Will AI take over active fund management?

AUTHOR:
Dr. Gareth Witten FMR., *Executive-in-Residence, Global Risk Institute*

GLOBAL RISK INSTITUTE

TECHNOLOGY DISRUPTION & INNOVATIONS

The invasion of artificial intelligence algorithms in many different fields from finance and autonomous cars to diagnosing and recommending treatment plans have dominated the public media and academic literature (see The Economist Special Report, 25 June 2016[1]). The number of conference papers or journals on artificial intelligence or machine learning has increased exponentially[2] and computer science academic papers and studies in artificial intelligence increased nine times between 1996 and 2015[3]. Furthermore, the entire academic database, Scopus, owned by the publishing house Elsevier, contained over 200,000 papers in the field of computer science that was indexed with the key term "artificial intelligence". Despite this increased interest in artificial intelligence by industry and academic experts over the past few years, some are beginning to question whether this trend is just another hype that will end in a whimper or whether this trend will have the stamina to transform certain industries entirely. Andrew McAfee and Eric Brynjolfsson from MIT share the latter view. In their book published in June 2017, Machine, Platform, Crowd: harnessing Our Digital Future[4], they present an optimistic view of how emergent technologies, including AI, are having a huge impact on our daily lives and careers. However, in a recently published article in the Harvard Business Review (Artificial Intelligence for the Real World[5]), Davenport and Ronanki surveyed 152 organizations regarding their artificial intelligence projects and showed that the projects started out with extremely ambitious goals but ended with cost over-runs and significantly less ambitious outcomes. The authors categorized the completed projects within these organizations into 3 types: automating business processes, gaining insight through data analytics and engaging with customers and employees. These were far from the lofty goals they started with. Furthermore, there are reports asking whether the gap between the theoretical underpinnings and implementation of AI is just too big for AI to make a further meaningful impact in the real-world. For example, in an article (Is AI Riding a One-Trick Pony?[6]) published in MIT Technology Review, the author questioned whether current AI applications are relying on three-decade old theory and whether this will limit the scope of innovation in multiple applications going forward. At a recent conference (Are We Ready For The Next Financial Crisis?) co-hosted by the Rotman School of Management and the Global Risk Institute in Financial Services, there were mixed views among professionals and academics on how technology, from cryptocurrency to artificial intelligence, will transform financial services.

The financial services industry, including investment management, have also experienced a dramatic increase in the use of artificial intelligence. The applications range from the development of robo-advisors, that attempts

1 *"The return of the machinery question. Special Report on Artificial Intelligence" The Economist June 25th, 2016. pg 3-16.*

2 *Antonio Regalado, "Google's AI Explosion in One Chart", MIT Technology Review, (Mar 25, 2017).*

3 *Alex Gray, "These charts will change how you see the rise of artificial intelligence", World Economic Forum, (Dec 18, 2017).*

4 *Andrew McAfee and Erik Brynjolfsson. "Machine, Platform, Crowd: Harnessing Our Digital Future" W. W. Norton & Company: 1 edition (Jun 27, 2017).*

5 *Thomas H. Davenport and Rajeev Ronanki, "Artificial Intelligence for the Real World" Harvard Business Review, January-February 2018.*

6 *James Somers. "Is AI Riding a One-Trick Pony?" MIT Technology Review. Nov/Dec 2017.*

to individualize the asset allocation decision, to the use of AI in portfolio construction and stock picking. There are mixed results at this early stage, however, a more cautious approach is suggested following the performance of a sample of these funds during the market correction of February. The Eurekahedge® index of funds, created in 2011, that use AI in investment management showed that these funds came horribly short during the first major equity market correction in 2 years. The AI index fell 7.3% in February compared to a 2.4% decline for the broader Hedge Fund Research Index in the US. This is a far cry from media headlines in 2017, for example, "Blackrock's Fink looks to Invest "better than humans"". To be fair to Blackrock's CEO, Larry Fink, he did say that pure AI-driven investing "is more of a myth than a reality". This illustrates the media's obsession with talking-up the role of AI in investing, which is far from the practical realities of the performance of these funds during the market correction. Any good investment analyst will profess that one of the biggest challenges when estimating the intrinsic value of a company (or other asset) is understanding the difference between what the market perceives the company to be worth (e.g., its stock price) and what its fundamental/intrinsic value (present value of future free cash flows) actually is. This difference is referred to as the margin of safety by Warren Buffett, which may be viewed as the risk of an investment: if the intrinsic value is below the stock price then the risk increases as the stock price moves closer to the intrinsic value (for a buy-and-hold strategy). The same can be said about the practical realities vs. the market's perceived ability for technology that use AI.

