LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Artificial intelligence, financial risk management and systemic risk

Jon Danielsson
Robert Macrae
Andreas Uthemann

SRC Special Paper No 13
November 2017

SRC Systemic Risk Centre

Special Paper Series

ISSN 2055-0375

This paper is published as part of the Systemic Risk Centre's Special Paper Series. The support of the Economic and Social Research Council (ESRC) in funding the SRC is gratefully acknowledged [grant number ES/K002309/1].

Jon Danielsson, Systemic Risk Centre and Department of Finance, London School of Economics and Political Science
Robert Macrae, Systemic Risk Centre, London School of Economics and Political Science
Andreas Uthemann, Systemic Risk Centre, London School of Economics and Political Science

Published by
Systemic Risk Centre
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form other than that in which it is published.

Requests for permission to reproduce any article or part of the Working Paper should be sent to the editor at the above address.

© Jon Danielsson, Robert Macrae and Andreas Uthemann, submitted 2017

Artificial intelligence, financial risk management and systemic risk*

Jon Danielsson, Robert Macrae and Andreas Uthemann
Systemic Risk Centre
London School of Economics

This version: December 2017

Abstract

Artificial intelligence (AI) is rapidly changing how the financial system is operated and we can expect it to increasingly take over core functions because of cost savings and operational efficiencies. AI will likely be very helpful to risk managers and microprudential authorities. It meanwhile has the potential to destabilise the financial system, creating new risks and amplifying existing ones due to procyclicality, endogenous complexity, optimisation against the system and the need to trust the AI engine.

*We thank the Economic and Social Research Council (UK) [grant number ES/K002309/1] and the Engineering and Physical Sciences Research Council (UK) [grant number EP/P031730/1] for their support. Updated versions of this paper can be downloaded from our website `www.riskresearch.org`.

1 Introduction

Artificial intelligence (AI) is rapidly changing how financial institutions are operated and regulated.[1] Some functions are naturally well suited for AI, like many tasks in risk management and day-to-day financial supervision. A different picture emerges when we look at the stability of the entire financial system where AI can amplify the types of risks that culminate in financial crises.

AI promises significant cost savings and increases in operational efficiency for risk managers and microprudential supervisors, both of which are mostly concerned with the day-to-day operations of financial institutions. This is because AI is particularly useful for controlling an existing system, one with abundant data and clearly understood risks, such as risk management and microprudential supervision. There is no need to control risk in the entire financial system or bank as a single optimisation problem, instead the we can focus on each sub-component of the system individually. Such local optimisation leads to an acceptable solution to the global problem, the control of risk for a bank or financial system. This means we can generally assume risk is exogenous and do not have to consider how it arises.

None of these apply to macroprudential regulations, focused on systemic financial risk. By controlling risk in each local area of banks separately, we may easily increase the very risk we are trying to contain because the most dangerous types of risk typically emerge at the intersection between apparently disparate parts of the system. It is necessary to model and control risk in the entire financial system as a single optimisation problem, challenging as the financial system is almost infinitely complex. Furthermore, financial crises are rare, less that every 42 years for OECD member countries, giving AI little historical data to train on. Assuming risk is exogenous will miss out on extreme outcomes, instead, policy makers need to consider the endogenous nature of financial risk.

The complex nature of the problem means that existing AI approaches are not sufficient if AI is to make significant inroads into macroprudential policy, significant improvements are needed. Such AI would need to demonstrate how it reasons. With day-to-day risk control this does not create many conceptual problems because we can substantiate the work of an AI engine by observing repeated outcomes. It does not matter how the engine came to its answer, and we can leave it to do its job mostly undisturbed. It is

[1] A 2017 Financial Stability Board study identifies how AI will affect the practice of finance, finding the impact of AI to be broadly positive.

different with macroprudential policy. Not only is the frequency of adverse outcomes that are to be controlled very low, the intermediate objectives are important and the cost of failure catastrophic. To succeed and be trusted, the engine needs to capture and explain endogenous risk.

Meanwhile, AI may be destabilising simply because of the way in which it operates. AI will favour best practices and standardised best-of-breed models that closely resemble each other, all of which, no matter how well-intentioned and otherwise efficient, will also increase monoculture and hence pro-cyclicality and systemic risk. Furthermore, while AI likely excels at managing known exogenous risk, it will be worse at handling the unquantifiable endogenous risk behind most financial crises. This would focus attention on the wrong part of the financial system, giving instability room to build up. AI likely will increase endogenous risk.

