray et al. (2024) demonstrate that a deep convolutional network trained solely on past monthly returns can predict the cross-section of next-month stock returns, closely mimicking the information available to chartists. The model achieves a 1% per month long–short spread, outperforming classic momentum and reversal strategies. The nonlinear patterns learned by the network are stable over time and persist even among the largest, most liquid stocks.

ML has also pushed the frontier of high-frequency and short-horizon return prediction. Traditional approaches, which rely on economic intuition to pre-select predictors, might overlook ephemeral signals. Chinco et al. (2019) tackle the challenge of identifying fleeting, short-lived return predictors in intraday stock data. They leverage LASSO regression to automatically select return predictors in an ultra-high-frequency setting. Analyzing minute-by-minute stock returns with thousands of candidate features, their LASSO procedure efficiently picks out a handful of short-lived but powerful return predictors—often the lagged returns of other stocks with news-related shocks. The LASSO improves out-of-sample fit and delivers a forecast-implied Sharpe ratio of 1.8 after trading costs. However, these predictors often last under 15 minutes and are not easily discovered through standard approaches.

The flexibility of AI-related methods also allows researchers to incorporate alternative data sources into return prediction. Studies have used ML to extract predictive signals from news articles, corporate disclosures, and other text corpora. For example, Glasserman et al. (2020) employ a supervised Latent Dirichlet Allocation (sLDA) model linking daily news text about S&P 500 firms to same-day stock returns. A methodological innovation in that study is a branch-and-bound algorithm to select the optimal number and composition of topics based on out-of-sample explanatory power. This approach mitigates the typical overfitting of off-the-shelf topic models and yields interpretable news topics that significantly explain return variation. In a related vein, Bybee et al. (2023) explore news text from The Wall Street Journal for risk factor construction: they identify 180 economic topics from news and then apply a sparse factor-picking algorithm (Instrumented PCA) to build a small set of "narrative" risk factors. These narrative factors serve as state variables in an ICAPM-style model and significantly improve the pricing of cross-sectional stock anomalies.

Beyond text, AI methods can process vast and unconventional data sources, potentially uncovering return predictors missed by traditional financial metrics. Obaid and Pukthuanthong (2022) construct a "Photo Pessimism" index using machine vision to classify the sentiment of news photographs. This daily sentiment measure from images is shown to predict aggregate market returns, demonstrating the broad reach of AI in capturing investor sentiment signals that are hard to quantify otherwise. In a similar setting, Jiang et al. (2023) transform stock-level historical price and volume data into images and apply convolutional neural networks to identify visually encoded price patterns predictive of future returns. Their approach outperforms canonical strategies, such as momentum, short-term reversal, and trend composites, achieving out-of-sample Sharpe ratios up to 7.2.

Conventional wisdom suggests that highly parameterized models risk overfitting noise, especially when predictors outnumber observations. However, Kelly et al. (2024) demonstrate that heavily parameterized models, even when the number of predictors exceeds the number of observations, can outperform sparse ones in terms of portfolio performance (e.g., Sharpe ratio).[24] ML-based return prediction also raises the question of whether the superior predictive ability corresponds to genuine risk premia or to mispricing exploitable by investors. Avramov et al. (2023) examine the realistic profitability of ML strategies after imposing trading frictions (e.g., no microcap stocks, short-sales constraints, transaction costs). They find that while the raw ML strategies deliver impressive returns, their performance declines once frictions are accounted for. Yet, even under realistic constraints, ML-driven strategies still outperform conventional factor-investing approaches, particularly in hard-to-arbitrage market segments and during volatile periods. The ML signals appear to aggregate many weak pieces of information into a tradable strategy that picks up mispricings in neglected stocks.

Further, despite their strong predictive performance, ML methods, especially deep learning models like convolutional neural networks, are often criticized for their lack of interpretability. This issue arises because these models learn complex, nonlinear representations through many layers of abstraction, making it difficult to trace how specific inputs lead to specific outputs. Bell et al. (2024) caution that while their methods achieve high predictive accuracy, these complex models remain largely “black boxes” in economic terms, as they do not inherently identify struc-

[24]Nevertheless, Berk (2023) argues that it is still an open question for future research whether more complex ML models can yield broader and more practically useful implications.

tural economic mechanisms behind the predictive patterns. Accordingly, Bell et al. (2024) turn to a "glass box" model, which is designed to be inherently interpretable by virtue of its structure, while still preserving much of the flexibility and scalability. Specifically, they propose an Explainable Boosting Machine, a type of interpretable additive model (e.g., Nori et al., 2019). In their application to US corporate bonds, the EBM achieved prediction performance on par with complex models while yielding clear economic narratives. For instance, the model revealed nonlinear effects: extreme macroeconomic uncertainty has a disproportionately large negative impact on bond returns, and the relationship between bond maturity and returns is hump-shaped rather than monotonic. It also uncovered heterogeneous effects across firm sizes and credit qualities. These findings validate that ML methods can balance complexity and interpretability.

In a further attempt to enhance interpretability, Cong et al. (2021a) translate ML models back into familiar terms. The authors introduce an "economic distillation" framework that improves interpretability without sacrificing the flexibility of deep reinforcement learning. Specifically, they use a polynomial-feature-sensitivity analysis to project the complex, nonlinear model onto a more interpretable linear feature space, allowing them to identify the most influential drivers of investment performance, including both standard fundamentals and higher-order terms. This method reveals which features consistently affect portfolio decisions and how their importance evolves over time.

4.2 The factor zoo and stochastic discount factor

A central question in asset pricing is to uncover the true factor structure driving asset returns: How many factors are there? What do they represent? How can they be identified from data? Over the past four decades, asset pricing research has focused on characterizing the stochastic discount factor (SDF) to explain differences in expected returns. Empirical failures of the CAPM and evidence from Banz (1981), Fama and French (1993), Carhart (1997), and others suggest that single-factor models are insufficient, leading to widespread adoption of multi-factor models.[25]

[25] Fama and French (1993) introduce a three-factor asset pricing model, Hou et al. (2015) move on to four factors, Fama and French (2015) extend to five factors, and Barillas and Shanken (2018) argue for a six-factor model.

This shift gives rise to an explosive increase in stock return anomalies—firm characteristics with unexplained return spreads. This "factor zoo" phenomenon has since drawn criticism for its statistical robustness and economic interpretation (Harvey et al., 2016; McLean and Pontiff, 2016; Hou et al., 2020; Feng et al., 2020).[26]

The factor zoo debate centers on identifying the true number of return-driving factors. Traditional models either impose a sparse factor structure or yield an excessive number of potential factors. They typically suffer from issues like high dimensionality, model misspecification, and persistent pricing errors. In response, a growing literature has turned to ML techniques to better handle complex data and uncover latent return structures in a data-driven yet economically grounded way.

Early work by Light et al. (2017) uses a partial least squares (PLS) framework to treat expected returns as a low-dimensional latent variable and aggregate many noisy firm characteristics into a single, high-signal ranking that outperforms PCA, Fama–MacBeth, and simple rank-averaging in cross-sectional return tests.

One influential example is the Instrumented Principal Component Analysis (IPCA) framework of Kelly et al. (2019) (KPS), which uses firm characteristics to instrument the time-varying factor loadings of latent factors. This approach allows the model to determine whether a characteristic's predictive power stems from exposure to priced risk factors (i.e., through betas) or represents unexplained alpha. They show that IPCA bridges firm characteristics and factor models through a semi-supervised learning approach, and that a small set of latent factors can explain a wide array of return anomalies. Gu et al. (2021) extend KPS by relaxing the linearity assumption and introducing a conditional autoencoder model that uses neural networks to allow for nonlinear and interactive relationships between characteristics and risk exposures. Empirically, the autoencoder dramatically outperforms traditional factor models like Fama-French or even linear IPCA in explaining return covariance and cross-sectional differences. Importantly, the study shows that most return predictability from firm characteristics disappears once proper nonlinear

[26] There is also a factor zoo for bond returns, as summarized by Dickerson et al. (2023).

conditional betas are accounted for, suggesting that many variables in the factor zoo are not true anomalies, but rather proxies for complex, time-varying risk exposures.

Further, Freyberger et al. (2020) propose a nonparametric group LASSO model to select return predictors and estimate nonlinear relationships between firm characteristics and expected returns. Using 62 firm characteristics, the authors find that only a small subset of characteristics (11–14) have true incremental explanatory power. The results show substantial time variation in the relevance of return predictors, with characteristics like momentum and size showing different dynamics across time. The authors emphasize the importance of model selection and accounting for nonlinearities to address the factor zoo problem. In another effort to "tame" the factor zoo, Feng et al. (2020) propose a framework for evaluating whether a newly proposed asset pricing factor contributes meaningfully beyond the large set of existing factors. They develop a method combining double-selection LASSO with two-pass (Fama-MacBeth) regressions to identify and control for relevant benchmark factors in a high-dimensional setting. Applied to a library of 150 proposed factors, their method shows that most new factors are statistically redundant, although some retain significant marginal explanatory power. The authors argue that using their approach, many factors could have been screened out at the time of their introduction, helping discipline the growth of the factor literature.

To extract economically meaningful factors, Lettau and Pelger (2020) introduce Risk Premium PCA (RP-PCA), which improves upon standard PCA by identifying factors that explain both return covariance and expected returns. The method is especially adept at detecting “weak” factors—those that affect only subsets of assets—and delivers a more accurate estimate of the SDF than PCA. Empirically, applying RP-PCA to anomaly portfolio returns yields a more parsimonious factor model that prices assets better out-of-sample than PCA factors do. Moreover, the authors show that RP-PCA can enhance other factor extraction methods: for instance, replacing the PCA step in Instrumented PCA or in the Bayesian shrinkage of Kozak et al. (2020) with RP-PCA leads to models with better Sharpe ratios and pricing performance. Shifting the focus to time series factors, He et al. (2023a) propose a Reduced-Rank Approach (RRA) to identify the effective number of priced factors by constraining the SDF to a low rank $r < K$ across a large set of candidate proxies. Empirically, a small number of composite factors capture most pricing information: one factor closely replicates the market, and five outperform the Fama-French five-factor model. Unlike PCA, RRA directly targets pricing errors, suggesting many anomalies are redundant.

ML methods have also been applied to improve the construction of test assets and factor portfolios themselves. Classic asset pricing tests often rely on sorting stocks into portfolios to create a manageable basis for the cross-section. However, simple sorts lose information and cannot capture multiple characteristics simultaneously. Bryzgalova et al. (2025) address this with Asset Pricing Trees (AP-Trees), a tree-based algorithm that learns how to partition stocks into optimally informative test assets. The AP-Trees method conditionally sorts stocks by characteristics in a decision-tree structure, choosing splits that maximize the improvement in SDF representation (i.e., the spanning of the SDF's variance). By doing so, it captures high-dimensional interactions and forms a small set of diversified portfolios that collectively span the space of asset payoffs much better than ad hoc sorts. The authors show that AP-Tree portfolios dramatically outperform standard sorted portfolios in out-of-sample Sharpe ratios and produce much smaller pricing errors when used to test models. Cong et al. (2025a) extend this idea with Panel Trees (P-Trees), an ML framework designed to construct test assets and estimate the SDF by clustering assets based on high-dimensional characteristics to span the mean–variance efficient frontier. Traditional approaches using pre-sorted portfolios (e.g., size-value portfolios) or standard factor models often fail to span the efficient frontier due to their ad hoc construction and inability to capture nonlinear, asymmetric interactions among characteristics. P-Trees address this by recursively splitting the cross-section of individual assets into characteristic-based clusters (leaf portfolios), with the objective of maximizing the Sharpe ratio of the resulting tangency portfolio. Empirically, using US equity returns and 61 firm characteristics, the P-Tree method achieves significantly higher Sharpe ratios than conventional methods, outperforming Fama-French and recent ML-based factor models in both in-sample and out-of-sample tests.

There are four longstanding challenges in estimating the SDF: (1) high dimensionality of the SDF, (2) unknown SDF functional form, (3) time-variant risk exposure for individual assets, and (4) low signal-to-noise in stock risk premia. ML methods allow researchers to relax functional form restrictions and include a vast array of state variables or characteristics when estimating the SDF. Chen et al. (2024a) design a model that integrates three neural network architectures: a feedforward network to capture the SDF's functional form, an LSTM network to extract economic state variables from macro time series, and a generative adversarial network to construct the most informative test assets by targeting unexplained pricing errors. By framing SDF estimation as a conditional GMM problem, their model emphasizes pricing accuracy over pure predictive power. Using over 12,000 stocks, 46 firm characteristics, and 178 macroeconomic variables, their approach outperforms standard benchmarks, including Fama–French models and deep learning predictors, achieving an out-of-sample Sharpe ratio of 2.6 and explaining up to 23% of expected returns on individual stocks.

Leveraging a large-scale transformer architecture (Vaswani et al., 2017) into an asset pricing model, Kelly et al. (2025) propose an Artificial Intelligence Pricing Model (AIPM). The motivation is that transformers enable cross-asset information sharing: instead of each asset's returns being modeled in isolation or only through common factors, a transformer-based SDF uses attention to condition each asset's expected return on the characteristics and returns of all other assets. Compared to traditional factor models and other ML methods, the AIPM achieves notably lower out-of-sample pricing errors, especially relative to neural networks without attention mechanisms. It also delivers higher out-of-sample Sharpe ratios by effectively capturing cross-asset dependencies. Furthermore, the study finds that increasing model complexity through deeper transformer architectures consistently enhances predictive accuracy, suggesting that highly parameterized AI models are particularly well-suited for asset pricing applications (Kelly et al., 2024).

4.3 Investment management

Beyond improving asset pricing models, ML techniques are increasingly being adopted to inform investment decisions and portfolio construction. This subsection reviews a growing body of literature that applies these ML techniques to improve investment decision-making across several dimensions: enhancing portfolio optimization and statistical arbitrage, forecasting fund performance, and extracting insights from qualitative disclosures.

Traditional portfolio management follows a two-step process: first estimating risk premia or minimizing pricing errors, then constructing portfolios to meet investment goals. This method suffers from estimation errors in the first step and misaligned objectives between the two steps. Additionally, the complex, high-dimensional, and nonlinear nature of financial data makes traditional econometric methods ineffective. To address these challenges, Cong et al. (2021a) introduce AlphaPortfolio, a deep reinforcement learning framework using LSTMs and Transformers to process asset characteristics and macro signals. The agent delivers Sharpe ratios above 2 and annual alpha exceeding 13%, outperforming standard ML strategies even under realistic trading constraints. To enhance interpretability, the authors apply an "economic distillation" tool to project the reinforcement learning policy onto interpretable factors like Tobin's Q and inventory changes.

Deep learning has also advanced statistical arbitrage. Guijarro-Ordonez et al. (2021) build a modular system that generalizes common statistical arbitrage strategies into three components: (1) portfolio construction using residuals from asset pricing models (e.g., Fama–French, PCA, IPCA), (2) signal extraction via a novel time-series model that combines convolutional neural networks and transformers to detect local and global patterns in residual price series, and (3) trading allocation using flexible neural networks optimized over economic objectives such as the Sharpe ratio. Their approach integrates these components into a global, constrained optimization routine tailored to real-world trading conditions. The authors show that their deep learning statistical arbitrage model significantly outperforms benchmark methods, achieving out-of-sample Sharpe ratios over 4 and annual returns above 20% while satisfying realistic trading constraints.

The key performance driver is the transformer-based signal extraction, which captures asymmetric short-term price patterns that are largely orthogonal to common risk factors and persistent across time.

ML techniques are also being used to evaluate and select investment funds and managers, essentially treating funds as assets whose returns can be predicted. Wu et al. (2021) is the first in this domain. They apply four supervised ML methods—LASSO, random forest, gradient boosting, and deep neural networks—to forecast future hedge fund performance using 17 fund-specific variables. They find that cross-sectional ML forecasts outperform time-series regression in both return and alpha. For US actively managed mutual funds, Li and Rossi (2020) use boosted regression trees to predict fund returns based on their stock holdings. Their model delivers over 7% annualized alpha, though most performance is attributable to static exposure to known anomalies. This supports the view that what appears as "skill" is often just systematic factor exposure. Similarly, Kaniel et al. (2023) employ feedforward neural networks to revisit the predictability of abnormal returns in mutual funds from 1980 to 2019. They find that fund flows and fund return momentum are robust predictors of future fund performance, especially during periods of high investor sentiment, while stock-level characteristics add little incremental value. The model achieves a large and persistent out-of-sample performance spread: the top decile of predicted funds outperforms the bottom by 191% annually, or 40 basis points per month. Furthermore, DeMiguel et al. (2023) apply gradient boosting and random forests to select portfolios of mutual funds and actually find positive net alpha (over 2% per year after fees) for investors. A key distinguishing feature of this paper is that it focuses on long-only, tradable mutual fund portfolios using only past data and evaluates out-of-sample net performance after all costs, thereby assessing realistic investor gains from active management. In contrast, Kaniel et al. (2023) study long-short fund portfolios, and most of the predictability in after-fee alpha comes from the short leg, which is not directly investable for most retail investors.

Another novel application of ML in investment management is deriving insights from qualitative information. For instance, mutual fund managers write quarterly shareholder letters discussing their outlook and strategy. Cao et al. (2023b) develop a deep learning method to quantify the sentiment managers express toward risk in these reports. They find that when managers voice negative sentiment about risk, their funds subsequently reduce risk exposure and perform better. A trading strategy that buys funds whose managers are pessimistic and sells those whose managers are complacent yields significant annual alpha of 1.5%. Similarly, in the realm of private equity, Fernández Tamayo et al. (2023) analyze nearly 400 private equity fundraising prospectuses using ML to assess whether qualitative textual information can predict fund performance. They find that quantitative metrics (such as past performance and fund size) predict fundraising success but do not forecast future fund returns. However, term frequency–inverse document frequency features derived from the investment strategy section of private placement memoranda predict fund outperformance with high accuracy. Gradient Boosting Machines achieve a 75% success classification rate, with a spread of 0.23 in total value to paid-in capital between top-performing funds and the average.

5 AI and Household Finance

Recent studies report that AI can outperform humans in a variety of decision-making tasks. For example, AI can substantially improve judges' jail-or-release decisions (Kleinberg et al., 2018a), loss estimates in insurance (Ding et al., 2020), and underwriting profitability (Jansen et al., 2023). Not surprisingly, household financial decision-making also benefits from the rise of advanced predictive methods in the new wave of AI. FinTech represents a leading application of AI in household finance and a disruptive technology with substantial value (Chen et al., 2019).[27] In this section, we review how AI contributes to financial inclusion and innovation in financial services, while also highlight emerging risks related to equity, discrimination, and data governance.

[27] The Financial Stability Board (FSB) defines FinTech as "technologically enabled financial innovation that could result in new business models, applications, processes, or products with an associated material effect on financial markets and institutions, and the provision of financial services." For comprehensive overviews of the literature on FinTech and FinTech lending, see Thakor (2020) and Berg et al. (2022).

5.1 Credit markets and financial inclusion

The rise of AI has significantly improved credit markets by enhancing predictive accuracy, expanding financial inclusion, and increasing operational efficiency. Different from traditional methods (e.g., Campbell and Dietrich, 1983; Schwartz and Torous, 1993), AI methods are capable of capturing complex, nonlinear relationships in borrower and macroeconomic data, resulting in better risk assessment and more accurate loan pricing. For example, Sadhwani et al. (2021) develop a deep learning model to analyze mortgage borrower behavior using over 120 million US mortgages from 1995 to 2014. Unlike traditional linear models, this model captures highly nonlinear relationships and complex interactions between risk factors, such as the way unemployment and credit scores jointly affect prepayment probabilities, or how the effect of house price appreciation on delinquency varies with local labor market conditions. The authors show that their deep learning model outperforms linear models in forecasts of loan-level and pool-level risks. From the lenders' perspectives, Zhou (2024) demonstrates AI's superior ability to process complex data and allocate efforts more efficiently in debt collection. In a randomized experiment, the author finds that AI-generated calling decisions lead to higher repayment rates with fewer calls. Even beyond conventional information, Chang et al. (2024) show that machine learning methods can predict loan delinquency by detecting borrowers' fleeting emotional cues. They find that happiness expressions are negatively associated with future loan delinquency, while fear expressions are positively associated with delinquency. They also show that the micro-expression metrics add incremental predictive power beyond internal credit scores.

These technological advances have been rapidly adopted by leading FinTech firms, leading to extended financial services to previously underserved entities. For example, Buchak et al. (2021) examine the launch of Yu'ebao, a money market fund distributed through Alipay, which allowed households to earn market-based returns on deposits. Yu'ebao's rapid adoption led to substantial deposit outflows from incumbent banks, supporting a bottom-up financial liberalization and improving household welfare. Liu et al. (2022) show that big tech loans are primarily offered to borrowers with limited access to other forms of credit in a study of a FinTech subsidiary of Ant

Group that leverages AI to lend to small and medium-sized enterprises. [28]

By incorporating advanced lending technologies that utilize alternative data and AI, FinTech firms can expand credit access to so-called "invisible primes"—creditworthy individuals who are frequently excluded by traditional lenders that rely heavily on traditional credit scores for underwriting decisions. For example, FinTech lenders can overcome traditional banks' data advantages by leveraging the rich, low-cost information embedded in borrowers' cashless payment records (Ghosh et al., 2022). Using Alipay's user data, Ouyang (2021) provide causal evidence on how digital payment data can facilitate financial inclusion at the household level. The author finds that increased adoption of cashless payments leads to higher credit access and larger credit lines, particularly benefiting older and less-educated borrowers.[29] In a US context, Di Maggio et al. (2022) show that a sophisticated underwriting ML algorithm approves 15–30% of applicants who would have been rejected by traditional models and offers them substantially lower interest rates. These benefits are most pronounced among borrowers from minority and immigrant-heavy regions. The improvements in credit inclusion stem both from the use of alternative data (e.g., education, employment, digital footprint) and the use of more advanced ML techniques.

Further complementing this evidence, an emerging literature uses predictive models to demonstrate that digital footprints can play a significant role in promoting financial inclusion. Berg et al. (2020) show that simple digital footprint variables, such as device type, operating system, and email provider, can predict consumer default risk with accuracy comparable to traditional credit bureau scores. Importantly, digital footprints help reallocate credit toward borrowers with favorable behavioral signals without increasing default rates, including those with low or no formal credit history. In a closely related study, Agarwal et al. (2019) apply AI-driven models

[28]Some evidence suggests that while FinTech lending boosts market efficiency, it does not significantly enhance financial inclusion. Fuster et al. (2019) find that FinTech lenders in the US mortgage market do not disproportionately serve credit-constrained or underserved borrowers—they tend to be used more by educated and older populations. Tang (2019a) suggest that P2P lending, a form of technology-based lending, does not primarily improve financial inclusion for the previously underserved. Instead, it shows that P2P platforms act as substitutes for banks by serving borrowers who already had access to bank credit but were displaced by a negative credit supply policy shock.

[29]While FinTech innovations have expanded credit access and promoted financial inclusion, they often raise concerns about overborrowing. Dong et al. (2024) argue that credit reporting regulation can effectively mitigate this issue.

trained on mobile and social footprint data from a mobile-only lender in India. They find that these alternative data sources outperform traditional credit scores in predicting loan defaults, particularly for financially excluded individuals.

5.2 AI-empowered financial advice

AI automation has steadily advanced across many sectors over the past few decades (e.g., Autor, 2015; Acemoglu and Restrepo, 2019). More recently, it has begun to transform traditionally high-skill professions, such as law, accounting, and financial services. Robo-advisors, as the most prominent example of automated financial advice, have grown rapidly.[30] Robo-advisors offer two main advantages over traditional advisors: (1) they use big data and algorithms to lower monetary, cognitive, and psychological transaction costs associated with human decision-making, and (2) they can democratize access to high-quality financial management for less-wealthy investors (D'Acunto and Rossi, 2023; Reher and Sokolinski, 2024). Therefore, robo-advice holds the potential to improve financial literacy, narrow wealth gaps, and enhance policy effectiveness.

