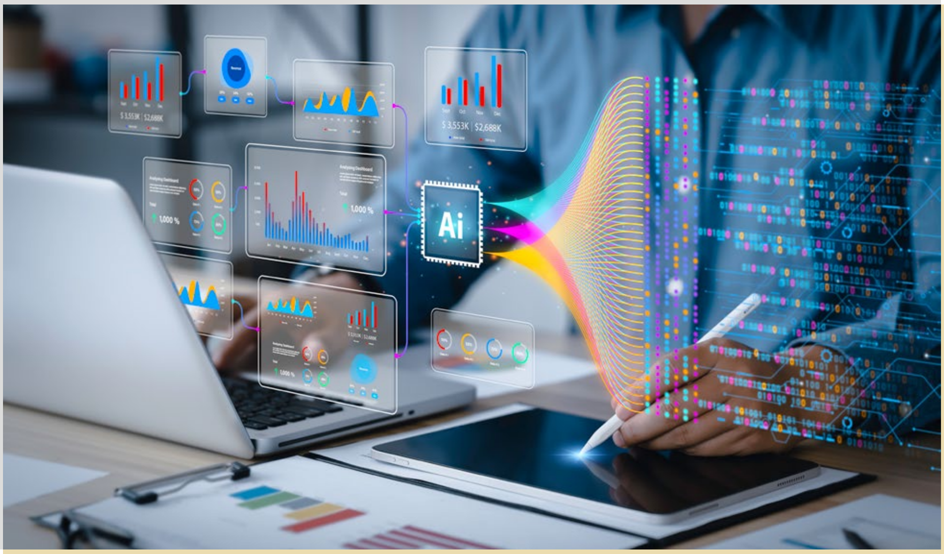
SFI Public Discussion Note

The Role of AI in Transforming Financial Practices



November 2024

Introduction

Prof. Markus Leippold
SFI Senior Chair and Professor of Financial Engineering at the University of Zurich

"I think old-fashioned intelligence works pretty well" (Munger, 2023).

Since OpenAI's introduction of ChatGPT in November 2022, talk of artificial intelligence (AI) has captured Wall Street's enthusiasm. Yet, even as AI surges to the forefront in corporate earnings calls, some esteemed investors remain skeptical, viewing its promises as mere hype.

Berkshire Hathaway's Warren Buffett, a leading source of investment wisdom, alongside his long-time collaborator, the late Charles Munger, juxtaposes artificial intelligence with "old-fashioned" intelligence. While others are increasingly fascinated by AI's capabilities, its rapid advancements, and its integration into the corporate sphere, Buffett stresses the enduring value of the human intellect. Can AI, he asks, genuinely surpass the depth and nuance of human thought and creativity?

As was true for the Industrial Revolution, the current era of AI is characterized by swift technological progress and by optimism about its implications for our social structures, economy, and daily lives. The introduction of mechanized production methods 200-some years ago marked a significant turning point in human history; AI represents another. As we explore the ability of machines to create, reason, and interact in ways previously thought to be exclusive to humans, we must also be mindful of the ethical, privacy, and security concerns that will accompany AI's widespread adoption.

In this *Public Discussion Note*, I hope to strike a balance between exuberance and caution. By contrasting the optimistic projections of the big-tech companies, who view AI as pivotal to market competition, with the skeptical outlook of investors such as Buffett and Munger, I will attempt to unravel the complexities of today's AI phenomenon. I will begin by discussing the technology itself and the potential for AI to reshape the economic landscape, and end by touching on ethical considerations, societal implications, and the challenges that must be met before its widespread application in finance.

With its *Public Discussion Note* series the Swiss Finance Institute (SFI) is actively promoting a well-founded discussion of topics relevant to the financial industry, politics, and academia. Furthermore, SFI disseminates its findings through research, publications, Master Classes, and conferences.

This *Public Discussion Note* is based on the research that my team and I have conducted in the areas of sustainability and AI over the past few years. I would like to acknowledge my co-authors, Julia Bingler, Chiara Colesanti-Senni, Glen Gostlow, Jingwei Ni, Tobias Schimanski, Dominik Stammbach, Saeid Vaghefi, and Tingyu Yu, as well as all the research assistants. These talented and motivated young researchers have been instrumental in the progress of this work. I hope their contributions will inspire future scholars to continue advancing in this important field.

A Brief History of AI

Early Visions (before 1950):
Since ancient times, civilizations have envisioned forms of artificial life, as seen in the myths of Pygmalion and the Golem. In the story of Pygmalion, a sculptor falls in love with a statue he carved from stone, which then comes to life. The Golem is a creature made of clay that, likewise, is brought to life through mystical means to serve its creator. Additionally, during the ancient and medieval periods, inventors across various cultures created mechanical devices that imitated human or animal actions. These automata also reflect early visions of artificial intelligence.

Philosophers such as René Descartes (born 1596) and Gottfried Leibniz (born 1646) played pivotal roles in thinking about the mind, intelligence, and the potential for machines to mimic human reasoning. Descartes questioned the nature of thought and consciousness, while Leibniz envisioned a universal language of reasoning, known as the *calculus ratiocinator*. This early conceptual framework, which represents logical statements and reasoning in a precise, mathematical form, has influenced modern symbolic logic and computation.

Birth of the Field (1950s):
Alan Turing's work in the 1940s and 1950s, especially in developing the Turing Test, laid the foundational concepts for evaluating machine intelligence. The Turing Test is a method for determining whether a machine can exhibit intelligent behavior indistinguishable from that of a human. Turing's earlier work also established fundamental principles that underpin modern computing. The Turing machine, for example, is a theoretical device that manipulates symbols on a strip of tape according to a set of rules, providing a simple yet powerful model for understanding what it means for a function to be computable.

