CAN LARGE LANGUAGE MODELS «EX. GPT» TRANSFORM THE FINANCE INDUSTRY?

Hosted by CDPQ on August 24, 2023

REPORT

from the "Bloomberg-CDPQ" event organized by FIAM.

From left to right: Dr. Ranjan Bhaduri, David Ouellet, Jean-François Bérubé, Claude Perron, Dr. Russ Goyenko, Shawn Edwards and Dr. Hossein Kazemi.

In partnership with

**SHAWN
EDWARDS**

SPEAKER: SHAWN EDWARDS' BLOOMBERG' CTO

EXECUTIVE SUMMARY
Edwards provided a deep dive into Large Language Models (LLMs), emphasizing their emergent behavior, world knowledge, and technical capabilities.

He introduced Bloomberg's journey with LLMs, highlighting the development of BloombergGPT and its superior performance in domain-specific tasks. Edwards also discussed the importance of prompt engineering and the potential risks associated with LLMs.

Present a comprehensive overview of the capabilities, applications, and potential of Large Language Models, particularly in the context of the financial industry.

Understanding LLMs: LLMs are probabilistic models that predict the next word.

- **Emergent Behavior:** LLMs, unlike their predecessors, exhibit true language understanding, showcasing reasoning capabilities.
- **World Knowledge:** LLMs possess a vast amount of information, exceeding the knowledge of an average adult human.

Technical Insights:
- **Coding assistant** (GitHub Copilot): A tool that generates code based on user descriptions, with surveys showing claims of boosting productivity by 30-40%.

- **LLM Differentiation:** Unlike previous Bloomberg AI models, LLMs are flexible and high-performing right from the start. They are:
  - **General:** Capable of handling a myriad of tasks without specialized training. (Build models that can be used for multiple tasks)
  - **Broad:** Equipped with extensive world knowledge
  - **Accessible:** Designed for direct interaction using language, not code. This democratizes their use, allowing domain experts to program systems without relying on programmers.

Historical Perspective:
- **Evolution:** While the concept of predicting the next word in a sequence isn't new, what sets LLMs apart is their ability to consider the entire context to generate a probability distribution of what the next word should be.

- **Scale and Development:** The rapid development in AI has been driven by scale. The computational power available today, combined with existing theories, has enabled the training of these massive models, even if it takes months on cloud servers.

Applications and Use Cases:
- **Document Processing:** LLMs excel in tasks like summarization, information extraction, document understanding, and text generation.
- **Code Generation:** This is significant as it bridges the gap between natural language and tool building.

- **API Integration:** Bloomberg is exploring the potential of teaching LLMs to interact with its own APIs.

Bloomberg's Journey with LLMs (BloombergGPT):
- **Development:** Two years were invested in building BBGPT, aiming to determine if domain specific data could enhance GPT's performance in that domain.

- **Training Data:** A balanced mix of publicly available data and financial data drawn from Bloomberg's existing data creation, collection, curation and research efforts was used to train BloombergGPT.

- **Results:** BloombergGPT performed well in general tasks but outshone in domain-specific tasks, suggesting that organizations with high-quality data have a significant advantage.

- **Domain-Specific Training:** Initially uncertain, Bloomberg found that training on domain-specific data yielded substantial benefits.

Prompt Engineering:
- **Importance:** Bloomberg is investing in prompt engineering to refine prompts, ensuring users can extract accurate information from GPTs more easily.

- **Best Practices:** It's recommended to provide context and examples in prompts. For instance, instructing GPT to search for news about the debt ceiling before asking related questions ensures more accurate responses. Start with high level prompt, and then dive deeper with context and examples.

Reliability and Best Practices:
- **Unreliable Methods**: Using long documents without providing context can lead to inaccurate outputs.
- **Reliable Methods:** Short input documents with clear context yield the best results.

Bloomberg's Future Plans with LLMs:
- **Operational Integration:** Bloomberg is in the early stages of exploring how BloombergGPT can be integrated into AI-enabled applications.

Potential Use Cases:
- Converting natural language text to Bloomberg Query Language (BQL),
- Creating a language interface for the Terminal,
- Building a Copilot for Analysts that can summarize, analyze trends, integrate research, and more.

Risks Associated with LLMs:
- **Traditional Risks:** These include concerns about the usage rights of model outputs, data confidentiality, record-keeping requirements, and the provenance of generated content.
- **Novel Risks:** New challenges arise, such as the risk of the model generating false or misleading information.

