QuantMinds

21st century finance: How will quants lead the way?

Insights into the key trends and disruptive technologies within quant finance

QuantMinds

Contents & introduction

Introduction

With QuantMinds International approaching fast, we at QuantMinds HQ are very excited to be talking to experts to discuss their upcoming presentations and find out more about their latest research.

In this eMagazine, we've selected a wide range of articles our experts provided for us to cover the key change-makers in the industry, starting with what's on everyone's mind: AI. Ioana Boier, Ph.D. shares her extensive experience in the field and looks at what the future holds in this arena. Speaking of the future, there are still a lot of unanswered questions regarding the IBOR transition, and Maurizio Garro took the lead to guide us through the next considerations to take into account. Meanwhile, ESG investing is in, but will it push smart beta strategies out? Hamza Bahaji takes on the challenge of innovating smart beta with ESG principles. And as life as a quant gets more complicated, Uwe Naumann seeks to simplify it with his take on differentiable scripting for computational finance. Finally, we hear from Alexander Lipton (curtesy of the *Quantitative Finance* journal), who tells us how the current financial system is being disrupted by cryptocurrencies' growing popularity.

We hope you'll find the insights you're looking for from our contributing experts. Thank you for reading and we hope to see you in December.

The QuantMinds team

Contents

QuantMinds

AI challenges in discretionary trading

Ioana Boier, Ph.D.

As AI and cryptocurrencies launch the finance sector into the realm of bold 21st century tech experiments, quant researcher Ioana Boier explores the impact of artificial intelligence on discretionary trading practices, and asks key questions about the future of AI applications in this sector.

As a computer science researcher with more than a decade of experience in quantitative finance, I often find myself reflecting on the relationship between computer science and financial modelling. Of course, computers have long been used to carry out financial computations and simulations, but through a recent paradigm shift, catalysed by machine learning (ML), computer science has been brought to the forefront in finance. ML has enabled us to move from "direct programming" based on explicit sets of instructions on how to do something to "indirect programming" or learning from data. For quant finance, this means shifting the focus from models governed by explicitly prescribed laws devised by quants to models with laws that are implicit in the patterns discovered by algorithms from data.

A long time in the making, this shift has mostly been confined to the secretive corners of systematic hedge funds that hired computer science researchers with backgrounds seemingly unrelated to finance such as signal processing, speech recognition, and cryptography. In a 2000 interview with Institutional Investor Magazine, Jim Simons spelled out the difference that has ignited this remarkable movement: "We don't start with models. We start with data. We don't have any preconceived notions."

Today, AI and cryptocurrencies have taken finance into the realm of bold 21st century tech experiments. For quants like myself, computer scientists once deemed "unicorns" on teams otherwise composed of mathematicians, physicists, and aconomists, it is exciting to witness how this shift to a data-driven approach is extending beyond high-frequency and systematic trading, to all styles of investing. I am particularly fascinated by the prospect of combining human and machine intelligence in support of discretionary trading. Because of its subjective nature, this style of investment that relies heavily on human experience and decision-making, may seem the least likely to benefit from AI. Yet, as Nassim Taleb points out in "Fooled by Randomness", people cannot make decisions without emotion. He goes on to say that emotions are, in fact, the shortcuts required to avoid lengthy mental optimisations over a large set of variables to arrive to decisions. What if AI could supplement the human brain in ways that help reduce reliance on emotions during decision making? Would this be feasible and if so, what are the main challenges to be addressed?

AI at present time mixes two main ingredients: data and models. Let's explore them next from the perspective of supporting human decisions.

Data

A data-driven approach entails collecting as much data as possible and feeding it to computer algorithms, i.e., models, that are *trained* on this data to identify patterns as well as recognise and predict their occurrence when new data comes along. Before any training can occur, however, the data not only needs to be collected, but it must also be cleaned and processed into a format suitable for use by ML algorithms.

Financial markets have traditionally been a rich source of numerical data: stock prices, interest rates, option volatilities, econometric measures of growth, inflation, etc. Traders have combined these with their own intuition, knowledge, and experience accumulated over time or by reading articles, speaking to various experts, and listening to the news. As the amount of information has become too overwhelming to keep track of in one's mind, and as ML techniques have surged in popularity, the notion of *alternative data* has made its way into the quant lexicon. Rather than listening to countless hours of news or reading a myriad pages of speech transcripts, corporate announcements, regulatory filings, or scouring social media, and attending industry events, there is now hope that valuable information can be extracted and distilled into manageable, essential bits automatically. Besides text and speech in natural language, alternative data examples include images, geolocation information, and sensor data. Using ML to process this type of data not only saves time but, if done well, it is likely to capture patterns and nuances that would otherwise escape human analysts. Nonetheless, converting alternative data into alpha is fraught with challenges such as finding the right type and amount of preprocessing before feeding it to ML models, lack of standardisation, and being able to search, retrieve, measure, and efficiently navigate ever increasing piles of it.

