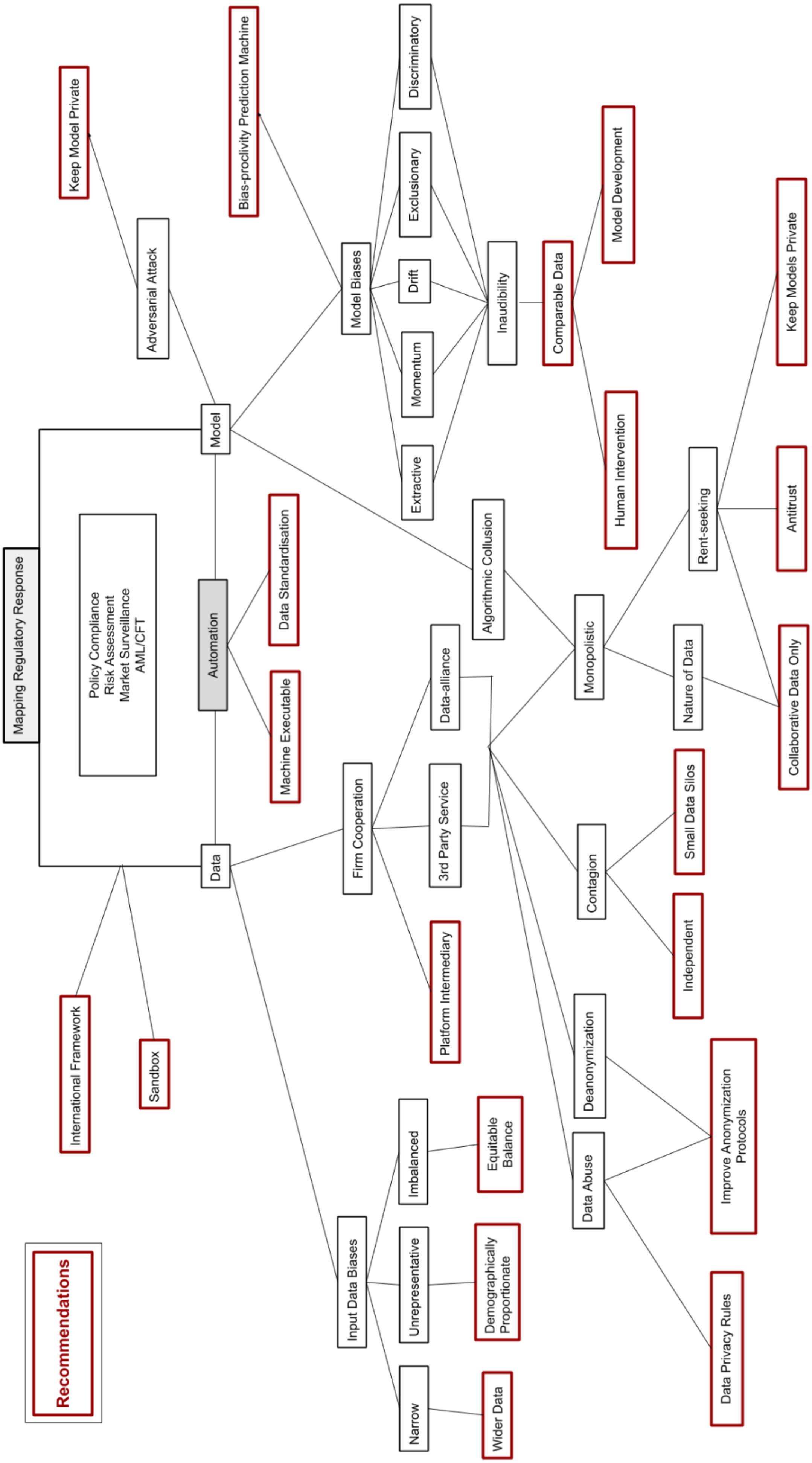
Financial Machine Learning Regulation

created: 18 Jan 2019; modified: 15 April 2019; status: in-progress; confidence: likely; importance: 9; by: Derek Snow

Regulatory Recommendations are Bolded.

