In Machines We Trust: Are Robo-Advisers More Trustworthy Than Human Financial Advisers?

Hui Chia

Melbourne Law School, Melbourne, Australia

Abstract

This article examines the use of artificial intelligence (AI) and deep learning, specifically, to create financial robo-advisers. These machines have the potential to be perfectly honest fiduciaries, acting in their client's best interests without conflicting self-interest or greed, unlike their human counterparts. However, the application of AI technology to create financial robo-advisers is not without risk. This article will focus on the unique risks posed by deep learning technology. One of the main fears regarding deep learning is that it is a "black box", its decision-making process is opaque and not open to scrutiny even by the people who developed it. This poses a significant challenge to financial regulators, whom would not be able to examine the underlying rationale and rules of the robo-adviser to determine its safety for public use.

The rise of deep learning has been met with calls for 'explainability' of how deep learning agents make their decisions. This paper argues that greater explainability can be achieved by describing the 'personality' of deep learning robo-advisers, and further proposes a framework for describing the parameters of the deep learning model using concepts that can be readily understood by people without technical expertise. This regards whether the robo-adviser is 'greedy', 'selfish' or 'prudent'. Greater understanding will enable regulators and consumers to better judge the safety and suitability of deep learning financial robo-advisers.

Keywords: Robo-adviser; financial advice; artificial intelligence; deep learning; Explainable AI.

Introduction

Greed—an issue that could be overcome with more financial regulation or an inescapable weakness of the human condition? Since the global financial crisis, the issue of perverse incentives has been a source of public anger and distrust towards the banking and financial investment sector.¹ Instead of being rewarded for acting in the client's best interests, as they are legally bound to do, financial advisers are rewarded when they sell financial products to clients, even if those financial products are not in the client's best interests.² Proposed solutions to the problem of perverse incentives for financial advisers have focused on changing the structure of financial incentives for human agents (e.g., limiting the use of incentive payments for advisers).⁴ This paper takes a decidedly different approach to the problem by posing the question, if we cannot trust human financial advisers to act in their client's best interests, should we trust a machine instead?

⁴ Stanford, "Sectoral Collective Agreement Could Combat Finance Industry Misconduct."



This work is licensed under a <u>Creative Commons Attribution 4.0 International Licence</u>. As an open access journal, articles are free to use with proper attribution. ISSN: 2652-4074 (Online)

¹ Zammit-Lucia, "Misaligned Incentives."

² Burke, "Impacts of Conflicts of Interest in the Financial Services Industry."

³ Beyer, "Do Financial Advisor Commissions Distort Client Choice?"

Unlike humans, a machine has no interest in material wealth, nor does it feel greed or temptation. A machine does not need to pay its mortgage or save for a vacation. A machine can be programmed such that it is 'rewarded' without the use of any financial resources at all.⁵ Thus, an artificially intelligent (AI) agent is an excellent candidate to perform a job that requires acting in someone else's interests, when there is no financial incentive to do so. Until recently, machines have been used to automate mostly manual tasks, but we are now at the dawn of a new era in AI technology⁶—a so-called fourth industrial revolution that will reshape every industry, including the financial sector.⁷ At the forefront is 'deep learning', the technology responsible for the recent leap in AI capabilities.⁸ Deep learning has shown remarkable progress in meeting or surpassing human-level performance in tasks typically thought to require human intelligence. The method is already used for the diagnosis of cancers, predictions of suitability for parole¹⁰ and many other high-stakes functions.

The great strength of deep learning lies in its ability to digest vast amounts of data that are used to identify patterns and make predictions. This is particularly well suited to the task of giving financial advice, with access to oceans of financial data and the ability to make informed estimates being a vital function of financial asset management. Already, several financial service firms are offering services using deep learning technology. With big investment companies in the research phase of developing robo-advisers, it is likely that such methods will become widely available in the next few years. Robo-advisers are expected to be available at a fraction of the cost of their human counterparts, making financial advice services accessible to low- and middle-income individuals. Along with the excitement and hype, there is also controversy and fear surrounding deep learning technology. One of the main criticisms of deep learning is that the technology is a "black box": no one knows or can explain exactly how deep learning agents arrive at their decisions. Even the developers who create these mechanisms cannot explain exactly how they work—but one thing is certain: it works with almost unnerving accuracy. As such, if we are to hand over decisions as important as financial investment to an AI agent, we need to know if we can trust the technology. This paper explores the question of whether we can trust deep learning agents to be better financial advisers than humans. Part 1 provides a primer on AI and delves into the unique characteristics of deep learning technology. Part 2 outlines current approaches to the regulation of robo-advisers, and Part 3 proposes how regulators and lawmakers might tackle the challenges of regulating deep learning robo-advisers.

Part 1: What Is 'Deep Learning'?

State Of Technology

The term 'robo-adviser' is loosely defined as being algorithms that automatically generate financial advice without direct human involvement. Across the sector, algorithms are commonly used in finance for asset allocation, portfolio management and robo-advisers that serve as chatbots, capable of answering common questions from customers. Robo-advisers that are widely used in the financial industry are mostly algorithms that are based on decision trees or heuristics and have been developed based on human knowledge. This technology has been referred to as 'old-style' AI, which is being outperformed by newer AI techniques based on deep learning. Page 19.

⁵ Sutton, Introduction to Reinforcement Learning.

⁶ McAfee, "The Dawn of the Age of Artificial Intelligence."

⁷ Marr, "The 4th Industrial Revolution is Here."

⁸ Parloff, "Why Deep Learning is Suddenly Changing Your Life."

⁹ Tibbetts, "The Frontiers of Artificial Intelligence."

¹⁰ Teich, "Management AI."

¹¹ Zhang, "Big Data Analysis Using Neural Networks."

¹² Wojcik, "Pefin, a Fintech Start-up, is Using AI to Offer Financial Advice."

¹³ Kavout, "Kavout Releases K Score for United Kingdom and Germany Stock Markets."

¹⁴ Newell, "Artificial Intelligence."

¹⁵ Flores, "AI Platform Claims it Can Advise at 1/20th of Cost."

¹⁶ The Economist, "For Artificial Intelligence to Thrive, it Must Explain Itself."

