Performance and Risk *with* AI, or Performance and Risk *of* AI

More than a choice of prepositions

NORTHFIELD | modeling financial markets worldwide

Dan diBartolomeo

PMAR Conference

May 2024

Introduction

- I'll begin by describing three overriding themes that are the keys to understanding both the potential benefits and failures of AI applications.

- The presentation will next describe the ways in which recent advances in artificial intelligence can be used in the analysis of investment performance and risk, and the methods by which this information can be communicated to relevant parties.
  - We will review early examples of these applications
  - We will review a sample query put to popular AI "large language models" such as ChatGPT, Microsoft Copilot, and Google Gemini.

- The last part of the presentation will discuss what can go wrong in the use of AI methods for the investment practitioners for analysis and communication.

The Fundamental Limitation of AI

- A half century ago in 1974 Amos Tversky (one of the seminal researchers in behavioral economics) had this to say:

"My colleagues study artificial intelligence.

Me, I study natural stupidity."

- To the extent that AI systems are built by imperfect humans, AI systems cannot themselves be foolproof.
  - Unfortunately, many of the early adopters of AI system have proven to be foolish.
  - There is a theoretical upper bound on the "robustness" of AI systems.

The Epistemology of Financial Information

We view the concept of information in three parts

- What we know because it is true by definition (e.g. 2 + 2 = 4)
- What we know because the factual evidence is overwhelming (the Earth is not flat)
- ***What we pretend to know because we don't want to admit we don't know.***

A classic example "pretending" was made obvious in the investigation of the massive Madoff fraud in which I was deeply involved.

- See Nobody Would Listen by Harry Markopolos or Bernie Madoff death: Whistleblower Dan diBartolomeo says lessons remain (afr.com)

The Consequences of Being Wrong Matter

- In the Nichomachean Ethics, Aristotle quotes another writer, Agathon in stating:

Even unto the Gods it is forbidden to change the past.

  - This key point makes clear that our *only purpose in studying the past* is to reduce the likelihood of making bad decisions for the future.

- For the effective use of AI systems, we also need to build in an understanding that different types of errors have different consequences.
  - **Pascal's conjecture on the existence of God** is a clear example.
  - While he could not provide a mathematical proof that God existed, he argued that a sensible person would believe in God anyway, as the consequence of believing in a non-existent God were mild. *The consequence of not believing in God, if God did exist, risked burning in Hell for eternity.*

A Stopped Clock is Right Twice A Day

Traditional Statistical Methods: State a hypothesis describing "linear cause-and-effect", *then* test its validity on empirical data to obtain unbiased estimates of the relationship.

Bayesian Statistical Methods: Augment statistical methods with consideration of error consequences to allow for biased estimates that are less likely to be very wrong.

Expert Systems: Computer systems that rapidly compute a solution to a problem based *on custom built numeric data sets that express the opinion of experts* on how the problem should be solved.

"Data Science" methods start by trying many different statistical procedures on empirical data until strong (but often non-linear) patterns in the data are observed. We then *invent a hypothesis in hindsight* that seems to explain the empirical result.

- See "Psedo-Mathematics and Financial Charlatanism", Notices of the American Mathematical Society (2014), Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance by David H. Bailey, Jonathan Borwein, Marcos Lopez de Prado, Qiji Jim Zhu :: SSRN

NORTHFIELD | modeling financial markets worldwide

The Leap to Unstructured Data

Once the idea that statistical relationships could be highly non-linear, analytical procedures were expanded to include "neural networks" *whose computations try to mimic the animal brains.*

Modern AI systems based very large multi-layer neural networks that have some fraction (e.g. 5%) of the complexity of primate brains.

This amount of "horsepower" allows for computations on unstructured data such as text and images

Natural language processing and "machine learning" of massive data sets (e.g. everything posted on the Internet).

"Generative" AI (e.g. ChatGPT) based on "Large Language Models" can be thought of as a giant-sized effort to "cut and paste" all text sharing words with the question posed into a "tokenized" data set.

The last step is to reorder the all information into what sounds like coherent text using databases of which words are most likely to follow other words (e.g. you don't want three consecutive verbs then eight nouns).

"Small language models" are the same process but built on a tightly supervised dataset (e.g. only articles published in *Journal of Performance Measurement*).