Table of Content


Mapping Regulatory Response
Policy Compliance
Risk Assessment
Market Surveillance
AML/CFT
Model
Automation
Data
Adversarial Attack
Keep Model Private
Model Biases
Bias-proclivity Prediction Machine
Discriminatory
Exclusionary
Drift
Momentum
Extractive
Inaudibility
Comparable Data
Model Development
Human Intervention
Algorithmic Collusion
Monopolistic
Rent-seeking
Keep Models Private
Antitrust
Collaborative Data Only
Nature of Data
Data Standardisation
Machine Executable
Firm Cooperation
Data-alliance
3rd Party Service
Platform Intermediary
Contagion
Small Data Silos
Independent
Deanonymization
Data Abuse
Improve Anonymization Protocols
Data Privacy Rules
Input Data Biases
Imbalanced
Equitable Balance
Unrepresentative
Demographically Proportionate
Narrow
Wider Data
International Framework
Sandbox
Recommendations

Executive Summary

> *Regulators should not insist that you can explain how your AI system works. I think that would be a complete disaster...You should regulate them based on how they perform.* - Geoffrey Hinton, Deep Learning Researcher, University of Toronto, Google Brain.

Machine learning (ML) is the process of finding patterns in data, whereas artificial intelligence (AI) is the process of using ML to take an action. These patterns are useful for predicting future trends and for categorising units of data. Unlike traditional statistical algorithms, ML relies on a lot of data and computing power. Consequently, the success of ML, and AI at large, lies in the hands of companies that (1) are the closest to the automatable user-activity interface, (2) possess the right to access or generate large and diverse datasets. This access to large datasets is dependent on the ability to use cloud-based services, public and private cloud infrastructure, and cloud-hostable data.

Regulatory challenges posed by AI

Biases and contagion

The main failures of AI systems can be divided into biases and contagion.

Even though ML decreases overall bias and errors compared to manual systems, new errors and biases may emerge. Bias can occur at the model and data level**.**

Input data

- **Narrow**: input-data bias cannot be fixed by simply excluding variables that are thought to be discriminative, like gender. Despite narrowing down the dataset by removing the variable(s) associated with the bias, the bias may persist in remaining variables. For instance, correcting a gender bias in a lending model by removing a gender variable from the dataset is not sufficient; the model will find correlations in the remaining variables that correlate with gender, such as height. Instead, all variables that could highlight biases should be included in the model to clearly identify gender-related bias. Once identified, the difference in outcome can be measured and corrected for manually.[1]

- **Unrepresentative**: Many applied ML models use unrepresentative input data, mainly because of the so-called *labelling bias*.[2] To correct for this bias, datasets should be demographically proportionate, i.e. all subpopulations should be represented in the dataset in equal proportion to the user demographic of that technology.

- **Imbalanced**: If a specific group of individuals appears more frequently than others in the training data, the model will optimise for those individuals to boost overall accuracy, referred to as *algorithm bias*. For instance, Asians and African Americans often have a high error rate because they are not represented equally in the datasets

[1] It should be noted that the company with the most diverse, widest and largest quantity of data, *ceteris paribus*, will produce the least biased model.

[2] Labelling bias: certain categories and groups in the dataset have been annotated more than others. ML algorithms need annotated data to train their models.

for the mere fact that they are minorities. A US dataset would as a result of demographic proportion, only include 1.7% Native Americans; consequently they would perform worse compared to other ethnicities due to a lack of equal representation.

An equally balanced datasets is a more extreme version of equal representation than demographic representativeness.

Model

ML is purely backward-looking: predictions are based on the historical dataset the model has been trained with. This has implications for 1) current groups, 2) current groups with a state change, 3) current groups with feedback loop drift, and 4) groups that have not yet been included in the model.

- **Extractive**: ML models have the ability to personalise recommendations. Consequently, most of these models turn out to be economically extractive as they are trained to seek and sell to consumer as much as possible without considering the customers' financial position and cash-flow.

- **Momentum**: ML models can be slow to update for changes in circumstance ("state changes"). There are countless anecdotes of emotional suffering, like the mother-to-be who lost her baby, but kept receiving baby related advertisements.[3]

- **Drift**: ML models can drift towards discrimination. As AI systems self-improve and learn, they may acquire new behaviours that have unintended consequences, i.e. *self-fulfilling bias*. For instance, Google Translate defaults to masculine pronouns, due to the ratio of what is found in the English corpora on the Internet (2:1 ratio). Each time Google translates a new document and publishes it on the net, the ratio nudges upwards. In this loop, given enough time, the masculine can completely overpower the feminine as the effect is exponential.

- **Exclusionary**: Any recent changes not incorporated in the training set will lead to inadequate results. A group that has always been excluded will never be included under the ML paradigm, because the model is unaware of the group's existence. As an example, if the women's suffrage movement and their increasing role in the global labor force occurred in the age of AI, a ML model might still be suspicious of lending out money to women, even after they earn a comparable level of money to their male counterparts. The model preserves old stereotypes as a result of training on historical data.