A final challenge to the use of AI in macroprudential financial regulation is that it gives malicious agents ample scope for optimising against the system. They have a considerable informational advantage over the AI engine, both because of its inherent rationality and also since its rules and objectives are public and change slowly. Meanwhile, any attacker only needs to solve a local optimisation problem while the AI engine has to solve the global problem. As soon as one agent succeeds, competitive pressure will ensure that many others follow and a systemic crisis may ensue. If we want to make the AI engine resilient to attacks it might be necessary to endow it with the ability to experiment and randomise its reactions, even give it some power over the rulebook, features that are unpalatable to the financial authorities.

2 Artificial intelligence, risk and endogeneity

A celebrated recent AI is Google's AlphaGo Zero (Silver et al., 2017), which, being only instructed with the rules of the game of Go, was able to learn over the span of three days how to conquer its predecessor that had earlier defeated the world champion. Go has been considered the most demanding game for AI to play because of the vast number of possible moves and the complexity of patterns involved. However, like all games of complete information, Go is an ideal use case for AI; the domain of the problem is precisely defined by explicit rules and the objectives of the opponent are known.

AI will not perform as well playing games where information is incomplete.[2] The AI engine will neither have exact knowledge of the types of agents it is

[2]Fudenberg and Levine (1998) and Shoham and Leyton-Brown (2008).

playing against, nor will it necessarily be aware of all of their possible moves. Consequently, the engine will not fully know at the start of the game the rules of the game, instead having to learn them during play. This is particularly challenging if the rules evolve or change during play.

When it comes to controlling the financial system, a useful concept to map out the boundaries of the ability of AI is Danielsson and Shin's (2002) classification of risk along a spectrum from exogenous to endogenous. *Exogenous risk* comes from outside the financial system, as an asteroid might hit the earth. *Endogenous risk* is created by the interaction of the entities that make up the financial system, each with their own abilities, biases, resources and objectives. Playing the lottery where the odds of winning are one in a million is exogenous risk, the financial impact of Donald Trump's presidency is endogenous risk. Exogenous risk is measurable and quantifiable and results in statistical distributions that we can use to exercise control. Endogenous risk is usually neither measurable nor quantifiable and does not lend itself to straightforward statistical representations, being consequently much more challenging to address with formal analytical tools. AI is ideally suited for dealing with exogenous risk. It finds endogenous risk much harder because it cannot train against unknown data.

While almost all risk in the financial system is created by the individuals that make up the system, and hence is endogenous, such risk is very hard to model. Consequently, most control processes assume risk is exogenous, like the risk methodology in the Basel III market risk regulations. Such assumptions are not problematic when involving the control of risk in repeated outcomes where each decision is inconsequential enough to be effectively exogenous, the cost of failure is low, and objectives are short term. An example is the day-to-day risk management of proprietary traders.

As we attempt to aggregate individual risk-taking up from the level of trading desks to large financial institutions and eventually the entire financial system, an assumption of risk being exogenous becomes progressively less realistic and more prone to disastrous mistakes. This is both because it is technically very hard to model the dynamic nature of the dependence structure between different assets and asset classes and also since investment decisions are both affected by risk measurements and change the measurements, creating endogenous risk. Consequently, the tendency is to ignore the technical difficulties and just use crude approximations, such as static correlation matrices and exogenous risk based methods. This implies that as the level of aggregation increases, AI becomes increasingly less useful because the underlying data and models are increasingly unreliable.

In other cases, events are much more unique, especially the most extreme, for most parts caused by unique sets of events arising from endogenous risk. Trying to control extreme risk with AI may well be impossible, because it is not sufficient to consider observable statistical relationships, it becomes necessary to identify the deep, and for most part latent, vulnerabilities in the financial system.

3 Financial policy

Artificial intelligence is set to have considerable influence on financial regulations, both microprudential regulations focused on the conduct of individual banks and their solvency, and macroprudential regulations concerned with the stability of the financial system.