The Dartmouth Conference (1956) gathered researchers who shared the belief that machine intelligence could be achieved and set the stage for future research. It is often cited as the formal birth of AI as a scientific discipline.

The Rollercoaster of Progress (1960s-1990s):
Due to its hype outpacing its actual progress, the field experienced periods of reduced funding and interest. These so-called "AI Winters" occurred primarily in the late 1970s to early 1980s, and again in the late 1980s to early 1990s. They were marked by skepticism about AI's potential to fulfill its promises, leading to reduced investment and a reevaluation of goals.

Despite the challenges, the development of expert systems represented a significant success for AI in commercial and industrial applications. These systems were able to mimic the decision-making abilities of human experts in specific domains, such as medical diagnosis and mineral exploration. In finance, expert systems have been used to estimate default losses, providing banks and financial institutions with sophisticated tools to assess credit risk and to predict the likelihood of borrowers defaulting on loans.

Resurgence and Transformation (1990s-Present):
The resurgence of AI has been fueled significantly by advances in machine learning and by the availability of the large datasets known as "big data." These developments have enabled AI systems to learn from examples, rather than relying solely on hard-coded rules.

Deep learning, a type of machine learning using neural networks, has led to significant breakthroughs in computer vision, speech recognition, and natural language processing, making possible such applications as real-time language translation, autonomous vehicles, and effective web searching.

Motivation

Reading Stanisław Lem's science fiction novel "Thus Spoke Golem" as a young person in the early 1990s sparked my fascination with the possibilities of AI. The narrative unfolds around a character named Golem, a machine of unprecedented intellect that engages in deep discussions with the world's leading scientists. While the concept felt purely speculative to me at the time, Lem's novel, written in the 1980s, eerily mirrors today's interactions with AI in sectors like finance. Lem's exploration of AI's capabilities, ethical implications, and potential to coexist with humans resonates now more than ever, as technologies such as OpenAI's ChatGPT and Google's Gemini become entwined with our daily operations.

Today, I am actively involved in the technological developments Lem once envisioned. Reflecting on Lem's novel underscores the importance of navigating AI's expanding role in our industry with careful thought and foresight. It reminds me of the delicate balance between leveraging technological advancements and considering their broader implications. This contemplation becomes particularly salient as we observe AI's transformative impact in medicine, biology, and meteorology.

For instance, AlphaFold, developed by Google DeepMind, has revolutionized biology by accurately predicting the 3D structures of proteins based solely on their amino acid sequences (Jumper et al., 2021). This breakthrough addresses the long-standing protein-folding problem, expediting drug discovery and enhancing our understanding of biological processes by predicting the structures of all 200 million proteins known to science in just one year. Given that, at a rough estimate, it takes one PhD student between four and five years to uncover a single protein structure, this task would otherwise have taken hundreds of millions of years (!) of human labor. Demis Hassabis and John M. Jumper, both from Google DeepMind, were awarded the Nobel Prize in Chemistry 2024 for their groundbreaking contributions to predicting protein folding. Similarly, GraphCast, another innovative model from Google DeepMind, has redefined weather forecasting by outperforming traditional methods with remarkable accuracy (Cookson, 2023; Lam et al., 2023). By analyzing extensive historical data with a graph neural network, GraphCast delivers superior predictions of weather conditions, from temperature to wind patterns to the landfalls of hurricanes. It also significantly reduces energy consumption by lowering the computational time needed from many hours to a minute.

Might we witness a comparable AlphaFold or GraphCast moment in the finance sector? Generative AI, as it unearths patterns and draws insights from vast datasets, has the potential to fundamentally transform financial analytics, risk assessment, and economic forecasting. Powerful AI tools could redefine our strategic decision-making process—emphasizing the need for readiness and adaptability in the finance industry to ensure that we achieve significant, responsible results.

Generative AI in the Financial Sector

Before we delve into its impact on the financial sector, I want to clarify what I mean by generative AI. AI and generative AI are related but distinct concepts; while all generative AI is AI, not all AI is generative.

AI is a broad term encompassing all computer systems designed to mimic human cognitive functions. These systems can be categorized as narrow or weak AI, designed to perform a specific task that previously required human intelligence, or as general or strong AI (AGI), which can apply intelligence across a wide range of tasks, as humans do.

The AI systems currently in the public spotlight, like ChatGPT, are examples of weak AI. They are highly advanced and capable of generating human-like text, but they are specialized for specific tasks and do not possess the wide-ranging capabilities of AGI. Many researchers and experts agree that complete AGI remains hypothetical, as no system has yet met the agreed-upon criteria. However, some people believe that achieving AGI is a matter of *when*, not *if* (Fry, 2022).

Generative AI refers to an AI system—whether weak or strong—that can generate new content or data similar, but not identical, to the data it was trained on. Training an algorithm involves presenting it with a large dataset and assigning it a series of tasks to solve, guided by a reward mechanism that provides feedback on its performance. This process allows the algorithm to refine its understanding of patterns and decision-making criteria, improving its ability to solve similar tasks in the future. The dataset may include text, images, music, voice, and other media. Generative AI models are designed to understand the underlying patterns and structures of this input data, allowing them to produce new, original outputs based on what they have learned.

The crucial feature of generative AI is its ability to create, offering significant potential in content creation, design, and even drug discovery. The cognitive and creative capabilities of generative AI are the subject of ongoing philosophical debate (Millière & Buckner, 2024). It has the potential, however, to not only replicate human tasks, but also to augment human creativity by generating new ideas, designs, and perspectives, offering innovative applications that extend beyond the present scope of AI's analytical and decision-making capabilities.