Extensive studies have documented that AI-empowered advisors provide incremental value to investors. D'Hondt et al. (2020) compare actual investor performance to that of robo-investors using a rich dataset of nearly 23,000 real investors. They find that AI-guided robo-advisors deliver the greatest benefits to low-income and less-educated investors, who tend to under-diversify and mismanage risk. Particularly during the financial crisis, AI-based strategies significantly outperformed human investors. Coleman et al. (2022) show that robo-analysts issue more balanced (less optimistic) stock recommendations and are less influenced by conflicts of interest. Their buy recommendations generate significantly higher long-run abnormal returns (4–5% annually) than those of human analysts. In a similar context, Cao et al. (2024b) find that AI-based stock analysts outperform human analysts in 54.5% of forecasts, particularly when processing large volumes of information. Further, Guo et al. (2022) provide the first empirical evidence on the impact of

[30] Deloitte (2016) defines a robo-advisor as "an online portfolio management solution that aims to invest client assets by automating client advisory." The global robo-advisory market size was valued at USD 8.39 billion in 2024. The market is projected to grow from USD 10.86 billion in 2025 to USD 69.32 billion by 2032 (Fortune Business Insights, 2025).

GenAI investment consultants (GAICs) using data from Ant Fortune's Zhi Xiaobao. The authors find that GAIC use improves investment decisions, reduces suboptimal redemption activity, and increases returns, particularly for experienced and risk-seeking investors.

One important mechanism through which robo-advisors improve investment performance is by mitigating behavioral biases, such as loss aversion, overconfidence, confirmation bias, and the disposition effect. This is particularly crucial because many biases act as investor-specific "fingerprints."[31] For example, Ouyang and Ouyang (2025) show that the disposition effect is highly persistent for the same individuals across both low-stakes trading simulations and their actual mutual-fund portfolios. They also find that extrapolative beliefs combined with realization preferences jointly account for the fixed disposition effect: contrarian investors show stronger disposition effects, and there is a sharp increase in selling precisely when a holding moves from a slight loss to a slight gain. This concept of a durable, predictable behavioral style is further reinforced by Han et al. (2020). They not only find a similar persistence for a contrarian trading tendency but also demonstrate that this style can be predicted using machine learning. Specifically, they show that behavioral data from a virtual trading experiment, analyzed with logistic regression, is significantly more powerful at predicting an investor's real-world trading style than traditional demographic data, as measured by Area Under the Curve (AUC). The combined evidence that investors display stable, predictable yet varied biases, such as the disposition effect and contrarian behavior, emphasizes the importance of robo-advisors delivering personalized nudges instead of generic, one-size-fits-all interventions.

Rossi and Utkus (2024) study the adoption of Vanguard's hybrid robo-advisor by over 55,000 US investors and find that it improves portfolio efficiency and risk-adjusted performance (higher Sharpe ratio). Such improvement is mainly driven by reducing poor diversification, low equity exposure, and excessive home bias. The largest gains accrue to initially unsophisticated investors. Using an international sample, D'Acunto et al. (2019) study the adoption and effects of a robo-

[31] Hwang et al. (2025) conduct AI-driven field interviews with 1,540 actual investors. They uncover thirteen recurrent mechanisms that together form actual investor behavior, with substantial heterogeneity both across and within investors.

advising portfolio optimizer introduced by an Indian brokerage firm. Comparing adopters and non-adopters, they find that robo-advising improves portfolio diversification, especially for previously underdiversified investors. Across the board, robo-advice reduces behavioral biases, including the disposition effect, trend chasing, and the rank effect.

Robo-advising is not necessarily without pitfalls. Uptake of such automated tools often requires the tool to grant a degree of autonomy to investors, such as allowing them to experiment with portfolio allocations and approve trades. This autonomy can reduce algorithm aversion—the common hesitation to fully rely on automated decision-making (Dietvorst et al., 2018; Stradi and Verdickt, 2025). At the same time, full control given to investors reintroduces suboptimal behavior such as that due to lack of self-control (Thaler and Shefrin, 1981) or overtrading (Barber and Odean, 2000). Some robo-advisors have been criticized for prioritizing firm profits over clients' best interests.[32]

This raises the question of whether and how human experts continue to offer distinct value in the era of automation. On the one hand, the primary function of financial advisors is to provide technically grounded investment recommendations, a task for which low-cost, automated "robo-advisors" already serve as a widely adopted alternative. On the other hand, human advisors also deliver complementary services that rely on "soft" interpersonal skills, such as offering emotional reassurance to help clients take compensated risks (Linnainmaa et al., 2018), mitigating behavioral biases like loss aversion (Calvet et al., 2023), and fostering trust in the financial system (Gennaioli et al., 2015). Although recent studies suggest that robo-advice may elicit similar behavioral responses (Linnainmaa et al., 2018; D'Acunto et al., 2019), it remains unclear whether certain client-facing functions, especially those requiring interpersonal engagement, can be replaced by automated tools. Modeling this trade-off, Huang and Ouyang (2025) formalize the key friction: human advisors can elicit subjective "soft information" but are prone to strategic bias, whereas unbiased AI advisors struggle to process unarticulated client needs due to architectural limitations, leading to a novel form of information loss.

[32]See, for example, "Should Retirees Use Robo Advisors?," Wall Street Journal, November 12, 2017.

Costello et al. (2020) investigate whether human discretion still adds value in an era of AI-driven credit decision-making. The authors conduct a randomized field experiment with trade creditors. They show that human discretion can improve portfolio outcomes by incorporating private information and competitive considerations, which AI models may not fully capture. In a related vein, Greig et al. (2024) investigate the role of human financial advisors in a hybrid robo-advisory platform, where investment portfolios are automated but clients are randomly assigned to human advisors for onboarding and support. The authors find that human advisors significantly improve investor retention, especially during market downturns, despite having no influence on portfolio allocation or performance. In the structural model, high-retention advisors are shown to enhance investor confidence as if the client had observed 20 years of return data (vs. 12 for low-retention), and deliver ongoing utility equivalent to 30 bps in annual portfolio returns. In economic value, investors would require a 9.6% lump-sum wealth transfer to be indifferent between high- and low-retention advisors.

Cao et al. (2024b) further confirm the complementing role of human analysts. They find that while AI-based stock analysts outperform human analysts in forecasts, human analysts perform better in contexts requiring institutional knowledge, such as firms with intangible assets, financial distress, or low liquidity. The authors find that combining AI and human forecasts—the "Man + Machine" model—outperforms either alone. This combination exhibits the greatest synergy in high-uncertainty, low-data environments.

With regard to managing debt, Chak et al. (2022) study the effectiveness of robo-advice in improving loan repayment decisions, particularly for vulnerable households with poor debt-management skills. In a randomized controlled trial with UK consumers, access to free robo-advice significantly improved repayment outcomes, and many users valued the tool beyond its monetary benefits, likely due to reduced cognitive effort. However, distrust of algorithms led to lower adoption and frequent overriding of advice, especially among vulnerable groups, potentially limiting its equity-enhancing potential. Nevertheless, robo-advice did not improve users' future unassisted decisions and even crowded out learning-by-doing, raising concerns about its long-term impact on financial capability.

5.3 Data and privacy

The advancement in AI has led to an explosion in the usefulness of data, shifting knowledge production from human judgment toward data-driven systems (Farboodi et al., 2022; Abis and Veldkamp, 2024; Liu et al., 2025). Data now play a critical role in driving innovation and shaping long-term economic outcomes, making it a cornerstone of the broader macroeconomy (e.g., Jones and Tonetti, 2020; Goldstein et al., 2021; Cong et al., 2021b). While this transformation creates new opportunities, it also raises important concerns about data privacy (Acquisti et al., 2016; Chen et al., 2021a; Goldfarb and Que, 2023). Data-driven credit provision enables financial inclusion while also intensifying the need for careful governance of user data, particularly when lenders gain persistent monitoring power through platform integration (Gambacorta et al., 2023). The increasing reliance of digital platforms and AI systems such as Amazon, ChatGPT, and Alipay on user data has further intensified these concerns.[33] In response, an emerging literature seeks to answer a central question: How does the design of data-sharing environments affect user behavior, welfare, and market efficiency via technology, incentives, and regulation?

There are fears that platforms may use their data and gatekeeping power not only to improve consumer targeting but also to increase sellers' market power and extract more surplus via managed advertising and product steering. Bergemann and Bonatti (2024) investigate this issue and find that precise consumer data enable digital platforms to steer users toward high-margin products, increasing seller surplus but reducing consumer privacy and potentially harming welfare. However, off-platform alternatives and privacy-enhancing regimes can limit this exploitation by restricting data access and curbing the platform's steering power. As a result, the impact of data precision on welfare critically depends on the strength of privacy protections and consumers' ability to avoid being fully tracked. In support of this view, Ahnert et al. (2025) study

[33]The change in consumer attitudes is reflected in the passage of several major data privacy regulations, such as the European Union's General Data Protection Regulation (GDPR) in 2018, California's Consumer Privacy Act (CCPA) in 2020, China's Personal Information Protection Law (PIPL) in 2021, as well as recent amendments to Japan's Act on the Protection of Personal Information (APPI).

how the design of digital payment systems affects the trade-off between transaction efficiency and data privacy in lending markets. They find that while digital payments improve efficiency, they expose users to data disclosure, reducing their informational rents. The value of transparent, user-friendly consent mechanisms in privacy regulation has also been highlighted by Matos and Adjerid (2022). They evaluate the impact of GDPR's enhanced consent requirements using a field experiment involving 33,629 households at a major European telecom provider. Contrary to concerns that stricter consent rules would reduce data sharing, they find that opt-in rates for data use increased, particularly for service-related information, while more sensitive data remained less frequently shared. One caveat with policy intervention is that there is a growing welfare gap between strong-willed consumers who benefit from data sharing and weak-willed consumers who suffer from it (Liu et al., 2023).[34] An opt-in/opt-out privacy regime (like GDPR) can reduce inequality but may lower overall welfare, depending on the severity of self-control problems.

From individual consumers' perspectives, Tang (2019b) investigate whether and how much online borrowers value the privacy of personal data. Using randomized controlled trials on a Chinese P2P lending platform, the author shows that borrowers value privacy and are less likely to complete loan applications when asked to disclose personal information such as social network ID and employer contact. A structural model estimates that applicants value the privacy of these data at 230 RMB ($33), or about 8% of the average loan's net present value. While this evidence indicates users' aversion to intrusive data requests, Chen et al. (2021b) uncover an interesting pattern using survey and behavioral data from Alipay. They find that despite nearly 85% of users expressing privacy concerns, users across all concern levels authorized data sharing with a similar number of third-party mini-programs. They further suggest that privacy concerns may emerge from increased digital engagement, rather than being innate.

[34]An example is open banking where data sharing is voluntary. A perverse effect may arise: high-quality borrowers tend to opt in, while others opt out, making non-participation a negative signal (He et al., 2023b). Privacy-conscious borrowers suffer from this signaling effect despite not sharing data.

6 AI and the Labor Market

The rapid rise of AI has sparked wide-ranging discussions about its implications for labor markets and macroeconomic outcomes. While AI is widely seen as a transformative general-purpose technology (GPT), its effects on labor markets remain complex and uneven. Some research highlights its potential to boost productivity and reduce skill gaps; others point to its role in widening wage inequality and reducing labor's share of income. Scholars have explored how AI interacts with task-level skills, reshapes occupational demand, affects innovation strategies within firms, and influences aggregate economic growth. This section reviews the emerging literature on the labor market implications of AI. We organize the discussion of the impact of AI on the labor market by its theoretical frameworks, empirical evidence, and macroeconomic consequences.

6.1 Theoretical foundations: automation vs. augmentation

The impact of the technological advancement on labor markets remains a subject of active debate. Some argue that the rise of automation—including technologies such as computer-controlled machinery, industrial robots, and AI—signals the possibility of large-scale job displacement (e.g., West, 2018). An opposing view contends that these advancements complement labor and raise output in ways that lead to higher demand for labor, higher employment, and wage growth (e.g., Bughin et al., 2017). A third perspective recognizes the dual nature of AI and similar innovations: they can simultaneously substitute for human labor in some tasks while complementing labor in others, with the net effect depending on the balance of these forces (e.g., Autor, 2015). Agrawal et al. (2019) argue that the overall impact of AI on labor demand is ambiguous and context-dependent, varying by task structure, organizational response, and availability of complementary skills.

Acemoglu and Restrepo (2018) develop a task-based framework that helps reconcile these views by analyzing how technology reassigns tasks between labor and capital. In this model, automation induces a displacement effect by substituting machines for workers in specific tasks, reducing both employment and the labor share of income. This displacement can be counteracted by three offsetting forces: productivity gains that increase demand for non-automated tasks, capital accumulation that boosts overall output, and deepening automation that improves task efficiency without expanding its scope. However, the most important counter force is the creation of new, labor-intensive tasks that restore demand for human labor. Yet, the transition is often slow and painful due to skill mismatches, slow worker reallocation, and training lags. Subsequent work by Acemoglu and Restrepo (2019) provides empirical evidence for this framework, showing that over the past three decades, automation in the United States accelerated while the creation of new tasks lagged. This led to a significant slowdown in labor demand growth and only modest gains in productivity. Their decomposition suggests that the mix of technologies adopted has increasingly favored displacement over reinstatement. Echoing this concern, Acemoglu and Restrepo (2020b) caution that much of contemporary AI development is focused on what they term "so-so automation"—automating tasks for marginal efficiency gains rather than pioneering truly new tasks—a direction that could yield meager productivity improvements yet still displace workers. Drawing on evidence from the diffusion of industrial robots, they show that such technologies reduce the labor share and disproportionately harm low- and middle-skill workers. Over time, the failure to generate new labor-intensive tasks has weakened both employment and productivity growth.

AI is often described as a GPT with broad transformative potential, and history offers analogies for its labor market trajectory. Like past GPTs such as the steam engine, electricity, and computing, AI's impact may unfold gradually and non-linearly, involving substantial complementary innovations and adjustments. A notable pattern observed with earlier GPTs is the "productivity paradox," wherein significant technological advances did not immediately translate into measured productivity growth. Brynjolfsson et al. (2021) formalize this idea in what they call the productivity J-curve: measured productivity may initially lag as firms invest in intangible complements (business process redesign, human capital, new business models), then surge once those investments pay off. They predict that as complementary investments accumulate, AI could eventually drive substantial productivity gains. Thus, current productivity data may understate AI's long-run potential, emphasizing the need for patient investment and supportive policy to unlock its broader economic benefits. This framework echoes Solow's paradox[35] from the information technology (IT) era and suggests that the current slow growth in productivity despite rapid AI progress is not inconsistent with a large future impact. Rather, it reflects implementation lags.

Taken together, the theoretical and historical perspectives highlight a tension between displacement and complementarity. Will AI follow the path of past technologies that ultimately created more jobs and tasks (albeit with lagged gains and skill biases), or will it primarily automate existing work to the benefit of a narrow elite? The literature suggests the outcome is contingent on choices—how firms and policymakers steer AI deployment. AI is often described as having a "dual nature": it can be used to automate tasks formerly done by humans, or to augment human productivity and even generate entirely new tasks and industries.

6.2 Micro-level evidence: AI exposure and task-level impact

Early empirical attempts to gauge AI's impact on jobs often produced alarmingly high estimates of automation potential. Frey and Osborne (2017) estimate that 47% of US jobs are at high risk of automation using experts' assessments. In contrast, Arntz et al. (2016) yield much lower estimates. They argue that only around 9% of jobs in OECD countries are highly susceptible to full automation once one accounts for the fact that many jobs include both automatable and non-automatable tasks. This stark difference highlights a crucial point: occupations are not monolithic, and focusing on the task level is essential for understanding AI's labor market impact.

A growing body of empirical work measures the "exposure" of occupations to AI and examines early labor market effects. With the task-based focus, Brynjolfsson et al. (2018) investigate how ML is reshaping the landscape of work and economic organization by treating occupations as bundles of tasks—some of which are more suitable for ML than others. They develop a mea-

[35] Robert Solow's comment on computers appears in his 1987 New York Times Book Review article: "...what everyone feels to have been a technological revolution, a drastic change in our productive lives, has been accompanied everywhere, including Japan, by a slowing-down of productivity growth, not by a step up. You can see the computer age everywhere but in the productivity statistics."

sure of “Suitability for Machine Learning” for over 18,000 tasks across nearly 1,000 occupations. They show that while nearly all occupations have some tasks that are automatable, very few are fully automatable, implying that ML adoption will lead to task-level reorganization rather than complete job replacement.

Building on the conceptual framework of Brynjolfsson et al. (2018) that describes “labor” via the bundle of skills or abilities, Felten et al. (2018) develop a data-driven method to systematically assess how advances in AI affect labor by linking specific AI capabilities to human occupational abilities. Unlike prior approaches that broadly categorize entire occupations as "at risk" of automation (e.g., Frey and Osborne, 2017), their method focuses on task-level and ability-level granularity by mapping AI advances to 52 distinct human abilities. The authors show that as AI technologies improved between 2010 and 2015, the skills needed for many jobs changed in the following years, suggesting that their method can help predict which job skills and which jobs might be affected as AI continues to advance. Felten et al. (2019) introduce an AI Occupational Impact (AIOI) measure that refines Felten et al. (2018)’s approach by dynamically linking AI progress to occupations through weighted ability mappings. They find that AIOI is not significantly associated with employment declines but is positively correlated with wage growth, particularly in occupations requiring high levels of software familiarity. A one standard deviation increase in AIOI leads to a 0.41 percentage point rise in wage growth overall, and a 0.61 point rise in high-software-use jobs, while showing no meaningful effects for low- or medium-software-use roles. Additionally, they raise concerns that the benefits of AI are concentrated in high-income occupations, indicating that AI may be contributing to labor market polarization and income inequality.

Similarly, Webb (2019) introduces a task-level measurement of occupational exposure based on the overlap between job task descriptions and the text of AI-related patents. He first validates this approach on past technologies (like software and robotics) and shows that occupations heavily exposed to those earlier innovations experienced greater declines in employment and wages, especially for routine and mid-skill roles. When applied to AI, the pattern of exposure looks very different from past automation waves. The method reveals that AI is most associated with high-skilled cognitive tasks such as programming, legal analysis, and scientific research. Accordingly, the occupations he identifies as most exposed to AI include highly educated, white-collar roles (e.g., engineers, physicians, financial analysts), unlike historical automation which primarily threatened manufacturing and other routine manual jobs.

Amid growing interest in GenAI, Eloundou et al. (2024) estimate their potential impact on the labor market by evaluating task-level exposure across 923 occupations. They define "exposure" as the potential for LLMs (with or without complementary software) to reduce task completion time by at least 50% while maintaining quality. They find that only 1.8% of jobs are highly exposed to LLMs alone, but this rises to over 46% when accounting for likely software integrations. High-exposure jobs tend to be high-skill, high-wage roles involving text, coding, or information processing. While full automation remains limited (under 2% of tasks), over 70% of tasks have components LLMs could assist with.

Motivated by the debate over whether AI will displace jobs or enhance productivity as an augmentation, Acemoglu et al. (2022) provide a demand-side perspective by investigating how firms adapt hiring in response to AI. They measure AI exposure by linking each establishment's occupational structure to three task-based indices in previous studies: the AIOI (Felten et al., 2019), the Suitability for Machine Learning index (Brynjolfsson et al., 2018), and AI exposure score (Webb, 2019). They show that establishments with higher AI exposure tend to increase AI-specific hiring and adjust the skill requirements in job postings, while reducing overall and non-AI hiring. This indicates that AI adoption is related to task substitution. However, these effects are not detectable at the industry or occupation level, suggesting that the broader employment and wage impacts of AI remain modest at this stage. On the heterogeneity of the effects of AI on hiring decisions, Lichtinger and Hosseini Maasoum (2025) find that AI adoption disproportionately reduces demand for junior workers, while leaving senior roles largely unaffected. This creates a seniority-biased technological shift with potential long-term consequences for career ladders, wage growth, and inequality. Nevertheless, its aggregate effects remain limited given the relatively small share of adopting firms.

On the supply side of the labor market, Gofman and Jin (2024) suggest that while big-tech firms benefit from hiring top AI talent, this may come at the cost of reduced innovation and startup activity in the broader economy by disrupting the academic knowledge pipeline essential for training future AI entrepreneurs. They document a "AI brain drain" effect—the departure of AI professors from academia to industry negatively affects the creation and funding of AI startups by their former students.

In addition to observational studies, a recent wave of randomized experiments has begun providing causal evidence on how AI tools affect worker productivity and performance in specific tasks. Brynjolfsson et al. (2025) study the deployment of a GPT-3-based assistant among over 5,000 customer service agents at a Fortune 500 firm, finding a 15% average productivity increase, with gains reaching 30% among less experienced workers. In contrast, high-skill agents benefit little, and in some cases, performance slightly declines. Kanazawa et al. (2022) report similar patterns among taxi drivers using an AI navigation tool: productivity rises 5% on average and 7% for low-skilled drivers, narrowing skill-based performance gaps. Noy and Zhang (2023) show that ChatGPT reduces task time by 40% and improves writing quality, particularly for lower-performing professionals. Peng et al. (2023) find that GitHub Copilot accelerates software development by 56%, especially benefiting less experienced and older programmers. In a similar field experiment, Cui et al. (2024) estimate that the usage of a GenAI coding assistant led to a 26% increase in weekly task completion based on experiments at Microsoft, Accenture, and a Fortune 100 electronics firm. These productivity benefits were concentrated among more junior developers and recent hires, consistent with the findings in Peng et al. (2023).[36] Jabarian and Henkel (2025) conduct a large-scale natural field experiment involving 70,000 job applicants randomly assigned to interviews with human recruiters, AI voice agents, or a choice between the

[36] A related study by Sarkar (2025) documents that an AI coding agent, agentic AI systems that perceive and act on their environment to autonomously pursue objectives, increases coding mode, software output by 39% relative to trend, with no increase in error or revert rates, suggesting short-run productivity gains without quality deterioration. Interestingly, unlike prior AI tools that mainly benefited less experienced workers, agent usage and acceptance rates are higher among experienced workers.

two. They find that automating interviews with AI increases job offers by 12%, job starts by 18%, and retention up to four months by 16–18%.

In the context of financial services, Liu (2024) shows that AI improves sales productivity using a large-scale randomized experiment with 11,000 insurance agents in China. However, AI use also led to increased adverse selection—more high-risk consumers purchased insurance—suggesting a crowding out of agents' own risk-assessment efforts. Choi and Xie (2025) provide some of the first empirical evidence on how GenAI is reshaping financial accounting through a survey of 277 accountants, a formal theoretical model, and proprietary field data from 79 small- and mid-sized firms. The authors find that AI users serve 55% more clients and reallocate time from routine tasks like data entry (reduced by 3.5 hours per week) to higher-value activities such as client communication and quality assurance. AI also improves reporting quality and shortens monthly close cycles without sacrificing accuracy. Survey and field experiment results suggest that AI is most effective when complemented by professional judgment, especially as experienced accountants better interpret and intervene based on AI-generated uncertainty signals.

While the above-mentioned studies highlight AI's potential to boost individual productivity and reduce within-occupation skill disparities, the evidence is limited to narrowly defined tasks in specific settings and therefore cannot be generalized to broader macroeconomic or labor market outcomes.