What I will do in the next section is to explain why this large difference between perception and reality exists: the last decade saw the shift from individual, sophisticated AI algorithms to ensembles of different algorithms. Although this shift has been significant and should not be underestimated, I will argue that AI only solves particular problem-types, which does not include true active fund management. A similar sentiment was echoed in a report[7] published in 2011 that attempted to represent the relevant financial network for systemic events and risk. The report argued that political and social networks may emerge to play a larger role in liquidity transactions and/or in the spread of rumours, which can ultimately influence market fear and greed and hence consensus valuation of markets.

What has driven the hype in AI?

The past decade or so has seen a significant shift from the application of sophisticated single algorithms to ensembles of diverse predictors. For example[8], in the past if one wanted to translate, say, French to English, then a single algorithm will be deployed that would use built-in rules of syntax and grammar. Computer scientists call these individual classifiers weak learners, which range from relatively sophisticated neural networks to simple heuristics or decision rules used in random forest algorithms[9]. More recently, the state-of-the-art Google Translate® software relies on ensembles of predictors that scan the internet for words, phrases, sentences and "learn" patterns. This results in a more natural sentence structure than the single sophisticated algorithms that often served as good entertainment. What is important to note is that for the ensemble as a whole to predict accurately, the classifiers within the ensemble must differ[10] – they must look at a unique feature set or assume different relationships between features. A natural question to ask is whether there is a general mathematics that can capture this logic. In other words, will it always be the case that when algorithms/predictors are combined that they will result in more accurate predictions?

7 *Kambhu, J., Weidman, S. & Krishnan, N. "New Directions for Understanding Systemic Risk" (National Academies Press, Washington DC, 2007). Also published as Econ. Policy Review. 13(2) 2007.*

8 *Scott E. Page, "The Diversity Bonus" (Princeton University Press, 2017).*

9 *Leo Breiman, "Random Forests", University of California, Berkeley.*

10 *Dietterich. T. G. "Ensembles in Machine Learning", in "First International Workshop on Multiple Classifier systems", edited by J. Kittler and F. Roli, (New York: Springer Verlag, 2000), 1-15.*

This concept is called the "wisdom of crowds" effect, named after James Surowiecki's book, The Wisdom of Crowds[11]. The idea is that a crowd of people make predictions that are more accurate than the individuals in the crowd. An example of the wisdom of crowds effect from his book is based on data of the weight of cattle collected by Sir Francis Galton from the West of England Fat Stock and Poultry Exhibition in 1906. Collectively, 787 people at the exhibition estimated the weight of a steer almost exactly. There are many other examples where this concept was found to work; groups tend to be more accurate predictors than individuals in predictive tasks. However, we do see that crowds, under certain conditions, can be "mad" as well. Examples of this include the many financial crises that we have experienced and will continue to experience in the future. In order to understand the conditions for the wisdom of crowds effect to hold, Scott Page from the University of Michigan came up with an equation to show how the wisdom of crowds work. He called the equation that describe this effect, the "diversity prediction theorem"[12].

Collective error = Avg. individual error – Prediction diversity

One can think of the "Avg. individual error" as smarts or talent, the "prediction diversity" as diversity and the collective error as the wisdom of crowds. There are two things to notice: firstly, this is a theorem and we will see how the equation will help us see how and why crowds make accurate prediction. Secondly, this theorem also pertains to AI algorithms, that is, for a collection of AI algorithms to be "wise" (better predictors) depends on "talent" and "diversity" and we will show how this equation can explain the current hype around AI. What Scott Page shows mathematically in his book, The Difference, is that the collective accuracy of a crowd depends in equal measure on the accuracy of its members and on their diversity. Moreover, a diverse crowd will always be more accurate than its average member. What this means for AI algorithms is that if we can combine "diverse" algorithms then ensembles will be better predictors than the single sophisticated algorithms initially used to translate French to English. For those of us who participate in the markets, this message suggests appropriate humility.