¹⁷ The Economist, "For Artificial Intelligence to Thrive, it Must Explain Itself."

¹⁸ Frankenfield, "What is a Robo Advisor and How Do They Work?"

¹⁹ Lansing, "AI and the Modern Wealth Manager."

²⁰ Winship, "Artificial Intelligence in Banking."

The term itself 'deep learning' refers to 'deep convolutional neural networks', hereafter abbreviated as 'deep NN'. ²¹ To clarify the meaning of several other key terms used in this paper, first note that 'artificial intelligence' is a very broad phrase that describes many types of technology that perform tasks typically thought to require human intelligence. ²² Similarly, 'machine learning' is a subset of AI, referring to a group of techniques that enables computer programs to learn a task without being explicitly programmed. ²³ Therefore, 'deep learning' is a subset of machine learning, referring to the technique of using deep convolutional neural networks (or deep NN) as a method that facilitates computer programs to learn a task without being explicitly programmed. ²⁴

'Neural networks' are a collection of many algorithms, individually called artificial neurons, organised hierarchically into layers to collectively form a network. They can be crafted from thousands of artificial neurons and organised into many layers—hence, the term 'deep'. The revolutionary feature of deep NN is an ability to continuously learn from data; that is, such networks can analyse enormous amounts of information and draw correlations between seemingly unrelated events. ²⁵ The more data deep NN are given, the better it becomes at making predictions. This is unlike 'old-style' AI, in which the algorithm initially improves with more data, but at a certain point feeding more into the system does not result in improved performance. ²⁶

An ability to digest vast amounts of data and make predictions based on historical trends makes deep NN well suited to the task of analysing financial markets and making learned projections.²⁷ Currently, there are a few financial service firms already offering this technology to advisers,²⁸ ²⁹ with many more developing similar AI services.³⁰ The application of deep learning to financial investment is also of keen interest in the wider research space.³¹ ³²

Distinctive Features Of Deep Neural Networks Versus 'Old-Style' AI

Since around 2016, deep NN and deep learning have been *the* buzzwords in technology—but is the hype deserved?³³ To understand what is special about deep neural networks, we need to have some contextual knowledge about AI technology generally. With such keen interest on the rise, one could easily be led to believe that neural networks are the only form of AI. This is not the case, as these systems represent one of many approaches to machine learning. Hence, another method of significance is 'rules-based machine learning', which is a family of AI techniques that is based on application of a defined knowledge base, formulated as 'if ... then ...' rules.³⁴ For example, 'IF stock price moves X% AND currency moves Y% THEN buy'.

Rules-based machine learning requires the acquisition and organisation of knowledge and reasoning, such as the discovery of causal relationships.³⁵ One key advantage of this technique is that the machine applies a defined set of rules in a comprehensible form to humans.³⁶ This means its decision-making process is transparent and open to human analysis, with the discovery of new rules adding to the body of knowledge.

Conversely, some disadvantages include the rules becoming unmanageably complicated for more complex applications of machine learning. As such a rules-based method is limited to a defined knowledge base, it can only apply the body of knowledge that is already known, given that rules-based machine learning does not handle uncertainty well.³⁷ With neural networks, there

²¹ Parloff, "Why Deep Learning is Suddenly Changing Your Life."

²² Copeland, "Artificial Intelligence."

²³ Schuld, "An Introduction to Quantum Machine Learning."

²⁴ Deng, "Deep Learning."

²⁵ Zhang, "Big Data Analysis Using Neural Networks."

²⁶ Zhang.

²⁷ Bussy, "Financial Services is a Natural Match for AI."

²⁸ Wojcik, "Pefin, a Fintech Start-up, is Using AI to Offer Financial Advice."

²⁹ Kavout, "Kavout Releases K Score for United Kingdom and Germany Stock Markets."

³⁰ McWaters, "The New Physics of Financial Services."

³¹ Sang, "Improving Trading Technical Analysis."

³² Aggarwal, "Deep Investment in Financial Markets Using Deep Learning Models."

³³ Kotecki, "Deep Learning's 'Permanent Peak' on Gartner's Hype Cycle."

³⁴ Fürnkranz, "A Brief Overview of Rule Learning."

³⁵ Tan, "A Brief History and Technical Review of the Expert System Research."

³⁶ Tan

³⁷ Hurdal, "Comparison of Rule-based and Neural Network Solutions."

is neither a defined body of knowledge nor rules of reasoning humans can comprehend.³⁸ Essentially, such networks find patterns and predict future outcomes based on past patterns. Precisely how neural networks arrive at a prediction cannot truly be known, but what is certain is that they have shown great success in doing so.

Most of the 'old-style AI' applications used in the finance industry would fall into the category of being rules-based systems.³⁹ Thus, the transition to deep neural networks is not merely an incremental improvement in technology, but a shift towards a fundamentally different technology with different advantages and disadvantages. From a legal or regulatory perspective, one must understand the inherent nature of deep NN to effectively regulate use of the technology. Deep neural networks are intrinsically different from rules-based machine learning, with the key distinction being that the reasoning process used by neural networks is a 'black box', which is not open to meaningful analysis.⁴⁰ Indeed, this poses a significant limitation that researchers must accept and work with if they wish to gain the benefits of the technology's superior performance.

It should be noted that research is being done to find methods of uncovering the decision-making process within neural networks.⁴¹ However, researchers have had little technical success in explaining some of their complex decisions.⁴²

Based on the current state of technology, it is highly probable that in the next few years fully automated deep NN financial advisers will be widely and commercially available. Due to the scalable nature of AI, these services can be offered at a fraction of the cost of human financial advisers. ⁴³ There is great potential benefit in this prospect, as it opens up financial advice services to low- and middle-income people who could otherwise not afford such assistance.

That said, there are tangible risks that must be acknowledged. The 'black box' component involved in deep NN creates fear around the fact that we cannot predict what actions the agent might take; if there are adverse results, there is equal concern that the deep NN agent cannot explain its own actions. The perceived lack of transparency of this technology will, then, pose a challenge for financial regulators if they are unable to question or probe the decision-making processes in which deep NN financial advisers engage.