- **Discriminatory**: The choice of error metric can favour certain group behaviours over others. For example, the root squared error (RMSE) metric penalises outliers more than the squared error (MSE) metric. Therefore, groups that exhibit more eradicate behaviour will be penalised. Additionally, the group with eradict behaviour is likely to have a certain demographic profile.

[3] See for instance: https://www.kidspot.com.au/birth/pregnancy/miscarriage/i-had-a-miscarriage-and-social-media-hurt-me-in-the-strangest-way/news-story/beaeb4be46b290dbb1e2aacae4caaa1d

Data abuse and privacy

- Data alliances help companies to relieve costs through data economies of scale. However, the emergence of data alliances makes it hard to hold individual parties liable for privacy breaches.
- Data has the tendency to accumulate at third party providers who are less incentivised to keep data safe. These entities might be incentivised to sell data to hedge funds, telemarketers and political campaigns to investigate and change subject behaviour.
- The quest for alternative data sources (e.g. social media, email, and point of sale data) to achieve greater personalisation can lead certain FIs to pursue extractive and manipulative data practices.
- Additional privacy concerns emerge when users are uniquely identifiable in digital identity systems. Furthermore, depending on the quality of anonymisation practices, anonymised data may easily be de-anonymised.

Monopoly formation

- Ineffective ‘democratisation’ of AI models and competitive data creates invisible monopolies. Larger companies are better positioned to take advantage of the supposed ‘democratisation’ of AI because they generally act as data gatekeepers.
- Data alliances are not always made public; firms can, as a result, collude to form data cartels.
- As a result of the monopolistic forces of data, the number of acquisitions will greatly increase, and multi-facetted collaboration will grow.[4]
- Algorithms can also adopt monopolistic behaviour. Recent research show that algorithms can collude without communicating with each other. After a few iterations these algorithms set prices between the nash price and monopoly price. They can look at the actions of the other algorithm and without concerted action increase their prices to extract value from customers. [5]

Auditability challenges

- Automated and customised ML models make individually developed models hard to audit, because each iteration of the model is unique.
- The ‘in-auditability’ of many of these models will trigger a general tendency for ML scapegoating. Automation decreases employee error and maximises executive error, which will drive executive’s tendency towards blaming the model.

[4] Data competition will force smaller firms, who do not have access to data-repositories, to compete by conceiving new algorithms and products leading to idea-stage acquisition, while traditional players will leverage their data, capital and existing customer base.

[5] This is a mathematical model, algorithmic collusion have so far not yet been identified in the real world; see for instance: https://voxeu.org/article/artificial-intelligence-algorithmic-pricing-and-collusion.

Regulatory arbitrage

- The ease with which companies can effectively translocate ML models developed in permissive regimes to less permissive regimes, creates opportunities for regulatory arbitrage.

Regulatory Recommendations

General remarks

- AI implemented in a company is nothing more than a function replicator or automaton; to this extent whatever the corporate motive is, the AI motive becomes. Thus, regulatory focus should be at the corporate decision level to ensure that AI enables a fair, stable and inclusive financial system.
- Automation will improve processes, without necessarily changing the nature of undertaken activities. This “non-novelty” justifies the adoption of a technology-neutral approach to regulation.
- Nevertheless, the number of activities and the quantity of automated activities will explode. Regulators will be required to switch from a rule-based to data centric approaches for both cost-cutting and accuracy purposes.

Biases mitigation strategies

- Model development automation[6], depth[7] and stacking[8] render ML inauditable in nature. As a result, biases are only identifiable using comparable data across different FIs. It then becomes an issue of least and most biased institutions. However, the potential of collective collusion among FIs might render the comparable approach obsolete.

- Instead of predicting actual biases, regulators can use FIs’ metadata[9] to create bias-proclivity prediction machines. Such a measure can help to flag institutions that need an extensive audit. For instance, in the case of lending companies, regulators can build up databases that include the ratios between application volume, approval and denial rates, average interest rates and data on complaints. These data can be fed into a ML model to assess the likelihood of bias without creating hard rules.

AI as part of the regulatory machinery

- Financial markets regulators can leverage ML internally to facilitate compliance supervision. A primary purpose for regulators would be to supplant costly compliance supervision with automated systems using an outcome comparison approach.