6.3 Distributional impacts: wages, inequality, and labor share

Korinek and Stiglitz (2018) argue that the effects of AI on labor markets hinge on two key factors: the speed of AI's economic integration and the factor bias of this innovation—whether it complements or substitutes human labor. They contend that much of AI progress is likely labor-replacing, posing challenges for income distribution rather than aggregate efficiency. Bond and Kremens (2023) confirms this view that AI could potentially exacerbate income inequality. They analyze whether automation resulted from AI leads to capital dominance, where capital returns outpace wages, and income concentrates among capitalists, while workers become impoverished.

Their empirical calibration suggests current US automation rates are below the critical threshold, implying the economy will broadly retain its structure even as automation advances. Nevertheless, a theoretical work by Ide and Talamas (forthcoming) shows that different types of AI impact labor allocation and inequality in distinct ways: non-autonomous AI tends to benefit less-skilled workers, while basic autonomous AI may displace them and favor high-skilled individuals, and advanced autonomous AI can benefit both groups.

If AI indeed complements high-skilled workers and substitutes for lower-skilled or routine workers, one would expect an increase in wage inequality—a continuation of the skill-biased technological change trend that has characterized the late 20th century, where wages for college-educated or highly skilled workers rose faster than for others. There is evidence that earlier waves of computerization and robotics contributed to job polarization, benefiting high-skill professionals and some low-skill service workers while hollowing out many middle-skill, routine jobs (Autor and Dorn, 2013). However, AI's impact might not map neatly onto this pattern. Unlike earlier technological shifts that primarily affected middle- and low-skilled jobs (Autor and Dorn, 2013; Kogan et al., 2023), exposure to AI is disproportionately concentrated in white-collar occupations (Webb, 2019; Eloundou et al., 2024), which means even relatively skilled workers (e.g. those in technical, analytical, or creative professions) could face new competition from AI performing part of their work. On the other hand, AI tends to serve as a tool in many of these occupations that amplifies human productivity rather than a full replacement. In the short run, this can increase the productivity and potentially the earnings of the affected professionals.[37] This suggests a skill-biased effect in favor of already high-paying jobs, which would increase wage inequality by pushing high incomes higher.

At the same time, as seen in the experiments where less-experienced workers improved the most (e.g., Kanazawa et al., 2022; Cui et al., 2024; Brynjolfsson et al., 2025), AI's role as a "skill equalizer" in micro settings suggests that AI could reduce certain skill gaps. If a technology enables a moderately skilled worker to perform like a highly skilled worker, the wage premium for

[37] Felten et al. (2019) find that recent AI exposure has been associated with higher wage growth for occupations that are intensive in software and technical skills.

the highly skilled could diminish. Some theoretical scenarios even posit that AI might substitute for aspects of high-skill jobs (for instance, writing code or legal briefs), potentially lowering demand for those high-skill workers or compressing wages at the top. Webb (2019) conjectures an intriguing distributional twist along these lines: since AI is targeting many upper-middle-wage cognitive tasks, it might erode some of the wage advantages in those occupations (say, technical specialists and middle managers), which could compress wages between the 90th percentile and median. However, he also argues the very top earners may be less affected, so the share of income going to the very top might keep rising. Similarly, Autor (2024) argues that AI has the potential to help restore the middle class by allowing individuals with lower skill levels to perform more complex tasks.

Empirical evidence on these distributional outcomes remains limited, as broad AI adoption is just beginning. In the case of automation technology, Moll et al. (2022) show in their model that automation contributes to stagnant wages for workers, leading to a concentration of income at the top. Empirically, the authors calibrate the model to US data since 1980 and find that automation of routine jobs accounts for much of the observed increase in top income inequality and stagnation in lower-income wages. However, evidence on the impact of AI is mixed. Georgieff (2024) examine the relationship between AI and wage inequality across 19 OECD countries using data from 2014–2018, prior to the widespread adoption of GenAI. They find no evidence that AI has affected wage inequality between occupations, but it does find some evidence that AI may reduce wage inequality within occupations, possibly by narrowing productivity differences among workers. Cornelli et al. (2023) investigate the relationship between AI investment and income inequality across 86 countries between 2010 and 2019. The findings show that AI investment is associated with increased real incomes and income shares for the richest decile, while the bottom decile experiences no income growth and a decline in income share. AI adoption is also linked to reduced employment rates, a shift toward high-skill and managerial jobs, and a decline in labor's share of income. These effects are especially pronounced in AI-intensive sectors like real estate and robotics. Nevertheless, the paper does not provide evidence of a causal relationship.

One salient trend in many economies over the last few decades is the declining share of national income going to labor (wages and benefits), with a corresponding rise in the capital share.[38] If AI and automation allow capital owners to replace workers, one would expect a decline in labor's share unless new labor-intensive opportunities emerge or workers have enough bargaining power to raise their wages in line with productivity. The task-based model of Acemoglu and Restrepo (2018) explicitly predicts that automation tends to reduce the labor share, while the creation of new tasks that employ labor can increase it. Recent empirical evidence raises important concerns. Acemoglu and Restrepo (2019) find that US industries with faster automation and slower new task creation saw labor's share fall more. International evidence also points in this direction. Graetz and Michaels (2018) report that while robot adoption from 1993 to 2007 significantly boosted productivity and even average wages, it did not increase labor's share of output; the gains from higher productivity were largely accrued by capital and higher-skilled labor, not broadly distributed to the workforce.

A recent forward-looking analysis by Acemoglu (2025) provides estimates of AI's medium-term impact under current trends. His task-based simulation suggests that over roughly the next decade, AI is likely to further reduce the labor share and contribute little to ameliorate inequality. The intuition is that only a limited set of tasks will be automated in a short timeframe, and those tend to be tasks performed by middle- or lower-skill workers, so without concerted efforts to create complementary tasks for these workers, the primary beneficiaries will be firms and highly skilled workers who build or work alongside AI.

6.4 Macroeconomic outcomes: productivity and employment

As a GPT, AI has the potential to significantly boost aggregate productivity and economic growth, but the magnitude and timing of its impact are subjects of active investigation. Thus far, despite rapid progress in AI capabilities, there has not been a clear inflection point in economy-wide pro-

[38] According to the Bureau of Labor Statistics (the statistics can be accessed at https://www.bls.gov/productivity/tables/), labor share has declined for non-farm workers from about two-thirds, 64.1% in the first quarter of 2001, to 55.8% in the first quarter of 2024.

ductivity statistics attributable to AI. Brynjolfsson et al. (2019) emphasize that we may currently be in the flat part of the productivity J-curve with AI—meaning that the economy is still absorbing the costs of AI adoption (e.g., reorganizing workflows, developing new business models, training workers) and has not yet reaped the full benefits.

Several recent studies have tried to estimate how much AI could contribute to growth in the coming years. Filippucci et al. (2024) build a micro-to-macro model that factors in the share of tasks in each sector that could be enhanced by AI, the pace of AI adoption by firms, and general equilibrium effects. Their simulations suggest that AI could raise US Total Factor Productivity (TFP) growth by approximately 0.25 to 0.6 percentage points per year over the next decade. This is a meaningful contribution. It implies AI alone won't deliver a new era of 5%+ annual GDP growth, but it could help lift the economy out of the sluggish productivity growth experienced in the 2010s.[39] By contrast, Acemoglu (2025) offers a more tempered outlook. His task-based macro model, using current estimates of AI task exposure, suggests no more than a 0.5–0.6% increase in TFP over 10 years due to AI—essentially an order of magnitude smaller impact than Filippucci et al. (2024)'s optimistic scenario. Furthermore, after accounting for the likely slowdown as AI moves to more complex tasks, Acemoglu pegs the 10-year TFP gain at under 0.53%. The gap between these forecasts highlights the immaturity of AI deployment today: Acemoglu's empirical calibration reflects the relatively modest role of AI in the current economy, whereas Filippucci's involves forward-looking assumptions about rapid improvements and adoption.

It is also possible that we are underestimating AI's future impact by focusing on the current generation of technologies. If AI breakthroughs continue, the scope of tasks susceptible to automation could broaden dramatically, potentially leading to larger productivity jumps. Furthermore, the eventual macro impact of AI may come less from direct labor replacement and more from enabling innovations that were previously unattainable. Eloundou et al. (2024) highlight that many of the occupations highly exposed to LLMs are in research and development-intensive fields, hinting at AI's role as a catalyst for innovation. If those innovation effects materialize, the

[39] Between 2010 and 2019, the average annual TFP growth in the US was approximately 0.8%, significantly lower than the 1.4% average from 1949 to 2010 (Gordon and Sayed, 2022).

long-run impact on productivity could be substantially higher than what one would predict from current task automation alone.

Fears of "AI-induced mass unemployment" have been common in public discourse, but economists generally view this as unlikely in the long run, barring extremely advanced AI. Historically, even as technology eliminates some jobs, new ones emerge and employment has grown with the population. The more pressing issue is the transition: certain occupations or regions can suffer significant job losses, and workers may face spells of unemployment or need to retrain for new careers. Evidence from the introduction of industrial robots illustrates this point. Acemoglu and Restrepo (2020a) find that in US commuting zones that saw greater robot adoption in manufacturing (1990–2007), there were larger declines in employment and wages, especially for blue-collar workers. These local adverse effects of automation were substantial: each additional robot per thousand workers reduced the employment-to-population ratio by about 0.4 percentage points, and there was no evidence that other industries in those regions absorbed the displaced workers in the short run.

Korinek and Stiglitz (2018) examine the phenomenon of technological unemployment. Two mechanisms are emphasized. First, persistent unemployment may arise when wages fail to adjust due to efficiency wage considerations. Firms may pay "fair" wages above the market-clearing level to motivate worker effort, and if the marginal productivity of labor falls below this wage, workers may be priced out of employment altogether. In extreme cases, workers might not survive on market wages, resulting in sustained joblessness without government support. Second, unemployment may occur as a transitional phenomenon when job displacement outpaces job creation or retraining. Slow adjustment, sticky wage norms, and weak reemployment policies can intensify this problem. Given that jobs also provide identity and social purpose—not just income—Korinek and Stiglitz (2018) argue that subsidizing employment may be more beneficial than simply providing transfers.

It is worth noting that so far, we have not observed a sharp uptick in overall unemployment attributable to AI. If anything, labor markets in advanced economies have been quite tight in recent years, even as AI technologies made remarkable advances. This underscores that macroeconomic factors (business cycle, monetary policy, pandemics, etc.) dominate short-run employment fluctuations, and any displacement from AI is likely gradual. Hampole et al. (2025) provide a nuanced explanation for the limited observed impact of AI on employment, despite its rapid diffusion and clear micro-level effects. They develop a task-based model and empirical framework to measure how AI adoption affects occupational labor demand, focusing on the distinction between mean task exposure (average substitutability across tasks) and dispersion in exposure (variation in which tasks are affected). They link firm-level AI implementations to specific tasks within occupations. They find that higher mean exposure to AI reduces employment in affected occupations, while greater dispersion in exposure increases it by enabling reallocation of worker effort toward non-automated tasks. These opposing forces dampen net job loss and explain why aggregate AI effects appear muted despite strong substitution at the task level.

7 Risk and Challenges

The integration of AI into financial systems holds transformative potential but also poses significant challenges. This section examines these concerns in depth, explores strategies for mitigating the associated risks, and outlines key regulatory and policy implications.

7.1 Methodological challenges of GenAI in finance research

The application of LLMs in financial systems as well as financial research introduces several important challenges that warrant careful consideration. We now discuss key limitations of LLMs that researchers should be aware of.

Hallucinations and errors: A foremost concern is that LLMs have a tendency to "hallucinate" or "confabulate," i.e., to produce confident but false or fabricated outputs (Ji et al., 2023; Zhang et al., 2023). This deficiency is well-documented as a fundamental problem, especially acute in domains like finance that demand accuracy.[40] Kang and Liu (2023) find that "off-the-shelf" LLMs experience serious hallucination behaviors in financial tasks. For example, a model might inaccurately summarize a financial report or invent a statistic, which, if relied upon in trading or risk management, could lead to substantial losses. It is possible to detect hallucinations by training the detector with both positive examples (correct statements) and negative examples (explicitly labeled incorrect statements) (Karbasi et al., 2025). Ji et al. (2023) suggest two hallucination mitigation techniques: (1) data-related methods, such as building or cleaning datasets and augmenting inputs with external knowledge; (2) modeling methods, such as adapting architectures, using planning, reinforcement learning, multi-task learning, and controllable generation. Further, post-processing—refining outputs from fluent models—can improve faithfulness. However, each approach involves trade-offs between faithfulness, fluency, data needs, and generalization. Zhang et al. (2023) review recent work toward addressing the problem and suggest similar mitigation strategies. They also introduce emerging approaches, including multi-agent collaboration (e.g., Li et al., 2024b), prompt engineering (e.g., chain-of-thought, system prompts), and leveraging internal model states. Some researchers in financial economics have made serious efforts to address this concern. For example, Fang et al. (2024) mitigate hallucinations by decomposing complex questions into sub-questions, providing precise definitions and counter-examples, requiring responses to include supporting text, reasoning, and confidence scores, and by separating extraction from classification tasks. They also validate their approach through word clouds, manual inspection, heuristic ("smell") tests, and time-series cross-validation.

Look-ahead bias and overfitting: Financial data are inherently temporal, so look-ahead bias is a chronic pitfall when applying LLMs. LLMs trained on vast internet text might inadvertently incorporate future information relative to a given prediction task (e.g., knowing how a past market event turned out). As financial economists learned from the “quant quake” and

[40] Lo and Ross (2024) discuss how hallucinations of LLMs in the financial advisory sector pose serious risks to investors and financial stability. A recent business example of this risk is Air Canada being held responsible for inaccurate information given to a customer by its chatbot, which was powered by LLMs (Garcia, 2024). Due to these reliability concerns, Stokel-Walker and Van Noorden (2023) emphasize the need for regulation, transparent usage, improved attribution systems, and safeguards against misuse, as these tools become more deeply embedded in research, education, and broader scientific communication.

multiple testing issues, models that look superb in-sample may crumble out-of-sample. The high flexibility of LLMs exacerbates this risk—they can fit noise as easily as signal. Developing robust evaluation frameworks is an active area of methodological research to ensure that AI-generated insights reflect genuine patterns rather than artifacts. To address this issue, Glasserman and Lin (2023) and Sarkar and Vafa (2024) introduce robust methods for isolating temporal information, such as systematic data-masking strategies.

Crucially, verifying model performance on data released after a stated "knowledge cutoff" is insufficient, as many closed-source models, like ChatGPT or Claude, are continuously updated through post-training fine-tuning, particularly with reinforcement learning from human feedback (RLHF), which can inadvertently leak future knowledge into past contexts (Ludwig et al., 2025). Examples of mitigating look-ahead bias in financial economic research include masking firm identifiers (e.g., Breitung and Müller, 2025), sub-sampling (e.g., Jha et al., 2024a), and altering dates in the data (e.g., Gao et al., 2025).[41] He et al. (2025) introduce chronologically consistent LLMS, ChronoBERT and ChronoGPT, which are trained exclusively on text available at each point in time to eliminate lookahead bias and training leakage. The authors build annual vintages of models from 1999 to 2024 that remain competitive with widely used open-weight models despite strict temporal data constraints. On language tasks, ChronoBERT and ChronoGPT outperform or match existing no-leakage models and even rival BERT. Real-time outputs from ChronoBERT and ChronoGPT achieve Sharpe ratios comparable to a much larger Llama model in predicting next-day stock returns from news.

Methodological replicability: One concern is the variability in outputs across different model versions, APIs, and even slight changes in prompt construction. This presents a significant barrier to research replicability and cumulative scientific progress. Mirzadeh et al. (2024) conduct a large-scale study across 25 models and show that small changes, such as altering numerical values or adding irrelevant clauses, cause significant performance drops (up to 65%), suggesting that current LLMs rely heavily on pattern-matching rather than genuine logical reasoning.

[41] In the context of finance-specialized LLMs, Rahimikia and Drinkall (2024) address look-ahead bias by pre-training separate annual models strictly on historical textual data.

Similarly, Ross et al. (2024) and Ouyang et al. (2024) show that LLMs' economic decision-making depends on how they are fine-tuned. As a baseline, researchers should eliminate stochasticity in generation by setting parameters such as temperature to zero. In addition, researchers introduce workarounds, such as feeding summaries of past interactions back into the prompt or using external databases. Nevertheless, this is not foolproof and can lead to errors or "forgotten" information that a real economic agent would remember.

Alignment and objective functions: Another challenge is the alignment of GenAI outputs with human objectives and constraints. For example, Fedyk et al. (2024) find that default ChatGPT-4 is biased—its responses disproportionately reflect the preferences of young, high-income individuals. They show that prompt engineering is needed to ensure AI outputs align closely with human responses. By design, LLMs predict text sequences; they do not inherently understand truthfulness, fairness, or legal compliance. If tasked with maximizing profit or prediction accuracy, an unaligned AI might generate strategies that conflict with ethical or regulatory norms—for example, recommending exploitative lending practices or manipulating market quotes. In finance, this concern is not merely hypothetical: a trading agent could conceivably learn to trigger other algorithms' stop-loss orders for profit (market manipulation) or output misleading advice that boosts short-term returns at the expense of clients. Regulators and researchers stress the need for transparency, human oversight, and accountability in AI systems to combat such risks (Aldasoro et al., 2024). In practice, this means any generative model used in decision-making should have constraints or post-processing checks (for instance, filtering outputs through compliance rules) to ensure it adheres to the "rules of the game." The extent to which current LLMs can be aligned with fiduciary duties or the public interest is an open question. Early evidence suggests simple prompt engineering is insufficient (e.g., Ouyang et al., 2024)—deeper solutions (like fine-tuning on expert-verified data or integrating causal reasoning modules) may be required to guarantee that AI recommendations serve intended outcomes.

Model interpretability: The black-box nature of AI poses a significant barrier to their adoption in finance, where understanding the rationale behind a decision is often as important as the decision itself. Complex deep learning models lack the explainability of traditional financial models or econometric approaches. This opacity is "massively exacerbated in GenAI" compared to prior AI (Oecd, 2023b).[42] Finance professionals and regulators are understandably wary of a model that, for example, flags a borrower as high-risk or executes a trade but cannot explain why in human-understandable terms. Moreover, it hampers learning: if an AI-driven fund underperforms, managers may struggle to adjust strategy because the model's reasoning is not transparent. There is active research into interpretability techniques (like prompt-based rationales, LLM self-explanation, or mapping attention weights to concepts), and even frameworks to evaluate an LLM's reasoning in economic tasks. A promising development by Lopez-Lira and Tang (2023) is an "interpretability framework to evaluate LLMs' reasoning and accuracy" in the context of return prediction. Such approaches could illuminate what information the model is relying on (e.g., macro news vs. firm fundamentals) and help ensure it conforms with financial theory and common sense. Nonetheless, achieving a level of interpretability acceptable to practitioners remains an important challenge. Another caveat is the validation of AI-based findings. The model might be systematically wrong in ways we don't immediately recognize. Thus, findings from AI-agent simulations should, whenever possible, be checked against human data or strong theoretical expectations. Many researchers (e.g., Hansen et al., 2024; Zarifhonarvar, 2024) have stressed that these models are best used as a supplement to, not a replacement for, traditional evidence.

Bias and data quality: Sources of bias in GenAI can arise from different phases of the machine learning pipeline, including data collection, algorithm design, and user interactions (Ferrara, 2023). In finance, this raises concerns about algorithmic bias and fairness in, for example, discriminatory outcomes in lending, credit scoring, or fraud detection (Favaretto et al., 2019; Gillis et al., 2021). If an LLM is trained on decades of financial text where certain groups were historically underserved or stereotyped, the model might unwittingly perpetuate those biases (e.g., suggesting higher interest rates for minority borrowers due to biased training examples). Bias is not limited to protected classes; models might also favor certain asset classes or geogra-

[42] However, Singh et al. (2024) argue that LLMs can enable more flexible, direct, and human-understandable explanations. For example, users can ask LLMs targeted questions about their reasoning or data insights.

phies if the data frequency is uneven. Data quality is another practical concern—financial text can be noisy (think of social media rumors) or unstructured (PDF reports), and mistakes in data preprocessing can propagate through an LLM's analysis. Biases in training data can translate into biased behavior. If an LLM learned from texts full of certain stereotypes or errors, its decisions will reflect those. For example, if most financial news in the training data was optimistic, the AI may systematically be too bullish in forecasts (Chen et al., 2024b). Empirical researchers should validate that AI-driven decisions do not have disparate impacts unless justified by sound risk factors. This intersects with regulation: in consumer finance, for example, laws require lenders to provide specific reasons for adverse decisions and to avoid discrimination. If a bank uses an LLM to help screen loan applications, it must ensure compliance by auditing the model's decisions for bias and explicability.

7.2 AI-induced risks and regulatory response

The current lack of a coherent and enforceable AI regulatory framework, along with the absence of well-developed models to evaluate AI's implications for financial stability, presents a complex and evolving challenge for regulators and financial institutions aiming to manage systemic risk in the age of AI. Recent work underscores the urgency of addressing these vulnerabilities by quantifying the macroeconomic value of proactive mitigation: allocating at least 1% of GDP annually could be justified to reduce AI-related existential risks.[43]

The broad integration of AI into the financial sector poses systemic risk through several unique mechanisms. These include the procyclical nature of AI, its capacity to rapidly execute and magnify financial outcomes, increased interdependence across institutions through AI-powered networks and reliance on third-party providers, and the emergence of new forms of "too big to fail" entities. We discuss them below in detail.

Procyclicality: AI-driven trading systems may amplify market swings by reacting in highly

[43] Jones (2025) develops a simple economic model to estimate the optimal share of GDP that should be allocated annually toward reducing AI-related existential risks. Calibrated to plausible assumptions, the model finds that even under conservative estimates, spending at least 1% of GDP annually over the next decade is typically justified. On average, optimal spending exceeds 8% of GDP.

synchronized and rapid ways to news or asset price fluctuations (International Monetary Fund, 2024). Empirical evidence shows that algorithmic tradings, while improve price discovery(e.g., Chaboud et al., 2014), tend to exhibit higher turnover and faster responses, which can deepen volatility during stressed market conditions (Kirilenko et al., 2017). Additionally, the use of similar training data and model architectures across financial institutions increases the risk of herding behavior, leading many firms to simultaneously adjust portfolios or risk forecasts based on correlated AI signals. While each firm may be acting rationally, their collective reliance on AI tools could reinforce market trends and create feedback loops that exacerbate asset bubbles or crashes.

Concentration of AI models and service providers: Many financial institutions now depend on a limited number of external AI platforms, data vendors, and foundational model providers. This concentration raises the possibility that an error or outage in one model—whether due to a design flaw, data bias, or cyber attack—could propagate quickly and simultaneously across a large number of firms. These dynamics reflect a new variant of the "too big to fail" problem, wherein foundational AI systems become critical infrastructure for the financial sector without being subject to robust regulatory oversight. The result is an increase in interdependencies and a reduction in system resilience, particularly given that many AI providers fall outside the traditional regulatory perimeter of banking and securities supervision (Lo and Ross, 2024; Bradford, 2023; Floridi, 2024).

Algorithmic collusion: Algorithmic collusion presents another emerging threat (Fish et al., 2024; Calvano et al., 2020; Oecd, 2023a; Dou et al., 2024). Advanced pricing algorithms may independently learn to sustain supra-competitive pricing without explicit communication between firms. Recent empirical work documents this phenomenon in the German retail gasoline market, where the adoption of algorithmic pricing tools led to higher prices and margins—particularly in oligopolistic settings (Assad et al., 2024). The delay in these effects, consistent with learning dynamics, and the greater likelihood of coordinated pricing behavior among adopters suggest that AI tools may enable tacit collusion over time. In securities trading, Dou et al. (2024) show that AI collusion can robustly emerge through two distinct algorithmic mechanisms: (1)

price-trigger strategies, where agents tacitly coordinate based on price movements, and (2) over-pruning bias, where algorithms systematically undervalue aggressive strategies due to noise-driven losses, leading to conservative, profit-maximizing behavior. These findings raise concerns for competition policy, particularly in markets where AI systems operate with limited transparency or human oversight. As AI agents take on increasingly autonomous roles in trading, pricing, and contracting, regulators face the challenge of attributing responsibility when socially undesirable equilibria emerge, even absent malign intent.