Part 2: Regulation Of Automated Financial Advisers

Part 2 gives a brief overview of how automated financial advice services in several financial markets are regulated, and further examines whether current financial regulations are equipped to handle the unique features of deep NN technology.

Regulation Of 'Old-style AI' Versus Deep Neural Networks

An important point to highlight regards the new and unique challenges that deep NN pose for regulators compared to the AI technology that has been in use until now. 'Old-style AI', or algorithms that are based on decision trees or decision rules, are amenable to scrutiny as regulators can assess the rationale of those rules according to conventional wisdom.

However, with deep NN, regulators cannot review the rationale or rules behind the algorithms, as neither is actually intelligible to humans. As discussed in Part 1, the internal processes for how neural networks arrive at their decisions currently remain a 'black box', where no method is presently available to ascertain whether there exists any underlying rationale.

In terms of assessing the performance of deep NN financial robo-advisers, only time can reveal the quality of their advice and whether that advice delivers positive financial results. This is not a particularly reassuring method for financial regulators to monitor the provision of financial advice to the public, as it could take years before problems with such advice become apparent, wherein financial losses or other harms may occur before a problem is identified.

From a regulatory perspective, the advanced and intuitive learning capacity of deep NN concurrently presents a notable degree of risk. Part of the potential benefit of applying deep learning to financial advice is for deep NN agents to discover new methods

³⁹ Seetharaman, "Financial Applications of Neural Networks."

³⁸ Hurdal.

⁴⁰ Knight, "The Dark Secret at the Heart of AI."

⁴¹ Norton, "Inside Darpa's Push to Make Artificial Intelligence Explain Itself."

⁴² Adadi, "Peeking Inside the Black-box."

⁴³ Flores, "AI Platform Claims it Can Advise at 1/20th of Cost."

of financial investment unknown to humans. In a research setting, new knowledge has already been uncovered ahead of human discovery.

One example regards AlphaGo Zero, Google's latest deep NN agent created to play the game 'Go'.⁴⁴ Given only the rules of the game, the agent learned to play the game purely through trial and error. It was neither given human knowledge, such as a game strategy, nor provided moves used by expert players.⁴⁵ What researchers found was that AlphaGo Zero quickly discovered from first principles well-known moves and strategies used by human players, after which it discarded these for new moves it had learned that were previously unknown.⁴⁶ Human expert Go players have since been analysing games of AlphaGo Zero playing against itself, with some describing its methods as 'amazing', 'strange' and 'alien'.⁴⁷ Some of the novel opening moves that AlphaGo Zero discovered, which seemed to go against the conventional wisdom of human Go players, are now being copied in professional tournaments.⁴⁸

It is possible that the application of deep NN to financial investing could follow a similar path, with the deep NN financial adviser discovering superior investment strategies than any contained within current human knowledge. This makes oversight of such an agent especially challenging, as a human financial adviser may not be able to assess their robot counterpart's advice except in hindsight. Even if this advice goes against conventional wisdom, that alone would not conclusively rule the advice as of poor quality, as part of the purpose of applying deep NN to financial advice is to discover superior investment methods yet unknown to humans.

Regulation of Automated Financial Advice

Most financial regulators have taken a technology-neutral approach in which regulations do not distinguish between the provision of AI- or human-based financial advice. While this may have been adequate when regulators were able to scrutinise the rationale and rules behind 'old-style AI', it may be insufficient to deal with the unique characteristics of deep NN.

United States

The United States (US) has been a major player in the development of AI over the past decade, with China in close second.⁴⁹ How lawmakers and regulatory bodies in these two jurisdictions approach deep NN technology will play a critical role in the industry's prospective development.

Generally, the US Government has taken a hands-off approach to regulating AI.⁵⁰ With the exception of some legislation targeted at self-driving cars, there has been little attempt to control the development of this technology.⁵¹ The most recent Congress report on AI emphasised the value the nation has placed on innovation and entrepreneurial spirit, including the US Government's reluctance to hamper technological development with regulation.⁵²

Regarding the regulation of automated financial advice, the country's approach is technology neutral, with providers of such advice being subject to the same obligations as human financial advisers under the Securities and Exchange Commission's Investment Advisers Act of 1940.⁵³ There are no signs that in the near term US regulators will attempt to curb application of deep NN technology to financial advice services.

China

The market for AI financial services in China is rapidly growing.⁵⁴ Several factors make automated financial advisers especially well suited to the Chinese investment market, including a large middle class, a general lack of traditional financial advisers and

⁴⁴ Hassabis, "AlphaGo Zero."

⁴⁵ Silver, "Mastering the Game of Go Without Human Knowledge."

⁴⁶ Silver, 357–358.

⁴⁷ Chan, "The AI That Has Nothing to Learn From Humans."

⁴⁸ Chan.

⁴⁹ Batson, "China May Become the World's Leader in AI."

⁵⁰ Hurd, "Rise of the Machines."

⁵¹ West, "The State of Self-driving Car Laws across the US."

⁵² Hurd, "Rise of the Machines."

⁵³ US Securities and Exchange Commission, "SEC Staff Issues Guidance Update and Investor Bulletin on Robo-advisers."

⁵⁴ Huang, "How Fintech is Shaping China's Financial Services?"

market conditions that favour active asset management—all of which make low-cost automated financial advisers very attractive.⁵⁵

Regulators in China have issued guidelines specifically addressing the provision of automated AI financial advisers that will come into force at the end of 2020.⁵⁶ Its main features are that providers must disclose the parameters of their respective AI model, divulge to customers any inherent flaws and risks of the algorithm, and have in place plans to address any system failures or instability to the market from the potential 'sheep-flock' effect of defects in their machines' systems.⁵⁷ The guidelines demonstrate that regulators in China are attentive to the emerging challenges posed by AI financial advisers; how this will be implemented to deep NN financial advisers precisely remains to be seen.

European Union

The approach in the European Union (EU) has also been technology neutral, with the same obligations applying to providers of financial advice, whether through automation or through human intervention.⁵⁸ There is no specific domestic legislation aimed at automated financial advice.