A Direct Example of AI as "Cut and Paste"

- I gave the following question to multiple AI systems (ChatGPT, Microsoft Copilot, Google Gemini):
  - *In 500 words describe the statistical significance of the relative performance of the Fidelity Magellan fund over the past ten years compared to S&P 500 ETFs"*

- This is the response from Google Gemini, but all were **similarly coherent but uninformative.**

Here's a comparison of the Fidelity Magellan Fund (FMAGX) and S&P 500 ETFs over the past ten years (as of May 2024):

- Performance: While FMAGX outperformed the S&P 500 in the first quarter of 2024, it has generally underperformed the S&P 500 over the past ten years.

- * Fees: S&P 500 ETFs typically have lower expense ratios than FMAGX. Lower fees can lead to higher returns over time.
- * Management: FMAGX is an actively managed fund, meaning a fund manager tries to beat the market. S&P 500 ETFs are passively managed, tracking the S&P 500 index.

- Actively managed funds often underperform their benchmarks after accounting for fees. Given FMAGX's historical underperformance and higher fees, S&P 500 ETFs may be a better choice for most investors. However, if you believe FMAGX's manager can outperform the market in the future, you may consider it. It's important to carefully consider your investment goals and risk tolerance before investing in any fund.

Some Commercial AI Applications to "Make My Job Easier"

Factset has announced a text generation system using an AI engine for *portfolio commentary* that describes firm-level attribution of investment performance.

- The "stock picking" aspect of a Brinson-style attribution.
- Includes of brief discussions of published information as to why a particular stock has outperformed or underperformed.
- Includes embedded links to related data tables and reference articles.

Clearwater Analytics uses text analytics in its services (LPx, LPx Clarity, and MLx) to convert prospectus, annual report documents into traditional databases, replacing very expensive manual conversion.

- Related processes convert contract documents for OTC derivatives.
- Once in a dataset investors can generate detailed reports on portfolio composition and performance that are comparable to performance and risk reporting for publicly traded assets.

A Sophisticated Example of AI Use for Active Managers

A new paper by Bai, Boyson, Cao, Liu and Wan (2024) proposes a new measure of information content (Human-AI Differences, HAID) that demonstrates predictive power for stock returns, liquidity, and analyst forecast accuracy.

The HAID method analyzes the words used in the "question and answer period" of corporate earnings conference calls using multiple Large Language Models (ChatGPT, Google Bard, etc.)

The differences between the exact words used by corporate executives and how those words are summarized by LLMs provides insight into how the call content is likely to be interpreted by financial analysts in their reports and recommendations.

This effort is exemplary of what sophisticated hedge funds have been doing for a while, but not talking about.

Conditional Risk Estimation with Text Analytics

The risk of investment portfolios is *always in the future*.

- Grammatically, we are in the subjunctive mood, the world of "what if"

Almost all risk systems focus on the past *while ignoring what we know of the present.*

- Starting in 1995, our risk assessment methods included adjusting the risk estimates daily using changes in option implied volatility. See diBartolomeo and Warrick (2005) for details.

In diBartolomeo, Mitra, and Mitra (*Quantitative Finance*, 2009) we introduce a method to use machine learning analysis of financial news text in place of option data

- ***News is the set of information that tells us how the present is different from the past,*** and hence is a vital element in forming expectations about possible future outcomes.
- Extends the analysis to all global financial assets on which options are not traded.
- Portfolio risk assessments that are adjusted daily for the content of news have been in client production since 2017 in our **Risk Systems That Read®** process.
- We currently analyze about 20,000 articles per day about companies, industries, and countries.

Using AI to Parameterize Expert Systems

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favour one over another
5	Strong importance	Experience and judgment strongly favour one over another
7	Very strong importance	Activity is strongly favoured and its dominance is demonstrated in practice
9	Absolute importance	Importance of one over another affirmed on the highest possible order
2, 4, 6, 8	Intermediate values	Used to represent compromise between the priorities listed above
Reciprocal of above non-zero numbers		If activity i has one of the above non-zero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i

One of the most widely adopted "expert system" methods in decision theory is the Analytic Hierarchy Process put forward by Tom Saaty at University of Pittsburgh in 1980.