SPEAKER DR. RUSS GOYENKO (MCGILL)

EXECUTIVE SUMMARY
Dr. Goyenko presented research suggesting AI's capability to interpret corporate executives' intentions with a significant degree of accuracy. The research utilized AI to predict earnings surprises based on the Management Discussion & Analysis (MD&A) section of corporate reports.

DR. RUSS GOYENKO

Presentation title: "Can Large Language Models Provide More Accurate Analysts' Forecast". The presentation is based on the research paper "Can AI read the Minds of Corporate Executives?" (source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4493166)

Key Insights: Large Language Models, LLM, LLMs, having undergone rigorous training to predict earnings surprises through the analysis of 10Q and 10K filings, coupled with an exhaustive examination of the entire EDGAR filing history, have emerged as formidable tools for making more accurate positive and negative earnings surprise predictions when compared to conventional analyst consensus estimates.

Economic validation: By leveraging the predictive capabilities of the LLM model, we constructed a straightforward Long-Short investment strategy. This strategy involves purchasing the top 20% of stocks predicted to have positive earnings surprises and shorting the bottom 20% of stocks expected to suffer the most negative surprises. Remarkably, this approach yields an annual market-risk adjusted Alpha of 6% on average.

Underlying Mechanism: The crux of this remarkable transformation lies in the wealth of information concealed within the seemingly mundane 10-Q and 10-K corporate filings. These documents contain an overwhelming amount of data, far beyond human capacity to absorb and contextualize for all publicly traded securities. The LLMs, however, have the unique ability to comprehend the nuances of these reports. They can sift through and process an immense volume of information, allowing them to continually refine their forecasting accuracy with every piece of data they "digest."

The Bigger picture:
This research underscores the undeniable truth: the integration of cutting-edge technology is no longer a choice but a necessity for making superior investment decisions. Embracing LLMs and their unparalleled analytical prowess is the path forward towards a new era of precision in financial forecasting and investment strategies.

Reporting presentation
Dr. Russ Goyenko presented findings from a research paper on AI's ability to "read the minds" of corporate executives. The crux of the research suggests that AI can indeed interpret the intentions of corporate executives with some investable degree of certainty.

Key Insights
Predicting Fundamentals Driving Returns: The research aimed to predict fundamentals known to influence returns. This was achieved by examining quarterly earning surprises against market expectations, specifically from the 10-Q and 10-K forms available on EDGAR. The Management Discussion & Analysis (MD&A) section of both Quarterly and Annual reports was a primary source of data.

Volume of Data:
The average MD&A section contains around 10,000 characters. Given the vast number of companies and their yearly reports, it's impractical for humans to read every MD&A. Interestingly, machines already account for 75% of EDGAR downloads.

Machine Learning Approaches:
Traditional machine learning models relied on a dictionary of words, which they compared against words in the MD&A sections. However, corporate management began to manipulate this system.

The research introduced a novel Large Language Model (LLM) method capable of processing filings of any length. These AI models were trained to predict the next 10-Q earnings surprises. Using a strategy of Long-Short (L-S) for the top and bottom 20% of predicted surprises the model outperformed the S&P 500 by 6.74% annually (6.01% on a risk-adjusted basis.

BERT Model Testing:
Both a generic BERT model and a fine-tuned BERT were tested for their ability to predict earnings surprises. The fine-tuned BERT, when used in a long-only strategy, outperformed the S&P 500 by a factor of 4 from 2004 to 2022.

MD&A Length as a Predictor:
The length of the MD&A section itself proved to be a significant predictor, almost as effective as a non-fine-tuned BERT model.
Reports that were excessively long tended to underperform. Conversely, overly short reports often overperformed. The research achieved a 60% success rate in predicting the direction of earnings surprises.

Additional Information - training Sample:
The model's training sample was updated annually. At the start of each year, the model was retrained using the latest data, and its weights were fixed for that year. This process was repeated at the beginning of subsequent years.

KEY POINTS FROM CDPQ & PSP LEADERS

Jean-François Bérubé from CDPQ, a global investment group active in the major financial markets, introduced the topic, emphasizing the transformative potential of Generative AI (GenAI), citing an often repeated comparison with the possibilities permitted by the invention of the personal computer. He provided a brief history of ML/AI, tracing its evolution from simple models to advanced systems like ChatGPT.

GenAI's value propositions include identifying investment opportunities, analyzing portfolio risks, and enhancing operational efficiencies.

- **Offensive:** Help find investment opportunities and emerging trends in the market
- **Defensive:** Help analyse the exposure of a portfolio to innovation risks
- **Digital:** Improve the operational side of the business. David (Ouellet) sees many opportunities on this front in the short run.

JEAN-FRANÇOIS BÉRUBÉ

David Ouellet from PSP Investments, a major player in the private sector and the world's largest farm Owner.