The general data protection laws that have come into force in the EU are likely to have a major effect on the development and provision of AI services in Europe, particularly concerning automated financial advice that uses deep NN technology. The EU's General Data Protection Regulations (GDPR) contains what has been termed the 'right to an explanation'. ⁵⁹ That is, where a purely automated decision is made that significantly affects a person's rights, that person is entitled to 'meaningful information about the logic involved'. ⁶⁰ This could be interpreted as requiring an explanation for how a deep NN agent arrived at a particular decision. ⁶¹ This 'right to an explanation' has been criticised as being based on misunderstanding of deep NN technology, and that it will have a chilling effect on AI innovation in the EU. ⁶² ⁶³

United Kingdom

The United Kingdom (UK) has similarly maintained a technology-neutral approach, where providers of financial advice have the same obligations, whether using automation or human financial advisors.^{64 65} The UK has not legislated specifically on deep NN technology but is enacting general data protection legislation that implements the protections outlined in the European GDPR.⁶⁶

One House of Lords report into AI highlighted the 'black box' issue in deep NN as one key problem that must be overcome if the technology is to become a trusted and integral part of society.⁶⁷ The report also emphasised that AI technology must be 'explainable' to the public, in that AI systems must be able to clearly state the information and logic used to arrive at their decisions.⁶⁸ Where this is not possible, such as with deep neural networks, the report recommends delaying use of AI technology that is not explainable:

In cases such as deep neural networks, where it is not yet possible to generate thorough explanations for the decisions that are made, this may mean delaying their deployment for particular uses until alternative solutions are found.⁶⁹

⁵⁵ Rogers, "Digital China."

⁵⁶ Xie, "China."

⁵⁷ Xie.

⁵⁸ Joint Committee of the European Supervisory Authorities, "Joint Committee Report," 11.

⁵⁹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC.

⁶⁰ Regulation (EU) 2016/679, arts 13(2)(f), 14(2)(g), 15(1)(h).

⁶¹ Kaminski, "The Right to Explanation, Explained.".

⁶² Wallace, "The Impact of the EU's New Data Protection Regulation on AI."

⁶³ Koerner, "GDPR: Boosting or Choking Europe's Data Economy?"

⁶⁴ Financial Conduct Authority, "Automated Investment Services."

⁶⁵ Sharma, "The Next Frontier."

⁶⁶ Data Protection Act 2018.

⁶⁷ Select Committee on Artificial Intelligence, "AI in the UK."

⁶⁸ Select Committee on Artificial Intelligence, 127–128.

⁶⁹ Select Committee on Artificial Intelligence, 128.

Australia

The Australian approach to the provision of financial advice is also technology neutral. The *Corporations Act 2001* imposes the same obligations on people who provide financial advice to retail clients, whether through a computer program or with human financial advisers. Where automated financial advice is provided through a computer without the direct involvement of a human agent, the person offering the technology is taken to be the person offering the financial advice. ⁷¹

Although providers of automated financial advice have the same legal obligations as humans, they are required to disclose additional information compared to human financial advisers to demonstrate that they have fulfilled their legal obligations. Australian financial regulators have given further guidance on what information and requirements are expected regarding computer programs used for automated financial advice.⁷² These include:

- having people within the business who 'understand the rationale, risks and rules behind the algorithms underpinning the digital advice' 73
- having people able to review the digital advice generated by algorithms⁷⁴
- having a documented test strategy that explains the testing of algorithms, including defect resolution and final test results⁷⁵
- regular review of a sample of the automated advice for compliance with the law by a suitably qualified person⁷⁶
- heightened scrutiny of the automated advice when changes to the algorithm are made.

Evidently, the unique characteristics of deep NN pose a challenge for regulators, particularly regarding the issue of 'black box' decision-making. Financial regulators in the abovementioned jurisdictions have not yet developed any specific regulations to address this challenge. While not specifically aimed at deep NN, the general data protection laws in Europe and the UK will have an effect on the application of this technology because it cannot be explained in the way that lawmakers are demanding.

Instead, what the public and regulators need is reassurance that there is a way for laypeople to have some understanding of deep NN agents, and that there is some level of predictability in their behaviour. This does not have to be achieved by demanding that individual decisions by a deep NN agent be explainable. As such, the remainder of this paper proposes an alternative solution through which such machines can meet societal expectations of transparency and predictability.⁷⁸

Part 3: Trust in Personality

One way deep NN agents can gain public trust, despite being unable to explain their decision-making process, is by disclosing their 'personality'. This section articulates the idea that deep NN agents can be understood as having 'personality traits'—ranging from greediness, selfishness or prudence—which can give the public and regulators alike an adequate understanding of how they will behave.

Other commentators have argued that it is not necessary to open the 'black box' of neural network technology. Notably, Wachter, Mittelstadt and Russell⁷⁹ proposed the concept of unconditional counterfactuals as a way of giving the public a meaningful explanation of how AI makes automated decisions. This paper situates disclosing a deep NN agent's 'personality traits' as another potential method for proposing a meaningful framework for understanding neural network technology, without needing to crack the 'black box' problem.

⁷⁰ Corporations Act 2001 (Cth) s 912A.

⁷¹ Corporations Act 2001 (Cth) s 961(6).

⁷² Australian Securities and Investments Commission, *Regulatory Guide* 255.

⁷³ Australian Securities and Investments Commission, para 255.62.

⁷⁴ Australian Securities and Investments Commission, para 255.61.

⁷⁵ Australian Securities and Investments Commission, para 255.74.

⁷⁶ Australian Securities and Investments Commission, para 255.110.

⁷⁷ Australian Securities and Investments Commission, para 255.112.

⁷⁸ Bostrom, "The Ethics of Artificial Intelligence."

⁷⁹ Wachter, "Counterfactual Explanations Without Opening the Black Box."

With regards to the fear surrounding AI decision-making, it can be posited that such apprehensions are misplaced when considering a deep NN agent's behaviour and its susceptibility to regulation. Just because individual decisions cannot be explained does not mean that the agents cannot be controlled. The behaviour of deep NN agents can be controlled by rewarding 'wanted' decisions and punishing 'unwanted' decisions. As AI pioneer Alan Turing⁸⁰ envisioned back in 1950:

we normally associate punishments and rewards with the teaching process. Some simple child-machines can be constructed or programmed on this sort of principle. The machine has to be so constructed that events which shortly preceded the occurrence of a punishment-signal are unlikely to be repeated, whereas a reward-signal increased the probability of repetition of the events which led up to it.