Fairness and bias: The rapid adoption of AI has raised concerns about the possibility that benefits from improved statistical modeling might not be equally distributed (e.g., Hardt et al., 2016; Kleinberg et al., 2016; Kleinberg et al., 2018b). Fuster et al. (2022) study how the adoption of ML techniques in credit markets impacts distributional consequences. They find that ML models improve default prediction accuracy but disproportionately benefit White non-Hispanic borrowers. Black and Hispanic borrowers are less likely to see gains and face increased within-group interest rate dispersion. These disparities are driven mostly by higher model flexibility rather than triangulation on restricted characteristics, suggesting that improved technology can unintentionally amplify inequality even when explicit discrimination is absent. However, focusing on a setting with no credit risk, Howell et al. (2024) find that lenders with more automated processes are substantially more likely to lend to Black-owned businesses. Automation enables smaller loans, broader geographic coverage (especially in underbanked areas), and reduces opportunities for human discrimination. They highlight that automation—not just algorithmic underwriting—as a key mechanism for reducing racial disparities in small business lending.

Model opacity: The above-mentioned risks are further compounded by the opacity and unreliability of some AI outputs. Many LLMs remain "black boxes," generating plausible-sounding but sometimes factually incorrect outputs (Blattner et al., 2021). In finance, such hallucinations may mislead users, especially when outputs are used to support high-stakes decisions such as credit approvals, investment allocations, or risk management. The danger of automation bias, i.e., blindly trusting AI-generated content, may lead to widespread misjudgment if inaccurate AI assessments are taken at face value (Lo and Ross, 2024). Moreover, as AI systems become increasingly embedded in routine operations, the line between human and machine responsibility may blur, complicating regulatory enforcement, liability attribution, and model governance.

Regulations: While regulatory responses are underway, they remain fragmented across jurisdictions. Broadly, governance frameworks follow three models: the market-driven approach in the United States, the state-driven model in China, and the rights-driven framework in the European Union (Bradford, 2023). Despite differences in emphasis, most regulatory bodies now converge on shared principles, including transparency, accountability, human oversight, and proportionality in risk management (Ala-Pietilä et al., 2020; China Daily, 2019). The EU's forthcoming AI Act imposes binding obligations on "high-risk" applications, including many in the financial sector. The United States has issued guidance through agencies such as NIST,[44] and regulators like the SEC and Federal Reserve are evaluating how to extend existing model risk management frameworks to AI. China has introduced national standards focused on algorithmic transparency and compliance with social objectives. International bodies such as the OECD and ISO/IEC 23894:2023 provide global benchmarks,[45] and financial regulators such as the FSB have emphasized the need for systemic risk monitoring, data auditability, and third-party provider oversight.

Several policy responses are warranted to address these emerging risks. As discussed by Floridi (2024), AI models used in financial decision-making should be subject to rigorous stress testing, independent audits, and ongoing validation procedures. Just as banks are required to conduct capital stress tests, AI systems must be evaluated for performance under adverse scenarios and rare shocks. Model transparency should be improved, including disclosure of training data sources, design assumptions, and known limitations. Regulators should address vendor concentration by requiring contingency planning for reliance on critical third-party AI providers. Competition policy may also play a role in ensuring that AI innovation is not monopolized by a

[44]See NIST website for details: https://www.nist.gov/itl/ai-risk-management-framework

[45]See the document "Information technology — Artificial intelligence — Guidance on risk management" at https://www.iso.org/standard/77304.html.

handful of firms whose failures could destabilize the system. Finally, market surveillance tools must be updated to detect AI-induced distortions, such as anomalous trading patterns or coordinated price movements, and safeguard mechanisms such as circuit breakers should be adapted for algorithmic contexts (Oecd, 2023a). Human oversight remains essential: high-stakes AI systems should be monitored or overrideable by qualified personnel to prevent cascading failures or unintended coordination.

7.3 Policy implications in labor market

AI's impact on labor markets is neither exogenous nor deterministic. Instead, it is contingent on policy and institutional responses that determine whether AI augments human labor, displaces workers, or exacerbates inequality.[46] The economic literature increasingly emphasizes that without deliberate policy intervention, the gains from AI are likely to be concentrated among capital owners and high-skilled elites, while many workers face stagnant wages, eroding bargaining power, and reduced labor market participation (Acemoglu, 2021; Morton, 2024).

The task-based framework offers a useful lens through which to assess AI's labor market effects. AI technologies can automate specific tasks within jobs rather than replacing entire occupations. When AI displaces routine or codifiable tasks, it can reduce demand for middle-skill workers—particularly in administrative, clerical, and certain analytical roles. Conversely, AI can augment workers engaged in non-routine cognitive or interpersonal tasks, raising their productivity and potentially their earnings. Early evidence suggests that such effects are unevenly distributed (Athey et al., 2024). While GenAI tools may benefit high-skilled professionals (e.g., programmers, writers, or lawyers), workers in lower-skilled service jobs may remain less affected due to the physical and social complexity of their roles. The result is likely to be a continuation, if not an amplification, of job polarization: shrinking demand for middle-skill occupations, growing employment at the high and low ends of the skill spectrum, and rising wage inequality.

[46] Korinek (2024) argues that AI will transform the economy by shifting value from human labor to reproducible factors like computation, ending the industrial age. More broadly, it outlines eight key policy challenges: inequality, human capital erosion, instability, macroeconomic shifts, antitrust, intellectual property, environmental impact, and global governance.

Moreover, the direction of AI innovation is not neutral. Firms often invest in automation technologies that lower costs and reduce reliance on human labor, particularly when private incentives do not align with social welfare. This bias toward labor-replacing AI, driven by cost-cutting motives, can lead to underinvestment in technologies that complement workers. As Acemoglu and Restrepo (2020b) and Acemoglu et al. (2023) argue, innovation policy should explicitly steer AI development toward augmentative applications—AI that enhances rather than replaces human capabilities. Governments can do this by subsidizing R&D in human-centric AI, adjusting tax codes that favor capital over labor, and prioritizing public investments in sectors with unmet labor demand, such as education, health care, and infrastructure (Korinek and Stiglitz, 2018; Mann, 2019).

Beyond steering innovation, labor market institutions must evolve to support workers through technological transitions. Expanding access to education and vocational training is essential, particularly in digital skills, data literacy, and human-AI collaboration. Retraining programs should be linked to labor market needs and complemented by income support to allow workers to transition without undue hardship. Strengthening collective bargaining institutions can ensure that workers share in the productivity gains from AI (Oecd, 2023c). Unions and worker councils could negotiate deployment plans for AI tools, advocate for transparency in algorithmic management systems, and secure commitments for job retraining or profit-sharing arrangements.

Distributional concerns also call for a more robust tax-transfer system. Progressive taxation on high incomes, capital gains, and automated value-added could redistribute some of the gains from AI to fund social protection and public services. Proposals such as universal basic income (UBI) aim to ensure a minimum standard of living amid potential declines in wage-based income (Rubin, 2024). Although still controversial, UBI and related policies are increasingly discussed as long-term strategies to manage structural changes in the labor market. Short of UBI, measures such as expanded unemployment insurance, wage subsidies, and portable benefits can provide flexibility and security for workers navigating AI-induced disruptions.

Furthermore, alternative models of ownership and value distribution may be necessary to prevent a winner-take-all outcome. These include employee equity stakes in AI-driven firms, data dividends for individuals contributing to model training, and cooperative or public ownership of foundational AI infrastructure. Historical parallels with the Industrial Revolution suggest that inclusive prosperity does not emerge automatically from technological change. Rather, it requires institutions that align innovation with broad-based welfare gains. As Morton (2024) emphasizes, the challenge is to build governance systems where the benefits of AI are not captured solely by the most powerful actors but are shared across society through deliberate policy design.

8 Discussion and Future Directions

The rise of AI in finance presents new opportunities and challenges that demand focused research. This section discusses five key areas for future work in financial economics: improving model interpretability, using LLMs as economic agents, designing effective human–AI collaboration, establishing causal inference, and assessing AI's long-term welfare and structural impacts.

8.1 Unpack the "black box"

A major limitation of AI, particularly ML methods, is their "black box" nature: combining or stacking multiple models often obscures how individual input variables influence the final prediction, resulting in poor interpretability. However, in finance research and real-world applications, it is critical to understand the rationale behind model outputs. An important direction for future work is to interpret and uncover the mechanisms behind the outputs of these models. Economists have long stressed that prediction alone is not enough—AI excels at prediction, whereas economics often demands understanding causal parameters and theory (Mullainathan and Spiess, 2017). Researchers have attempted to bridge this gap. For example, Bell et al. (2024) introduce Explainable Boosting Machines (EBMs). Unlike black-box models such as neural networks or ensemble methods, EBMs reveal functional relationships (e.g., nonlinearities, asymmetries) through their additive structure, enabling researchers to understand and visualize the role of each predictor.

Lopez-Lira and Tang (2023) propose a novel two-step interpretability framework that makes their financial predictions more transparent, allowing researchers to systematically investigate the reasoning behind LLMs' predictive decisions. Further, Ludwig and Mullainathan (2024) demonstrate how ML can be used to surface unexpected relationships. In their case, an ML algorithm revealed that judges' incarceration decisions appeared to rely on defendants' facial characteristics, a pattern that then became a testable hypothesis. Such approaches help open the black box by using AI to identify candidate mechanisms that economists can subsequently analyze with traditional empirical techniques.

Opening the black box is also crucial for algorithmic accountability and regulation. In high-stakes financial decisions such as lending or insurance underwriting, regulators face a dilemma: complex ML models are often too opaque to fully audit, yet requiring simplistic models can be inefficient. Recent work by Blattner et al. (2021) tackles this by proposing an algorithmic auditing framework. Instead of demanding full transparency (often impossible with thousands of parameters), they suggest using simplified surrogate models, "explainers", targeted at the aspects of the black-box model most relevant to regulatory objectives (such as fairness or risk sensitivity). Future research can develop better measures of an AI model's economic rationale, for example, devising tests to see if a stock-picking algorithm is implicitly trading on known risk factors or exploiting market inefficiencies, or if a robo-advisor's recommendations align with classical portfolio theory.

As discussed in Section 2 and 4, AI's strength in prediction has led to powerful forecasting models. A promising future direction lies in the development of large-scale multimodal datasets that integrate diverse data types into unified modeling frameworks, such as structured financial variables, textual information from news and regulatory filings, social media sentiment, satellite imagery, and network or geospatial data. By combining these sources, AI models can uncover early and often overlooked signals of value, risk, or market sentiment that would be invisible in traditional datasets. Most importantly, these models must move beyond predictive accuracy alone and "open the black box" by examining the partial effect of each information source on the model's output.

Furthermore, policy decisions in finance and economics hinge on causality (what causes a price move, or what is the impact of a new lending algorithm on credit access?). Mullainathan and Spiess (2017) emphasize that ML tools "solve a different problem" than traditional econometrics, and using them in economics requires adapting them to answer causal questions. On the analytical side, there is exciting progress in combining AI with causal inference techniques. One direction is using AI to enhance causal discovery. For example, Han (2024) propose using LLMs to search for new instrumental variables through narratives and counterfactual reasoning, similar to how a human researcher would. Another is the development of causal ML algorithms that directly target treatment effect estimation.

8.2 LLMs as economic agents

Current studies all point to the possibility that GenAI can serve as a versatile new kind of economic agent—one that can be utilized, analyzed, and replicated in ways human agents cannot. If GenAIs can reliably play the role of economic agents, rather than relying solely on theoretical proofs or limited lab experiments with human subjects, economists could employ a rich simulated society of AI agents to test policies or market designs. Likewise, regulatory agencies might use AI agents to stress-test a trading strategy by populating a market with hundreds of AI-driven traders (e.g., Fish et al., 2024), examine auction designs by having AI bidders with different risk aversions (e.g., Horton, 2023), model mutual fund redemption (e.g., Anand et al., 2025), and study depositor behavior in panic-driven bank runs (e.g., Kazinnik, 2026). In the context of household finance, Huang and Ouyang (2025) use multi-round, LLM-driven simulations to validate a model of investor-advisor conversations, demonstrating how frictions like investor impatience or the AI's lack of memory impact the quality of financial advice. The cost efficiency and scalability of AI simulations are a huge advantage, particularly in areas such as financial market stability testing or policy impact analysis: one can run thousands of trial scenarios with slight tweaks to see how outcomes change, which is infeasible with human subjects. Furthermore, using LLMs as economic agents might help incorporate qualitative knowledge (narratives, sentiments, cultural context) into quantitative economic forecasting. Because these models learn from text, they carry a form of common sense and contextual reasoning that pure numerical models lack. This could bridge the gap between qualitative economic insights (often found in analyst reports or news commentary) and formal quantitative analysis.

Nevertheless, the potential of LLM agents is balanced by clear limitations. These models are imperfect proxies for human cognition (Horton, 2023). They sometimes exhibit illogical or extremist behavior not grounded in reality, and they reflect the biases in their training data. Therefore, one research direction is to validate and refine LLM-based simulations. Empirical tests could compare LLM agent outcomes with real human data to identify where the model deviates. There is also scope for developing better-aligned AI agents (through fine-tuning or novel architectures) that more faithfully represent economic actors' behavior (e.g., Ouyang et al., 2024). Another promising avenue, highlighted by Manning et al. (2024), is to incorporate structural causal models into the simulation frame. By guiding LLM agents with an underlying economic structure (e.g., utility functions or budget constraints), they find that when allowed to interact within a causal model, the LLM could generate and test hypotheses that it could not articulate on its own. This suggests that LLMs may possess latent economic knowledge that careful experimentation can reveal. Going forward, using LLMs as economic agents will likely become a valuable tool for theorists and empiricists alike—a way to conduct rapid, low-cost experiments in areas from household finance (simulating consumer financial decisions) to corporate finance (simulating negotiations or managerial decisions), thereby generating new insights and sharpening research questions to be later verified in the real world.

8.3 Human–AI complementarity

As AI systems become more capable, an important question is which tasks in finance should remain human-led and how to optimally integrate AI with human expertise. Rather than a wholesale automation of financial decision-making, many scholars envision hybrid human–AI systems where each party does what it excels at (e.g., Costello et al., 2020; Greig et al., 2024; Cao et al., 2024b; McLaughlin and Spiess, 2024). Indeed, there is evidence that naïve combinations of humans and algorithms often underperform—algorithmic assistants "again and again fail to improve human decisions" (McLaughlin and Spiess, 2024). One reason is that humans may mistrust or misinterpret algorithmic recommendations, or conversely, over-rely on them in ways that negate human expertise. Another reason could be that social image concerns constitute a meaningful barrier to effective AI adoption (Almog, 2025). Future research could address how to design AI that truly augments human judgment in finance.

McLaughlin and Spiess (2024) provide a blueprint for this by formalizing the interaction of a recommendation algorithm with a human decision-maker. Using a causal framework, they classify human responses to AI advice, for example, distinguishing between users who would follow any recommendation versus those who only follow advice when it confirms their prior inclination. This allows the algorithm to adapt by only intervening when it can genuinely change a human's decision for the better. They demonstrate this approach in a hiring experiment and show it can improve decisions by achieving better human–AI complementarity. Such frameworks could be applied to financial settings. For example, robo-advisors that adjust their suggestions based on an investor's tendency to heed or ignore advice, or credit underwriting models that identify when to defer to a human loan officer's discretion. Huang and Ouyang (2025) develop a framework where the choice between a human and an AI advisor depends on the nature of the investor's uncertainty. They model the interaction with an AI as an optimal stopping problem, showing that communication primarily helps investors resolve uncertainty about their own objectives rather than about external fundamentals. Furthermore, the human element can be a liability rather than an asset in some contexts, and AI can serve as a substitute where trust in human discretion fails.[47] In this sense, trust capital becomes a key dimension along which AI and human agents interact, and the optimal combination remains an open question.

[47] Using the Wells Fargo scandal as an exogenous shock, Yang (2025) finds that a loss of trust in traditional banks leads to increased household adoption of FinTech mortgage lenders. This shift occurs because FinTech lenders are perceived as less likely to deceive borrowers due to reduced human involvement and more transparent, standardized processes.

One lesson from early studies is that AI may be especially valuable as a training and decision-support tool for less-experienced humans (e.g., Brynjolfsson et al., 2025; Peng et al., 2023; Cui et al., 2024). The AI system effectively disseminates the best practices of top performers to new staff, helping them climb the learning curve faster. This suggests that in a financial context, junior analysts or retail investors might benefit the most from AI guidance (e.g., summarizing earnings reports or suggesting portfolio rebalancing), whereas seasoned experts might rely on their own experience. Future work should explore such heterogeneity: Which tasks (trading, risk management, financial advising, etc.) see the greatest improvement from human–AI collaboration, and for whom? For example, Zhong (2024) finds that in multi-layered decision-making processes involving humans and AI, optimal outcomes are achieved by placing higher-quality technologies (those with lower type-1 error rates) closer to the final decision layer, acting as safeguards against earlier mistakes. Moreover, trust and transparency are critical in these hybrid systems. Research could explore design features that foster trust and reduce algorithm aversion (e.g., Dietvorst et al., 2018; Stradi and Verdickt, 2025).

8.4 Causal inference

Going forward, an important direction for research is developing strategies or collecting data to isolate the causal effect of AI technologies on firms, markets, and households. On the empirical front, some studies have started to establish causality using firm-level data and rigorous identification strategies (e.g., Babina et al., 2024; Babina et al., 2023a; Cheng et al., 2025), and other researchers are starting to treat the rollout of AI as a natural experiment. For instance, field studies and quasi-experiments have examined AI adoption within companies (Brynjolfsson et al., 2025; Kanazawa et al., 2022; Noy and Zhang, 2023; Peng et al., 2023; Cui et al., 2024). Echoing the call for more granular data on AI (Seamans and Raj, 2018; Frank et al., 2019), task-level data on how participants in financial markets use AI can shed light on the specific mechanisms through which it influences financial decision-making.[48] These will help provide new insights into old questions

[48] For example, analysts (see survey evidence by Christ et al. (2024)), households, and retail investors (e.g., Cheng et al., 2025). There is not yet research using individual-level data.

like: How much incremental profit does robo-advising causally provide to retail investors? How does the adoption of AI credit scoring expand credit access for underserved households? And what is the effect of AI-driven financial advice on household savings behavior?

8.5 AI in the long run

Finally, a crucial set of future research questions revolves around the long-run structural implications of AI in financial economics. The current AI hype might be another bubble paralleling historical examples like the Dot-Com Bubble, Telecom Bubble, Chinese Tech Bubble, Cryptocurrency Boom, and the recent Tech Stock Bubble (Floridi, 2024). Brynjolfsson et al. (2024) call for a comprehensive research agenda to understand these long-run effects, ranging from economic growth to labor markets and inequality.[49] One fundamental issue is the impact of AI on inequality. If AI automates tasks that were previously done by lower-skill workers (for example, automating routine credit underwriting or customer service jobs in banking), it could depress wages and opportunities for those workers, widening the skill gap. On the other hand, AI could lower costs and improve access to financial services for traditionally underserved groups (for instance, fintech apps using AI might offer cheap banking services to unbanked households). The net effect on inequality is an open empirical question. Research like Korinek and Stiglitz (2018) has voiced concerns that, without intervention, AI's benefits might accrue primarily to those who control the technology (large firms or high-skill workers) and exacerbate wealth gaps. Future research will likely use macro-finance models and calibration to assess how AI-driven productivity gains are distributed across capital and labor, and whether policies (training, redistribution, competition policy) are needed to ensure broad-based gains. An open question to theorists is how to model AI: researchers lack a settled, formal framework to represent AI within economic models. Perhaps incorporating data-as-an-asset thinking into AI modeling is a promising direction (e.g., Jones and Tonetti, 2020; Abis and Veldkamp, 2024).

[49] As the research agenda on AI advances, it will raise critical policy challenges across labor, taxation, education, social insurance, macroeconomics, antitrust, intellectual property, environmental sustainability, and global governance (Brynjolfsson et al., 2024)

AI is also poised to alter market structure and competition in finance. On one side, AI tools can level the playing field, for example, cloud-based AI services allow even small fintech startups to access world-class algorithms, potentially spurring competition. On the other side, AI might have scale effects that favor big players: large financial institutions have more data to feed AI models and more resources to develop them. Recent evidence from China shows a nuanced picture: cloud computing technology tended to decrease industry concentration by enabling new entrants, but AI adoption increased concentration by disproportionately benefiting large incumbent firms (Lu et al., 2024). In other words, AI might be creating "winner-take-most" dynamics in certain markets, as bigger firms leverage algorithms to cement their advantages. Going forward, research should examine whether this pattern holds in other contexts, for instance, do big asset managers gain an edge from AI that smaller firms cannot match? Do "Big Tech" firms extending into finance (with their AI prowess) threaten to dominate financial services? Most importantly, what role can regulation or data-sharing policies play in mitigating the scale advantages of large firms while preserving the efficiency gains of AI adoption in financial systems?

Another structural dimension is market dynamics and financial stability. Widespread use of AI in trading and asset allocation could change how markets behave. On one hand, AI might improve market efficiency, quickly arbitraging away mispricings and incorporating news faster, which could reduce volatility. On the other hand, if many actors rely on similar black-box models, markets could become more brittle; correlated errors or herding by algorithms might lead to flash crashes or systemic risks. An important direction for finance research is to study these systemic effects, for example, using agent-based models (potentially with AI agents) to see if certain algorithmic strategies amplify tail risks in asset prices (e.g., Dou et al., 2024). Likewise, in banking, if AI-driven credit models become ubiquitous, a model error or an unforeseen shock (say, all models too optimistically assess a certain loan segment) could have system-wide consequences. Regulators and researchers are starting to consider algorithmic stress tests analogous to traditional bank stress tests to assess how an AI-heavy financial system would respond to extreme scenarios. Work like Blattner et al. (2021) underscores the need for new regulatory tools when financial decisions are driven by opaque algorithms.

Finally, an additional long-term consideration is the evolution of financial services and institutions under the influence of AI. Will we see new financial products tailored by AI to individual consumers (truly personalized portfolios, insurance, credit terms)? How might AI alter the traditional roles of financial intermediaries? For example, if AI can directly connect savers and borrowers through decentralized finance platforms, does the importance of banks wane or change (e.g., Sockin and Xiong, 2023a; Sockin and Xiong, 2023b)? And what are the welfare implications of these shifts? There may be trade-offs between efficiency and other values: an AI-optimized market might be hyper-efficient, but perhaps at the cost of reduced human oversight or new forms of exclusion.

Long-term welfare analysis of AI in financial economics will extend beyond standard efficiency metrics to consider fairness, inclusion, and stability. As Brynjolfsson et al. (2024) note, truly transformative AI will require economists to develop new indicators and models to track its impact. In sum, the long-run questions span macro, micro, and regulatory domains. By addressing these questions, future research will help ensure that the AI revolution in financial economics leads to broadly shared prosperity and a stable financial system.

References

Abis, Simona and Laura Veldkamp (2024). "The changing economics of knowledge production". *The Review of Financial Studies* 37.1, pp. 89–118.

Acemoglu, Daron (2021). *Harms of AI.* Tech. rep. National Bureau of Economic Research.

— (2025). "The simple macroeconomics of AI". *Economic Policy* 40.121, pp. 13–58.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo (2022). "Artificial intelligence and jobs: Evidence from online vacancies". *Journal of Labor Economics* 40.S1, S293–s340.