3.1 Day-to-day risk and microprudential regulations

The efforts of the microprudential authorities are spent on detailed rules of how a regulated bank should behave, what it can do, what it cannot do and what it should do, codified in the *rulebook*. The supervisor exhaustively monitors compliance with the rulebook in various ways, ranging from on-site inspections to analysing reported data. The authority has almost complete access to the internal information held by banks and considerable power to change bank behaviour. The main focus is on day-to-day risk, the underlying assumption being that so long as each individual activity has limited risk, with the appropriate processes in place, the policy objectives are met.

The microprudential authorities deal mostly with risk rather than uncertainty because the focus is on vast numbers of small issues, implying distributions tend to be well estimated. The technical challenges embedded in the underlying optimisation problem can be solved as a sequence of local optimisations. Endogeneity can typically be ignored, and the problem of the microprudential regulation of all institutions can be solved as a sequence of smaller problems, one position or institution at a time. Microprudential regulations are an ideal domain for AI since they requires the evaluation of a vast quantity of objective and factual data against an equally vast body of well-defined rules with explicit objectives. Endogeneities are modest and can typically be handled by updating the rulebook. Indeed, AI has already spawned a new field called regulatory technology — *regtech*.[3]

[3]Arner et al. (2016).

It is not all that hard to translate the rulebook of a supervisory agency, now for most parts in plain English, into a formal computerised logic engine which could constitute the core of the supervisory AI engine.[4] Such work is already being undertaken, for example on behalf of the UK Financial Conduct Authority,[5] where the rulebook bot can be queried for compliance issues, usually answering more accurately and rapidly than human supervisors. This will give the regulators the ability to validate their rules for consistency and give banks an application programming interface (API) to validate practices against regulations.

Within financial institutions, risk managers play a similar microprudential role. Their primary focus is also the control of day-to-day exogenous risk, not endogenous risk. They are inherently focused on relatively short time horizons where data scarcity is typically not an issue and distributions that can be reasonably well estimated. The risk managers' problem is however far from trivial as a good solution requires the consideration of many aspects of risk relating to counterparty, liquidity, volatility, fungibility, settlement, regulatory controls, fraud and operational concerns. Each requires the application of different techniques and approaches. Data may also present a practical problem. In the large majority of cases, however, the underlying problem of bank risk management can be approximately solved as a sequence of local problems, providing that shared exposures are identified appropriately. This greatly simplifies the technical challenges and makes risk management well suited for AI.

AI should make increasing inroads into risk management, heading towards the establishment of an integrated AI risk management engine with full knowledge of risk, positions, counterparties, the humans making risk decisions and all aspects of day-to-day risk. It could perform risk management and investment functions such as recommending position limits, evaluating performance and advising on risk concentrations. Its development will involve the progressive reduction of back office, middle office and then front office roles, leading to significant cost savings.

Once we have supervisory and risk management AI engines up and running, they will presumably be very efficient at ensuring compliance because both sides will have very similar knowledge representations and data structuring requirements. The end result will be a much improved risk management and microprudential regulation process. Costs will be significantly lower, mistakes fewer and risk better allocated.

[4]See e.g. Willis Towers Watson (2017).

[5]See the work of the Governance, risk & compliance technology centre (2017).

There is still some time to go before the risk management and supervisory AI engines become a practical reality, but as there are obvious advantages and no obvious technical problems it seems inevitable that they will. The main brakes to development are likely to be political, social and legal, not technical. The various authorities may not want to coordinate on data compatibility or API interfaces. They may even see not doing so as creating a competitive advantage for their domestic financial industry.

3.2 Systemic risk

The macroprudential authorities are concerned with the stability of the entire financial system, and in particular systemic risk, the potential for a major financial crisis to adversely affect the real economy, as defined by the IMF, BIS and the FSB in 2009. The macroprudential problem is much harder than the microprudential problem. To begin with, systemic crises are not frequent. Studying the IMF-World Bank crisis database,[6] we find that a systemic crisis only happens once every 42 years for OECD countries. If anything, that is an overestimate as the database includes relatively inconsequential events such as the stock market crash in October 1987. There are very few events to train a machine on if crises are that uncommon. To complicate matters, the structure of the financial system will be very different from one crisis to the next, so that each event will in many ways be unique.