- Automated supervision can be used for sentiment analysis over policy documentation, automated risk assessments, market surveillance, and anti-money laundering and counter terrorism financing (AML/CTF).

[6] https://autokeras.com/

[7] https://www.nature.com/articles/nature14539

[8] https://link.springer.com/article/10.1023/b:mach.0000015881.36452.6e

[9] In this context, metadata describes the descriptive measures of the activity, rather than granular transaction data.

- Regulators can establish themselves as a data-platform intermediary to improve their data collection procedures and the automation of supervision. They can do so by creating a collaborative service where FIs share non-competitive data.

The risk of transparency

- Regulators should understand how *adversarial attacks*[10] and *reverse engineering*[11] can undermine future regulatory automation activities.
- Competitive data and most ML models should not be made publicly transparent as a way to mitigate adversarial attacks, rent-seeking, contagion, discrimination, blame-shifting and privacy concerns.

A participative regulatory process

- Together with institutions, regulators have to redevelop ways for FIs to share their collaborative-compliant information. Banks have already established their own collective that employs AI to spot fraudulent transactions on shared databases, such as EarlyWarning.
- Companies should focus on sharing compliant and fraternal data, i.e. collaborative data, instead of sharing competitive data, to decrease market failure. Sharing competitive data will generally exacerbate rent-seeking behaviour.[12]

International cooperation

- Differences in data standards and regulation are apparent across regions, this challenges institutions that operate globally. International frameworks are needed to manage common systemic issues that would have an impact across jurisdictions.
- Individual countries can test out different regulations; country-wide implementation can in itself serve as a pseudo-sandbox for partnering countries.
- The introduction of cross-country regulatory frameworks and a simple network analysis can help regulators combat regulatory arbitrage.
- Regulations should be broadly written to prevent firms from engaging in regulatory shopping.[13] Consequently, narrow regulations could constrain innovation and national competitiveness.

Futureproofing and flexibility

- Data regulations formulated in the coming years will have long lasting effects on financial markets. A review clause should be added to data regulation to ensure that it is flexible and future proof.
- AI calls for scientific reactive as well as preventative measures, such as adaptive sandbox strategies. Many small AI trials are needed to investigate the outcome and downfalls of automating intelligence before reliable preventative measures can be

[10] Adversarial attacks refer to an ‘attacking’ model used to find loopholes in a ‘supervisory’ model.
[11] Reverse engineering is defined as the recreation of a model based from inputs and feedback.
[12] However, there is an argument for an optimal level of competitive data sharing that benefits the consumer.
[13] The author recognises that FIs are not always inclined to play regulatory arbitrage for reasons relating to trustworthiness and reputation.

prescribed. Additional preventative measures include compliant data submissions (by FIs), the application of models to compare activity outcome datasets across FIs (on the regulator side), and the development of prediction machines (bias-proclivity, monopoly formation, etc).

Introduction

Some are convinced that to shape the future of financial regulations, regulators need in-depth knowledge of artificial intelligence (AI). That is not entirely true. Regulators instead need to collaborate with risk engineers to focus on the contagion, discriminative and systematic errors posed by companies using machine learning (ML) ("ML-activated companies"). For the most part regulators should remain technology-neutral while seeking to understand the behaviour and activity outcomes of ML-activated firms rather than audit their algorithms.

Notwithstanding, it is important that regulators pick up a certain knowledge in AI to understand AI-related activities that may require regulatory oversight, as well as for the introduction of AI in the regulatory processes. Ideally the monitoring and supervision of automated systems would be dealt with by regulators' own automated systems. The reason for this is the expected explosion of regulatory requirements due to increased automation. Regulators are expected to progressively develop automated intelligent systems to ease the internal burden that accumulates in the ML age. This article considers the potential of a new form of adapted regulation focused on data and comparative compliance. By being early in setting up internal AI and data science units, regulators would come to get in-depth first-hand understanding as to where these technologies tend to fail and to delineate between what is hype and what is reality.

-

Section 1: Regulatory Challenges

Regulatory Implications of Self-driving Businesses

In the future, smart and hard rules would drive a lot of a company's decision-making. It is not hard to imagine a large portion of company processes being driven by AI (colloquially dubbed a self-driving business). Using the analogy of AI as a company can help us understand the coming dangers of AI. Why? Because we have been facing the same issues from companies for millennia.

At its core a company, like an automated model, is a collection of supervised agents incentivised to pursue a goal. At first, the company draws a circle around the AI system and then the AI becomes the circle; in this circle the AI system has the same objectives as the company. This objective function is programmed into the core of the AI system. Therefore, like for a company, it is not worth regulating away AI as a generator of wealth.