The collective accuracy of an ensemble of AI algorithms depends on the accuracy of each predictor and their diversity. How does this work in our translation example? Let's begin with random forest algorithms that have a collection of random decision tree predictors. If a sample of data is used to determine accuracy then the algorithm separates out the accurate decision trees. Furthermore, the predictors are trained on subsets of data, which guarantees accuracy ("talent"). This is not so easily applied to people – we can't easily eliminate inaccurate people – therefore, AI algorithms have a built-in accuracy advantage. These algorithms also build in diversity, the second component required for the ensemble of predictors to be "wise". This is done in what is called "bagging" and "boosting": "bagging" trains predictors on randomly drawn subsets of data and determines the size of those subsets. In essence, this gives predictors the ability to "learn" from different data. Bagging ensures diversity by adding predictors that are accurate when the ensemble makes mistakes. For example, let's say you train an AI algorithm to detect fraudulent transactions. Say you have 100,000 transactions and we have 40,000 cases for which we know the outcome. We might then generate a set of random predictors and only keep those predictors that classified more than, say, 60% of the 40,000 cases correctly. We then have all the predictors collectively classify the 100,000 transactions and what we might typically find is that the predictors correctly predict about, say, 80,000 transactions. The remaining 20,000 transactions can then be classified as the challenging cases. In order to perform boosting, a new set of random predictors are created, and those predictors are trained on the challenging cases. Adding the predictors that classify the challenging cases (more than 50% of the time) to the original predictors creates a diverse ensemble of predictors that leads to better accuracy. It's the "wisdom of crowds" effect that explains why AI has become so accurate at solving problems such as facial recognition, fraud detection etc.

11 *James Surowiecki, "The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations" (New York: Doubleday and Company, 2004).*

12 *Scott E. Page, "The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies" (Princeton University Press, 2008).*

What is active fund management?

In the previous section we explained what we believe is driving the hype in AI. What we will see in this section is that the types of problems that AI algorithms can solve are complicated but far from the complexity of the real world. In particular, financial markets are complex adaptive systems[13] and being an active fund manager means believing in both market efficiency and inefficiency (just not at the same time!). The 2013 the Noble Prize in Economics was awarded to Robert Shiller from Yale University (a proponent for market inefficiency) and Fama and Hansen from the University of Chicago (proponents for market efficiency). If markets are perfectly efficient then there is no reason to try to beat the market through active management. Most of the time, the market prices stocks correctly (wisdom of crowd effect) but we know that there are times when the market acts in a manic way that leads to market crashes. In the previous section we presented a theory that gave two conditions for a crowd to be "wise" and we showed how this is applied to translation algorithms that use AI. However, its not that simple for this to apply to crowds of investors in a market, as you can imagine. Let's firstly examine theories to explain "wise" crowds (market efficiency) and then look at how these conditions may be violated and result in "mad" crowds. Andrei Shleifer described 3 basic theories to explain market efficiency[14]. According to Michael Mauboussin[15], the first theory assumes that investors are rational, which means that they make normatively acceptable choices based on expected utility theory and correctly update their views based on new information. The second is to assume a small set of rational investors who use arbitrage to remove pricing errors. The final theory is the wisdom of crowds, the idea that a group of diverse individuals can come up with an efficient price. Market inefficiency occurs when the conditions for market efficiency is violated but this is practically not that straightforward to determine because the market is a complex adaptive system (composed of diverse, interconnected components/agents that are adapting). Examples of complex systems are pervasive: the immune system consists of many different immune cells that are interconnected, diverse and adapt to various threats; ant colonies consist of different types of diverse functions that are interconnected and adapt to an environment and our brains consists of many different neurons that are interconnected and adapt. Research of many different complex adaptive systems reveal that they do have common attributes, for example, they are inherently unpredictable, exhibit fat tail events and display emergent behaviour where the sum of the parts is greater than each of the underlying components/agents. An example of emergent phenomenon is consciousness from the network of neurons in the brain or various crises/bubbles that we see in financial markets. "Tipping points", "thresholds" and "regime shifts" are all terms that have been used in the media to describe the flip of a complex adaptive system from one state to another.