The underlying policy objective of macroprudential regulation cannot be met by solving a sequence of local optimisation problems. Instead, it becomes necessary to solve a global problem, particularly challenging because the endogenous nature of systemic crises tends to result in seemingly unconnected parts of the financial system revealing previously hidden connections. Vulnerabilities spread and amplify through opaque channels, often in areas in which confidence is supported by ill-thought-through assumptions rather than in places known to create risk. This global problem is hard because the financial system is for all practical purposes infinitely complex and any entity, human or AI, can only hope to capture a small part of that complexity. The combination of sparse data, complex structure, uncertain and changing rules with high degrees of endogeneity make systemic risk an exceptionally difficult and quite possibly intractable challenge for AI.

[6] Laeven and Valencia 2012.

3.3 Looking for danger in all the wrong places

The reason risk management and regulatory systems are so well-suited for AI is because their focus is on exogenous risk with endogenous risk only a minor consideration. If something is well described by the notion of exogenous risk it is unlikely to be very dangerous from the point of view of systemic risk. An example is the stock market and we are well placed to manage the risk arising from it. If the US stock market were to go down by $200 billion today it would probably have a minimal systemic impact because it is a known risk. Even the largest stock market crash in history, on October 19, 1987, with a downward move of about 23%, implying losses in the US of about $600 billion, or $1.2 trillion in today's dollars and global losses exceeding $3 trillion in today's dollars, had little impact on financial markets and practically no impact on the real economy.

Endogenous risk captures the danger we do not know is out there until it is too late, and any macroprudential AI will have to address this. In the financial crisis of 2008, US subprime mortgages played a key role. What is however surprising is how small the losses in this market segment were. The overall subprime market was less than $1 trillion, and if half of the mortgage holders had defaulted with assumed recovery rates of 50%, the ultimate losses would have amounted to less than $250 billion. And that is an extreme scenario, actual losses were smaller. Still the mere threat of such an outcome managed to bring the financial system to its knees. A major reason is that these subprime mortgages were structured into collateralised debt obligations, CDOs, often with embedded liquidity guarantees. The problem was not the subprime mortgages per se, it was the usage to which they were put. While this information was available in fractured forms in the databases of the various financial institutions and supervisory agencies, nobody noticed the systemic implications of the maturity mismatches and liquidity guarantees until it was too late.

The human regulators at the time did miss the danger. Could AI have done any better? Unlikely. If there are no observations on the consequences of subprime mortgages put into CDOs with liquidity guarantees, there is nothing to train on. It is conceivable that an appropriately instructed AI would have become concerned in 2007 by scanning the global financial system for generic maturity mismatches and liquidity guarantees, noting that the CDOs were vulnerable to even small changes in correlated subprime mortgage defaults. The AI engine could have figured out the mapping between house prices, mortgage defaults and default correlations, the factors that determine prices of CDOs. It could have also noted the fragility of the structured credit

products to the evaporation of liquidity.

However, even if it is conceivable that AI could have made each step individually, the likelihood of putting all the pieces together is quite remote. This, however, is necessary for the chain of vulnerabilities to be discovered. We are asking for a lot, not only of AI but also of the national financial authorities who would have to allow such intrusive international supervision.

The ability to successfully scan the financial system for systemic risk hinges on where the vulnerability lies. Financial crises are driven by common factors well-founded in economic theory. Yet, the underlying details are usually unique to each event. After each crisis, regulators and financial institutions learn, adapt processes, and tend not to repeat exactly the same mistakes. When we examine the details of past crises it is both clear that each had unique aspects, and that most of these were missed at the time of crisis. Indeed it is almost definitional that each crisis triggers a sudden and painful re-evaluation of previously comfortable assumptions.

Here, the systemic danger emanating from an AI engine working for the financial authorities is that it will focus on the least important types of risk, those that are readily measured while missing out on the more dangerous endogenous risk. In effect, it will automate and reinforce the adoption of mistaken assumptions that are already a central part of current crises. In doing so, it will make the resulting complacency even more likely to build up over time.

While human risk managers and supervisors can also miss endogenous risk, they are less likely to do so as they have historical, contextual and institutional knowledge, reason well with theoretical concepts and consequently have some tools to handle it in a way that AI may not.

3.4 Artificial intelligence is procyclical

A main driver of financial instability is the procyclicality so inherent in the financial system. In boom times, market participants are especially willing to take risk, and it is easy to do because most constraints on risk, such as bank capital, do not bind very hard when times are good. This amplifies the financial cycle on the way up. When the cycle is trending down, actors become increasingly risk-averse. But this is also exactly when constraints begin to bind sharply, further amplifying the downwards movements. See e.g. Brunnermeier and Pedersen (2008) and Daníelsson et al. (2011) for an example.