During the 1950s, early research work in the field of AI explored the creation of machines that learned through trial and error. In the 1990s, Sutton and Barto⁸² brought together principles from trial-and-error machine learning and theories of psychology to conceive the discipline of 'reinforcement learning'. Essentially, this field deals with the study of the nature of learning and the general concept of learning through positive reinforcement of desired behaviour and negative punishments for undesired behaviour. Sa

Reinforcement learning as a method of machine learning is now at the forefront in the design of AI agents capable of understanding through trial and error. ⁸⁴ For example, Google's AlphaGo, the first computer program to defeat a world champion at Go in 2016, was trained using this technique. ⁸⁵ Google's AI research lab DeepMind continues to push the capabilities of their deep NN agents through reinforcement learning. ⁸⁶

This paper proposes that the conceptual framework of 'reinforcement learning'—which sees an agent learn through reward and punishment—is the most appropriate framework through which AI developers can communicate to laypeople about deep NN agents. This is because:

- 1. reinforcement learning is at a level abstract enough that it will not become obsolete as deep learning techniques progress
- 2. it is a concept intuitive enough to be understood by people without any expertise in computer science.

Reinforcement Learning

Reinforcement learning is a method of machine learning by which an AI agent learns to 'take sequences of actions in an environment in order to maximize cumulative rewards'. 87 The agent learns to make better decisions through a system of positive rewards and negative rewards (punishments). 88 The deep neural network can be conceived of as an "agent", which interacts with its environment. The positive rewards and punishments are signals that the environment sends to the agent in response to the agent's actions. The goal of the agent is to take the actions that lead to maximal cumulative rewards.

Such a system seems conceptually simple, but it belies a highly complex decision-making process of balancing competing rewards. An AI developer can control an agent by setting the positive and negative rewards for particular states. For example, research has demonstrated that by manipulating the scheme of rewards and punishments, developers could control whether their machine exhibited purely competitive behaviour or purely cooperative behaviour, or variations of both actions. 89

Three key traits that will be discussed below are how a developer can control whether the agent is "greedy" or "prudent", how a developer can control the agent's appetite for risk-taking and how a developer can assign punishments for illegal actions.

83 Sutton, 7–8.

⁸⁰ Turing, "Computing Machinery and Intelligence," 457.

⁸¹ Sutton, Introduction to Reinforcement Learning, 16–17.

⁸² Sutton, 3–5.

⁸⁴ Francois-Lavet, "An Introduction to Deep Reinforcement Learning," 99–102.

⁸⁵ Silver, "Mastering the Game of Go with Deep Neural Networks and Tree Search."

⁸⁶ Castro, "Introducing a New Framework for Flexible and Reproducible Reinforcement Learning Research."

⁸⁷ Francois-Lavet, "An Introduction to Deep Reinforcement Learning," 6, 15–16.

⁸⁸ Sutton, Reinforcement Learning, An Introduction, 6–7.

⁸⁹ Tampuu, "Multiagent Cooperation and Competition with Deep Reinforcement Learning."

This is best illustrated with an example of how an AI developer can control the traits of an agent, through rewards and punishments. 90 Take for example an agent developed to drive a race car around a race track. We can set a positive reward of 20 if the race car agent records the fastest lap time, a positive reward of 10 for crossing the finish line, a neutral reward of zero for remaining on track, a penalty of minus 2 for being off-track, and a penalty of minus 10 for crashing. The agent will seek to maximise its reward by completing the race as fast as possible whilst staying on track.

The developer can control how greedy or prudent the agent will be in weighing the reward of a fast lap time versus the risk of punishment for going off-track or crashing. A race car agent who is greedy will try to make turns as fast as possible, which may result in a lower overall reward if it crashes or goes off-track. A more prudent race car agent will slow down at turns, which may yield lower rewards in the short term but higher rewards in the long term.

Developers can control how "greedy" an agent is by adjusting what is termed the "discounting factor".⁹¹ The discounting factor is the proportion by which the value of a reward in the future is diminished, meaning that the agent places greater value on more immediate rewards than rewards further in the future. By modifying the discounting factor, a developer can control how greedy or prudent the agent is. For example, Sun et al.⁹² discusses the ability of developers to create non-greedy agents by modifying the reward function with diminishing return.

An AI developer can also control the behaviour of the agent with regards to whether the agent is risk-averse or risk-seeking, by the relative rewards and punishments. ⁹³ For example, if we set the reward for the race car agent recording the fastest lap time to 100, and the punishment for crashing or going off-track to minus 10, the agent will engage in more risky driving behaviour, because the rewards for a fast lap time far outweigh the punishments for crashing or for not completing the lap.

If instead we set the reward for the fastest lap time to 20, and the punishment for crashing or going off-track to minus 10, then the race car agent will drive in a more conservative manner, as the potential reward of a fast lap time is balanced against the risk of crashing or going off-track.⁹⁴ As such a developer can control the agent's appetite for risk, by controlling the relative value of the positive rewards versus negative punishments.

In the context of a Deep NN agent giving financial advice, the same principles apply as in our race car agent example. The Deep NN agent can be controlled with regards to how "greedy" or "prudent" it is, in balancing the pursuit of immediate financial gains versus long term growth of a portfolio. The Deep NN agent can also be controlled in terms of risk appetite in how much financial loss it is willing to risk for financial gains, in the relative values of rewards for financial gain, versus punishment for financial losses.

An important role of punishment is that it can also be a method of curbing illegal or unethical behaviour by the Deep NN agent. Actions which are illegal can be assigned negative rewards, so that the agent learns to avoid taking illegal actions. There is however the challenge that including violations of financial laws and regulations as punishments for the Deep NN agent is incredibly complex in real-world application. Unlike games like chess and Go, there is no simple list of rules for what constitutes lawful conduct in the provision of financial advice. Laws are open to context and interpretation, and an AI developer faces significant difficulty in how to incorporate the complexities of corporate law and ethical business conduct.