Let's think of an arbitrageur who wants to buy an asset cheap and enjoys excess return and then delivers a proper price[15]. Markets must be inefficient enough to encourage active managers to participate and at the same time the participation of active managers creates efficiency. Active investor also creates liquidity in the market. One crucial factor (or lack thereof) is diversity (correlated believes and models) that results in market inefficiency. This, in of itself, is what makes exploiting that inefficiency challenging. AI algorithms work well when markets are efficient but comes short when market inefficiencies occur. This follows from another result from studies in complex adaptive systems that is both intuitive and profound. Stephen Wolfram[16] came up with an explanation for what causes fundamental randomness: he showed using cellular-automata model that certain interdependent rules can result in randomness. That means that you don't need random event after random event, just a complex adaptive system (with diverse, interconnected and adapting agents) that have interdependent rules. This implies that for a complex adaptive system like financial markets the distribution of outcomes today may not be the same as the distribution

13 *John H. Miller and Scott E. Page, "Complex Adaptive Systems: An Introduction to Computational Models of Social Life" (Princeton, NJ: Princeton University Press, 2007).*

14 *Andrei Shleifer, "Inefficient Markets: An Introduction to Behavioral Finance" (Oxford, UK: Oxford University Press, 2000).*

15 *Michael Mauboussin, "Animating Mr. Market", Credit Suisse, (2015).*

16 *Stephen Wolfram, "A New Kind of Science", (2002).*

of outcomes in the future (called nonstationarity). If we therefore assume stationarity in a nonstationary world then we will be in for a shock. AI algorithms and robo-advisors, for example, typically make these assumptions. Furthermore, to get a better sense of the distribution of data to use as input into AI algorithms requires more data than expected but also some sense-making of the context of the data (these are necessary conditions but not sufficient).

Three questions to ask of your algorithm (including AI) in your investment process?

Whether humans or AI algorithms will do better at a particular investing task depend on many factors; including whether the task being automated is familiar or unique, whether there is sufficient data and of sufficient quality to both verify and validate relationships and whether the task it is attempting to automate is complicated or complex. Let's examine each question and then provide a matrix of tasks with examples in investments that can best be solved with humans or without humans. This matrix is meant to serve as a decision-making guideline for any executive who wants to implement an AI project in or across his business.

ARE THE TASKS FAMILIAR OR UNIQUE?

When a task is routine/familiar it is straightforward to see that such tasks can be automated using appropriate algorithms. Whether there is little or lots of good data available, computers can mine the data sets for appropriate relationships. For example, a fund tracking a market index could easily balance the task of rebalancing and minimizing the tracking error between a fund and an index. These funds are typically called index trackers. These tasks may be complicated (we will distinguish between complex and complicated tasks later) but can be solved and can be automated. These automation algorithms work in all "states of the world". In the example above, tracking an index happens when the market is up or down. In contrast, unique tasks are decision-making processes that are unique to either a particular state of the world or over all states. For example, a smart beta strategy is a unique tilt on certain factors over all states of the world. These strategies attempt to harvest particular risk premia over time. Another example of a unique strategy is a tilt (for whatever reason) away, say, from US tech firms at a particular point in time. This strategy may only work over specific states of the world. Notice that the latter strategy has some element of timing/insight whereas the former strategy takes into account the potential benefits over all states.

DO YOU HAVE SUFFICIENT GOOD QUALITY DATA?

For tasks where lots of good quality data is available, algorithms that do routine tasks will likely do better than humans. However, no amount of good quality data may provide some perspective on a unique task (e.g., arbitrage opportunity) over a particular state of the world. In this situation, humans provide broader "contextual intelligence" than AI algorithms. Relationships between existing long-term metrics may change leaving some strategies perplexed and managers losing out. This may provide some insight for the task but it may be mostly luck that rewards or punishes the bet. We may, however, have a unique task over all states of the world where lots of good quality data help provide some insight into the task. For example, if there is persistence of a risk premium (e.g., Fama-French factors) that delivers alpha over a long time period.

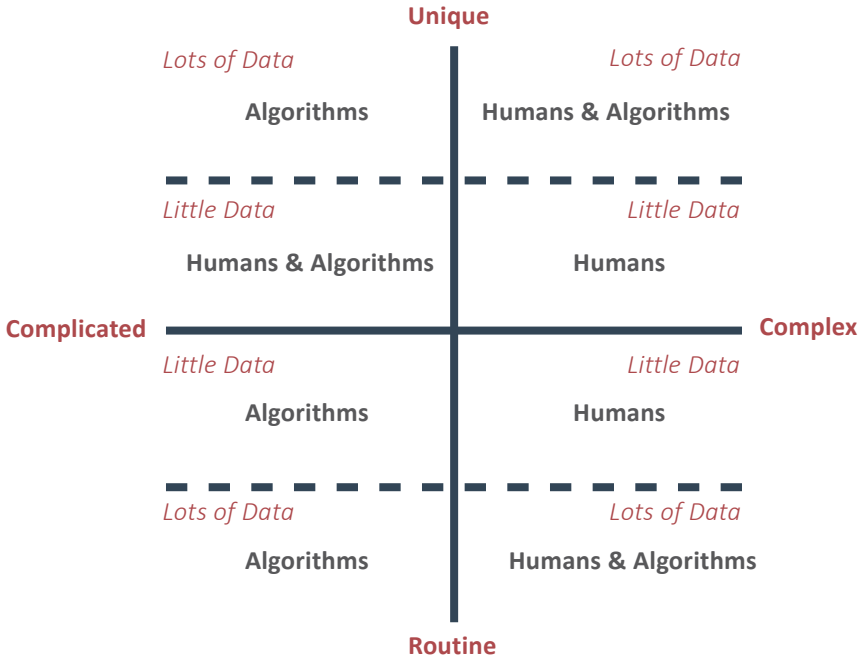
IS THE TASK COMPLICATED OR COMPLEX?