The degree to which we react in a procyclical manner is to a considerable extent determined by how similar our perceptions and objectives are. Diverse views and objectives dampen the impact of shocks and act as a stabilising force, reducing systemic risk. Increased homogeneity in beliefs and actions amplifies systemic risk. Financial regulations and standard risk management practices inevitably push towards homogeneity. As control processes become more quantitative and sophisticated they tend to become more procyclical because data-driven risk estimates are at their lowest before a crisis and their highest immediately after.

With increasing sophistication, and particularly with AI, we inevitably see more homogeneity because it favours best practice and standardised best-of-breed models that closely resemble each other. After all, there is usually only one optimal solution suggested by any given dataset, and with increasing sophistication it will be approached more closely by all participants as noted by Watkins (2008).

All of this, no matter how well-intentioned and otherwise efficient, also increases pro-cyclicality and hence systemic risk. Regulatory coordination of AI engines will lead to further amplification as data exchangeability and equivalence standards will require use of AI engines with standard APIs and hence standardised measures of risk. We may consequently expect AI to increase systemic risk, even when AI is only used within the isolated context of risk management and microprudential supervision.

We may also expect crises to develop orders of magnitude more rapidly because decisions will be made on millisecond or nanosecond timescales rather than over weeks or days. The potential has already been illustrated by various "flash crash" events, though the work of the Foresight group has shown that as yet none has been close to systemic proportions, (Beddington et al., 2013).

3.5 Trusting the engine

If AI is to make inroads into policymaking beyond microprudential policy, it becomes important to correctly and exhaustively specify its objectives, both intermediate and ultimate, to prevent undesirable outcomes. Suppose I tell the machine to minimise $f(x)$. My true objective function is $U(x, z) = f(x)+z$, but either I am, ex-ante, unaware of z or it is simply too complicated to spell out. The AI engine might opt to minimise $f(x)$ but at the cost of maximising z. A human regulator with identical initial objectives will find out along the way that z also matters and update its objective function

accordingly. But what about the machine? There is an widely repeated story about a US naval AI being tested in a wargame. When the AI found that a convey was moving too slow for its taste, it solved the problem by sinking the slowest moving ships.

We have frequently seen the adverse consequences of ignoring important factors in past crises. During the Great Depression, the Federal Reserve was focused on moral hazard and inflation, ignoring the danger from deflation and failing banks. Similarly, the central banks before 2007 were primarily concerned with the immediate objectives of monetary policy, neglecting financial stability.

Even so, the human decision maker has well-known strategies for coping with unforeseen contingencies. As the presence and importance of hitherto ignored factors becomes apparent, she can update the objectives, making use of established political processes to impose checks and balances on the way such decisions are made. While AI might be able to do the same, we would have to trust it to make decisions in line with the objectives of its human operators.

This question of trust is fundamental. The longer we leave an AI engine successfully in charge of some policy function, the more it becomes removed from human understanding and the more we need to rely on trust. Eventually, we might come to the point where neither its understanding of the economic system, nor possibly even its internal data representations, will be intelligible to its human operators.

Paradoxically, as trust in an AI engine increases so does the possibility of a catastrophic outcome when, eventually, the machine is forced to reason about an unforeseen contingency. While AI will come up with some course of action, its analysis and conclusions might not agree with our human objectives. The consequences could be disastrous, perhaps a Minsky moment. This might not necessarily be the case. But we have no obvious way of entering into a dialogue with it in the same way a financial stability committee would consult with its experts. We might be forced to take its reasoning on faith, an outcome that is unlikely to be acceptable to the financial authorities.

The issue of trust is more relevant for macroprudential policy than risk management and microprudential supervision. The latter mostly execute low-level functions with clear objectives and limited damages in case of failure. With macroprudential policy, the underlying problem is highly complex, the objectives are ill-defined and the cost of failure potentially catastrophic, all characteristics that make AI not only less suitable but also more dangerous.

3.6 Learning by experiment

If we want to make full use of the abilities of AI to learn about its economic environment and discover successful policies, we need to allow it to experiment with different policy options. Only by trying out seemingly inferior actions will it be able to learn about the consequences of these alternative options. When learning about its environment, the AI engine will have to solve the classical trade-off between the exploitation of policies known to be successful and exploration of new courses of action. In practice, the engine will tweak its algorithms and see how its counterparts, algorithmic or human, react to these experiments. This is how AI systems such as AlphaGo Zero and others learn.