There is one antithetical aspect of financial data: it can be enormous and sparse at the same time. The enormous part is obvious; we just discussed a plethora of sources producing data at an unprecedented pace and scale. The sparsity is more nuanced, and it becomes salient when one tries to apply ML techniques developed in non-financial contexts to solve financial problems. The success of state-of-the-art methods largely depends on availability of large training sets containing millions of samples where many are "alike" to facilitate the learning of patterns. Macro financial measurements are rarely in the millions and not often “alike”. Datasets of daily data (i.e., measured once a day, like the closing price of a stock) go back at best several decades rendering the size of a full dataset into the thousands, instead of the millions. Moreover, among these thousands of samples, only a few, if any, are illustrative of special occurrences, aptly referred to as black swans or rare events. Yet, the juiciest chunks of alpha concentrate around such events, and finding them before anyone else does is what the investment craft is all about.

This brings us back to the humans versus the machines: people, children even, can learn from a few examples, but machines are not fully there yet. *Zero-shot* and *few-shot* ML tackle this issue, but such methods are yet to mature. In the meantime, *synthetic data generation* aims to fill the gap by augmenting the inputs to data-hungry learning algorithms. In principle, data simulated according to a prescribed dynamics or distribution can be made arbitrarily large. The crux of the problem is to ensure such distributions are as close as possible to the unknown empirical ones. This can become a chicken-and-egg problem as generative models meant to capture data distributions require large training sets to subsequently be able to generate even larger ones.

> “Converting alternative data into alpha is fraught with challenges such as finding the right type and amount of preprocessing before feeding it to ML models, lack of standardisation, and being able to search, retrieve, measure, and efficiently navigate ever increasing piles of it.

Models

It’s a standing stereotype that the most vexing trading style for a quant to tend to is that of buying and selling financial assets “like potatoes”. Typically, this means that the trader relies mostly on her views and intuition, and less on quant models and their predictions. Nonetheless, when studied up close, these are exactly the discretionary patterns that are most fascinating to observe and challenging to support with useful models. Bringing AI into this context raises important questions of explainability, causality, and responsibility in high-stakes decision making. Let’s consider them one by one.

Explainability

The text completion function on our mobile phones has become uncannily good at predicting the next word that would best complete a message in progress. Unlike words that we may deem as ill-fitting in a given context, financial predictions churned by black box algorithms may not be as intuitive to debug by us humans. This impedes the establishment of trust, especially in early stages of developing new algorithms. We accept that human intuition is itself not always explainable, knowing that there is a substantial amount of activity in our unconscious brains that contributes to the formation of our “gut feelings". We value and often make decisions based on intuition, which we view as our superpower. The question is how can we harness AI into a complementary superpower that we would trust to an equal extent? How much explainability and in what form is required to get us to that point?

Causality

Good decisions can be made for the wrong reasons. For instance, in computer vision, images may be correctly tagged by algorithms, without a true understanding of their contents. A well-known example is that of wolves being correctly distinguished from dogs based on the presence of snow in the training images of wolves. Most likely, such a system would mis-classify an image of a dog in winter or that of a wolf in a summer landscape because the connections inferred during training did not disentangle the true differentiating features. In finance, a market move could be correctly predicted based on spurious or incorrect associations in the data. When these associations break, disaster may ensue as we've seen happen more than once.

Can AI do better? Machine learners today are narrowly focused on correlations and are far from being able to capture causal structures. As David Hume put it in his Treatise of Human Nature, causation is crucial as it allows us to imagine and do things we've never seen before. In other words, we have the ability to generalise through causal reasoning.

> "I believe that the next generation of ML techniques used in finance must get better at finding causal links instead of mere correlations to be useful in practice.