Impact of Generative AI in the Financial Sector

Generative AI is already significantly impacting the financial sector, with implications for productivity, operational efficiency, risk management, and the creation of new business models. According to a McKinsey report, generative AI could add between USD 200 billion and USD 340 billion annually to the banking industry, equivalent to 9 to 15 percent of its operating profits (McKinsey & Company, 2023). This value is derived mainly from increased productivity, with the most significant absolute gains expected in the corporate and retail sectors.

Through generative AI, financial institutions can enhance their service operations and improve the customer's experience. For example, according to McKinsey & Company (2023), Goldman Sachs uses an AI-based tool to automate test generation, which previously was labor-intensive. Citigroup has used generative AI to assess the impact of new US capital rules.

Maufe and Brown (2023) identify five practical uses for generative AI in the financial services industry, all of which increase employee productivity. These include financial document search and synthesis, enhanced virtual assistants, capital markets research, software development, and personalized financial recommendations.

However, the International Monetary Fund (2023) has noted that deploying generative AI in the financial sector raises several risks. These include embedded bias, privacy shortcomings, opaqueness in the decision-making process, robustness issues, cybersecurity threats, and potential impacts on broader financial stability. The International Monetary Fund warns that excessive reliance on generative AI could increase contagion risk and build new systemic risks into the financial sector.

Despite these concerns, the financial sector is actively exploring the capabilities of generative AI, with competitive pressures fueling the rapid adoption of applications that promise gains in efficiency, cost savings, and improved risk management.

Generative AI technology offers a range of potential uses within finance, including:

- **Crafting forecasts and budgets.** A KPMG survey found that 83% of the participants employed AI in financial planning (KPMG, 2024), relying on it for predictive analytics, the generation of scenarios, and the provision of budgetary insights.

- **Creating financial reports and presentations.** Generative AI significantly reduces the time and effort needed to produce regular financial reports by automatically populating templates with new data and by deriving insights from financial and business data.

- **Extracting market intelligence.** Using generative AI's expansive language models, organizations can tap into public datasets to glean market insights, craft competitive intelligence, and uncover customer trends through tailored analyses that may be either region-specific or persona-based.

- **Deriving strategic insights from data.** By using generative AI to analyze data relevant to customer relationship management (CRM) or enterprise resource planning (ERP), firms can gain valuable strategic insights to inform their financial decisions, to address pricing and performance challenges, or to enhance other business operations.

- **Automating contract management.** Generative AI streamlines contract creation, focusing on non-standard terms, identifying revenue-related clauses, and documenting accounting treatments.

- **Enhancing anomaly detection and fraud prevention.** Through identifying data irregularities and monitoring ongoing transactions, generative AI offers a robust mechanism for detecting errors and preventing fraud, thereby safeguarding the financial integrity of an organization.

- **Personalized investment advice.** Generative AI can tailor investment recommendations to an individual's risk profile and financial goals, enhancing the client's experience.

- **Fraud detection.** By analyzing large datasets of transactions, generative AI can identify fraudulent patterns, improving financial security.

- **Market analysis and risk management.** Likewise, given the appropriate datasets, generative AI can identify market trends and assess potential risks, aiding investment decisions.

This wide variety of applications in finance raises several key points of discussion within the profession. First are ethical considerations, specifically ways in which transparency can be maintained. New standards and protocols are needed to ensure that AI-generated financial content does not perpetuate biases or misinformation. Second is the impact of automation on our employees. As AI displaces traditional roles, our workforce must learn new skills and be prepared to adapt. Third is the regulatory landscape. As generative AI technologies advance, and financial regulations become inadequate or obsolete, the regulatory frameworks must adapt to ensure both innovation and consumer protection. Last is the continued development of innovative products and services to meet emerging market demands and provide financial institutions with a competitive advantage.

Practical Applications

Let's examine some practical applications of generative AI and explore their role in driving advancements within the financial sector. These examples from Environmental, Social, and Governance (ESG) initiatives demonstrate how generative AI can transform a specific area of finance, highlighting its potential to revolutionize the entire industry.

Case Study 1: Identifying Green Innovation

In a recent paper, my co-author and I delve into how generative AI plays a crucial role in deciphering the dense language of green patents (Leippold and Yu, 2024). Understanding the technical intricacies of these patents can be daunting for someone wanting to evaluate a firm's sustainability initiatives. Generative AI models surmount this challenge by turning complex descriptions and specialized terminology into concise summaries. The process involves the AI systematically "reading" the patent, interpreting its crucial elements, and rearticulating these in simple terms. This application of generative AI performs multiple functions that are critical to the financial sector.

First, by distilling technical jargon into straightforward language, generative AI grants broader access to the details of green innovation. It extends the patent's information to a diverse audience, including financial analysts, investors, and policymakers, who may not possess the deep technical expertise otherwise needed to understand it. This democratization of information is crucial, as it lets a broader range of stakeholders make informed decisions.

Second, the summaries provided by AI enhance analytical efficiency. Financial analysts, operating in environments where speed and accuracy are paramount, benefit immensely from being able to quickly comprehend a patent's significance without getting bogged down in the details. Here, AI not only saves time, it also enhances the quality of the analysis, ensuring that financial decisions are based on a thorough understanding of the potential impact and value of a firm's innovations.

Third, the AI summaries facilitate market-wide analysis. Analysts can more easily compare sustainability efforts across various firms, sectors, and industries when they have a clear understanding of each patent's core improvements. This capability is invaluable for potential investors, ensuring that their decisions are grounded in a full understanding of the commitment of different firms toward green innovation.