Acemoglu, Daron, David Autor, and Simon Johnson (2023). "AI needs to be more 'pro-worker.' These 5 policies can help". *MIT Sloan Ideas Made to Matter.*

Acemoglu, Daron and Pascual Restrepo (2018). "Artificial intelligence, automation, and work". In: *The economics of artificial intelligence: An agenda.* University of Chicago Press, pp. 197–236.

— (2019). "Automation and new tasks: How technology displaces and reinstates labor". *Journal of economic perspectives* 33.2, pp. 3–30.

— (2020a). "Robots and jobs: Evidence from US labor markets". *Journal of political economy* 128.6, pp. 2188–2244.

— (2020b). "The wrong kind of AI? Artificial intelligence and the future of labour demand". *Cambridge Journal of Regions, Economy and Society* 13.1, pp. 25–35.

Acikalin, Utku, Tolga Caskurlu, Gerard Hoberg, and Gordon M Phillips (2022). "Intellectual property protection lost and competition: An examination using large language models". *Tuck School of Business Working Paper* 4023622.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016). "The economics of privacy". *Journal of economic Literature* 54.2, pp. 442–492.

Adams, Jonathan J, Min Fang, Zheng Liu, and Yajie Wang (2025). "The rise of ai pricing: Trends, driving forces, and implications for firm performance". *Journal of Monetary Economics*, p. 103875.

Agarwal, Sumit, Shashwat Alok, Pulak Ghosh, and Sudip Gupta (2019). “Financial inclusion and alternate credit scoring: Role of big data and machine learning in Fintech”. *Indian School of Business.*

Agarwal, Sumit, Wenlan Qian, Yuan Ren, Hsin-Tien Tiffany Tsai, and Bernard Yin Yeung (2025). “The real impact of FinTech: Evidence from mobile payment technology”. *Management Science (forthcoming).*

Agrawal, Ajay, Joshua S Gans, and Avi Goldfarb (2019). “Artificial intelligence: the ambiguous labor market impact of automating prediction”. *Journal of Economic Perspectives* 33.2, pp. 31–50.

Ahmadi, A, AMBRUS KECSKÉS, RONI MICHAELY, and PA NGUYEN (2023). “Producing AI Innovation and Its Value Implications”.

Ahnert, Toni, Peter Hoffmann, and Cyril Monnet (2025). “Payments and privacy in the digital economy”. *Journal of Financial Economics* 169, p. 104050.

Ajith, Anirudh, Mengzhou Xia, Alexis Chevalier, Tanya Goyal, Danqi Chen, and Tianyu Gao (2024). “Litsearch: A retrieval benchmark for scientific literature search”. *arXiv preprint arXiv:2407.18940.*

Ala-Pietilä, Pekka, Yann Bonnet, Urs Bergmann, Maria Bielikova, Cecilia Bonefeld-Dahl, Wilhelm Bauer, Loubna Bouarfa, Raja Chatila, Mark Coeckelbergh, Virginia Dignum, et al. (2020). *The assessment list for trustworthy artificial intelligence (ALTAI).* European Commission.

Aldasoro, Iñaki, Leonardo Gambacorta, Anton Korinek, Vatsala Shreeti, and Merlin Stein (2024). *Intelligent financial system: how AI is transforming finance.* Bank for International Settlements, Monetary and Economic Department.

Alderucci, Dean, Lee Branstetter, Eduard Hovy, Andrew Runge, and Nikolas Zolas (2020). “Quantifying the impact of AI on productivity and labor demand: Evidence from US census microdata”. In: *Allied social science associations—ASSA 2020 annual meeting.*

Almog, David (2025). “Barriers to AI Adoption: Image Concerns at Work”. *arXiv preprint arXiv:2511.18582.*

Amel-Zadeh, Amir, Jan-Peter Calliess, Daniel Kaiser, and Stephen Roberts (2020). "Machine learning-based financial statement analysis". *Available at SSRN 3520684*.

Anand, Kartik, Sophia Kazinnik, Agnese Leonello, and Ettore Panetti (2025). "Ex Machina: Financial Stability in the Age of Artificial Intelligence". *Available at SSRN 5543139*.

Arntz, Melanie, Terry Gregory, and Ulrich Zierahn (2016). "The risk of automation for jobs in OECD countries: A comparative analysis".

Assad, Stephanie, Robert Clark, Daniel Ershov, and Lei Xu (2024). "Algorithmic pricing and competition: empirical evidence from the German retail gasoline market". *Journal of Political Economy* 132.3, pp. 723–771.

Athey, Susan, Herman Brunborg, Tianyu Du, Ayush Kanodia, and Keyon Vafa (2024). "LABOR-LLM: Language-Based Occupational Representations with Large Language Models". *arXiv preprint arXiv:2406.17972*.

Athey, Susan and Guido W Imbens (2019). "Machine learning methods that economists should know about". *Annual Review of Economics* 11.1, pp. 685–725.

Audrino, Francesco, Jessica Gentner, and Simon Stalder (2024). "Quantifying uncertainty: a new era of measurement through large language models".

Autor, David (2024). *Applying AI to rebuild middle class jobs*. Tech. rep. National Bureau of Economic Research.

Autor, David H (2015). "Why are there still so many jobs? The history and future of workplace automation". *Journal of economic perspectives* 29.3, pp. 3–30.

Autor, David H and David Dorn (2013). "The growth of low-skill service jobs and the polarization of the US labor market". *American economic review* 103.5, pp. 1553–1597.

Avramov, Doron, Si Cheng, and Lior Metzker (2023). "Machine learning vs. economic restrictions: Evidence from stock return predictability". *Management Science* 69.5, pp. 2587–2619.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson (2024). "Artificial intelligence, firm growth, and product innovation". *Journal of Financial Economics* 151, p. 103745.

Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson (2023a). *Firm investments in artificial intelligence technologies and changes in workforce composition*. Vol. 31325. National Bureau of Economic Research.

Babina, Tania, Anastassia Fedyk, Alex Xi He, and James Hodson (2023b). "Artificial intelligence and firms' systematic risk". *Available at SSRN*.

Baik, Bok, Alex G Kim, David S Kim, and Sangwon Yoon (2024). "Vocal delivery quality in earnings conference calls". *Journal of Accounting and Economics*, p. 101763.

Banz, Rolf W (1981). "The relationship between return and market value of common stocks". *Journal of financial economics* 9.1, pp. 3–18.

Bao, Yang, Bin Ke, Bin Li, Y Julia Yu, and Jie Zhang (2020). "Detecting accounting fraud in publicly traded US firms using a machine learning approach". *Journal of Accounting Research* 58.1, pp. 199–235.

Barber, Brad M and Terrance Odean (2000). "Trading is hazardous to your wealth: The common stock investment performance of individual investors". *The journal of Finance* 55.2, pp. 773–806.

Barillas, Francisco and Jay Shanken (2018). "Comparing asset pricing models". *The Journal of Finance* 73.2, pp. 715–754.

Barrios, John Manuel, John L. Campbell, Ryan G. Johnson, and Y. Christine Liu (2024). *Signals or Smoke? The Determinants and Informativeness of Corporate Artificial Intelligence (AI) Disclosures.*

Bartik, Alexander, Arpit Gupta, and Daniel Milo (2024). "The costs of housing regulation: Evidence from generative regulatory measurement". *Available at SSRN 4627587*.

Bastianello, Francesca, Paul H Décaire, and Marius Guenzel (2024). "Mental Models and Financial Forecasts". *Marius, Mental Models and Financial Forecasts (October 30, 2024).*

Beckmann, Lars, Heiner Beckmeyer, Ilias Filippou, Stefan Menze, and Guofu Zhou (2024). "Unusual Financial Communication: ChatGPT, Earnings Calls, and Financial Markets". *Olin Business School Center for Finance & Accounting Research Paper* 2024/02.

Bell, Sebastian, Ali Kakhbod, Martin Lettau, and Abdolreza Nazemi (2024). *Glass Box Machine Learning and Corporate Bond Returns*. Tech. rep. National Bureau of Economic Research.

Beraja, Martin, David Y Yang, and Noam Yuchtman (2023). "Data-intensive innovation and the state: Evidence from AI firms in China". *The Review of Economic Studies* 90.4, pp. 1701–1723.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri (2020). "On the rise of fintechs: Credit scoring using digital footprints". *The Review of Financial Studies* 33.7, pp. 2845–2897.

Berg, Tobias, Andreas Fuster, and Manju Puri (2022). "Fintech lending". *Annual Review of Financial Economics* 14.1, pp. 187–207.

Bergemann, Dirk and Alessandro Bonatti (2024). "Data, competition, and digital platforms". *American Economic Review* 114.8, pp. 2553–2595.

Berk, Jonathan (2023). "Comment on'The Virtue of Complexity in Return Prediction'". *Available at SSRN 4410125*.

Bertomeu, Jeremy, Yupeng Lin, Yibin Liu, and Zhenghui Ni (2024). "Voluntary Disclosure, Misinformation, and AI Information Processing: Theory and Evidence". *Misinformation, and AI Information Processing: Theory and Evidence (November 19, 2024)*.

— (2025). "The Impact of Generative AI on Information Processing: Evidence from the Ban of ChatGPT in Italy". *Journal of Accounting and Economics*, p. 101782.

Bian, Bo, Qiushi Huang, Ye Li, and Huan Tang (2025). "Data as a networked asset". *Available at SSRN*.

Bianchi, Daniele, Matthias Büchner, and Andrea Tamoni (2021). "Bond risk premiums with machine learning". *The Review of Financial Studies* 34.2, pp. 1046–1089.

Bick, Alexander, Adam Blandin, and David J Deming (2024). *The rapid adoption of generative ai*. Tech. rep. National Bureau of Economic Research.

Bini, Pietro, Lin William Cong, Xing Huang, and Lawrence J Jin (2025). "Behavioral Economics of AI: LLM Biases and Corrections". *Available at SSRN 5213130*.

Blankespoor, Elizabeth, Joe Croom, and Stephanie M Grant (2024). "Generative AI and Investor Processing of Financial Information". *Available at SSRN 5053905*.

Blankespoor, Elizabeth, Ed deHaan, and Christina Zhu (2018). "Capital market effects of media synthesis and dissemination: Evidence from robo-journalism". *Review of Accounting Studies* 23, pp. 1–36.

Blattner, Laura, Scott Nelson, and Jann Spiess (2021). "Unpacking the black box: Regulating algorithmic decisions". *arXiv preprint arXiv:2110.03443*.

Bommasani, Rishi, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. (2021). "On the opportunities and risks of foundation models". *arXiv preprint arXiv:2108.07258*.

Bond, Philip and Lukas Kremens (2023). "Income and inequality under asymptotically full automation". *Available at SSRN 4466558*.

Borgschulte, Mark, Marius Guenzel, Canyao Liu, and Ulrike Malmendier (2025). "CEO stress, aging, and death". *The Journal of Finance* 80.6, pp. 3401–3442.

Bradford, Anu (2023). *Digital empires: The global battle to regulate technology*. Oxford University Press.

Breitung, Christian and Sebastian Müller (2025). "Global Business Networks". *Journal of Financial Economics* 166, p. 104007.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei (2020). *Language Models are Few-Shot Learners*.

Brynjolfsson, Erik (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. Vol. 236. WW Norton Company.

Brynjolfsson, Erik, Anton Korinek, and Ajay Agrawal (2024). *The Economics of Transformative AI: A Research Agenda*.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond (2025). "Generative AI at work". *The Quarterly Journal of Economics*, qjae044.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock (2018). "What can machines learn and what does it mean for occupations and the economy?" In: *AEA papers and proceedings*. Vol. 108. American Economic Association, pp. 43–47.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson (2019). "Artificial intelligence and the modern productivity paradox". *The economics of artificial intelligence: An agenda* 23.2019, pp. 23–57.

— (2021). "The productivity J-curve: How intangibles complement general purpose technologies". *American Economic Journal: Macroeconomics* 13.1, pp. 333–372.

Bryzgalova, Svetlana, Markus Pelger, and Jason Zhu (2025). "Forest through the trees: Building cross-sections of stock returns". *The Journal of Finance* 80.5, pp. 2447–2506.

Bubb, Ryan and Emiliano M Catan (2022). "The party structure of mutual funds". *The Review of Financial Studies* 35.6, pp. 2839–2878.

Buchak, Greg, Jiayin Hu, and Shang-Jin Wei (2021). *FinTech as a financial liberator*. Tech. rep. National Bureau of Economic Research.

Buehlmaier, Matthias MM and Toni M Whited (2018). "Are financial constraints priced? Evidence from textual analysis". *The Review of Financial Studies* 31.7, pp. 2693–2728.

Bughin, Jacques (2020). "Artificial intelligence, its corporate use and how it will affect the future of work". In: *Capitalism, Global Change and Sustainable Development*. Springer, pp. 239–260.

Bughin, Jacques, Eric Hazan, Paris Sree Ramaswamy, Washington DC, Michael Chu, et al. (2017). "Artificial intelligence the next digital frontier".

Bustamante, Maria Cecilia, Julien Cujean, and Laurent Frésard (2019). "Knowledge cycles and corporate investment".

Bybee, J Leland (2023). "The ghost in the machine: Generating beliefs with large language models". *arXiv preprint arXiv:2305.02823*.

Bybee, Leland, Bryan Kelly, and Yinan Su (2023). "Narrative asset pricing: Interpretable systematic risk factors from news text". *The Review of Financial Studies* 36.12, pp. 4759–4787.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello (2020). “Artificial intelligence, algorithmic pricing, and collusion”. *American Economic Review* 110.10, pp. 3267–3297.

Calvet, Laurent E, Claire Celerier, Paolo Sodini, and Boris Vallee (2023). “Can Security Design Foster Household Risk-Taking?” *The Journal of Finance* 78.4, pp. 1917–1966.

Campbell, Tim S and J Kimball Dietrich (1983). “The determinants of default on insured conventional residential mortgage loans”. *The Journal of Finance* 38.5, pp. 1569–1581.

Campello, Murillo, Lin William Cong, and Luofeng Zhou (2023). “AlphaManager: A Data-Driven-Robust-Control Approach to Corporate Finance”. *Available at SSRN 4590323*.

Cao, Sean, Yichen Cheng, Meng Wang, Yusen Xia, and Baozhong Yang (2024a). “Visual information and ai divide: Evidence from corporate executive presentations”. *SSRN Electronic Journal*.

Cao, Sean, Wei Jiang, Junbo Wang, and Baozhong Yang (2024b). “From man vs. machine to man+machine: The art and AI of stock analyses”. *Journal of Financial Economics* 160, p. 103910.

Cao, Sean, Wei Jiang, Baozhong Yang, and Alan L Zhang (2023a). “How to talk when a machine is listening: Corporate disclosure in the age of AI”. *The Review of Financial Studies* 36.9, pp. 3603–3642.

Cao, Sean, Baozhong Yang, and Alan L Zhang (2023b). “Managerial risk assessment and fund performance: Evidence from textual disclosure”. *Available at SSRN 4060307*.

Cao, Sean Shun, Wei Jiang, Lijun (Gillian) Lei, and Qing (Clara) Zhou (2024c). “Applied AI for finance and accounting: Alternative data and opportunities”. *Pacific-Basin Finance Journal* 84, p. 102307.

Cao, Yi, Long Chen, Jennifer Wu Tucker, and Chi Wan (2025). “Can generative AI help identify peer firms?” *Review of Accounting Studies*, pp. 1–43.

Cao, Yi and Jia Zhai (2022). “A survey of AI in finance”. *Journal of Chinese Economic and Business Studies* 20.2, pp. 125–137.

Cardinaels, Eddy, Stephan Hollander, and Brian J White (2019). "Automatic summarization of earnings releases: attributes and effects on investors' judgments". *Review of Accounting Studies* 24, pp. 860–890.

Carhart, Mark M (1997). "On persistence in mutual fund performance". *The Journal of finance* 52.1, pp. 57–82.

Carlson, Murray, Adlai Fisher, and Ron Giammarino (2004). "Corporate investment and asset price dynamics: Implications for the cross-section of returns". *The Journal of Finance* 59.6, pp. 2577–2603.

Chaboud, Alain P, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega (2014). "Rise of the machines: Algorithmic trading in the foreign exchange market". *The Journal of Finance* 69.5, pp. 2045–2084.

Chak, Ida, Karen Croxson, Francesco D'Acunto, Jonathan Reuter, Alberto G Rossi, and Jonathan M Shaw (2022). *Improving household debt management with robo-advice*. Tech. rep. National Bureau of Economic Research.

Chang, Anne, Xi Dong, Xiumin Martin, and Changyun Zhou (2023). "AI democratization, return predictability, and trading inequality". *Available at SSRN 4543999*.

Chang, Xiangyu, Lili Dai, Lingbing Feng, Jianlei Han, Jing Shi, and Bohui Zhang (2024). "A good sketch is better than a long speech: evaluate delinquency risk through real-time video analysis". *Review of Finance* 29.2, pp. 467–500.

Chen, Binkai, Dongmei Guo, Junjie Xia, and Zirun Zhang (2025a). "The Transformative Role of Artificial Intelligence and Big Data in Banking". *Available at SSRN 5120308*.

Chen, Hui, Yuhan Cheng, Yanchu Liu, and Ke Tang (2023a). "Teaching economics to the machines". *Available at SSRN* 4642167.

Chen, Jian, Guohao Tang, Guofu Zhou, and Wu Zhu (2025b). "ChatGPT and Deepseek: Can They Predict the Stock Market and Macroeconomy?" *arXiv preprint arXiv:2502.10008*.

Chen, Joanne and Brandon Yueyang Han (2024). "Corporate AI Backfire and Potential Remedies". *Available at SSRN*.

Chen, Long, Patrick Bolton, Bengt R Holmström, Eric Maskin, Christopher A Pissarides, A Michael Spence, Tao Sun, Tianshu Sun, Wei Xiong, Liyan Yang, et al. (2021a). "Understanding big data: Data calculus in the digital era".

Chen, Long, Yadong Huang, Shumiao Ouyang, and Wei Xiong (2021b). *The data privacy paradox and digital demand*. Tech. rep. National Bureau of Economic Research.

Chen, Luyang, Markus Pelger, and Jason Zhu (2024a). "Deep learning in asset pricing". *Management Science* 70.2, pp. 714–750.

Chen, Mark A. and Joanna Xiaoyu Wang (2024). "Displacement or Augmentation? The Effects of AI on Workforce Dynamics and Firm Value". *The Effects of AI on Workforce Dynamics and Firm Value*.

Chen, Mark A, Qinxi Wu, and Baozhong Yang (2019). "How valuable is FinTech innovation?" *The Review of financial studies* 32.5, pp. 2062–2106.

Chen, Shuaiyu, T Clifton Green, Huseyin Gulen, and Dexin Zhou (2024b). "What Does ChatGPT Make of Historical Stock Returns? Extrapolation and Miscalibration in LLM Stock Return Forecasts". *arXiv preprint arXiv:2409.11540*.

Chen, Shuaiyu, Lin Peng, and Dexin Zhou (2024c). "Wisdom or Whims? Decoding Investor Trading Strategies with Large Language Models". *Decoding Investor Trading Strategies with Large Language Models (June 19, 2024)*.

Chen, Wilbur, Terrence Tianshuo Shi, and Suraj Srinivasan (2024d). "The Value of AI Innovations". *Working paper*.

Chen, Xi, Yang Ha Cho, Yiwei Dou, and Baruch Lev (2022a). "Predicting future earnings changes using machine learning and detailed financial data". *Journal of Accounting Research* 60.2, pp. 467–515.

Chen, Yifei, Bryan T Kelly, and Dacheng Xiu (2022b). "Expected returns and large language models". *Available at SSRN 4416687*.

Chen, Yiting, Tracy Xiao Liu, You Shan, and Songfa Zhong (2023b). "The emergence of economic rationality of GPT". *Proceedings of the National Academy of Sciences* 120.51, e2316205120.

Cheng, Qiang, Pengkai Lin, and Yue Zhao (2025). “Does Generative AI Facilitate Investor Trading? Early Evidence from ChatGPT Outages”. *Journal of Accounting and Economics*, p. 101821.

China Daily (2019). *China issues principles of next-generation AI governance.*

Chinco, Alex, Adam D Clark-Joseph, and Mao Ye (2019). “Sparse signals in the cross-section of returns”. *The Journal of Finance* 74.1, pp. 449–492.

Choi, Jung Ho and Chloe Xie (2025). “Human+ AI in accounting: Early evidence from the field”.

Christ, Margaret H, MJ Kim, and Michael A Yip (2024). “Survey Evidence on the Determinants and Consequences of Artificial Intelligence Use in Sell-Side Equity Research”. *Available at SSRN 5071836.*

Cockburn, Iain M, Rebecca Henderson, Scott Stern, et al. (2018). *The impact of artificial intelligence on innovation.* Vol. 24449. National bureau of economic research Cambridge, MA, USA.

Coleman, Braiden, Kenneth Merkley, and Joseph Pacelli (2022). “Human versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations”. *The Accounting Review* 97.5, pp. 221–244.

Collins, Katherine M, Albert Q Jiang, Simon Frieder, Lionel Wong, Miri Zilka, Umang Bhatt, Thomas Lukasiewicz, Yuhuai Wu, Joshua B Tenenbaum, William Hart, et al. (2024). “Evaluating language models for mathematics through interactions”. *Proceedings of the National Academy of Sciences* 121.24, e2318124121.

Cong, Lin William, Guanhao Feng, Jingyu He, and Xin He (2025a). “Growing the efficient frontier on panel trees”. *Journal of Financial Economics* 167, p. 104024.

Cong, Lin William, Tengyuan Liang, Xiao Zhang, and Wu Zhu (2025b). “Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information”. *Management Science* 71.12, pp. 10727–10739.

Cong, Lin William, Ke Tang, Jingyuan Wang, and Yang Zhang (2021a). “AlphaPortfolio: Direct construction through deep reinforcement learning and interpretable AI”. *Available at SSRN* 3554486.

Cong, Lin William, Danxia Xie, and Longtian Zhang (2021b). "Knowledge accumulation, privacy, and growth in a data economy". *Management science* 67.10, pp. 6480–6492.

Cookson, J Anthony, William Mullins, and Marina Niessner (2024). "Social media and finance". *Available at SSRN 4806692.*

Cornelli, Giulio, Jon Frost, and Saurabh Mishra (2023). *Artificial intelligence, services globalisation and income inequality.* Tech. rep. Bank for International Settlements.

Costello, Anna M, Andrea K Down, and Mihir N Mehta (2020). "Machine+ man: A field experiment on the role of discretion in augmenting AI-based lending models". *Journal of Accounting and Economics* 70.2-3, p. 101360.

Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz (2024). "The effects of generative ai on high skilled work: Evidence from three field experiments with software developers". *Available at SSRN 4945566.*

Da, Zhi, Xinyi Peng, Yongxin Xu, and Gaoping Zheng (2025). "Artificial Intelligence and the Law of One Price: Evidence from the Chinese AH Premium". *Available at SSRN 5234282.*

D'Acunto, Francesco and Alberto G Rossi (2023). "Robo-advice: Transforming households into rational economic agents". *Annual Review of Financial Economics* 15.1, pp. 543–563.

Dedrick, Jason, Vijay Gurbaxani, and Kenneth L Kraemer (2003). "Information technology and economic performance: A critical review of the empirical evidence". *ACM Computing Surveys (CSUR)* 35.1, pp. 1–28.

Dell, Melissa (2025). "Deep learning for economists". *Journal of Economic Literature* 63.1, pp. 5–58.

Deloitte (2016). *The Expansion of Robo-Advisory in Wealth Management.* Tech. rep. Deloitte.