Nevertheless, at this stage where Deep NN technology is still in its early phase, the emphasis should be on disclosure of whether the AI developer has managed to incorporate laws and regulations, and if so to what degree. This disclosure would assist regulators in using their resources efficiently, focusing their attention on Deep NN agents that are at higher risk of giving financial advice that is illegal or unethical.

⁹⁰ Based on the example used to explain reinforcement learning in Dettmers, "Deep Learning in a Nutshell."

⁹¹ Dettmers

⁹² Sun, "Designing Non-greedy Reinforcement Learning Agents with Diminishing Reward Shaping."

⁹³ Pan, "Risk Averse Robust Adversarial Reinforcement Learning."

⁹⁴ Pan, 2

⁹⁵ Francois-Lavet, "An Introduction to Deep Reinforcement Learning."

Disclosing "Personality Profiles" For Deep NN Financial Advisors

This paper contends that AI developers should disclose the following three "personality traits" of Deep NN financial advisors, so that people without technical expertise are able to understand significant aspects of the agent's behaviour in a financial investment setting:

- 1. Greedy versus prudent: this trait measures how much the agent will prefer short term gain versus long term gains
- 2. Risk appetite: this trait measures how much risk of loss the agent is willing to take versus prospect of gain
- 3. Ethical behaviour: this trait indicates whether laws and regulations have been incorporated into the reward and punishment system of the agent. If not, then regulators will be alerted to the higher risk of the agent giving advice that may be illegal or unethical.

AI developers should give a plain language description of these 3 basic personality traits of the Deep NN agent. This should be disclosed before the Deep NN agent's services are offered to the public, and updated descriptions should be disclosed if the AI developers make changes to the agent's system. Financial regulators could undertake independent review of the performance of the Deep NN agent to assess whether the agent's "personality profile" matches the description as disclosed by the AI developers.

The goal of disclosing the Deep NN agent's "personality profile" is to allow a person without any expertise in AI technology to have a meaningful understanding of the likely behaviour of the agent. This allows consumers to make informed choices regarding which Deep NN agent is suitable for them. It also assists regulators in identifying Deep NN agents that pose a higher risk to the public, enabling regulators to use their resources in a targeted fashion.

Conclusion

Deep neural network agents have the potential to be the perfect financial advisers: one that has no self-interest and operates on reliable parameters of "reward" and "punishment". Compared to human financial advisers who will always feel temptation to act in their own self-interests.

However, like any powerful new technology there are good reasons for us to remain cautious. As Deep NN agents take over increasingly sophisticated tasks with serious social consequences, the community needs reassurance that Deep NN agents will adhere to legal and ethical codes of conduct.

The solution proposed here is a framework for AI developers to communicate the essential traits of Deep NN agents to the public. Understanding Deep NN through the concept of reinforcement learning and personality traits, provides a common language for people with and without computer science expertise to communicate about this technology. The ability for lay people to understand at a basic level how the Deep NN agent will act, allows the public and regulators to respond based on knowledge and understanding, instead of fear and suspicion. If from this understanding Deep NN can gain the trust of the public, the "black box" issue can be overcome and society can reap the benefits of affordable and innovative financial advice services.

Bibliography

Primary

Australian Securities and Investments Commission. Regulatory Guide 255: Providing Digital Financial Product Advice to Retail Clients. Canberra: ASIC, 2016.

Corporations Act 2001 (Cth).

Data Protection Act 2018.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, Art 13–15, 22.

Secondary

Adadi, Amina and Mohammed Berrada. "Peeking Inside the Black-box: A Survey on Explainable Artificial Intelligence (XAI)." *IEEE Access* 6 (2018): 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