After using generative AI to distill complex green patent texts into clear, concise summaries, my co-authors and I created a BERT-based language model, called ClimateBERT, to refine our understanding further (Webersinke et al., 2022). BERT stands for Bidirectional Encoder Representations from Transformers. Developed by Google, BERT better understands the context of a word by considering both the preceding and following words in the sentence. We trained ClimateBERT to recognize the nuances of sustainability terminology, allowing for deep contextual analysis of the simplified texts produced by generative AI.

Analyzing these texts over time, ClimateBERT can identify trends in a firm's discourse on sustainability. This type of analysis highlights shifts in the firm's emphasis on green innovation, providing early signals of strategic changes that could affect its long-term commitment to sustainability, as well as its financial performance. By combining generative AI and ClimateBERT, we thus reach a deeper understanding of how the firm engages with green innovation.

Finally, to understand the broader impacts of green innovation, we use our data to create metrics that show how sustainability initiatives influenced market-adjusted cumulative abnormal returns around key events, such as Trump's 2016 election victory, Biden's 2020 win, the outbreak of the Russia-Ukraine war, and the announcement of the Inflation Reduction Act. Our work offers financial analysts and investors a powerful tool to better understand and mitigate the risks related to green innovation and sustainability.

Case Study 2: Analyzing Corporate Sustainability

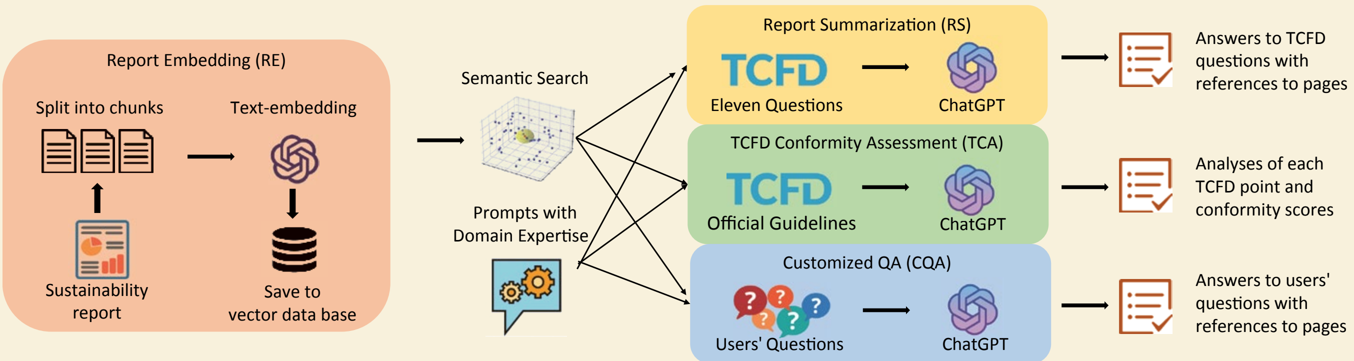
My co-authors and I developed an AI tool called CHATREPORT (Ni et al., 2023), based on our previous work on CHATCLIMATE (Vaghefi et al., 2023), to address the dense and voluminous data found in corporate sustainability reports. These reports detail a company's environmental impact, governance structures, and social initiatives; they serve as a critical lens through which investors, policymakers, and the broader public can assess the company's commitment to sustainable practices. However, their complexity, and the sheer volume of data they contain, make manual analysis a daunting task. The challenge is further magnified when the analysis must adhere to specific regulatory frameworks or guidelines. Traditionally, only a few entities have been equipped with the resources needed for thorough, detailed analysis of these reports, creating a significant barrier to transparency. CHATREPORT democratizes the process, letting a broad range of stakeholders engage with and act upon these reports.

At the heart of CHATREPORT is a sophisticated pipeline, illustrated in Figure 1, that transforms the intricate data of corporate sustainability reports into actionable insights. A critical aspect of our methodology involves integrating domain expertise directly into the AI development loop. The process begins with extracting the raw text of a report, normalizing the data formats, and removing any irrelevant information. To understand and analyze the content, the pipeline uses advanced large language models (LLMs) that are fine-tuned with domain-specific knowledge. Users and other human experts create and refine the analytical prompts, ensuring that the AI's outputs are technically accurate and contextually relevant. The tool identifies key themes, assesses compliance with guidelines like those of the Task Force on Climate-related Financial Disclosures (TCFD), and generates insights in a user-friendly format.

Each step in the CHATREPORT pipeline is designed to ensure that the data is processed with high precision and is aligned with the latest sustainability standards and practices, resulting in a powerful tool for stakeholders needing to make informed decisions.

A common challenge with AI-generated content, as discussed below, is the risk of "hallucination," in which the AI produces incorrect or misleading information not supported by the data. CHATREPORT tackles this issue through its traceability features. Each piece of AI-generated content is directly traceable to specific sections of the source documents, allowing the user to manually verify the accuracy of the information. Moreover, CHATREPORT is highly user-centric, offering features that allow users to tailor the analysis to their needs. Through a dynamic question-answering module, users can query the AI about particular aspects of a report. The AI then retrieves and processes the relevant data, providing customized insights.