Today, AI/ML-based solutions are being leveraged by financial institutions (FIs) for a variety of use cases – such as customer segmentation for improved marketing, cross- and up-selling, campaign management, client-facing chatbots, creditworthiness evaluation and credit score prediction, augmented products recommendations and personalised financial advice, investment and portfolio management, algorithmic trading, trade strategising and execution, dynamic portfolio rebalancing and capital optimisation.[14] The real question is: Should we regulate FIs that are pursuing these technologies differently from FIs that don't? And if so, how? Liability has often been the main point of contention when discussing AI regulation.

This paper argues that, there is no reason to not to hold the user of the technology in charge of any purported injustices and damage inflicted. Although AI can, like the industrial revolution, be a catalyst for new type of activities, as long as AI is programmed to imitate humans, it would mostly be an amplifier of existing technologies. Seen through the analogy of electricity, AI is not the discovery of electricity (cognitive power) but rather the cheapening of electricity. The real issues for regulators, thus, arise from the shift of responsibilities.

Biases and Contagion

Two trends will run parallel in the era of automation and optimisation: (1) misconduct will occur on a much larger scale and only visible through machine actions and behaviour changes; (2) operations conducted by self-driving agents will reduce conduct risks, resulting in proportionally fewer employee errors and more executive errors.

In other words, future automation does mean fewer conflicts of interest and better performance at a lower cost, but it comes at a cost of greater contagion risk as AI demands an increasing interconnectedness across domestic and cross-border systems. Additionally, despite an absolute decrease in overall bias and errors, new errors and biases will inevitably pop up and the next section deals with this.

[14] https://github.com/firmai/business-machine-learning: GitHub repository created to illustrate some of the business applications using ML.

Systematic Bias and Errors (SBE)

ML is purely backward looking, which means that changes in the environment, like the addition of a new previously non-viable group, will not be incorporated in the training data without human intervention. As an example, if the women's suffrage movement and their increasing role in the global labour force were to happen in the age of AI, an AI model might still be suspicious of lending out money to women even after they earn a comparable level of money to their male counterparts. The model simply persists old stereotypes as a result of training on historical data. Thus, without introducing this new segment manually into the training data the model would never seek and select for women due to algorithmic boundaries excluding such selection. In an ideal world, competitive firms would correct this bias themselves: it would be in their own interest to employ humans or develop alternative ML models approaching Artificial General Intelligence (AGI) that can correct for error in algorithmic decision-making. As a result, humans would still play an active role in the future, at least until the ML community develops real-time causal analysis to swiftly identify changes in the environment. It is also possible that a decision-making process informed by models may never be able to mechanically provide explanations for decisions or to self-correct for biases built into the design, making the symbioses between human and machine all the more important.

Different Types of Biases: Model and Data Level

Biases can occur at the model and data level. As an example, let's consider some implications for fair lending practices. Loan acceptance are primarily driven by the use of alternative data, which include social media, behavioural and educational data.

SBE as an input

1) Narrow: can we achieve a fully unbiased dataset? This is very unlikely. For instance, correcting a gender bias in a lending model is not as easy as removing a gender variable from the dataset. The reason being that the model will find correlations in the remaining variables. Height is one of multiple variables that correlate with gender so, by simply removing gender, height would be given more prominence in the model and the bias will persist. A counter-intuitive and opposite approach could partially correct for this bias: all variables that could highlight biases should be included.[15]

2) Unrepresentative: many applied ML models use unrepresentative input data, mainly because of the so-called *labelling bias*.[16] To correct for this bias, datasets should be demographically proportionate, i.e. all subpopulations should be represented in the dataset in equal proportion to the user demographic of that technology.

3) Imbalanced: if a specific group of individuals appears more frequently than others in the training data, the program will optimise for those individuals to boost overall accuracy ("algorithm bias"). For instance, Asians and African Americans have a high error rate, because they are not represented well in the datasets for the mere fact

[15] Note this is not the equivalent of data in error. Data with errors should be fixed before it is included. Once all bias relevant data is included the model can be adapted to marginally remove biases and discriminatory decision making.

[16] Labelling bias: certain categories and groups in the dataset have been annotated more than others. ML algorithms need annotated data to train their models.

that they are minorities even though they are demographically proportionally represented in the model. To best explain this phenomenon, a American dataset would as a result of demographic proportion, only include 1.7% Native Americans, and it is almost a certainty that these models, due to a lack of representation would perform worse compared to other ethnicities.