DeMiguel, Victor, Javier Gil-Bazo, Francisco J Nogales, and André AP Santos (2023). "Machine learning and fund characteristics help to select mutual funds with positive alpha". *Journal of Financial Economics* 150.3, p. 103737.

Demirer, Mert, Andrey Fradkin, Nadav Tadelis, and Sida Peng (2025). "The Emerging Market for Intelligence: Pricing, Supply, and Demand for LLMs". *Supply, and Demand for LLMs (December 15, 2025).*

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.*

Di Maggio, Marco, Dimuthu Ratnadiwakara, and Don Carmichael (2022). *Invisible primes: Fintech lending with alternative data.* Tech. rep. National Bureau of Economic Research.

Dickerson, Alexander, Christian Julliard, and Philippe Mueller (2023). "The corporate bond factor zoo". *UNSW Business School Research Paper Forthcoming.*

Dietvorst, Berkeley J, Joseph P Simmons, and Cade Massey (2018). "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them". *Management science* 64.3, pp. 1155–1170.

Ding, Kexing, Baruch Lev, Xuan Peng, Ting Sun, and Miklos A Vasarhelyi (2020). "Machine learning improves accounting estimates: Evidence from insurance payments". *Review of accounting studies* 25.3, pp. 1098–1134.

Dong, Xi, Yan Li, David E Rapach, and Guofu Zhou (2022). "Anomalies and the expected market return". *The Journal of Finance* 77.1, pp. 639–681.

Dong, Yanfei, Jiayin Hu, Yiping Huang, Han Qiu, and Yingguang Zhang (2024). "The use and disuse of FinTech credit: When buy-now-pay-later meets credit reporting". *Available at SSRN 4783923.*

Dou, Winston Wei, Itay Goldstein, and Yan Ji (2024). "Ai-powered trading, algorithmic collusion, and price efficiency". *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.*

Dyck, I.J. Alexander, Karl V. Lins, Lukas Roth, Mitch Towner, and Hannes F. Wagner (2025). "Controlled Firms, Preferences, and Environmental Sustainability".

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G Rossi (2019). "The promises and pitfalls of robo-advising". *The Review of Financial Studies* 32.5, pp. 1983–2020.

D'Hondt, Catherine, Rudy De Winne, Eric Ghysels, and Steve Raymond (2020). "Artificial intelligence alter egos: Who might benefit from robo-investing?" *Journal of Empirical Finance* 59, pp. 278–299.

Ecker, Frank, Xitong Li, Yilan Li, and Fan Wu (2026). "How Stock Market Participants Use Generative Artificial Intelligence: Evidence from User-Platform Interaction Data". *Journal of Accounting Research.*

Eisfeldt, Andrea L and Gregor Schubert (2025). "Generative AI and Finance". *Annual Review of Financial Economics* 17.

Eisfeldt, Andrea L, Gregor Schubert, and Miao Ben Zhang (2023). *Generative AI and firm values.* Tech. rep. National Bureau of Economic Research.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock (2024). "GPTs are GPTs: Labor market impact potential of LLMs". *Science* 384.6702, pp. 1306–1308.

Erel, Isil, Léa H Stern, Chenhao Tan, and Michael S Weisbach (2021). "Selecting directors using machine learning". *The Review of Financial Studies* 34.7, pp. 3226–3264.

Fama, Eugene F (1970). "Efficient capital markets". *Journal of finance* 25.2, pp. 383–417.

Fama, Eugene F and Kenneth R French (1993). "Common risk factors in the returns on stocks and bonds". *Journal of financial economics* 33.1, pp. 3–56.

— (2015). "A five-factor asset pricing model". *Journal of financial economics* 116.1, pp. 1–22.

Fang, Hanming, Ming Li, and Guangli Lu (2024). "Decoding China's Industrial Policies". *Available at SSRN.*

Farboodi, Maryam, Dhruv Singal, Laura Veldkamp, and Venky Venkateswaran (2022). *Valuing financial data.* Tech. rep. National Bureau of Economic Research.

Favaretto, Maddalena, Eva De Clercq, and Bernice Simone Elger (2019). "Big Data and discrimination: perils, promises and solutions. A systematic review". *Journal of Big Data* 6.1, pp. 1–27.

Fedyk, Anastassia, James Hodson, Natalya Khimich, and Tatiana Fedyk (2022). "Is artificial intelligence improving the audit process?" *Review of Accounting Studies* 27.3, pp. 938–985.

Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier (2024). "Chatgpt and perception biases in investments: An experimental study". *Available at SSRN 4787249.*

Felten, Edward W, Manav Raj, and Robert Seamans (2018). “A method to link advances in artificial intelligence to occupational abilities”. In: *AEA Papers and proceedings*. Vol. 108. American Economic Association, pp. 54–57.

— (2019). “The occupational impact of artificial intelligence: Labor, skills, and polarization”. *NYU Stern School of Business*.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (2020). “Taming the factor zoo: A test of new factors”. *The Journal of Finance* 75.3, pp. 1327–1370.

Fernández Tamayo, Borja, Reiner Braun, Florencio Lopez-de Silanes, Ludovic Phalippou, and Natalia Sigrist (2023). “Limited Partners versus Unlimited Machines; Artificial Intelligence and the Performance of Private Equity Funds”. *Available at SSRN*.

Ferrara, Emilio (2023). “Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies”. *Sci* 6.1, p. 3.

Fetzer, Thiemo, Peter John Lambert, Bennet Feld, and Prashant Garg (2024). “AI-generated production networks: Measurement and applications to global trade”.

Filippucci, Francesco, Peter Gal, and Matthias Schief (2024). *Miracle or Myth? Assessing the macroeconomic productivity gains from Artificial Intelligence*. Tech. rep. OECD Publishing.

Fish, Sara, Yannai A Gonczarowski, and Ran I Shorrer (2024). “Algorithmic collusion by large language models”. *arXiv preprint arXiv:2404.00806* 7.

Floridi, Luciano (2024). “Why the AI hype is another tech bubble”. *Philosophy & Technology* 37.4, p. 128.

Fortune Business Insights (2025). *Robo-Advisory Market Size, Share & Industry Analysis*. https://www.fortunebusinessinsights.com/robo-advisory-market-109986.

Frank, Morgan R, David Autor, James E Bessen, Erik Brynjolfsson, Manuel Cebrian, David J Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro, et al. (2019). “Toward understanding the impact of artificial intelligence on labor”. *Proceedings of the National Academy of Sciences* 116.14, pp. 6531–6539.

Frankel, Richard, Jared Jennings, and Joshua Lee (2022). "Disclosure sentiment: Machine learning vs. dictionary methods". *Management Science* 68.7, pp. 5514–5532.

Frey, Carl Benedikt and Michael A Osborne (2017). "The future of employment: How susceptible are jobs to computerisation?" *Technological forecasting and social change* 114, pp. 254–280.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber (2020). "Dissecting characteristics nonparametrically". *The Review of Financial Studies* 33.5, pp. 2326–2377.

Furman, Jason and Robert Seamans (2019). "AI and the Economy". *Innovation policy and the economy* 19.1, pp. 161–191.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther (2022). "Predictably unequal? The effects of machine learning on credit markets". *The Journal of Finance* 77.1, pp. 5–47.

Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery (2019). "The role of technology in mortgage lending". *The Review of Financial Studies* 32.5, pp. 1854–1899.

Gabaix, Xavier, Ralph SJ Koijen, Robert Richmond, and Motohiro Yogo (2024). "Asset embeddings". *Available at SSRN 4507511.*

Gambacorta, Leonardo, Yiping Huang, Zhenhua Li, Han Qiu, and Shu Chen (2023). "Data versus collateral". *Review of Finance* 27.2, pp. 369–398.

Gambacorta, Leonardo, Fabiana Sabatini, and Stefano Schiaffi (2025). *Artificial intelligence and relationship lending*. Tech. rep. Bank of Italy, Economic Research and International Relations Area.

Gao, Jiacui, Wei Xiong, and Hao Yuan (2025). "Structured Beliefs and Fund Performance: An LLM-Based Approach".

Garcia, Marisa (2024). *What Air Canada Lost In Remarkable 'Lying' AI Chatbot Case.*

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny (2015). "Money doctors". *The Journal of Finance* 70.1, pp. 91–114.

Georgieff, Alexandre (2024). *Artificial intelligence and wage inequality.* Tech. rep. OECD Publishing.

Ghosh, Pulak, Boris Vallee, and Yao Zeng (2022). “Fintech lending and cashless payments”. In: *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*.

Giglio, Stefano, Bryan Kelly, and Dacheng Xiu (2022). “Factor models, machine learning, and asset pricing”. *Annual Review of Financial Economics* 14.1, pp. 337–368.

Gillis, Talia, Bryce McLaughlin, and Jann Spiess (2021). “On the fairness of machine-assisted human decisions”. *arXiv preprint arXiv:2110.15310*.

Glasserman, Paul, Kriste Krstovski, Paul Laliberte, and Harry Mamaysky (2020). “Choosing news topics to explain stock market returns”. In: *Proceedings of the First ACM International Conference on AI in Finance*, pp. 1–8.

Glasserman, Paul and Caden Lin (2023). “Assessing look-ahead bias in stock return predictions generated by gpt sentiment analysis”. *arXiv preprint arXiv:2309.17322*.

Gofman, Michael and Zhao Jin (2024). “Artificial intelligence, education, and entrepreneurship”. *The Journal of Finance* 79.1, pp. 631–667.

Goldfarb, Avi and Verina F Que (2023). “The economics of digital privacy”. *Annual Review of Economics* 15.1, pp. 267–286.

Goldfarb, Avi, Bledi Taska, and Florenta Teodoridis (2023). “Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings”. *Research Policy* 52.1, p. 104653.

Goldman Sachs (2023). *Generative AI Could Raise Global GDP by 7 Percent*.

Goldstein, Itay, Chester S Spatt, and Mao Ye (2021). “Big data in finance”. *The Review of Financial Studies* 34.7, pp. 3213–3225.

Gordon, Robert J and Hassan Sayed (2022). *A new interpretation of productivity growth dynamics in the pre-pandemic and pandemic era US economy, 1950-2022*. Tech. rep. National Bureau of Economic Research.

Graetz, Georg and Guy Michaels (2018). “Robots at work”. *Review of economics and statistics* 100.5, pp. 753–768.

Greig, Fiona, Tarun Ramadorai, Alberto G Rossi, Stephen P Utkus, and Ansgar Walther (2024). "Human Financial Advice in the Age of Automation". *Available at SSRN 4301514*.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2020). "Empirical asset pricing via machine learning". *The Review of Financial Studies* 33.5, pp. 2223–2273.

— (2021). "Autoencoder asset pricing models". *Journal of Econometrics* 222.1, pp. 429–450.

Guijarro-Ordonez, Jorge, Markus Pelger, and Greg Zanotti (2021). "Deep learning statistical arbitrage". *arXiv preprint arXiv:2106.04028*.

Guo, Yue, Siliang Tong, Tao Chen, Subodha Kumar, and Shumiao Ouyang (2022). "The Impact of Generative Artificial Intelligence on Individual Manual Investment Decisions: Empirical Evidence from Mutual Funds". *SSRN Electronic Journal*.

Hampole, Menaka, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller (2025). *Artificial intelligence and the labor market*. Tech. rep. National Bureau of Economic Research.

Han, Lina, Xuan Luo, and Shumiao Ouyang (2020). "Investor's Responses to Market Fluctuations: Evidence from Experiment and Real Trading".

Han, Sukjin (2024). "Mining causality: Ai-assisted search for instrumental variables". *arXiv preprint arXiv:2409.14202*.

Hansen, Anne Lundgaard, John J Horton, Sophia Kazinnik, Daniela Puzzello, and Ali Zarifhonarvar (2024). "Simulating the Survey of Professional Forecasters". *Available at SSRN*.

Hansen, Anne Lundgaard and Sophia Kazinnik (2023). "Can chatgpt decipher fedspeak". *Available at SSRN*.

Hardt, Moritz, Eric Price, and Nati Srebro (2016). "Equality of opportunity in supervised learning". *Advances in neural information processing systems* 29.

Harvey, Campbell R, Yan Liu, and Heqing Zhu (2016). "… and the cross-section of expected returns". *The Review of Financial Studies* 29.1, pp. 5–68.

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun (2019). "Firm-level political risk: Measurement and effects". *The Quarterly Journal of Economics* 134.4, pp. 2135–2202.

Hau, Harald, Yi Huang, Chen Lin, Hongzhe Shan, Zixia Sheng, and Lai Wei (2024). "FinTech credit and entrepreneurial growth". *the Journal of Finance* 79.5, pp. 3309–3359.

He, Ai, Dashan Huang, Jiaen Li, and Guofu Zhou (2023a). "Shrinking factor dimension: A reduced-rank approach". *Management science* 69.9, pp. 5501–5522.

He, Songrun, Linying Lv, Asaf Manela, and Jimmy Wu (2025). "Chronologically Consistent Large Language Models". *arXiv preprint arXiv:2502.21206*.

He, Zhiguo, Jing Huang, and Jidong Zhou (2023b). "Open banking: Credit market competition when borrowers own the data". *Journal of financial economics* 147.2, pp. 449–474.

Heston, Steven L (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options". *The review of financial studies* 6.2, pp. 327–343.

Hirshleifer, David, Lin Peng, Qiguang Wang, and Xiaoyan Zhang (2025). "AI, Opinion Ecosystems, and Finance". *SSRN Electronic Journal*.

Hoberg, Gerard and Asaf Manela (2025). *The Natural Language of Finance*. Working Paper. USC Marshall School of Business Research Paper Sponsored by iORB.

Hoberg, Gerard and Gordon Phillips (2016). "Text-based network industries and endogenous product differentiation". *Journal of political economy* 124.5, pp. 1423–1465.

Hobson, Jessen L, William J Mayew, and Mohan Venkatachalam (2012). "Analyzing speech to detect financial misreporting". *Journal of Accounting Research* 50.2, pp. 349–392.

Horton, John J (2023). *Large language models as simulated economic agents: What can we learn from homo silicus?* Tech. rep. National Bureau of Economic Research.

Hou, Kewei, Chen Xue, and Lu Zhang (2015). "Digesting anomalies: An investment approach". *The Review of Financial Studies* 28.3, pp. 650–705.

— (2020). "Replicating anomalies". *The Review of financial studies* 33.5, pp. 2019–2133.

Howell, Sabrina T, Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong (2024). "Lender automation and racial disparities in credit access". *The Journal of Finance* 79.2, pp. 1457–1512.

Hu, Allen and Song Ma (2025). "Persuading investors: A video-based study". *The Journal of Finance* 80.5, pp. 2639–2688.

Huang, Allen H, Hui Wang, and Yi Yang (2023a). "FinBERT: A large language model for extracting information from financial text". *Contemporary Accounting Research* 40.2, pp. 806–841.

Huang, Jing and Shumiao Ouyang (2025). "Soft Information, Hard Decisions: AI Advising". *Working Paper.*

Huang, Linmei, Bill Qiao, and Dexin Zhou (2024a). "Are Memes a Sideshow: Evidence from WallStreetBets". *Available at SSRN 4785481.*

Huang, Qian, Jian Vora, Percy Liang, and Jure Leskovec (2023b). "Mlagentbench: Evaluating language agents on machine learning experimentation". *arXiv preprint arXiv:2310.03302.*

Huang, Yi, Sung Kwan Lee, and Bernard Yeung (2024b). "Learning from AI-finance: a selected synopsis". *The Journal of Finance and Data Science* 10, p. 100152.

Humlum, Anders and Emilie Vestergaard (2025). *Large language models, small labor market effects.* Tech. rep. National Bureau of Economic Research.

Hwang, Byoung-Hyoun, Don Noh, and Sean Seunghun Shin (2025). "How Investors Pick Stocks: Global Evidence from 1,540 AI-Driven Field Interviews". *Available at SSRN 5844543.*

Ide, Enrique and Eduard Talamas (forthcoming). "Artificial intelligence in the knowledge economy". *Journal of Political Economy.*

International Monetary Fund (2024). *Global Financial Stability Report, October 2024: Artificial Intelligence and the Future of Financial Markets.*

Jabarian, Brian and Luca Henkel (2025). "Voice AI in Firms: A Natural Field Experiment on Automated Job Interviews". *Firms: A Natural Field Experiment on Automated Job Interviews (August 18, 2025).*

Jansen, Mark, Hieu Nguyen, and Amin Shams (2023). "Rise of the Machines: The Impact of Automated Underwriting". *Management Science.*

Jha, Manish, Hongyi Liu, and Asaf Manela (2025). "Does finance benefit society? A language embedding approach". *The Review of Financial Studies*, hhaf012.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang (2024a). *ChatGPT and corporate policies*. Tech. rep. National Bureau of Economic Research.

— (2024b). "Harnessing Generative AI for Economic Insights". *arXiv preprint arXiv:2410.03897*.

Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung (2023). "Survey of hallucination in natural language generation". *ACM computing surveys* 55.12, pp. 1–38.

Jiang, Jingwen, Bryan Kelly, and Dacheng Xiu (2023). "(Re-) Imag (in) ing price trends". *The Journal of Finance* 78.6, pp. 3193–3249.

Jones, Charles I (2025). *How Much Should We Spend to Reduce AI's Existential Risk?* Tech. rep. National Bureau of Economic Research.

Jones, Charles I and Christopher Tonetti (2020). "Nonrivalry and the Economics of Data". *American Economic Review* 110.9, pp. 2819–2858.

Kakhbod, Ali, Leonid Kogan, Peiyao Li, and Dimitris Papanikolaou (2024). "Measuring Creative Destruction". *Available at SSRN 5008685*.

Kanazawa, Kyogo, Daiji Kawaguchi, Hitoshi Shigeoka, and Yasutora Watanabe (2022). *AI, skill, and productivity: The case of taxi drivers*. Tech. rep. National Bureau of Economic Research.

Kang, Haoqiang and Xiao-Yang Liu (2023). "Deficiency of large language models in finance: An empirical examination of hallucination". *arXiv preprint arXiv:2311.15548*.

Kaniel, Ron, Zihan Lin, Markus Pelger, and Stijn Van Nieuwerburgh (2023). "Machine-learning the skill of mutual fund managers". *Journal of Financial Economics* 150.1, pp. 94–138.

Karbasi, Amin, Omar Montasser, John Sous, and Grigoris Velegkas (2025). "(Im) possibility of Automated Hallucination Detection in Large Language Models". *arXiv preprint arXiv:2504.17004*.

Kazinnik, Sophia (2026). "Bank Run, Interrupted: Modeling Deposit Withdrawals with Generative AI". *Management Science*.

Kelly, Bryan, Semyon Malamud, and Kangying Zhou (2024). "The virtue of complexity in return prediction". *The Journal of Finance* 79.1, pp. 459–503.

Kelly, Bryan T, Boris Kuznetsov, Semyon Malamud, and Teng Andrea Xu (2025). *Artificial Intelligence Asset Pricing Models.* Tech. rep. National Bureau of Economic Research.

Kelly, Bryan T, Seth Pruitt, and Yinan Su (2019). "Characteristics are covariances: A unified model of risk and return". *Journal of Financial Economics* 134.3, pp. 501–524.

Kim, Alex, Maximilian Muhn, and Valeri Nikolaev (2023a). "Bloated disclosures: can ChatGPT help investors process information?" *arXiv preprint arXiv:2306.10224*.

— (2023b). "From transcripts to insights: Uncovering corporate risks using generative ai". *arXiv preprint arXiv:2310.17721*.

Kirilenko, Andrei, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun (2017). "The flash crash: High-frequency trading in an electronic market". *The Journal of Finance* 72.3, pp. 967–998.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018a). "Human decisions and machine predictions". *The quarterly journal of economics* 133.1, pp. 237–293.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan (2018b). "Algorithmic fairness". In: *Aea papers and proceedings*. Vol. 108. American Economic Association, pp. 22–27.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan (2016). "Inherent trade-offs in the fair determination of risk scores". *arXiv preprint arXiv:1609.05807*.

Klette, Tor Jakob and Samuel Kortum (2004). "Innovating firms and aggregate innovation". *Journal of political economy* 112.5, pp. 986–1018.

Kogan, Leonid, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller (2023). *Technology and labor displacement: Evidence from linking patents with worker-level data.* Tech. rep. National Bureau of Economic Research.

Korinek, Anton (2023). "Generative AI for economic research: Use cases and implications for economists". *Journal of Economic Literature* 61.4, pp. 1281–1317.

— (2024). *Economic policy challenges for the age of AI.* Tech. rep. National Bureau of Economic Research.

Korinek, Anton and Joseph E Stiglitz (2018). “Artificial intelligence and its implications for income distribution and unemployment”. In: *The economics of artificial intelligence: An agenda.* University of Chicago Press, pp. 349–390.

Korinek, Anton and Donghyun Suh (2024). *Scenarios for the Transition to AGI.* Tech. rep. National Bureau of Economic Research.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2020). “Shrinking the cross-section”. *Journal of Financial Economics* 135.2, pp. 271–292.

Krockenberger, Vanessa, Anthony Saunders, Sascha Steffen, and Paulina Verhoff (2024). “Covenantai-new insights into covenant violations”. *Available at SSRN 4640653.*

Lee, S., Y. Huang, T. Zhang, Y. Zhou, and B. Yeung (2025). “Learning from Stock Markets: Economic Impacts of Low-Cost, Open-Source AI Innovation”. *Working Paper.*

Leippold, Markus, Qian Wang, and Wenyu Zhou (2022). “Machine learning in the Chinese stock market”. *Journal of Financial Economics* 145.2, pp. 64–82.

Lettau, Martin and Markus Pelger (2020). “Factors that fit the time series and cross-section of stock returns”. *The Review of Financial Studies* 33.5, pp. 2274–2325.

Li, Bin and Alberto G Rossi (2020). “Selecting mutual funds from the stocks they hold: A machine learning approach”. *Available at SSRN 3737667.*

Li, Edward Xuejun, Zhiyuan Tu, and Dexin Zhou (2024a). *The Promise and Peril of Generative AI: Evidence from GPT-4 as Sell-Side Analysts.* SSRN Working Paper.

Li, Feng et al. (2010). “Textual analysis of corporate disclosures: A survey of the literature”. *Journal of accounting literature* 29.1, pp. 143–165.

Li, Junyou, Qin Zhang, Yangbin Yu, Qiang Fu, and Deheng Ye (2024b). “More agents is all you need”. *arXiv preprint arXiv:2402.05120.*

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan (2021). “Measuring corporate culture using machine learning”. *The Review of Financial Studies* 34.7, pp. 3265–3315.

Li, Kai, Feng Mai, Rui Shen, Chelsea Yang, and Tengfei Zhang (2026). “Dissecting corporate culture using generative AI”. *The Review of financial studies* 39.1, pp. 253–296.

Lichtinger, Guy and Seyed Mahdi Hosseini Maasoum (2025). "Generative AI as Seniority-Biased Technological Change: Evidence from US Resume and Job Posting Data". *Available at SSRN*.

Light, Nathaniel, Denys Maslov, and Oleg Rytchkov (2017). "Aggregation of information about the cross section of stock returns: A latent variable approach". *The Review of Financial Studies* 30.4, pp. 1339–1381.

Linnainmaa, Juhani, Brian Melzer, Alessandro Previtero, and Stephen Foerster (2018). "Financial advisors and risk-taking". *Unpublished manuscript*.

Liu, Ernest, Song Ma, and Laura Veldkamp (2025). "Data sales and data dilution". *Journal of Financial Economics* 169, p. 104053.

Liu, Lei, Guangli Lu, and Wei Xiong (2022). *The big tech lending model*. 30160. National Bureau of Economic Research.

Liu, Nelson F, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang (2024). "Lost in the middle: How language models use long contexts". *Transactions of the Association for Computational Linguistics* 12, pp. 157–173.