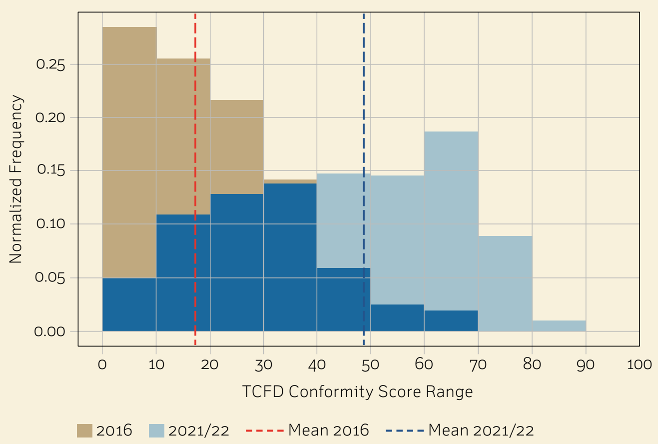
Figure 1: The CHATREPORT Pipeline



Note: The pipeline for CHATREPORT automates the analysis of corporate sustainability reports through four main modules. The Report Embedding (RE) module splits the report into text chunks, embedding them into a vector space for semantic search based on TCFD recommendations. The Report Summarization (RS) module retrieves relevant report sections and prompts a large language model (LLM) to summarize disclosures according to TCFD's eleven recommended aspects. The TCFD Conformity Assessment (TCA) module evaluates the report's adherence to TCFD guidelines, generating conformity scores and analysis paragraphs. The Customized Question Answering (CQA) module allows users to pose specific questions, with the LLM providing answers referencing the appropriate sections of the report. This comprehensive pipeline enhances the transparency and efficiency of sustainability report analysis.

Source: Ni et al. (2023)

Figure 2: Density Plot of TCFD Conformity Scores

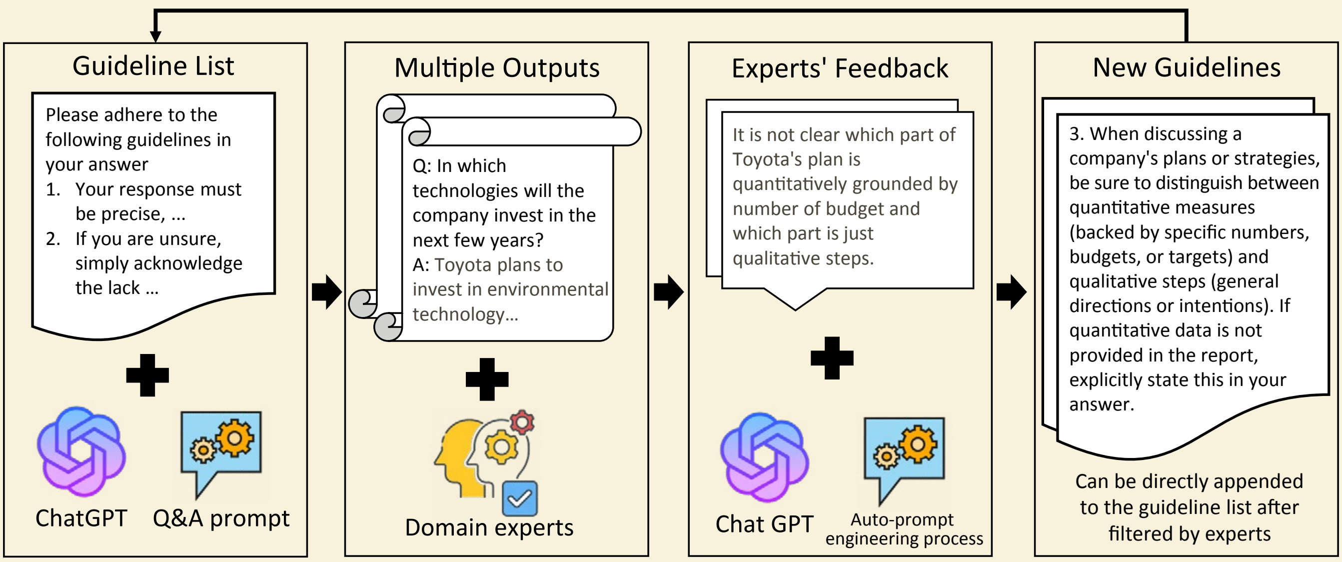


Note: This figure illustrates the distribution of TCFD conformity scores for a sample of corporate sustainability reports from companies listed on the NYSE, for the reporting years 2016 and 2021/22. The TCFD conformity scores are calculated based on how well the reports adhere to TCFD guidelines. The density plot shows a noticeable improvement in the average conformity scores between the years 2016 and 2021/22, indicating that companies are increasingly aligning their sustainability disclosures with the TCFD recommendations.

Source: Ni et al. (2023)

In our paper, for example, we assessed over 10'000 corporate sustainability reports for conformity with the TCFD reporting requirements. Figure 2 visually presents the scores for two reporting periods: before the TCFD, shown in brown, and (some years) after the TCFD, shown in light blue. Our data reveal that the level of TCFD conformity increased substantially between the two periods. Before the formal inauguration of the TCFD, in 2016, the average conformity score was 17%. In 2021/22, it was 49%. This trend highlights the growing emphasis on transparent and comprehensive climate-related financial disclosures. In further analyses, we identified trends and outliers, facilitating a discussion on the effectiveness of current practices and on potential areas for improvement.

Figure 3: Automatic Prompt Engineering Pipeline



Note: CHATREPORT incorporates users' and other domain experts' feedback to refine and improve the prompts used by the large language model (LLM). Initially, the domain experts provide feedback on specific answers generated by the LLM. This feedback is then transformed into general guidelines through an automated prompt-engineering process. The AI integrates the new guidelines into its templates, enhancing the quality and precision of its future outputs. This iterative process leads to more accurate, detailed, and critical analysis of sustainability reports.