PSP paused tech innovations to develop their own PSPGPT, which is now used daily by around a third of their team. David mentions the use of an AI tool "ChatPSP", during meetings to enhance discussions by prompting questions and insights. Worth mentioning – they use the model to challenge themselves and reduce the risk of behavioural biases creeping into their final decision.

DAVID OUELLET

While board members might be keen on increasing the budget for GenAI projects, David emphasizes the importance of prioritizing data cloud storage and accessibility as they are foundational to the success of future GenAI projects. He also mentioned leveraging external services like BloombergGPT remains a likely venue. PSP is currently building many Proof-of-Concepts for the Front Office to help increase the efficiency of their PMs.

**DR. HOSSEIN
KAZEMI**

**DR. RANJAN
BHADURI**

EXECUTIVE SUMMARY
Dr. Kazemi underscored the importance of finance professionals understanding GenAI and he mentioned CAIA's efforts to introduce a new Professional designation – called FDP "Financial Data Professional", a designation already present in 40 countries.

PANEL DISCUSSION
Ranjan Bhaduri moderated discussions / questions on …

AI ETHICS AND ESG CONCERNS
David Ouellet: Emphasizes a balanced approach, considering the importance of weighing both risks and potential benefits when deciding how to use this technology.

Shawn Edwards: Highlights the dangers of relying on statistical models built on biased data, which can inadvertently perpetuate biases. He expresses concerns about malicious use of AI models, such as influencing elections, creating undetectable phishing emails, and producing deepfakes. He underscores the importance of finding ways of authenticating AI vs. human outputs.

Dr. Hossein Kazemi: points out the risks posed by unsophisticated users who don't understand the complexities of AI, drawing a parallel to uninformed investors during the financial crisis.

SELF-LEARNING SYSTEMS LIKE ALPHA GO AND LLMS
Dr. Bhaduri: AlphaGo was taught the game Go by examining games by human champions, and famously became the world champion. However, AlphaGo Zero, which was self-learning, beat AlphaGo 100 to 0. AlphaGo Zero had creative moves that were completely unconventional. With LLMs, we are trying to learn from human patterns … is this limiting the upside potential?

Dr. Russ Goyenko: Advocates for designing AI "portfolio managers" that consistently outperform peers in all scenarios. He hints at having conducted specific research on this topic.

AI'S ROLE IN INVESTMENT COMMITTEES

Jean-François Bérubé: Notes that some AI models are already used to enhance decision-making processes.

David Ouellet: Mentions the use of an AI tool, "ChatPSP", during meetings to enhance discussions by prompting questions and insights. Worth noting, they use the model to challenge themselves and reduce the risk of behavioural biases creeping into their final decisions.

AI IN EDUCATION

Dr. Hossein Kazemi: Discusses the use of a tool that estimates the likelihood of content being AI-generated. He also mentions an innovative approach where students feed textbook information to AI and then have the AI quiz them.

AI IN INVESTMENT RESEARCH

Jean-François Bérubé: Shares that they are developing a tool to assist equity research analysts, especially in reviewing and summarizing earnings call transcripts in large volumes and at a fast pace, which would require considerable resources if performed by employees.

Shawn Edwards: Talks about identifying trends in MD&A topics across different companies.

THE CHALLENGES WITH SPARSE-DATA FIELDS

Dr. Russ Goyenko (McGill): Emphasizes the limitations of large models when working with sparse data, suggesting that relying solely on monthly price data would require over a millennium for effective model training.

Are you applying the Turing Test ? …

OPEN Q&A SUMMARY

The future of Jobs in Finance

Shawn Edwards shares that a significant portion of their news stories were already computer-generated, allowing journalists to delve deeper into stories. He also mentions the evolution of their global data teams from clerks to model trainers. However, he notes that coders using tools like Copilot tend to produce less secure and more error-prone code.

Dr. Hossein Kazemi believes it's premature to predict the exact changes but acknowledges the inevitability of shifts in the job landscape.

Using GenAI for Operational Efficiencies

Jean-François Bérubé states they are adopting GenAI to support operational efficiencies at the Front Office.

Feeding Large Documents to LLMs

Shawn Edwards suggests breaking down large documents into smaller chunks and feeding them to the model iteratively, then combining the outputs.

The Risks of Unsophisticated Investors Using AI

Dr. Hossein Kazemi (CAIA) Asserts that there's a high likelihood of a financial crisis stemming from unsophisticated investors using AI tools. He envisions a scenario where a few dominant AI investment services could inadvertently direct a large number of investors in the same direction, amplifying risks.