SBE as a model

ML is purely backward-looking: predictions are based on the historical dataset the model has been trained with. This has implications for 1) current groups, 2) current groups with a state change, 3) current groups with feedback loop drift, and 4) groups that have not yet been included in the model.

- **Extractive**: ML models have the ability to personalise recommendations. Consequently, most of these models turn out to be economically extractive as they are trained to seek and sell to consumer as much as possible without considering the customers' financial position and cash-flow.

- **Momentum**: ML models can be slow to update for changes in circumstance ("state changes"). There are countless anecdotes of emotional suffering, like the mother-to-be who lost her baby, but kept receiving baby related advertisements.[17]

- **Drift**: ML models can drift towards discrimination. As AI systems self-improve and learn, they may acquire new behaviours that have unintended consequences, i.e. *self-fulfilling bias*. For instance, Google Translate defaults to masculine pronouns, due to the ratio of what is found in the English corpora on the Internet (2:1 ratio). Each time Google translates a new document and publishes it on the net, the ratio nudges upwards. In this loop, given enough time, the masculine can completely overpower the feminine as the effect is exponential.

- **Exclusionary**: Any recent changes not incorporated in the training set will lead to inadequate results. A group that has always been excluded will never be included under the ML paradigm, because the model is unaware of the group's existence. As an example, if the women's suffrage movement and their increasing role in the global labor force occurred in the age of AI, a ML model might still be suspicious of lending out money to women, even after they earn a comparable level of money to their male counterparts. The model preserves old stereotypes as a result of training on historical data.

- **Discriminatory**: The choice of error metric can favour certain group behaviours over others. For example, the root squared error (RMSE) metric penalises outliers more than the squared error (MSE) metric. Therefore, groups that exhibit more eradicate behaviour will be penalised. Additionally, the group with eradict behaviour is likely to have a certain demographic profile.

[17] See for instance: https://www.kidspot.com.au/birth/pregnancy/miscarriage/i-had-a-miscarriage-and-social-media-hurt-me-in-the-strangest-way/news-story/beaeb4be46b290dbb1e2aacae4caaa1d

Progressive and Precipitous Contagion

Systematic issues precipitates and grows due to the monopolistic tendency of AI. Data wants to be shared to the point where all firms take the same information into account in assessing customer adequacy leading to certain groups being permanently excluded from market participation.

AI poses a new source of systemic risk. As algorithms get increasingly shared among multiple institutions, errors resulting from model miscalibration, such as miscalculation of credit risk) can quickly spread across all participating institutions. Like algorithmically driven flash-crashes[18], self-reinforcing models can lead to precipitous contagion having immediate consequences.

Data Abuse and Privacy

Data Alliances, Third Parties and Liabilities

Data alliances help companies to relieve costs through data economies of scale. The emergence of data alliances can be driven by corporate incentives or imposed by regulators. Such regulatory measures force a set of organisations to share their user data and have been implemented in several jurisdictions.[19]

The UK has been one of the first jurisdictions to adopt open banking, which requires that incumbent institutions share customers' financial data with third parties at the request of the customer. This push started in 2016, with a report by the Competition and Markets Authority that found "older and larger banks do not have to compete hard enough for customers' business, and smaller and newer banks find it difficult to grow"[20]. Across different parts of the world, governments are considering radical changes to their data-openness regimes. Australia, Singapore, Canada and Iran, among others, are actively considering some form of the open banking regulatory model, often mirroring the steps taken by the EU and the UK. These data regulations often extend beyond financial services and affect many different industries collectively. For example, the *G20's Anti-Corruption Working Group has identified cross-sector open data as a priority to advance public-sector transparency and integrity[21]*. Regulators should promote this form of collaborative data sharing as opposed to competitive data sharing.

In the United States, data-sharing alliances are more ad hoc than mandated, with individual banks building bilateral relationships with data aggregators. Regulators have not signalled an intention to implement regulatory frameworks to foster data-sharing across institutions similar to the ones established in the UK (Open Banking) and the EU (PSD2). However, the US Congress has been listening to testimony from large technology companies such as

[18] An extremely rapid decline in the price of one or more commodities or securities, typically one caused by automated trading.

[19] One caveat however: in creating this regulation, regulators should be aware of potential asymmetries that might occur. With large scale financial data redistribution, large technology firms can access financial data and use it alongside a wealth of other personal data, and gain a headstart in developing new AI applications for customers' finances. In such a scenario, financial institutions should also be given the reciprocal ability to access non-financial data from third parties (e.g. technology companies).