Source: Ni et al. (2023)

Finally, as Figure 3 shows, CHATREPORT is built to learn and adapt over time. The system continuously incorporates feedback from users and other human experts to refine its analytical models and prompts. This iterative process ensures that the system consistently improves in accuracy, reliability, and relevance. To emphasize transparency, my co-authors and I developed CHATREPORT as an open-source project; it can be accessed on https://reports.chatclimate.ai/. By making our methodologies, datasets, and tools publicly available, we encourage collaboration and invite developers and researchers from around the globe to contribute to advancing sustainable finance analytics.

The power of CHATREPORT lies in its ability to drastically speed up the process of analysis. With this AI tool, relevant information can be extracted from thousands of Corporate Sustainability Reports—up to 10'000—in just a few hours. CHATREPORT quickly condenses vast amounts of data into concise executive summaries, enabling faster, data-driven decision-making on a scale previously unattainable.

A Glimpse into the (Very) Near Future

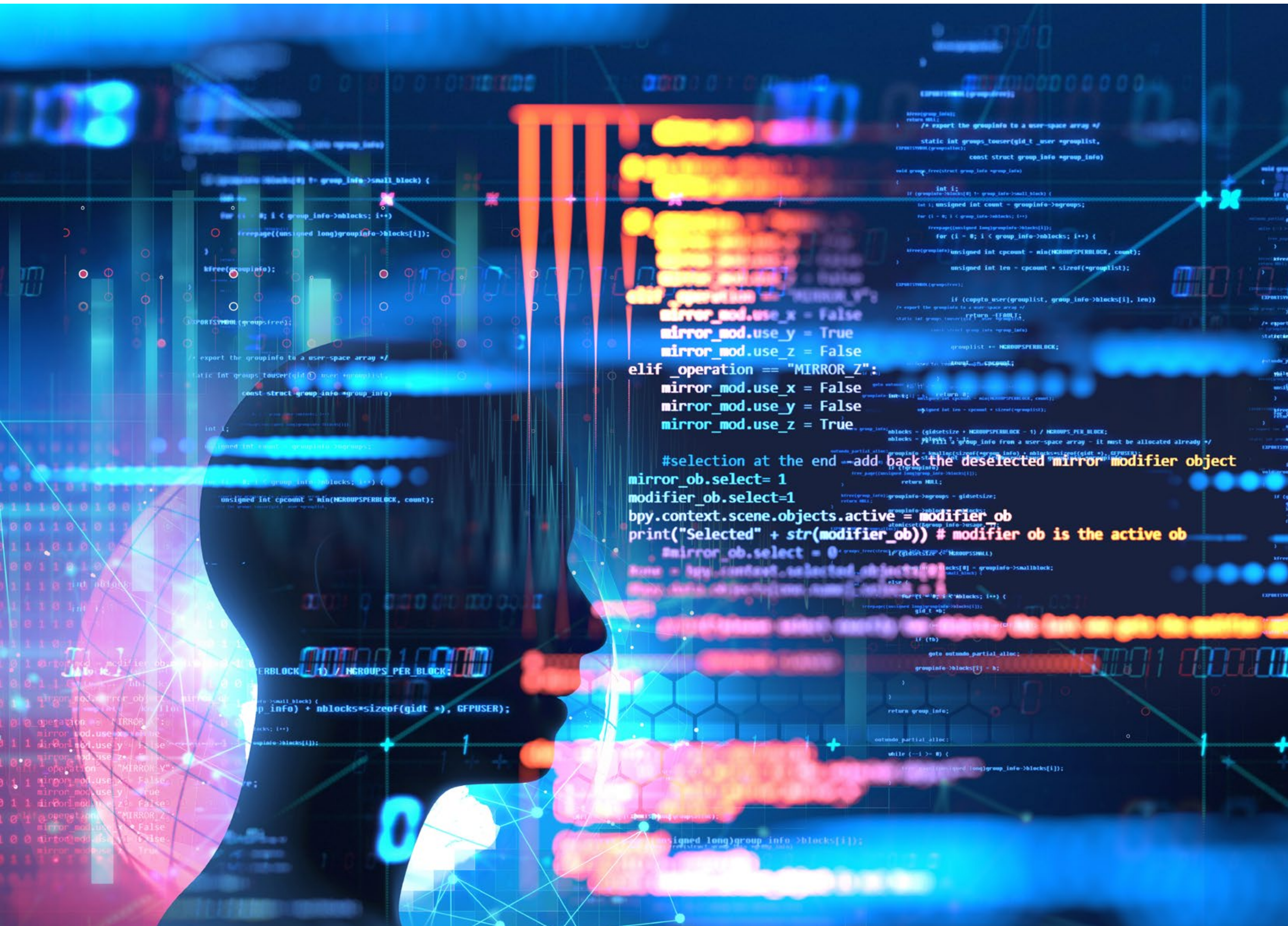
Generative AI could soon dramatically change how investments and financial decisions are analyzed. Recently, I tested an approach for automated fact-checking reminiscent of the Socratic method. In ancient Greece, Socrates engaged his interlocutors in relentless questioning to expose contradictions in their thoughts, encouraging a more profound exploration of underlying issues. This philosophical heritage is echoed in the Mediator-Advocate framework employed by Leippold et al. (2024) to fact-check climate claims.

Transferring this tool into finance, we would have multiple LLMs, termed "AI-Analysts," each analyzing a specific facet of financial data—such as market trends, regulatory impacts, economic indicators, or corporate performance metrics—and proposing potential investment strategies. These arguments would then be integrated by a "Chief AI-Investment Manager." To form a cohesive assessment, this Manager would ask follow-up questions to the AI-Analysts, similar to how Socrates guided his pupils to a deeper understanding through dialectic. This process ensures that many different perspectives are considered, contesting and complementing each other and ultimately producing a more robust conclusion.