Most tasks we encounter we think is complex but it is actually complicated. By complicated we mean a task that may be extremely difficult to solve or take a long time to solve because the choices in the task interact18. For example, renovating your house by moving a doorway will impact other choices like covering-up a window or affecting the electrical system etc. When the task has many choices and many combinations of choices and these choices interact, then the task is complicated. Darwinian evolution is an example of a complicated task. In contrast, a complex task is when a system contains interconnected, interdependent, diverse and adapting agents where what

one agent does impacts the utility of another agent. For example, firms in a market operate in a complex system where the introduction of a product line by one firm may impact what other firms do. The same is true for active and passive investments. AI algorithms may be able to harvest risk premia or find the "optimal" asset allocation but we know that risk premia do not persist and what is "optimal" today may not be tomorrow. We also know that certain managers who proclaim to be active managers may be closet indexers.

Let's now introduce a matrix that will provide a framework to help understand which tasks require algorithms (including AI) and which tasks require human judgment.

in a complex system, human interventions are required to make contextual sense of the changes. In the extreme opposite quadrant, having little or lots of data will be the domain where most algorithms will thrive above humans. The upper left quadrant is for tasks that are unique but complicated and this quadrant may be the domain of some algorithms. For example, trading algorithms in a bank that exploit very specific arbitrage opportunities. However, having little data may require both human judgment and AI algorithms. The bottom right quadrant is for tasks that are routine and complex. These tasks may be investment strategies that exploit behavioural biases in the market. Exploiting the "predictable irrationality" behaviour of the market, according to Daniel Kahneman,



The matrix above has "Unique" vs. "Routine" tasks on the y-axis and "Complicated" vs. "Complex" tasks on the x-axis. We then divide each quadrant into two: "Little" or "Lots of Data" where either little or lots of good quality data is available. Note that we are assuming here that the data is of good quality - that may not be the case in the real world. When tasks have little data, the task is unique and complex then humans will perform better. If there are lots of data then algorithms may "learn" patterns but when those patterns change, like they inevitably do

may require lots of good data. Little data may require both human judgment and algorithm intervention.

What the matrix above provides is a classification of tasks that may require only humans, only AI algorithms or both humans and AI algorithms. It can be used by any executive to determine whether to balance the cost of an AI project with the potential impact on her business. This will help minimize, as described by the Davenport and Ronanki survey above, the cost on lofty AI projects that may have

little or no business impact. Schoemaker and Tetlock[17] mentioned that "AI still lacks a broad intelligence of the kind humans have that can cut across domains. Human experts thus remain important whenever contextual intelligence, creativity, or broad knowledge of the world is needed." They argue that when tasks are familiar and there are lots of data available then computers are likely to be better than humans. However, when tasks are unique and when data overload is not an issue for humans, then humans will likely be better than computers. In this paper we went slightly further to provide a matrix that partition contexts when AI will perform better than humans and when human judgement will still be required. True successful active fund management will still be the primary domain of humans in the future.

17 *P.J.H. Schoemaker and P.E. Tetlock, "Building a More Intelligent Enterprise", MIT Sloan Management Review, (Spring 2017).*

© 2020 Global Risk Institute in Financial Services (GRI) . This "Will AI take over active fund management" is a publication of GRI. This "Will AI take over active fund management" is available at www.globalriskinstute.org. Permission is hereby granted to reprint the "Will AI take over active fund management" on the following conditions: the content is not altered or edited in any way and proper attribution of the author and GRI is displayed in any reproduction. All other rights reserved.