- Aggarwal, Saurabh and Somya Aggarwal. "Deep Investment in Financial Markets Using Deep Learning Models." *International Journal of Computer Applications* 162 no 2 (2017): 40–43. http://doi.org/10.5120/ijca2017913283
- Batson, Andrew, Virgilio Bisio, Elsa Kania, Lotus Ruan and Jeff Cao. "China May Become the World's Leader in AI. But at What Cost?" *ChinaFile*, July 30, 2018. http://www.chinafile.com/conversation/china-may-become-worlds-leader-ai-what-cost
- Beyer, Max, David de Meza and Diane Reyniers. "Do Financial Advisor Commissions Distort Client Choice?" *Economics Letters* 119, no 2 (2013): 117–119. https://doi.org/10.1016/j.econlet.2013.01.026
- Bostrom, Nick and Eliezer Yudkowsky. "The Ethics of Artificial Intelligence." In *The Cambridge Handbook of Artificial Intelligence*, edited by Keith Frankish and William M. Ramsey, 316–334. New York: Cambridge University Press, 2014.
- Burke, Jeremy, Angela Hung, Jack Clift, Steven Garber and Joanne Yoong. *Impacts of Conflicts of Interest in the Financial Services Industry. RAND Working Paper Series WR 1076* (Social Science Research Network, Rochester, NY, February 2015)
- Bussy, Eric. "Financial Services is a Natural Match for AI." *Financial Executives International*, September 27, 2018. https://www.financialexecutives.org/FEI-Daily/September-2018/Financial-Services-is-a-Natural-Match-for-AI.aspx
- Castro, Pablo Samuel and Marc G. Bellemare. "Introducing a New Framework for Flexible and Reproducible Reinforcement Learning Research." *Google AI Blog*, August 27, 2018. http://ai.googleblog.com/2018/08/introducing-new-framework-for-flexible.html
- Chan, Dawn. "The AI That Has Nothing to Learn From Humans." *The Atlantic*, October 20, 2017. https://www.theatlantic.com/technology/archive/2017/10/alphago-zero-the-ai-that-taught-itself-go/543450/
- Copeland, BJ. "Artificial Intelligence: Definition, Examples, and Applications." *Encyclopedia Britannica*, n.d. https://www.britannica.com/technology/artificial-intelligence
- Deng, Li and Dong Yu. "Deep Learning: Methods and Applications." Foundations and Trends in Signal Processing 7, no 3–4 (2014): 197–387. http://dx.doi.org/10.1561/2000000039
- Dettmers, Tim. "Deep Learning in a Nutshell: Reinforcement Learning." *NVIDIA Developer Blog*, September 8, 2016. https://devblogs.nvidia.com/deep-learning-nutshell-reinforcement-learning/
- Financial Conduct Authority. "Automated Investment Services: Our Expectations." May 21, 2018.
 - https://www.fca.org.uk/publications/multi-firm-reviews/automated-investment-services-our-expectations
- Flores, Adrian. "AI Platform Claims it Can Advise at 1/20th of Cost." *Independent Financial Adviser*, October 10, 2018. https://www.ifa.com.au/news/26072-ai-platform-claims-it-can-advise-at-1-20th-of-cost
- Francois-Lavet, Vincent, Peter Henderson, Riashat Islam, Marc G. Bellemare and Joelle Pineau. "An Introduction to Deep Reinforcement Learning." *Foundations and Trends in Machine Learning* 11, no 3–4 (2018): 219–354. https://doi.org/10.1561/2200000071
- Frankenfield, Jake. "What is a Robo Advisor and How Do They Work?" *Investopedia*, July 31, 2019. https://www.investopedia.com/terms/r/roboadvisor-roboadviser.asp
- Fürnkranz, Johannes and Tomáš Kliegr. "A Brief Overview of Rule Learning." In *Rule Technologies: Foundations, Tools, and Applications*, edited by Nick Bassiliades, Georg Gottlob, Fariba Sadri, Adrian Paschke and Dumitru Roman, 54–69. Switzerland: Springer, 2015.
- Hassabis, Demis and David Silver. "AlphaGo Zero: Learning from Scratch." *DeepMind* (blog), October 18, 2017. https://deepmind.com/blog/alphago-zero-learning-scratch/
- Huang, Elton, James Chang and Vivian Ma. *How Fintech is Shaping China's Financial Services?* (PricewaterhouseCoopers Hong Kong, 2018). https://www.pwchk.com/en/research-and-insights/how-fintech-is-shaping-china-financial-services.pdf
- Hurd, Will and Robin L. Kelly. *Rise of the Machines: Artificial Intelligence and Its Growing Impact on US Policy* (Committee on Oversight and Government Reform, US House of Representatives, September 2018). https://www.hsdl.org/?abstract&did=816362
- Hurdal, Brian and Jerry J. Hajek. "Comparison of Rule-based and Neural Network Solutions for a Structured Selection Problem." *Transportation Research Record*, no 1399 (1993): 1–7.
- Joint Committee of the European Supervisory Authorities. *Joint Committee Report on the Results of the Monitoring Exercise on 'Automation in Financial Advice'* (Joint Committee of the European Supervisory Authorities, September 2018). https://esas-joint-committee.europa.eu/Publications/Reports/JC%202018%2029%20-%20JC%20Report%20on%20automation%20in%20financial%20advice.pdf
- Kaminski, Margot E. "The Right to Explanation, Explained." *Berkeley Technology Law Journal* 34, no 1 (2019): 189–218. https://doi.org/10.15779/Z38TD9N83H

Kavout. "Kavout Releases K Score for United Kingdom and Germany Stock Markets." *PR Newswire*, April 18, 2019. https://www.prnewswire.com/news-releases/kavout-releases-k-score-for-united-kingdom-and-germany-stock-markets-300834867.html

- Knight, Will. "The Dark Secret at the Heart of AI." MIT Technology Review, April 11, 2017.
 - https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/
- Koerner, Kevin. "GDPR: Boosting or Choking Europe's Data Economy?" *Deutsche Bank Research*, June 13, 2018. https://www.dbresearch.com/PROD/RPS_EN-
 - PROD/PROD000000000470381/GDPR %E2%80%93 boosting or choking Europe%E2%80%99s data economy%3 F.xhtml
- Kotecki, James. "Deep Learning's 'Permanent Peak' on Gartner's Hype Cycle." Medium, August 17, 2018.
 - https://medium.com/machine-learning-in-practice/deep-learnings-permanent-peak-on-gartner-s-hype-cycle-96157a1736e
- Lansing, Nicholas. AI and the Modern Wealth Manager: How Artificial Intelligence is Creating a Personalized Investing Experience (Forbes Insights, 2018).
- Marr, Bernard. "The 4th Industrial Revolution is Here: Are You Ready?" *Forbes*, August 13, 2018. https://www.forbes.com/sites/bernardmarr/2018/08/13/the-4th-industrial-revolution-is-here-are-you-ready/#5433b290628b
- McAfee, Andrew and Erik Brynjolfsson. "The Dawn of the Age of Artificial Intelligence." *The Atlantic*, February 14, 2014. https://www.theatlantic.com/business/archive/2014/02/the-dawn-of-the-age-of-artificial-intelligence/283730/
- McWaters, Jesse. *The New Physics of Financial Services: How Artificial Intelligence is Transforming the Financial Ecosystem* (World Economic Forum, August 2018). https://www.weforum.org/reports/the-new-physics-of-financial-services-how-artificial-intelligence-is-transforming-the-financial-ecosystem/
- Newell, Richard. *Artificial Intelligence: Smart Advantage* (IPE, June 2018). <u>www.ipe.com/reports/special-reports/top-400-asset-managers/artificial-intelligence-smart-advantage/10025005.fullarticle</u>
- Norton, Steven and Sara Castellanos. "Inside Darpa's Push to Make Artificial Intelligence Explain Itself." *Wall Street Journal* (blog), August 10, 2017. https://blogs.wsj.com/cio/2017/08/10/inside-darpas-push-to-make-artificial-intelligence-explain-itself/
- Pan, Xinlei and John Canny. "Risk Averse Robust Adversarial Reinforcement Learning: Extended Abstract." ACM Computer Science in Cars Symposium, Munich, Germany, 2018. https://cscs.mpi-inf.mpg.de/
- Parloff, Roger. "Why Deep Learning is Suddenly Changing Your Life." *Fortune*, September 28, 2016. http://fortune.com/ai-artificial-intelligence-deep-machine-learning/
- Rogers, Jonathan. "Digital China: March of the Machines." *Global Finance Magazine*, February 15, 2018. https://www.gfmag.com/magazine/february-2018/digital-china-march-machines
- Sang, Chenjie and Massimo Di Pierro. "Improving Trading Technical Analysis with TensorFlow Long Short-term Memory (LSTM) Neural Network." *Journal of Finance and Data Science* 5, no 1 (2019): 1–11. https://doi.org/10.1016/j.jfds.2018.10.003
- Schuld, Maria, Ilya Sinayskiy and Francesco Petruccione. "An Introduction to Quantum Machine Learning." *Contemporary Physics* 56, no 2 (2015): 172–185. https://doi.org/10.1080/00107514.2014.964942
- Seetharaman, Krishna. "Financial Applications of Neural Networks." *Aspire Systems*, February 22, 2018. https://blog.aspiresys.com/banking-and-finance/financial-applications-neural-networks/
- Select Committee on Artificial Intelligence. "AI in the UK: Ready, Willing and Able." Report of Session 2017–19. House of Lords, April 16, 2018. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf
- Sharma, Raul, Rosalind Fergusson, Joy Kershaw, Valeria Gallo and Peter Evans. *The Next Frontier: The Future of Automated Financial Advice in the UK* (Deloitte, 2017).
 - https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/financial-services/lu-future-automated-financial-advice-uk-15052017.pdf
- Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, et al. "Mastering the Game of Go with Deep Neural Networks and Tree Search." *Nature* 529, no 7587 (2016): 484–489. https://doi.org/10.1038/nature16961
- Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, et al. "Mastering the Game of Go Without Human Knowledge." *Nature* 550, no 7676 (2017): 354–359. https://doi.org/10.1038/nature24270
- Stanford, Jim. "Sectoral Collective Agreement Could Combat Finance Industry Misconduct." *Centre for Future Work*, November 5, 2018. https://www.futurework.org.au/sectoral collective bargaining would help combat finance
- Sun, Fan-Yun, Yen-Yu Chang, Yueh-Hua Wu and Shou-De Lin. "Designing Non-greedy Reinforcement Learning Agents with Diminishing Reward Shaping." In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 297–302. New York: ACM, 2018. https://doi.org/10.1145/3278721.3278759
- Sutton, Richard S. and Andrew G. Barto. *Introduction to Reinforcement Learning*. 1st ed. Cambridge: MIT Press, 1998. ———. *Reinforcement Learning, An Introduction*. 2nd ed. Cambridge: MIT Press, 2018.