Experimentation with financial regulations, however, poses serious challenges. The AI engine will most likely be forced to follow predetermined rulebooks and level-playing-field considerations that sharply limit its the ability to experiment. Furthermore, some of the experiments that an AI might want to try out might, a priori, look too risky from its human operator's perspective. Finding the right parameters to control risk-taking by the machine and solving the optimal exploration versus exploitation trade-off will prove challenging. These problems may constitute a natural barrier for the idea of a autonomously learning AI policy engine.

3.7 Institutional setting

In many applications of AI what matters is how well the engine meets clearly defined objectives. Driving a car, winning in poker, defeating the Go world champion. This does not extend to the financial system. There are many stakeholders, each with their own set of preferences. The intermediate objectives, processes and constraints, can be as important as the ultimate objectives of financial stability and the efficient provision of financial services. A policy authority may want to ensure that financial services are provided to the most vulnerable segments of society, and typically the least profitable to banks, while the political leadership might want credit to be channeled to small and medium-sized enterprises. There is a large number of such intermediate objectives that are continually shifting, and often in conflict with each other.

Just as the financial system is composed of a number of different types of entities, such as insurance companies, very large systemically important banks, small banks, pension funds, asset managers and sovereign wealth funds, just

to name a few, so is the official sector fragmented into multiple national and international agencies, each cooperating and competing with each other. They may deliberately withhold information and impede cooperation in order to enhance the competitiveness of their own domestic financial industry. They might also do so to in order to enhance their own influence or because of political pressures they face at home. Each has a narrowly defined remit, while their domains often overlap resulting in turf fights.

Meanwhile, financial regulations are public information. High-level rules are decided on by governments and international institutions, and most rules involve extensive consultation processes. This implies that rules change slowly. The global body of international banking regulations has gone through three revisions, with decades between them, Basel I in 1992, Basel II and 2008 and Basel III coming in 2019. How would AI operate in such an environment? Does it take the rules decided on by the human regulators as given without any power to influence? That seems unlikely, as AI will increasingly inform the human regulators and become yet another input into the decision-making process.

National authorities may well limit data sharing and API interfaces for competitive reasons and even insist on incompatible AI engines and APIs. This would sharply limit the ability of any AI to function properly within the international regulatory environment.

The problem of the institutional setting might not be too difficult for the supervisory and risk management AI engines, as the problems are small and self-contained, and the intermediate objectives clear. When it comes to macroprudential policy, however, the challenges of the institutional environment become much more important.

3.8 What AI can do for policymakers

AI has the potential to be very useful for financial policymakers concerned with the overall operation of the financial system, its contribution to the economy and the risk arising from it.

1. It will be of considerable benefit to the microprudential authorities. The rulebook can be optimised and the supervisory process be made more robust and cost-effective;

2. AI could be instructed to scan for vulnerabilities meeting generic criteria, such as extreme maturity mismatches coupled with liquidity guar-

antees or the widespread use of trading strategies that could become disastrously harmonised.

3. It could scan the literature for new research, advising senior policy makers of promising new ideas;

4. It might be able to replace some applied research.[7] AI could even take over much of the model writing function, guided by high-level theories;

5. Finally, it could provide recommendations to the policy authority, based on its theoretical understanding of the system and provide conditional forecasts of its own behaviour.

In order to do many of these things it will have to justify and explain its reasoning, which remains a significant challenge. If it cannot justify its reasoning, advice is likely to be rejected.

4 Optimisation against the system

An AI engine working on the behest of the macroprudential authority might have a fighting chance if the structure of the financial system remained static or evolved in an exogenously determined stochastic manner. The problem is then simply is one of sufficient data and computational resources. But does not. The structure of the financial system is not static, instead it continually evolves because of the endogenous interactions of the agents that make up the system.