Such a tool would improve financial decision-making by leveraging AI's computational speed, as well as by reducing biases for one type of financial analysis over another. Moreover, as financial markets increase in complexity, the scalability of AI systems would ensure that decision-makers could handle the growing data demands efficiently.

A pivotal feature of such a system would be its ability to track and record the progression of arguments among the AI-Analysts, providing an audit trail of how their decisions were reached. Such traceability is crucial for enhancing trust among stakeholders and for meeting the regulatory requirements of the financial industry, which demand thorough documentation of decision-making processes, especially in areas like risk assessment and compliance.

While integrating AI systems into finance promises a shift toward more dynamic, real-time decision-making, this potential comes with many theoretical and practical challenges, such as ensuring data accuracy, maintaining source credibility, and mitigating the risks of algorithmic biases.



Current Limitations of Generative AI

Despite its potential, generative AI technology is plagued by significant limitations that need urgent attention. Before it can revolutionize the financial sector, issues such as hallucination, lack of verifiability, and vulnerability to adversarial attacks need to be addressed.

Hallucination and inaccuracy
In Schimanski et al. (2024), my co-authors and I critically analyze the accuracy and verifiability of responses generated by LLMs. In complex question-answering and content-generation tasks, the phenomenon known as "hallucination" is common. Hallucination in AI refers to the model generating information that, although plausible, is detached from the actual source material and is entirely unfounded. This issue is critical in sectors like finance, healthcare, and law, where the accuracy and reliability of information are paramount. Our study highlights how LLMs can often misattribute sources or fabricate content, severely undermining users' trust in these automated systems.

To address these challenges, we propose enhancing the process of extracting text from the relevant sources and improving the attribution of AI-generated content. However, such technical fixes are only part of the solution. The implications of unchecked inaccuracy are vast, with significant repercussions possible in financial decision-making, patient care, and legal proceedings. Research is urgently needed to refine these models and to develop a robust ethical framework for their use.

Lack of verifiability
Generative search engines, including well-known platforms such as Bing Chat, NeevaAI, perplexity.ai, and YouChat, are engineered to respond to users' queries by generating answers complete with in-line citations. However, Liu et al. (2023) find that only about 52% of the sentences produced by these search engines are adequately supported by citations. This figure is alarmingly low, considering the importance of reliable information in today's digital age. Furthermore, the study shows that only 75% of the citations actually support the statements they appear to validate. This gap between appearance and reality misleads users, who rely on seemingly credible but fundamentally unsupported or incorrectly cited information.

This research highlights a crucial aspect of AI development: trustworthiness. Despite these generative search engines' sophisticated output, the prevalence of unsupported claims and citation inaccuracies significantly detracts from their reliability. AI training must be refined and strategies developed to enhance the verifiability of the information these systems provide.

Vulnerability to adversarial attacks
Generative AI is also susceptible to adversarial attacks that manipulate AI systems to intentionally misinterpret or misrepresent data. This weakness is particularly critical in some types of financial analysis, where such manipulations can severely skew the analytical outcome. In Leippold (2023), for example, I investigated "sentiment spin" attacks, in which a sophisticated model like GPT-3 can be used to modify the sentiment of financial texts, changing the negative sentiment of a series of sentences, for example, to neutral or positive. Such attacks pose severe risks to the reliability of automated systems.

Hu et al. (2024) detail various other adversarial strategies that exploit the vulnerabilities of generative search engines. Their research shows how generative search engines, despite their advanced capabilities, can be misled by queries specifically crafted to induce errors. This susceptibility highlights the urgent need for improvement in AI security measures.

The implications of adversarial attacks are not confined to finance. AI systems can also be manipulated to produce biased or misleading outputs in other sectors, such as healthcare, politics, and pharmaceuticals. Regulation of generative AI should focus on creating frameworks that ensure transparency, accountability, and fairness. As the use of AI expands, regulators must implement strict guidelines to promote ethical practices and to prevent misuse. Such regulations will protect users from the effects of adversarial attacks and will foster trust in AI systems, paving the way for their broader acceptance and integration into society.

Although some frameworks do exist, such as the EU AI Act, there is still a long way to go in crafting comprehensive and effective regulations. Formulating such regulations requires a nuanced approach, taking into account the rapid pace of technological change and the diverse and expanding applications of AI. It is essential to maintain an ongoing dialogue among academics, industry leaders—particularly from the tech sector—and policymakers.

Conclusion

Generative AI presents both unprecedented opportunities and great challenges. In this *Public Discussion Note*, I have outlined its transformative potential for the financial industry, from enhancing decision-making and risk assessment to revolutionizing corporate sustainability practices. However, the journey toward fully integrating AI into finance is fraught with complexities, including the problems of hallucination, verifiability, and vulnerability to adversarial attacks, which could undermine the trust in and reliability of AI systems.

The substantial investment already made in AI underscores the high stakes involved. According to Goldman Sachs (2024), while the potential for significant productivity gains and operational efficiencies is clear, there are contrasting views on AI's overall economic impact. Experts like MIT's Daron Acemoglu, among others, express skepticism, citing concerns that the technology may not justify the high costs, which may not decline as expected (Acemoglu, 2023). On the other hand, many optimistic voices believe that AI will eventually deliver substantial returns and economic benefits, and that its most transformative applications have yet to emerge.