Tampuu, Ardi, Tambet Matiisen, Dorian Kodelja, Ilya Kuzovkin, Kristjan Korjus, Juhan Aru, Jaan Aru and Raul Vicente. "Multiagent Cooperation and Competition with Deep Reinforcement Learning." *PLOS ONE* 12, no 4 (2017): e0172395. https://doi.org/10.1371/journal.pone.0172395

- Tan, Haocheng. "A Brief History and Technical Review of the Expert System Research." *IOP Conference Series: Materials Science and Engineering* 242 (2017): 1–5. https://doi.org/10.1088/1757-899X/242/1/012111
- Teich, David A. "Management AI: Bias, Criminal Recidivism, and the Promise of Machine Learning." *Forbes*, January 24, 2018. https://www.forbes.com/sites/tiriasresearch/2018/01/24/management-ai-bias-criminal-recidivism-and-the-promise-of-machine-learning/
- The Economist. "For Artificial Intelligence to Thrive, it Must Explain Itself." February 15, 2018.

 https://www.economist.com/science-and-technology/2018/02/15/for-artificial-intelligence-to-thrive-it-must-explain-itself
 Tibbetts, John H. "The Frontiers of Artificial Intelligence: Deep Learning Brings Speed, Accuracy to the Life Sciences."

BioScience 68, no 1 (2018): 5–10. https://doi.org/10.1093/biosci/bix136

- Turing, Alan. "Computing Machinery and Intelligence." *Mind LIX*, no 236 (1950): 433–460. https://doi.org/10.1093/mind/LIX.236.433
- US Securities and Exchange Commission. "SEC Staff Issues Guidance Update and Investor Bulletin on Robo-advisers," press release 2017-52, February 23, 2017. https://www.sec.gov/news/pressrelease/2017-52.html
- Wachter, Sandra, Brent Mittelstadt and Chris Russell. "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR." *Harvard Journal of Law and Technology* 31, no 2 (2018): 841–887. https://doi.org/10.2139/ssrn.3063289
- Wallace, Nick and Daniel Castro. *The Impact of the EU's New Data Protection Regulation on AI* (Center for Data Innovation, March 2018). http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf
- West, Darrell and Jack Karsten. "The State of Self-driving Car Laws across the US." *Brookings* (blog), May 1, 2018. https://www.brookings.edu/blog/techtank/2018/05/01/the-state-of-self-driving-car-laws-across-the-u-s/
- Winship, Todd. "Artificial Intelligence in Banking: You Ain't Seen Nothing Yet." *International Business Times*, July 25, 2017. https://www.ibtimes.co.uk/artificial-intelligence-banking-you-aint-seen-nothing-yet-1631901
- Wojcik, Natalia. "Pefin, a Fintech Start-up, is Using AI to Offer Financial Advice. Just Don't Call it a 'Robo Advisor." *CNBC*, September 9, 2017. https://www.cnbc.com/2017/09/08/fintech-start-up-pefin-uses-a-i-to-offer-financial-advice.html
- Xie, Natasha. "China: Guiding Opinions on Asset Management Business—Key Provisions and Observations." *Conventus Law*, May 11, 2018. http://www.conventuslaw.com/report/china-guiding-opinions-on-asset-management/
- Zammit-Lucia, Joseph. "Misaligned Incentives Will Lead Us Right Back into Another Financial Crisis." *The Guardian*, August 29, 2013. https://www.theguardian.com/sustainable-business/financial-crisis-credit-agencies-incentives
- Zhang, Yi, Quan Guo and Jianyong Wang. "Big Data Analysis Using Neural Networks." *Advanced Engineering Sciences* 49, no 1 (2017): 9–18. https://doi.org/10.15961/j.jsuese.2017.01.002