Liu, Xing (2024). "Artificial Intelligence and Customer Acquisition in Retail Financial Services: Experimental Evidence from Insurance Distribution".

Liu, Zhuang, Michael Sockin, and Wei Xiong (2023). *Data privacy and algorithmic inequality*. Tech. rep. National Bureau of Economic Research.

Lo, Andrew W and Jillian Ross (2024). "Can chatgpt plan your retirement?: Generative ai and financial advice". *Available at SSRN*.

Lopez-Lira, Alejandro and Yuehua Tang (2023). "Can chatgpt forecast stock price movements? return predictability and large language models". *arXiv preprint arXiv:2304.07619*.

Loughran, Tim and Bill McDonald (2016). "Textual analysis in accounting and finance: A survey". *Journal of Accounting Research* 54.4, pp. 1187–1230.

— (2020). "Textual analysis in finance". *Annual Review of Financial Economics* 12.1, pp. 357–375.

Lu, Fangzhou (2025). "Robo-Advising Meets Large Language Models: Educating Investors on Alpha and Beta of Mutual Funds and Stocks".

Lu, Fangzhou, Lei Huang, and Sixuan Li (2023). "ChatGPT, generative AI, and investment advisory". *Available at SSRN 4519182*.

Lu, Yao, Gordon M Phillips, and Jia Yang (2024). *The impact of cloud computing and ai on industry dynamics and concentration*. Tech. rep. National Bureau of Economic Research.

Ludwig, Jens and Sendhil Mullainathan (2024). "Machine learning as a tool for hypothesis generation". *The Quarterly Journal of Economics* 139.2, pp. 751–827.

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan (2025). *Large language models: An applied econometric framework*. Tech. rep. National Bureau of Economic Research.

Lv, Linying (2024). "The Value of Information from Sell-side Analysts". *arXiv preprint arXiv:2411.13813*.

Manela, Asaf and Alan Moreira (2017). "News implied volatility and disaster concerns". *Journal of Financial Economics* 123.1, pp. 137–162.

Mann, Roberta F. (2019). "I Robot: U Tax? Considering the Tax Policy Implications of Automation". *McGill Law Journal* 64.4, pp. 763–807.

Manning, Benjamin S., Kehang Zhu, and John J. Horton (2024). *Automated Social Science: Language Models as Scientist and Subjects*.

Matos, Miguel Godinho de and Idris Adjerid (2022). "Consumer consent and firm targeting after GDPR: The case of a large telecom provider". *Management Science* 68.5, pp. 3330–3378.

Mayew, William J and Mohan Venkatachalam (2012). "The power of voice: Managerial affective states and future firm performance". *The Journal of Finance* 67.1, pp. 1–43.

McCarthy, J, ML Minsky, and N Rochester (1955). "A Proposal For The Dartmouth Summer Research Project On Artificial Intelligence".

McKinsey Global Institute, McKinsey & Company (2023). *The Economic Potential of Generative AI: The Next Productivity Frontier*.

McLaughlin, Bryce and Jann Spiess (2024). "Designing Algorithmic Recommendations to Achieve Human-AI Complementarity". *arXiv preprint arXiv:2405.01484*.

McLean, R David and Jeffrey Pontiff (2016). "Does academic research destroy stock return predictability?" *The Journal of Finance* 71.1, pp. 5–32.

Meincke, Lennart, Karan Girotra, Gideon Nave, Christian Terwiesch, and Karl T Ulrich (2024). "Using large language models for idea generation in innovation". *The Wharton School Research Paper Forthcoming* 9, p. 2024.

Mihet, Roxana, Kumar Rishabh, and Orlando Manuel Da Costa Gomes (2025). "Is it AI or data that drives firm market power?" *Journal of Monetary Economics*, p. 103878.

Mirzadeh, Iman, Keivan Alizadeh, Hooman Shahrokhi, Oncel Tuzel, Samy Bengio, and Mehrdad Farajtabar (2024). "Gsm-symbolic: Understanding the limitations of mathematical reasoning in large language models". *arXiv preprint arXiv:2410.05229*.

Mo, Hongwei and Shumiao Ouyang (2025). "(Generative) AI in Financial Economics". *Journal of Chinese Economic and Business Studies*, pp. 1–79.

Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo (2022). "Uneven growth: automation's impact on income and wealth inequality". *Econometrica* 90.6, pp. 2645–2683.

Morton, Fiona Scott (2024). "Can Massive Technological Progress Hurt Workers? A Review of Power and Progress by Daron Acemoglu and Simon Johnson". *Journal of Economic Literature* 62.4, pp. 1671–1681.

Mullainathan, Sendhil and Jann Spiess (2017). "Machine learning: an applied econometric approach". *Journal of Economic Perspectives* 31.2, pp. 87–106.

Murray, Scott, Yusen Xia, and Houping Xiao (2024). "Charting by machines". *Journal of Financial Economics* 153, p. 103791.

Nagel, Stefan (2021). *Machine learning in asset pricing*. Princeton University Press.

Nilsson, Nils J (2014). *Principles of artificial intelligence*. Morgan Kaufmann.

Nori, Harsha, Samuel Jenkins, Paul Koch, and Rich Caruana (2019). "Interpretml: A unified framework for machine learning interpretability". *arXiv preprint arXiv:1909.09223*.

Novy-Marx, Robert and Mihail Z Velikov (2025). *AI-powered (finance) scholarship*. Tech. rep. National Bureau of Economic Research.

Noy, Shakked and Whitney Zhang (2023). "Experimental evidence on the productivity effects of generative artificial intelligence". *Science* 381.6654, pp. 187–192.

Obaid, Khaled and Kuntara Pukthuanthong (2022). "A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news". *Journal of Financial Economics* 144.1, pp. 273–297.

Oecd (2023a). *Algorithmic competition, OECD competition policy roundtable background note.* https://www.oecd.org/daf/competition/algorithmic-competition-2023.pdf.

— (2023b). *Generative Artificial Intelligence in Finance.* Tech. rep. 9. Paris: OECD Publishing.

— (2023c). *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market.* Paris: OECD Publishing.

Ouyang, Qinglin and Shumiao Ouyang (2025). "The Fixed Disposition Effect". *Available at SSRN 5243056.*

Ouyang, Shumiao (2021). "Cashless payment and financial inclusion". *Available at SSRN 3948925.*

Ouyang, Shumiao, Hayong Yun, and Xingjian Zheng (2024). "AI as Decision-Maker: Ethics and Risk Preferences of LLMs". *Available at SSRN 4851711.*

Pástor, L'uboš and Pietro Veronesi (2009). "Technological revolutions and stock prices". *American Economic Review* 99.4, pp. 1451–1483.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer (2023). "The impact of ai on developer productivity: Evidence from github copilot". *arXiv preprint arXiv:2302.06590.*

Press, Ori, Andreas Hochlehnert, Ameya Prabhu, Vishaal Udandarao, Ofir Press, and Matthias Bethge (2024). "CiteME: Can Language Models Accurately Cite Scientific Claims?" *Advances in Neural Information Processing Systems* 37, pp. 7847–7877.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever (2021). *Learning Transferable Visual Models From Natural Language Supervision.*

Rahimikia, Eghbal and Felix Drinkall (2024). "R e (Visiting) Large Lan guage Models in Finance". *Available at SSRN.*

Reher, Michael and Stanislav Sokolinski (2024). "Robo advisors and access to wealth management". *Journal of Financial Economics* 155, p. 103829.

Rock, Daniel (2019). "Engineering value: The returns to technological talent and investments in artificial intelligence". *Available at SSRN 3427412.*

Ross, Jillian, Yoon Kim, and Andrew W. Lo (2024). *LLM economicus? Mapping the Behavioral Biases of LLMs via Utility Theory.*

Rossi, Alberto G and Stephen Utkus (2024). "The diversification and welfare effects of robo-advising". *Journal of Financial Economics* 157, p. 103869.

Rubin, Cathy (2024). "Can Universal Basic Income Save Us From The Destabilization Of AI And Automation?" *Forbes.*

Sadhwani, Apaar, Kay Giesecke, and Justin Sirignano (2021). "Deep learning for mortgage risk". *Journal of Financial Econometrics* 19.2, pp. 313–368.

Sarkar, Suproteem K (2025). "AI Agents, Productivity, and Higher-Order Thinking: Early Evidence From Software Development". *Available at SSRN 5713646.*

Sarkar, Suproteem K and Keyon Vafa (2024). "Lookahead bias in pretrained language models". *Available at SSRN.*

Schubert, Gregor (2025). "Organizational Technology Ladders: Remote Work and Generative AI Adoption". *Available at SSRN.*

Schwartz, Eduardo S and Walter N Torous (1993). "Mortgage prepayment and default decisions: A Poisson regression approach". *Real Estate Economics* 21.4, pp. 431–449.

Seamans, Robert and Manav Raj (2018). *AI, labor, productivity and the need for firm-level data.* Tech. rep. National Bureau of Economic Research.

Serafeim, George (2024). "Climate Solutions, Transition Risk, and Stock Returns".

Shaffer, Matthew and Charles CY Wang (2024). "Scaling Core Earnings Measurement with Large Language Models". *Available at SSRN.*

Sheng, Jinfei, Zheng Sun, Baozhong Yang, and Alan L Zhang (2024). “Generative AI and asset management”. *Available at SSRN 4786575*.

Si, Chenglei, Diyi Yang, and Tatsunori Hashimoto (2024). “Can llms generate novel research ideas? a large-scale human study with 100+ nlp researchers”. *arXiv preprint arXiv:2409.04109*.

Singh, Chandan, Jeevana Priya Inala, Michel Galley, Rich Caruana, and Jianfeng Gao (2024). “Rethinking interpretability in the era of large language models”. *arXiv preprint arXiv:2402.01761*.

Sockin, Michael and Wei Xiong (2023a). “A model of cryptocurrencies”. *Management Science* 69.11, pp. 6684–6707.

— (2023b). “Decentralization through tokenization”. *The Journal of Finance* 78.1, pp. 247–299.

Stokel-Walker, Chris and Richard Van Noorden (2023). “What ChatGPT and generative AI mean for science”. *Nature* 614.7947, pp. 214–216.

Stradi, Francesco and Gertjan Verdickt (2025). “Algorithm Aversion, Appreciation, and Investor Return Beliefs”. *The University of Auckland Business School Research Paper (forthcoming)*.

Tang, Huan (2019a). “Peer-to-peer lenders versus banks: substitutes or complements?” *The Review of Financial Studies* 32.5, pp. 1900–1938.

— (2019b). “The value of privacy: Evidence from online borrowers”. *Available at SSRN*.

Thakor, Anjan V (2020). “Fintech and banking: What do we know?” *Journal of financial intermediation* 41, p. 100833.

Thaler, Richard H and Hersh M Shefrin (1981). “An economic theory of self-control”. *Journal of political Economy* 89.2, pp. 392–406.

Tian, Minyang, Luyu Gao, Shizhuo Zhang, Xinan Chen, Cunwei Fan, Xuefei Guo, Roland Haas, Pan Ji, Kittithat Krongchon, Yao Li, et al. (2024). “Scicode: A research coding benchmark curated by scientists”. *Advances in Neural Information Processing Systems* 37, pp. 30624–30650.

Trinh, Trieu H, Yuhuai Wu, Quoc V Le, He He, and Thang Luong (2024). “Solving olympiad geometry without human demonstrations”. *Nature* 625.7995, pp. 476–482.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017). "Attention is all you need". *Advances in neural information processing systems* 30.

Webb, Michael (2019). "The impact of artificial intelligence on the labor market". *Available at SSRN 3482150*.

Webb, Michael, Nick Short, Nicholas Bloom, and Josh Lerner (2018). *Some facts of high-tech patenting*. Tech. rep. National Bureau of Economic Research.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. (2022). "Chain-of-thought prompting elicits reasoning in large language models". *Advances in neural information processing systems* 35, pp. 24824–24837.

West, Darrell M (2018). *The future of work: Robots, AI, and automation*. Brookings Institution Press.

Wu, Lynn, Bowen Lou, and Lorin M Hitt (2025a). "Innovation strategy after IPO: How AI analytics spurs innovation after IPO". *Management Science* 71.3, pp. 2360–2389.

Wu, Wenbo, Jiaqi Chen, Zhibin Yang, and Michael L Tindall (2021). "A cross-sectional machine learning approach for hedge fund return prediction and selection". *Management Science* 67.7, pp. 4577–4601.

Wu, Zongxiao, Yizhe Dong, Yaoyiran Li, and Baofeng Shi (2025b). "Unleashing the power of text for credit default prediction: Comparing human-written and generative AI-refined texts". *European Journal of Operational Research*.

Yang, Chih-Hai (2022). "How artificial intelligence technology affects productivity and employment: firm-level evidence from Taiwan". *Research Policy* 51.6, p. 104536.

Yang, Keer (2025). "Trust as an entry barrier: Evidence from fintech adoption". *Journal of Financial Economics* 169, p. 104062.

Yoon, So Hye (2025). "Selling Fast or Selling Junk: Is iBuying Sustainable?"

Zarifhonarvar, Ali (2024). "Experimental Evidence on Large Language Models". *Available at SSRN 4825076*.

Zhang, Yiming (2024). *Do Mutual Funds Benefit from the Adoption of AI Technology?* https://ssrn.com/abstract=4871159.

Zhang, Yue, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, et al. (2023). "Siren's song in the AI ocean: a survey on hallucination in large language models". *arXiv preprint arXiv:2309.01219*.

Zheng, Xiang (2025). "How can innovation screening be improved? A machine learning analysis with economic consequences for firm performance". *Journal of Financial and Quantitative Analysis*, pp. 1–57.

Zhong, Hongda (2024). "Optimal Integration: Human, Machine, and Generative AI". *Machine, and Generative AI (July 18, 2024)*.

Zhou, Tianyu, Pinqiao Wang, Yilin Wu, and Hongyang Yang (2024). "FinRobot: AI Agent for Equity Research and Valuation with Large Language Models". *arXiv preprint arXiv:2411.08804*.

Zhou, Yijun (2024). "Artificial Intelligence and Debt Collection: Evidence from a Field Experiment". *Available at SSRN 4905228*.

9 Figure

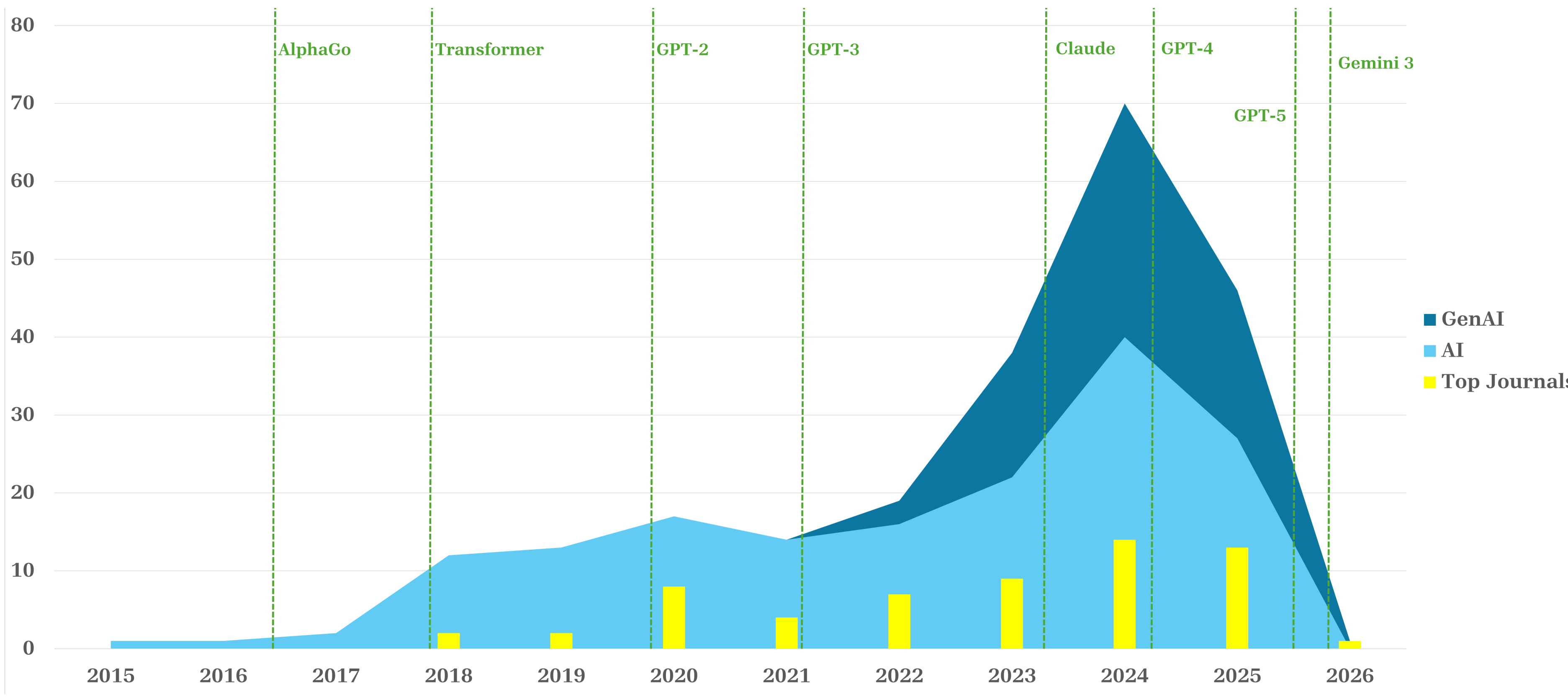

Figure 1: Number of AI-related Papers in Financial Economics Research from 2015 to January 2026. This figure illustrates the number of academic papers related to Artificial Intelligence (AI) (excluding generative AI) and generative AI (GenAI) each year from 2015 to January 2026 in the field of financial economics (see Table 1). The figure is based on a sample of AI-related papers from our manual screening and does not represent an exhaustive list of studies. The top journals are: American Economic Review (AER), Quarterly Journal of Economics (QJE), Journal of Political Economy (JPE), Econometrica, Review of Economic Studies (RES), Journal of Finance (JF), Review of Financial Studies (RFS), Journal of Financial Economics (JFE), Journal of Accounting and Economics (JAE), Journal of Accounting Research (JAR), The Accounting Review (TAR), Review of Finance (RF), Journal of Financial and Quantitative Analysis (JFQA), and Management Science (MS). The vertical dotted lines represent important events related to AI: AlphaGo (March 2016), the Transformer paper "Attention Is All You Need" (June 2017), GPT-2 (February 2019), GPT-3 (June 2020), Claude (March 2023), GPT-4 (March 14, 2023), GPT-5 (August 7, 2025), and Gemini 3 (November 18, 2025).

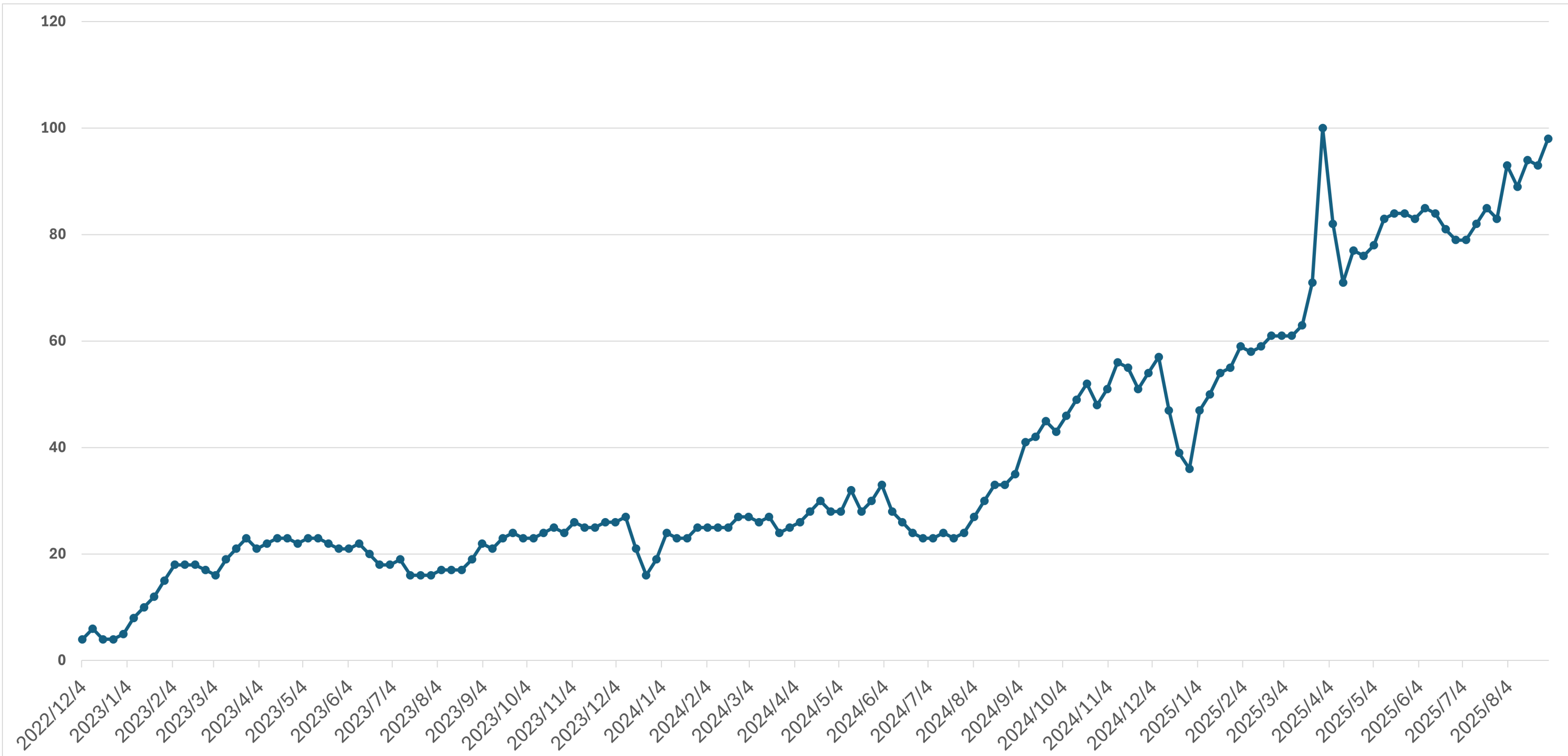

Figure 2: Search interest in "ChatGPT" on Google from December 2023 to August 2025. The numbers in the figure represent global search interest in "ChatGPT" relative to the highest point on the chart during the period from December 2023 to August 2025. A value of 100 indicates peak popularity for the term; a value of 50 means it was half as popular.