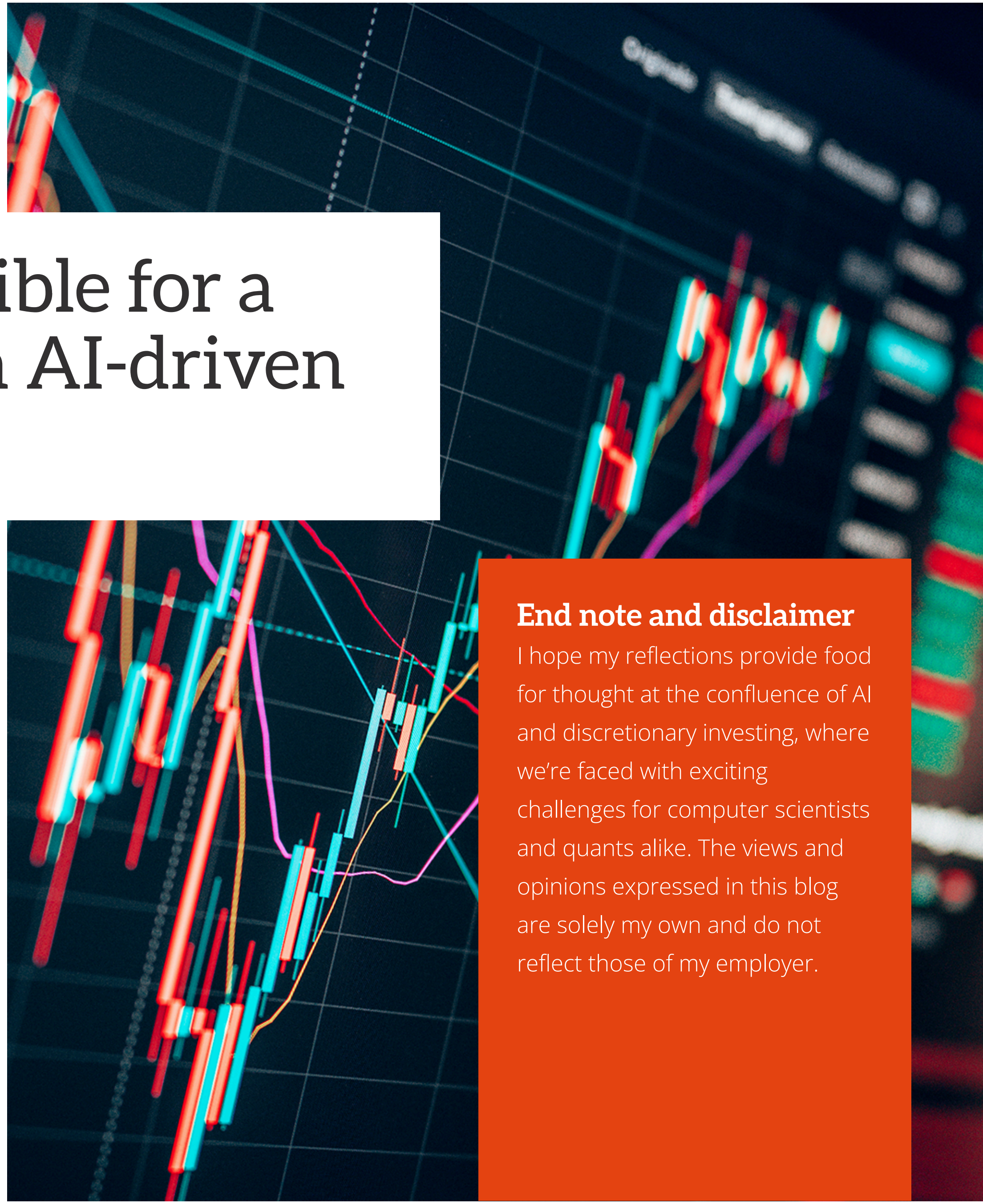
Responsibility

Who should be responsible for the choice of input data being used to train models? Should we leave it to humans or to machines to figure it out? To what extent should it be one or the other?

The assignment of responsibility question in AI arises in several ways, two of which I address here. The first is closely tied to the notion of feature engineering and the boundary between data and models: who should be responsible for the choice of input data being used to train models? Should we leave it to humans or to machines to figure it out? To what extent should it be one or the other? This drives the dialectic of shallow versus deep models which is neither new, nor specific to AI: a linear regression model could be viewed as shallow, but the work required to set up the regressors in a meaningful way could be very complex. I am not referring to regressing one given variable (e.g., housing prices) on another variable (e.g., square footage), but rather about having to find the right regressors (e.g., square footage doesn't account for the age of the house or improvements made), with the best predictive power over time. This type of feature engineering can be rather complex in the case of financial markets, where many interactions and events can contribute to any one data point. So, finding the set of variables that really matter can be like finding the proverbial needle in a haystack. Hence, it is not unreasonable to ponder the alternative: dumping all available data into a deep model and letting it do the combined work of learning the relevant features and using them to solve the problem at hand. Of course, this solution is not only hampered by the need for lots of training data, but also by the lack of standardisation across dimensions and data sources.

Who should be held responsible for a disastrous outcome due to an AI-driven or AI-made decision?

The second aspect is closer to the traditional view of responsibility and pertains to who should be held responsible for a disastrous outcome due to an AI-driven or AI-made decision? In investment finance, this comes down to loss of fortunes. In more general scenarios, loss of opportunity (such as being denied a loan due to hidden bias in the models), sustaining injury or even loss of life (in the case of medical diagnostics or self-driving vehicles) have sparked debate on this topic. A major takeaway so far has been that the burden of proof weighs heavier on AI-driven decisions than it is on those made by humans, despite statistical evidence that AI-caused accidents occur less frequently than accidents caused by humans. So how is progress to be made? In my opinion, we're back to establishing trust so that 'human plus machine' can do better than either of them alone.



End note and disclaimer

I hope my reflections provide food for thought at the confluence of AI and discretionary investing, where we're faced with exciting challenges for computer scientists and quants alike. The views and opinions expressed in this blog are solely my own and do not reflect those of my employer.

QuantMinds

Do you want to be involved in a future QuantMinds event or eMagazine?

Get in touch for speaking opportunities:
Sophie.Milton@informa.com
Get in touch for sponsorship opportunities:
Josef.Lanjri@informa.com