[20] https://www.gov.uk/government/news/cma-paves-the-way-for-open-banking-revolution

[21] http://www3.weforum.org/docs/WEF_New_Physics_of_Financial_Services.pdf

Facebook, Google and Twitter on the topic of privacy and data security, which could lead to the emergence of new rules.

The recent intensive harvest of data has led to the emergence of powerful third-parties that act as data service providers. These third-party data providers are a disguised form of data-alliance and will soon become critical in developing AI systems due to the data-hungry models underpinning performance. Companies will, as a result, increasingly rely on currently unregulated stakeholders. Third-party data providers will allow companies to plug into an existing data lake and model zoo in exchange of their user data. These companies that use third-party data providers will immediately outperform companies who rely on self-generated data and models. These third-party data providers, being the entities with all lucrative data, should become the prime targets for future regulation.

Third party data aggregators open up many points of concern that have to be addressed to avoid market failure and make it increasingly difficult to identify the source or entity that should be held liable. As companies increasingly rely on and get themselves locked into these mutualised services, significant innovations would become the only distinctive feature in market competition. When FIs use multiple models to inform their decision making its called *ensemble* models and where they use models as inputs to other models it is called *stacking*. Unlike an automobile where the component supplier that delivered faulty break-pads can easily be identified, with ML stacked or ensemble models, it will be near to impossible to identify the actor that has overstepped. Hence regulators may decide to hold the final provider liable for damages inflicted. However, this but this hasty remedy could be to the competitive detriment of the client interfacing firm.

Alternative and Paywall Data

Alternative data is any non-traditional data source that might provide additional insights over that of traditional sources. Alternative data can be intrusive and simultaneously legal. Data provider *Return Path* specializes in ‘volunteered’ (normally by agreeing to unread terms and conditions) personal email data covering approximately 70% of the worldwide total email accounts. By collating this with purchase email receipt data for around 5000 retailers, it offers analytics around purchase behaviour and consumer preferences. Similar transaction data is also available from *Slice Intelligence* and *Superfly Insights*. Are these methods to the benefit of the consumer or do they only provide extractive and manipulative power to retailers and hedge funds?

Alternative data can also be obtained from applications, like those created to help consumers manage their finances – such apps track spending patterns and advise their clients. These apps typically gain access to bank, investment, and retirement accounts, loan and insurance details, bills and rewards data and even payment transactions. This data is then anonymised, aggregated and sold to third parties. Firms in this category prefer to remain out of the media spotlight. *Envestment Yodlee* is a prime example, they have partnered with 12 of the 20 largest US banks and tracking around 6 million users. They sell credit- and debit-card transactions data investors and research firms, which mine the information for clues about consumer trends.

Data Privacy

The volume of data required to effectively develop AI raises data privacy concerns as consumer data is increasingly shared without informed consent. Further, AI’s ability to analyse data and identify users may in fact increase the sensitivity of data that was previously considered sufficiently anonymous. Data privacy, security privacy and data-protection regulations (e.g. GDPR) are placing new limitations on the collection, transmission and storage of personal data to address these risks. While fostering consumer protection, these regulations may impact competitive dynamics: data partnerships become increasingly difficult to manage as parties are held to stricter requirements.

A further issue is the use of ML models to de-anonymise user data. Crime scene investigation is a useful analogy to understand the notion of *deanonymisation*: Prior to the 1980s, criminals were not particularly concerned to leave their DNA at a crime scene, but after the development of DNA-kits a lot of offenders, previously unknown, have been de-anonymised. Likewise, AI methods are being developed to de-anonymise data to establish user identity.[22] The current lax way data anonymisation is treated, is to the detriment of user privacy.

Customisation and Personalisation

ML will bring customer customisation to another level. Marcos de Prado from Connell University and AQR Capital Management argues that future investment funds will be set up as a one fund per client model to satisfy individual needs of customers[23]. In the past, personal financial management apps were restricted to describing a customer's financial situation; they were unable to provide actionable insights and recommendations.

The next generation of these services (e.g. Clarity Money, MoneyLion) are using AI to offer mass advice and customisation to help improve customers' financial positions, such as refinance a loan, consolidate credit card debt or cancel certain recurring payments. For instance, Credit Karma, which has found success as a lead generator for loans, raised $500 million in March 2018 to build financial adviser tools and extend their control of customer experiences. Similar tools are being built for corporate clients (e.g. institutional investor dashboards).