Agents working within the financial system typically have an incentive to increase the system's complexity in a way that is very hard to detect by others. There are many ways to do so, for example by creating new types of financial instruments that have the potential to amplify risk across apparently distinct parts of the system. These agents may want to do so particularly in areas where they think the controllers are not paying attention. Consequently, the problems facing the financial supervisors are harder than those typically encountered in games of incomplete information. Not only are the rules unknown but they have a consistent tendency to evolve in a manner hostile to the interests of the supervisor. Any rule that restricts risk taking must be continually defended against new channels of risk transfer that attempt to profit by circumventing or attenuating it. The rules of the game evolve in

[7] Chakraborty and Joseph (2017).

response to players' behaviour rendering their motivation and action space is endogenous. This implies that the complexity of the financial system itself is endogenous.

AI can of course track changes in the structure of the system. But to do so effectively, it needs high-level reasoning to understand what the changes to the system are material and what they imply, based on data that will initially be very limited. In order to do so, the AI engine would need to reason about the objectives of financial regulations, interpret these objectives and possibly even adjust them in light of the high-level objectives of financial policy and theories of financial instability. That not only creates issues of trust as discussed in Section 3.5, it also implies AI that is much more able than current incarnations.

Meanwhile, a large number of well-resourced economic agents have strong incentives to take very large risks with the potential to deliver them large profits, and they will disregard the potential for significant collateral damage to their financial institutions and the system at large. This is exactly the type of activity risk management and supervision aim to contain. These agents are *optimising against the system*, aiming to undermine control mechanisms in order to profit, and will do so by identifying areas where the controllers are not sufficiently vigilant. The "malicious agents" have an inherent advantage over those who are tasked with keeping them in check. They can construct their trades so that they cross the silos inherent in control processes. There will be many agents simultaneously engaged is such activities. While the AI engines might catch most, even almost all, it only takes one slipping through the cracks, provided it is large enough.

Even worse, a large number of malicious agents with shared exposures and risk concentrations, perhaps across multiple jurisdictions will be hard to identify. The consequence of multiple agents locally optimising can create serious endogenous risk if they find similar solutions. The resulting homogeneity in beliefs and actions can then give rise to spontaneous and possibly disastrous co-ordination of behaviour. Each agent might be relatively small, and if they are engaged in activities not been seen before, the national AI engine may not realise the danger. A global AI engine might be required to properly identify such risks, but such AI is unlikely to be created as noted in Section 3.7.

Each malicious agent only has to solve a small local problem, looking for unsupervised niches in the financial system. Their computational burden is much lower than that of the authority. Meanwhile, the regulator has to consider the vastly harder global problem. The significantly higher dimen-

sionality of this problem makes it intractable even when abstracting from the much lower computational budget regulators have compared to the aggregate computational resources they are facing.

Of course, human supervisors face the same problem. The market has a good idea of their objectives and ways of thinking which makes their behaviour predictable. But AI supervisors will likely be more rational and hence easier to predict than their current human counterparts. Paradoxically, known rationality in strategic settings often constitutes a vulnerability. This is especially relevant when it is common knowledge that the objective of the AI engine is to prevent the system from collapsing. While the human regulators will have the same ultimate objectives as AI, their reactions may be harder to predict, both day-to-day and especially under the extreme stress of financial crises. This lack of predictability is further amplified by the complex social structure that conditions their behaviour.

The problem of predictability is inherent in many applications of AI. For example, with self-driving cars, human drivers knowing that the AI will respond rationally can safely exploit this knowledge to get better positioning in traffic. However, this competition is simply about relative advantages and the cost of failure is local and small, and such drivers do not much influence the rules of the game beyond their immediate environs. Drivers do not build new roads that are designed to be ever-tougher for driverless cars to navigate, but financial market participants are both able to, and strongly incentivised to do exactly this. This makes optimisation against the financial system much more dangerous and harder to prevent.

AI systems are frequently tested against malicious agents, and designers of AI systems have developed a number of strategies for coping. Three avenues have proven to be particularly fruitful. Keeping the AI engine opaque so that outside agents do not know how it reasons. Continually evolving the engine so agents cannot learn how it operates. And finally, experimenting against outside users, aiming to learn how they reason and how to undermine their attacks.

These strategies are likely to be of limited use in financial supervision. Most of the rulebook is necessarily public information and the intermediate objectives change slowly and in a transparent manner, as noted in Section 3.7. Market participants will be to a considerable extent aware of how the regulatory AI engine operates and makes its decisions, while the AI engine has limited flexibility in how it can respond. This obviously also applies to human regulators but they have some institutional flexibility in how to address it. For the foreseeable future it seem unlikely that any financial authority

would be willing to grant AI similar autonomy.