AHP has an extensive academic literature in financial applications such as asset allocation (both institutional and retail) and choosing between term and whole life insurance.

- The AHP process defines a problem in terms of a set of multiple-choice questions and answers that are deemed relevant to the required decision.
- The opinions of "experts" are summarized in a series of spreadsheets containing numeric values that algebraically map responses to questions to the relative attractiveness of available choices.
- We are currently updating our process to take input from LLM models to supplement the views of our staff experts.

Here is a fun video (2003) of how we use AHP for retail investor asset allocation.

- https://www.northinfo.com/videos/Northfield_AHP.mp4

NORTHFIELD | modeling financial markets worldwide

Performance and Risks of AI

In its most extreme form, some academic AI experts have made arguments that someday AI will achieve sentience comparable to humans and create the potential for robot machines to rebel against humanity.

While that scenario seems far-fetched for our lifetime, there is no doubt that use of any AI application has the potential to go *very wrong, very quickly.*

Very rigorous processes for the governance of all aspects of AI usage are clearly required.

Several sets of regulations have already been enacted that impact the use of AI by financial institutions.

- Particular regulatory focus on the impact on retail clients.
- GPDR (2018), USNIST (2023), EU (2023), SEC (2023), Biden Executive Order (2023)
- The release of SEC Rule 34-97990 is expected in mid-2024.

Problems of Non-Robustness

As previously mentioned, neural network methods on which AI is based allow for highly non-linear patterns to be observed between variables in a data set.

These complex patterns may be real but may arise from simple coincidence obtained by "overfitting".	Using higher thresholds for statistical significance is a simple step (i.e. T > 3 rather T > 2).	Does your AI output seem consistent with common sense and any applicable economic theory?	*AI systems can "hallucinate" generating text output based on non-existent reference material.*
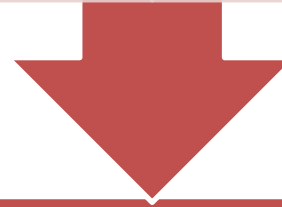
The complexity of non-linear relationships often make AI results "non-explainable", the exact opposite of the intent of investment performance attribution.

NORTHFIELD | modeling financial markets worldwide

A *Network* (1976) Warning About Invalid Training Data

Most training of LLMs is driven by on the volume and popularity of seemingly related content without better evaluation of the credibility of the sources.

Professional news services (e.g. Dow Jones) are probably more credible than a TikTok video espousing "conspiracy theories".

There can be tremendous bias even in professional sources.

Ample evidence that some of the polarization of geopolitics around the world arises from media firms customizing their news coverage for a particular audience. The NY Times did not cover World War One until US troops were in combat.

https://www.bing.com/videos/riverview/relatedvideo?&q=network+i%27m+mad+as+hell+monologue&qpvt=network+i%27m+mad+as+hell+monologue&mid=56BFBE858119B03CA8BA56BFBE858119B03CA8BA&&FORM=VRDGAR

More Resources on AI

- This listing is copied from a presentation by Artificial Intelligence Risk, Inc. where the key founders come from asset management backgrounds. www.aicrisk.com

Online Videos and Classes

Andrew Ng, Deep Learning (Coursera)

Cassie Kozyrkov, Intro to Machine Learning (YouTube)

Cassie Kozyrkov, Data Science for Beginners (YouTube)

Books

Ethical Machines by Reid Blackman (easy read)

The Age of AI by Henry Kissinger et al

Human in the Loop Machine Learning by Robert Monarch

Regulatory Information

https://www.mintz.com/ (sign up at "insights center")

https://commission.europa.eu/law/law-topic_en

https://www.nist.gov/itl/ai-risk-management-framework

Blogs/Podcasts

https://STAYblog.substack.com

https://lennyspodcast.com

https://openai.com/blog

NORTHFIELD | modeling financial markets worldwide

Conclusions

LLM artificial intelligence systems offer the promise of great benefits in terms of making routine operational tasks involving text more efficient.

There are a very limited number of commercial applications using AI methods to enhance the analysis of investment performance and risk.

AI systems have inherent limitations of non-robustness and the massive volume dominance of historical data over forward-looking information.

Please listen regularly to:

Jimmy Buffett - Volcano (Audio) - YouTube

NORTHFIELD | modeling financial markets worldwide