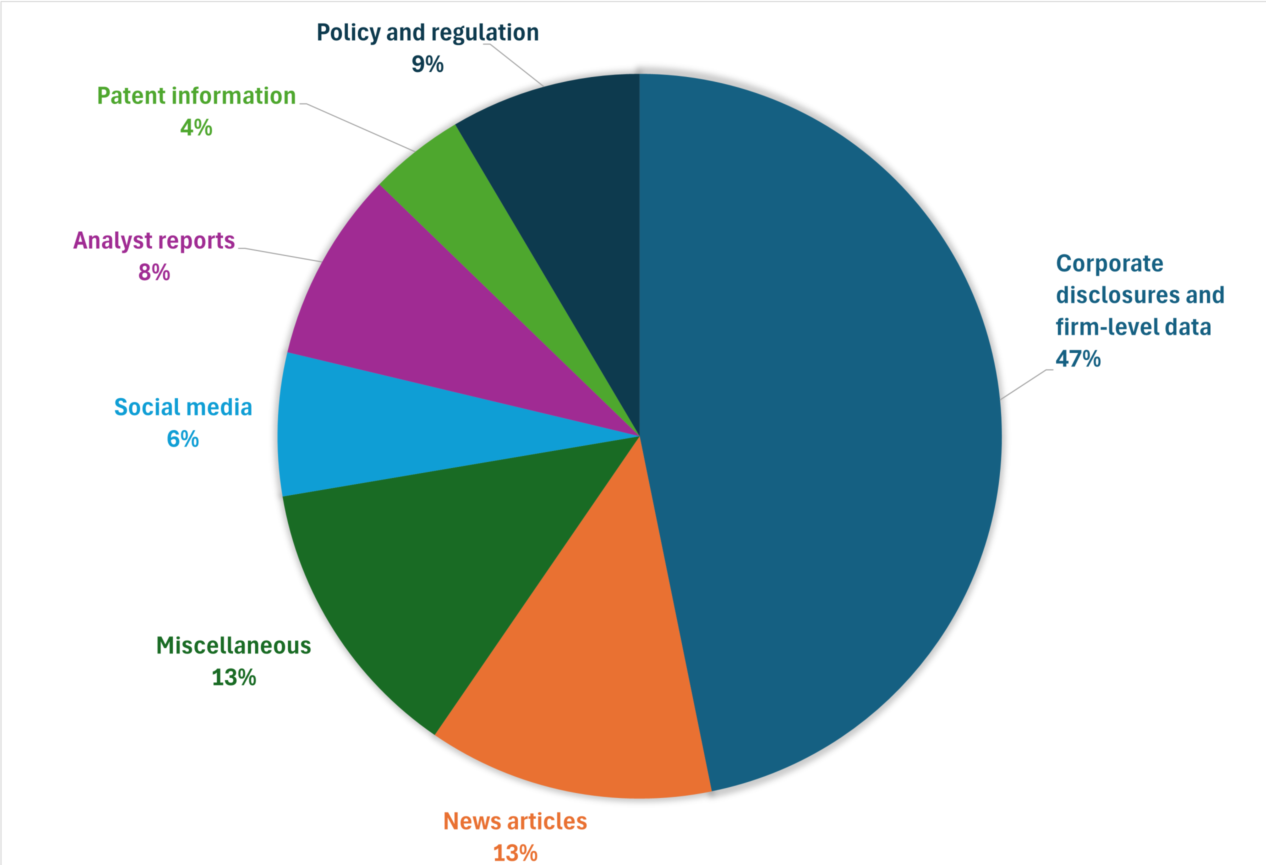

Figure 3: Distribution of GenAI Inputs Across Studies. This pie chart shows the distribution of input types used in recent academic studies that apply generative AI as an analytical tool in financial economics research. The chart is based on a total of 38 papers. One paper can use multiple types of information. Additional details are provided in Table 5.

10 Tables

Table 1: Number of AI-related Papers in Financial Economics Research from 2015 to January 2026. This table reports the number of academic papers related to Artificial Intelligence (AI) and generative AI (GenAI) each year from 2015 to January 2026 in the field of financial economics. The table is based on a sample of AI-related papers through our manual screening and does not represent the exhaustive list of studies. The top journals are: American Economic Review (AER), Quarterly Journal of Economics (QJE), Journal of Political Economy (JPE), Econometrica, Review of Economic Studies (RES), Journal of Finance (JF), Review of Financial Studies (RFS), Journal of Financial Economics (JFE), Journal of Accounting and Economics (JAE), Journal of Accounting Research (JAR), The Accounting Review (TAR), Review of Finance (RF), Journal of Financial and Quantitative Analysis (JFQA), and Management Science (MS).

Panel A: Trends in AI-related Financial Economics Research

Year	Total AI	(of which GenAI)	(of which Top Journals)
2015	1	0	0
2016	1	0	0
2017	2	0	0
2018	12	0	2
2019	13	0	2
2020	17	0	8
2021	14	0	4
2022	19	3	7
2023	38	16	9
2024	70	30	14
2025	46	19	13
2026	1	1	1
Total	234	69	60

Panel B: Top Journal Publication of AI-Related Financial Economics Research

Journal	Count
American Economic Review	3
Quarterly Journal of Economics	3
Journal of Political Economy	3
Econometrica	1
Review of Economic Studies	1
The Journal of Finance	6
The Review of Financial Studies	12
Journal of Financial Economics	13
Journal of Accounting and Economics	3
Journal of Accounting Research	2
The Accounting Review	1
Review of Finance	1
Journal of Financial and Quantitative Analysis	1
Management Science	10

Table 2: Major Large Language Models (LLMs) as of January 2026. This table reports the key specifications of major LLMs available as of January 2026.

Model (provider)	**Release date**	**Knowledge cutoff**	**Max token**	**Parameters**	**Cost (per 1M tokens)**
GPT-5.2 (OpenAI)	Dec 11, 2025	Aug 31, 2025	400k	Undisclosed	$1.75 In / $14.00 Out
GPT-5.1 (OpenAI)	Oct 15, 2025	Sep 30, 2024	400k	Undisclosed	$1.25 In / $10.00 Out
Gemini 3 (Google)	Nov 18, 2025	Jan 2025	2M	$\sim$2T (Est.)	$2.00 In / $12.00 Out
Gemini 2.5 Pro (Google)	June 12, 2025	Sep 2024	2M	$\sim$1.5T (Est.)	$1.50 In / $9.00 Out
Claude Opus 4.5 (Anthropic)	Nov 24, 2025	Mar 2025	200k	Undisclosed	$5.00 In / $25.00 Out
Claude Sonnet 4.5 (Anthropic)	Oct 20, 2025	Jan 2025	200k	Undisclosed	$3.00 In / $15.00 Out
Llama 4 Maverick (Meta)	May 14, 2025	Aug 2024	1M	400B (17B Active)	$0.18 In / $0.54 Out
Llama 4 Scout (Meta)	Apr 05, 2025	Jul 2024	128k	70B	$0.10 In / $0.30 Out
Qwen3-Max (Alibaba)	Sep 05, 2025	May 2025	262k	1T+ (MoE)	$1.20 In / $6.00 Out
Qwen3-72B (Alibaba)	Jul 15, 2025	Mar 2025	128k	72B	$0.40 In / $1.20 Out
DeepSeek-V3.2-Exp (DeepSeek)	Sep 29, 2025	Dec 2024	128k	671B (37B Active)	$0.28 In / $0.42 Out
DeepSeek-V3 (DeepSeek)	Dec 26, 2024	Jul 2024	128k	671B (37B Active)	$0.14 In / $0.28 Out
Grok 4.1 (xAI)	Nov 16, 2025	Oct 2025	256k	Undisclosed	$3.00 In / $15.00 Out
Grok 4 (xAI)	Aug 22, 2025	Jun 2025	128k	Undisclosed	$2.00 In / $10.00 Out

Table 3: Comparison between Traditional Textual Analysis and Large Language Models (LLMs) in Financial Economics.

Feature	Traditional Textual Analysis	LLMs
Primary Goal	Quantify simple text features (word counts, sentiment) for prediction.	Capture deep semantic meaning, context, and generate richer text representations.
Key Techniques	Bag-of-Words, Dictionary Methods, TF-IDF, Latent Dirichlet Allocation (LDA).	Transformers (BERT, RoBERTa, GPT), Fine-tuning, Embeddings.
Data Usage	Small, domain-specific datasets; manual feature engineering.	Massive pretraining on diverse corpora; fine-tuning with minimal labeled data.
Core Capability	Word frequency and topic extraction; ignores syntax and context.	Contextual understanding, nuanced sentiment detection, text generation.
Use cases	Sentiment indexing (10-Ks, news), topic analysis, basic risk flags.	Earnings call analysis, sentiment prediction, question answering, report summarization.
Strengths	Simple, interpretable, low computation cost; robust for coarse tasks.	High predictive power, domain adaptability, captures subtle language signals.
Weaknesses	Ignores context, static vocabulary, limited for complex language.	High computational cost, black-box nature, potential hallucinations.
Scalability	Fast and cheap; easy to deploy on large datasets.	Slower and costlier; scalable with cloud/GPU optimization.
Challenges	Updating vocabularies, shallow insights, context blindness.	Interpretability, regulatory concerns, model updating and bias management.

Table 4: **Comparison between Traditional AI/ML and Generative AI/LLMs in Financial Economics.**

Feature	Traditional AI/ML	Generative AI / LLMs
Primary Goal	Prediction, classification, pattern recognition, clustering from existing data.	Content generation (text, code, images, synthetic data), summarization, translation, complex instruction following, reasoning (emerging).
Key Techniques	Regression (Linear, Lasso, Ridge), Decision Trees, Random Forests, Support Vector Machines, Neural Networks, Clustering.	Transformers (e.g., GPT architecture), Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs).
Data Used	Primarily structured numerical data (e.g., firm characteristics, market data), but also unstructured data (text, images) for specific tasks.	Primarily trained on vast amounts of unstructured text and code, increasingly multimodal (images, audio).
Core Capability	Learning statistical relationships and making predictions based on learned patterns.	Understanding and generating human-like language/content, capturing context and semantics, performing zero-shot or few-shot tasks.
Use cases	Credit scoring, fraud detection, algorithmic trading (rule-based/predictive), risk factor identification, return prediction (factor models).	Enhanced fraud detection, robo-advising (conversational), sentiment analysis from news/social media, narrative asset pricing, report generation, code assistance, synthetic data generation.
Key Risks/Challenges	Bias in training data, overfitting, model risk, interpretability ("black box" for complex models), data privacy.	Hallucinations (generating plausible but false information), embedded bias from web-scale data, data privacy (training data leakage), prompt engineering dependency, high computational cost, explainability challenges, potential for misuse (disinformation).

Table 5: Inputs to Generative AI from Recent Applications in Financial Economics. This table summarizes recent applications of generative AI as analytical tools in financial economics. It highlights the primary structured and unstructured inputs used to prompt generative models, along with the key findings of these studies.

Authors (year)	GenAI input	Findings
Acikalin et al. (2022)	Patent documents	Weakened IP protection spurs competition but may deter innovation for resource-constrained firms.
Audrino et al. (2024)	Financial news articles	LLM-based uncertainty indices significantly forecast macroeconomic variables, asset returns, and fund flows.
Bartik et al. (2024)	Municipal zoning codes (the full text of local housing regulations)	Housing regulations are multidimensional: some cities use value-capture regulations that allow development while extracting surplus, while others adopt exclusionary regulations that restrict density and entrench racial and economic segregation.
Bastianello et al. (2024)	Analyst reports	Differences in attention allocation drive forecast disagreement and asset pricing anomalies.
Beckmann et al. (2024)	Earnings call Q&A sessions	Unusual financial communication correlated with lower announcement returns and higher volatility.
Blankespoor et al. (2024)	40,000 financial questions posed to a GenAI chatbot	Substantial gap in GenAI adoption between sophisticated and novice investors.

(The table is continued on the next page.)

Table 5 (continued)

Breitung and Müller (2025)	Business descriptions from 10-K filings and annual reports	Global business networks constructed with GPT-based embeddings can predict stock return and M&A.
Chang et al. (2023)	Earnings call transcripts	Generative AI narrows the information gap between retail investors and short sellers, reducing performance disparities.
Chen et al. (2022b)	News articles (multiple languages)	LLM-based sentiment and return forecasting models significantly outperform conventional approaches in both U.S. and international equity markets.
Chen and Wang (2024)	Patent texts	Functional AI innovations have heterogeneous effects on firm-level employment and value; augmenting AI boosts productivity and hiring, while displacing AI cuts costs but not productivity.
Chen et al. (2024b)	Historical stock return time series and price charts	ChatGPT mimics well-known human biases in financial prediction, overweighting recent return trends.
Chen et al. (2024c)	Social media posts	Sentiment extracted via LLMs predicts future stock returns.

(The table is continued on the next page.)

Table 5 (continued)

Chen et al. (2025b)	Wall Street Journal headlines	ChatGPT identifies good news signals that significantly predict future stock market returns, outperforming traditional sentiment methods and other language models such as DeepSeek.
Cong et al. (2025b)	News articles, SEC filings, patent documents	Textual Factors improve forecasting of macroeconomic variables and asset pricing exposures.
Dyck et al. (2025)	Public information on family owners (philanthropy, advocacy, green investments)	Family-controlled firms are not cleaner on average than widely held firms; only when owners have strong environmental preferences and abatement costs are low do emissions decline.
Fang et al. (2024)	Chinese government policy documents	Local government policies boost firm entry, with effectiveness dependent on implementation quality; excessive imitation leads to inefficiencies like overcapacity and protectionism.
Fetzer et al. (2024)	Harmonized System product descriptions	AIPNET, an AI-generated production network, highlights upstream shifts and onshoring trends.
Gabaix et al. (2024)	Institutional portfolios and earnings call transcripts	Asset embeddings learned from portfolios explain stock valuations and return comovement better than firm characteristics.

(The table is continued on the next page.)

Table 5 (continued)

Gao et al. (2025)	Semi-annual and annual mutual fund reports	GPT-4-extracted structured beliefs about the economy, policy, and markets predict future index returns and fund trading outcomes; countercyclical policy beliefs enhance return predictability and drive fund outperformance.
Hansen and Kazinnik (2023)	FOMC communications	GPT models outperform traditional NLP methods in classifying policy stances and identifying monetary shocks.
Huang et al. (2023a)	Analyst reports, earnings call transcripts, corporate social responsibility (CSR) reports	FinBERT, a finance-specific LLM, outperforms traditional dictionaries, classical ML, deep learning models, and generic BERT in sentiment and ESG classification, with stronger performance in negative sentiment detection.
Huang et al. (2024a)	Meme images on Reddit's WallStreetBets	Meme usage spurs higher engagement on social media and influences the linguistic tone toward fanaticism, trust, rebelliousness, and infectiousness, encouraging retail investors to hold or double down on losing stocks temporarily.
Jha et al. (2024a)	Earnings call transcripts	ChatGPT investment score provides incremental predictive power beyond traditional measures.

(The table is continued on the next page.)

Table 5 (continued)

Jha et al. (2024b)	Conference call transcripts	Managerial expectations extracted from earnings calls predict macro and firm-level outcomes.
Jha et al. (2025)	Book excerpts	Finance sentiment varies persistently across countries and predicts long-run GDP and credit growth.
Kakhbod et al. (2024)	10-K filings and USPTO patent text	The paper constructs a firm-level measure of innovation displacement exposure (IDE) that predicts lower future profit growth, especially for firms vulnerable to disruptive innovations by major innovators.
Kim et al. (2023a)	MD&A and earnings call transcripts	LLM-generated summaries outperform full documents in explaining market reactions.
Kim et al. (2023b)	Earnings call transcripts	GPT-based risk measures (political, climate, AI) outperform bigram-based proxies in predicting stock volatility and capturing firm-specific risk exposure; AI risks become significant post-2021 and influence firm behavior and economic outcomes.
Krockenberger et al. (2024)	SEC filings	CovenantAI improves covenant violation detection.

(The table is continued on the next page.)

Table 5 (continued)

Li et al. (2026)	Analyst reports, earnings calls, employee reviews	Cause-effect links between corporate culture and firm outcomes.
Li et al. (2024a)	Earnings press releases and financial statements	GPT-4 underperforms human analysts in forecasting earnings; its accuracy depends on textual ranking consistency, alignment with key financial metrics, and limitations from the model's knowledge cutoff.
Lopez-Lira and Tang (2023)	News headlines	ChatGPT-4 predicts stock returns without financial training; trading on its sentiment signals yields substantial abnormal returns, especially for complex or negative news. Predictive power improves with model size and erodes over time as adoption grows.
Lu et al. (2023)	Wall Street Journal news and Chinese policy announcements	ChatGPT portfolios generate significant alphas based on political and policy news.
Lv (2024)	Analyst reports	Textual information in analyst reports explains more return variation than numerical forecasts.
Novy-Marx and Velikov (2025)	Structured financial research data	LLMs autonomously generate finance research papers.

(The table is continued on the next page.)

Table 5 (continued)

Serafeim (2024)	10-K business descriptions	Firm-level exposure to climate solutions quantified from corporate disclosures.
Shaffer and Wang (2024)	SEC filings (10-K reports)	Firm exposure to climate solutions quantified from 10-K business descriptions, showing firms hedge transition risks and exhibit lower expected returns.
Zhou et al. (2024)	Financial statements, stock prices, news, and filings	FinRobot framework enables high-quality and automated equity research.

Table 6: Recent Applications of AI Models in Asset Pricing. This table summarizes recent research that applies machine learning techniques to problems in asset pricing. We report machine learning techniques and the research questions for each paper.

Authors (Year)	Method	Research Question
Avramov et al. (2023)	Neural Networks, Generative Adversarial Network, Instrumented Principal Component Analysis, Conditional Autoencoder	Do machine learning models generate economically significant stock return forecasts under real-world constraints and frictions?
Bell et al. (2024)	Explainable Boosting Machine, compared with OLS, LASSO, Random Forests, XGBoost	Can interpretable machine learning models improve prediction of corporate bond returns while preserving transparency and economic interpretability?
Bianchi et al. (2021)	Partial Least Squares, Penalized Linear Regression, Boosted Regression Trees, Random Forests, Extremely Randomized Trees, Shallow and Deep Neural Networks, Group-Ensembled Neural Networks, Principal Component Analysis	Can machine learning methods improve the prediction of Treasury bond excess returns using yield and macroeconomic data?
Bryzgalova et al. (2025)	Asset Pricing Trees, Mean-Variance Efficient Spanning, Dual Shrinkage Estimation	Can a new method for portfolio construction produce test assets that span the Stochastic Discount Factor better than traditional sorted portfolios?

(The table is continued on the next page.)

Table 6 continued.

Bybee et al. (2023)	Latent Dirichlet Allocation for textual topic modeling, Sparse Instrumented Principal Component Analysis for factor estimation	Can narrative themes extracted from business news serve as economically meaningful risk factors in an Intertemporal Capital Asset Pricing Model framework?
Cao et al. (2023b)	Deep neural networks.	Can deep learning extract forward-looking risk assessments from mutual fund disclosures, and do these assessments predict future risk-taking and performance?
Chen et al. (2023a)	Transfer Learning using Neural Networks.	Can misspecified economic models still improve machine learning-based financial forecasting? How can theory and data be combined for better performance and robustness?
Chen et al. (2024a)	Feedforward Neural Network; Recurrent Neural Network; Generative Adversarial Network.	Can a deep learning model trained under the no-arbitrage condition estimate a general stochastic discount factor that explains all stock returns?
Chinco et al. (2019)	LASSO.	Can machine learning methods like LASSO uncover short-lived, sparse, and non-intuitive cross-stock return predictors that traditional economic intuition would miss?

(The table is continued on the next page.)

Table 6 continued.

Cong et al. (2021a)	Deep Reinforcement Learning, Transformer Encoder, LSTM, Cross-Asset Attention Networks.	Can reinforcement learning directly optimize portfolio strategies using raw asset-level data, outperforming supervised learning and traditional methods?
Cong et al. (2025a)	Panel Tree (P-Tree), Boosted Panel Tree, Random Panel Forest.	Can we use tree-based models to systematically construct test assets that span the efficient frontier and reveal the underlying structure of the stochastic discount factor?
DeMiguel et al. (2023)	Elastic Net, Gradient Boosting, Random Forests.	Can nonlinear machine learning methods select long-only mutual fund portfolios with positive out-of-sample alpha net of all costs?
Dong et al. (2022)	Elastic Net, Forecast Combination, Principal Components, Partial Least Squares.	Do long-short anomaly portfolios, commonly used in cross-sectional studies, help predict aggregate US market excess returns over time?
Feng et al. (2020)	Double-Selection LASSO.	Can a new factor explain asset prices after accounting for a large set of existing known factors? How can we perform valid inference in this high-dimensional setting?

(The table is continued on the next page.)

Table 6 continued.

Fernández Tamayo et al. (2023)	Term Frequency-Inverse Document Frequency vectorization of text; Lasso Regression; Gradient Boosting.	Can qualitative information from private equity prospectuses predict fundraising success and future fund performance more effectively than traditional quantitative measures?
Freyberger et al. (2020)	Nonparametric Adaptive Group LASSO.	Which firm characteristics provide incremental predictive power for expected returns in a high-dimensional setting, and how do nonlinear models compare to linear benchmarks?
Glasserman et al. (2020)	Supervised Latent Dirichlet Allocation.	Can supervised topic models identify news topics that explain contemporaneous stock returns and help interpret market movements?
Gu et al. (2020)	Elastic Net, Principal Components Regression, Partial Least Squares, Random Forests, Boosted Trees, Neural Networks.	Can machine learning methods outperform traditional linear models in predicting equity risk premiums across stocks and portfolios?

(The table is continued on the next page.)

Table 6 continued.

Gu et al. (2021)	Conditional Autoencoder (nonlinear factor model with neural networks).	Can a nonlinear conditional factor model based on autoencoders better explain cross-sectional variation in stock returns than linear models like IPCA and Fama-French?
Guijarro-Ordonez et al. (2021)	Deep Learning, Convolutional Neural Network + Transformer.	Can deep learning techniques improve statistical arbitrage strategies by better modeling residual return dynamics and allocating capital more effectively?
He et al. (2023a)	Reduced-Rank Asset Pricing.	Can a reduced-rank factor model identify the few factors that truly explain expected returns, and are additional factors beyond the Fama-French five necessary?
Jiang et al. (2023)	Convolutional Neural Networks.	Can deep CNNs trained on images of price and volume data extract return-predictive patterns that outperform standard time-series signals?
Kaniel et al. (2023)	Feedforward Neural Network.	Can we predict mutual fund abnormal returns using modern machine learning methods, and which characteristics matter most?

(The table is continued on the next page.)

Table 6 continued.

Kelly et al. (2019)	Instrumented Principal Component Analysis.	Can latent factors and time-varying risk exposures be estimated via firm characteristics, and do they explain returns?
Kelly et al. (2024)	High-dimensional linear models, Random Feature Neural Networks, Ridgeless Regression, Ridge Shrinkage.	Does increasing model complexity (more parameters than observations) improve return prediction and portfolio performance?
Kelly et al. (2025)	Linear and Nonlinear Transformer-Based Stochastic Discount Factor.	Can transformer architectures with attention mechanisms and high complexity improve SDF estimation?
Kozak et al. (2020)	Bayesian SDF Estimation with L2 (and L1) Penalization.	Can a Bayesian estimator with an economically motivated prior robustly recover the SDF in a high-dimensional space of characteristic-based predictors? How sparse is the true SDF?
Leippold et al. (2022)	OLS, LASSO, Elastic Net, Partial Least Squares, GBRT, Random Forest, VASA, Neural Networks	How well do machine learning methods predict stock returns in the Chinese market, and which signals are most relevant?
Lettau and Pelger (2020)	Risk-Premium Principal Component Analysis.	Can incorporating mean return information into factor extraction improve asset pricing model performance?

(The table is continued on the next page.)

Table 6 continued.

Li and Rossi (2020)	Boosted Regression Trees.	Can fund performance be predicted in real time from characteristics of the stocks they hold? How does BRT compare to linear models and univariate sorts?
Light et al. (2017)	Partial Least Squares (PLS) estimator.	How can information from many firm characteristics be efficiently aggregated to estimate expected stock returns when true betas or risk factors are unobservable?
Murray et al. (2024)	Convolutional Neural Network.	Can machine learning models, using only past returns, predict the cross-section of future stock returns? Does this challenge the weak form of the efficient market hypothesis?
Obaid and Pukthuanthong (2022)	Convolutional Neural Networks with Transfer Learning (Google Inception v3).	Can investor sentiment extracted from news photos predict market returns and trading activity, and how does it interact with text-based sentiment?
Wu et al. (2021)	LASSO, Random Forest, Gradient Boosting, Deep Neural Network.	Can machine learning methods using return-based fund characteristics improve cross-sectional hedge fund return prediction and selection?