The more AI moves into financial regulations, the easier it becomes for malicious agents to optimise against the system. They will have detailed knowledge of the objectives of financial regulations and its main control processes. Their work is further helped by the AI engine's inherent rationality. If we want to make AI more resilient against attacks it might be necessary to give it power over the rulebook with the ability to alter the rules and allow it to experiment. It will have to be given the option to randomise its reactions. These features will mostly be unpalatable to the financial authorities.

5 Conclusion

Artificial intelligence will be of considerable benefit to bank risk managers and microprudential supervisors. Their objectives are clear, there is plenty of data to train on and and exogenous risk is more important than endogenous risk. The AI engine can mostly be trusted to do its job subject to the monitoring of output. We will get more coherent rules and automatic compliance, all with much lower costs than under current arrangements. The main obstacle to the creation of a risk management/supervisory AI is legal, political and social, not technological.

This does not extend to financial stability, where AI will most likely miss out on the most dangerous threats by focussing on exogenous risk at the expense of endogenous risk. Learning will be slow because systemic events are rare and unique. Meanwhile, the efficiency of the microprudential/risk management engine has the unfortunate side effect of increasing homogeneity in risk estimation and responses, increasing pro-cyclicality and systemic risk.

The macroprudential AI engine must be able to explain its decisions in human terms in order to be trusted, though much of the usefulness of AI comes from its ability to find data representations that are inherently non-transparent to the human mind. The AI engine will be more rational than human regulators and coupled with the need for transparent financial regulations this will hand malicious agents, intent on optimising against the system, an advantage. At the same time, the computational problem facing a macroprudential AI engine will always be much tougher than that of those who seek to undermine it.

Ultimately, the increased use of artificial intelligence in financial policy may result in us becoming very good at managing day-to-day risk at the expense of tail risk. Lower volatility and fatter tails.

References

International Monetary Fund, Bank for International Settlements and Financial Stability Board (2009). Report to G20 finance ministers and governors. Guidance to assess the systemic importance of financial institutions, markets and instruments: Initial considerations. Technical report.

Arner, D. W., J. Barberis, and R. P. Buckley (2016). Fintech, regtech and the reconceptualization of financial regulation. *Northwestern Journal of International Law and Business Forthcoming*.

Beddington, J., P. Bond, D. Cliff, K. Houstoun, O. Linton, C. Goodhart, and J.-P. Zigrand (2013). Financial stability and computer-based trading. In *The Future of Computer Trading in Financial Markets: An International Perspective*, pp. 60–85. Foresight, UK Government Office for Science.

Brunnermeier, M. and L. Pedersen (2008). Market liquidity and funding liquidity. *Review of Financial Studies 22*, 2201–2238.

Chakraborty, C. and A. Joseph (2017). Machine learning at central banks. Staff working paper no. 674. www.bankofengland.co.uk/research/Pages/workingpapers/2017/swp674.aspx, Bank of England.

Danielsson, J. and H. S. Shin (2002). Endogenous risk. In *Modern Risk Management — A History*. Risk Books. `http://www.RiskResearch.org`.

Daníelsson, J., H. S. Shin, and J.-P. Zigrand (2011). Balance sheet capacity and endogenous risk. `http://www.RiskResearch.org`.

Financial Stability Board (2017). Artificial intelligence and machine learning in financial services Market developments and financial stability implications. Technical Report November, Financial Stability Board (FSB).

Fudenberg, D. and D. Levine (1998). *The theory of learning in games*. MIT press.

Governance, risk & compliance technology centre (2017). Platform research. http://www.grctc.com/platform-research/.

Shoham, Y. and K. Leyton-Brown (2008). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Silver, D., J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis (2017). Mastering the game of go without human knowledge. *Nature*. www.nature.com/nature/journal/v550/n7676/full/nature24270.html.

Watkins, C. (2008, Sept). Selective breeding analysed as a communication channel: Channel capacity as a fundamental limit on adaptive complexity. In *2008 10th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, pp. 514–518.

Willis Towers Watson (2017). Digitising financial services regulation: are we there yet? www.willistowerswatson.com/-/media/WTW/PDF/Insights/2017/06/digitizing-financial-services-regulation-are-we-there-yet.pdf.

Systemic Risk Centre

The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Tel: +44 (0)20 7405 7686
systemicrisk.ac.uk
src@lse.ac.uk