As a researcher in this evolving field, I see both the potential and the challenges that lie ahead for its application in finance. The future of generative AI in finance is not predetermined, but will be shaped by our actions. We have an unparalleled opportunity to harness AI to drive innovation, enhance sustainability, and improve efficiency within the financial sector. However, reaching these goals demands careful and thoughtful implementation to mitigate the risks posed by unchecked exuberance.

References

Acemoglu, D. (2023). Harms of AI. In *The Oxford Handbook of AI Governance* (pp. 660–706). Oxford University Press eBooks. https://doi.org/10.1093/oxfordhb/9780197579329.013.65

Cookson, C. (2023, November 14). AI outperforms conventional weather forecasting methods for first time. *Financial Times.* https://www.ft.com/content/ca5d655f-d684-4dec-8daa-1c58b0674be1

Fry, H. [Host] (2022). *DeepMind: The podcast – The road to AGI* [Audio podcast]. Google. https://deepmind.google/discover/the-podcast/the-road-to-agi/

Golman Sachs (2024). *Gen AI: Too much spend, too little benefit?* https://www.goldmansachs.com/intelligence/pages/gs-research/gen-ai-too-much-spend-too-little-benefit/report.pdf

Hu, X., Li, X., Chen, J., Li, Y., Li, Y., Li, X., Wang, Y., Liu, Q., Wen, L., Yu, P. S., & Guo, Z. (2024). Evaluating robustness of generative search engine on adversarial factual questions. arXiv.org. https://arxiv.org/abs/2403.12077

International Monetary Fund (2023). *Generative artificial intelligence in finance: Risk considerations.* https://www.imf.org/-/media/Files/Publications/FTN063/2023/English/FTNEA2023006.ashx

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. a. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

KPMG (2024). *Why finance should lead the adoption of generative AI.* https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2024/generative-ai-finance.pdf

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science 382*(6677), 1416–1421. https://doi.org/10.1126/science.adi2336

Leippold, M. (2023). Sentiment spin: Attacking financial sentiment with GPT-3. *Finance Research Letters 55*(B), 103957. https://doi.org/10.1016/j.frl.2023.103957

Leippold, M., Vaghefi, S. A., Stammbach, D., Muccione, V., Bingler, J., Ni, J., Colesanti-Senni, C., Wekhof, T., Schimanski, T., Gostlow, G., Yu, T., Luterbacher, J., & Huggel, C. (2024). *Automated fact-checking of climate change claims with large language models.* arXiv.org. https://arxiv.org/abs/2401.12566

Leippold, M., & Yu, T. (2024). *Are green innovations priced? Evidence beyond patents.* Swiss Finance Institute Research Paper. http://dx.doi.org/10.2139/ssrn.4391444

Liu, N. F., Zhang, T., & Liang, P. (2023). Evaluating verifiability in generative search engines. arXiv.org. https://arxiv.org/abs/2304.09848

Maufe, Z., & Brown, T. (2023, October 3). *Five generative AI use cases for the financial services industry.* Google Cloud. https://cloud.google.com/blog/topics/financial-services/five-generative-ai-use-cases-financial-services-industry

McKinsey & Company (2023). *Capturing the full value of generative AI in banking.* https://www.mckinsey.com/industries/financial-services/our-insights/capturing-the-full-value-of-generative-ai-in-banking#/

Millière, R., & Buckner, C. (2024). *A philosophical introduction to language models — Part I: Continuity with classic debates.* arXiv.org. https://arxiv.org/abs/2401.03910

Munger, C. (2023, May 6). Answer [Q&A Session]. Berkshire Hathaway Annual Meeting, Omaha, NE, United States.

Ni, J., Bingler, J., Colesanti-Senni, C., Kraus, M., Gostlow, G., Schimanski, T., Stammbach, D., Vaghefi, S. A., Wang, Q., Webersinke, N., Wekhof, T., Yu, T., & Leippold, M. (2023). *CHATREPORT: Democratizing sustainability disclosure analysis through LLM-based tools* [Conference proceedings]. 2023 Conference on Empirical Methods in Natural Language Processing, Singapore. https://doi.org/10.18653/v1/2023.emnlp-demo.3

Schimanski, T., Ni, J., Kraus, M., Ash, E., & Leippold, M. (2024). Towards faithful and robust LLM specialists for evidence-based question-answering. arXiv.org. https://arxiv.org/abs/2402.08277

Vaghefi, S. A., Stammbach, D., Muccione, V., Bingler, J., Ni, J., Kraus, M., Allen, S., Colesanti-Senni, C., Wekhof, T., Schimanski, T., Gostlow, G., Yu, T., Wang, Q., Webersinke, N., Huggel, C., & Leippold, M. (2023). ChatClimate: Grounding conversational AI in climate science. *Communications Earth & Environment 4*(1). https://doi.org/10.1038/s43247-023-01084-x

Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2022). *ClimateBert: A pretrained language model for climate-related text.* arXiv.org. https://doi.org/10.48550/arXiv.2110.12010

Swiss Finance Institute

With support from its founders—the Swiss banking industry, the Swiss Confederation, and leading Swiss universities—the Swiss Finance Institute (SFI) competitively promotes world-class research and teaching in banking and finance in Switzerland. By combining academic excellence with practical experience SFI contributes to the strengthening of the Swiss financial center.

Editor and contact
Dr. Cyril Pasche
Senior Director Publications and Topic Development
+41 22 379 88 25
cyril.pasche@sfi.ch

swiss:finance:institute

Walchestr. 9, CH-8006 Zurich, T +41 44 254 30 80
info@sfi.ch, www.sfi.ch

Version 10/2024