A plethora of issues should concern companies about customisation. These tailored ML models incorporate individual's search record, but a sudden change in life circumstances may not appear on their radar and can lead to harmful advertising. One telling example is the case where companies like Google and Facebook kept trying to advertise infant products to mothers who have miscarried. [24] The same risk exists for financial ad targeting. A client may appear to the model as wealthy via his/her search results; the momentum of these previous research may persist while the client suffers economic disaster. In the worst-case scenario, financial advertising, no longer tailored to the client's reality can send him/her down a spiral of debt.

Monopoly Formation

Yeah, OpenAI was about the democratization of AI power. So that's why OpenAI was created...to reduce the probability that AI power would be monopolized – Elon Musk, OpenAI Founder.

Modern data alliances are as good as mergers. ML models are reasonably homogenised; hence data becomes the crucial ingredient. Data is the secret sauce to obtain future efficiency and profit gains. If data is shared it effectively leads to larger companies.

[22] https://arxiv.org/pdf/1801.05534

[23] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3197726

[24] https://www.kidspot.com.au/birth/pregnancy/miscarriage/i-had-a-miscarriage-and-social-media-hurt-me-in-the-strangest-way/news-story/beaeb4be46b290dbb1e2aacae4caaa1d

Depending on the setup, these data alliances need not be made public. Hence it can go under the radar compared to legal mergers. This is leading to a phenomenon that can be dubbed *monopoly by data*. AI is poised to create data-driven connections across industries and borders to deliver value. A large group of merged AIs can easily lead to a excessive concentration of market power. This large body of users enlarges the risk of contagion from massive data breaches which can lead to mass-manipulation campaigns. Cambridge Analytica and their psyops campaigns were a prelude of this.[25] Any AI monopoly will be able to execute mass-scale behavioural analysis leading to magical customisation and ease of use leading to a better quality product/service, more customers, and the ability to charge a higher price, and the positive feedback loop continues until excessive data ownership unwittingly trap consumers like a moth in a bath.

The democratisation of AI as a means to remove the monopolistic forces of AI is a futile attempt, and will in fact just do the opposite (i.e. advance monopolies). Data is the fuel and ML models are the furnace. You can democratise the furnace, but without the fuel it is useless. The recent advances in *open data* has allowed some smaller firms to compete against the larger powers, however, open data protocols have still been mostly to the benefit of large strategic players. The reason being that the open data initiatives are mostly sourced from public or non-profit entities instead of being scalped from large rent-seeking companies who generate and store a torrent of proprietary data. Big companies are competitively positioned to swiftly incorporate additional data into their established models and to use it to capitalise on existing clients and networks. Large players are those that have the privileged position to stand around the fire while it is being stocked by small and large firms alike. Data has become the invisible rent-seeking tool hiding in the *cloud's* pearly gates of heaven strictly guarded by Saint *Don't Be Evil* and friends.

Customer networks get entrenched as FIs learn your preferences, current state and general behaviour. This deepens further with the breadth of products/services on offer. Other factors include the quality and price of the offering and whether or not they also serve your friends and family (network effect). All of these factors benefit from a firm that is large (economies of scale), and firms that have existed for a long time, so as to benefit from bigger and more granular data sets to identify customer level preferences and behaviour. Therefore, the real question regulators have to ask is whether the policies they introduce helps to fairly distribute this customer network advantage to new entrants without undermining the incumbents, and more generally the economy. There would be no economically competitive benefit for widespread model and data democratisation without more radical network democratisation like mass corporate open data schemes.

Algorithms can also adopt monopolistic behaviour. Recent research show that algorithms can collude without communicating with each other. After a few iterations these algorithms set prices between the Nash price and monopoly price. They can look at the actions of the other algorithm and without concerted action increase their prices to extract value from customers.[26]

Auditability Challenges

For regulators, the centralisation of AI could prima facie ease the efforts of direct audits. Companies generally have two types of AI systems: (1) those provided by third parties and/or (2) internally developed AI systems. If the bulk of AI implementations are delivered by a handful of third-party service providers, then regulators might have enough resources to

[25] https://www.theguardian.com/news/series/cambridge-analytica-files

[26] This is a mathematical model, algorithmic collusion have so far not yet been identified in the real world; see for instance: https://voxeu.org/article/artificial-intelligence-algorithmic-pricing-and-collusion.

thoroughly investigate the underlying models for systematic biases, errors, and contagion risk.

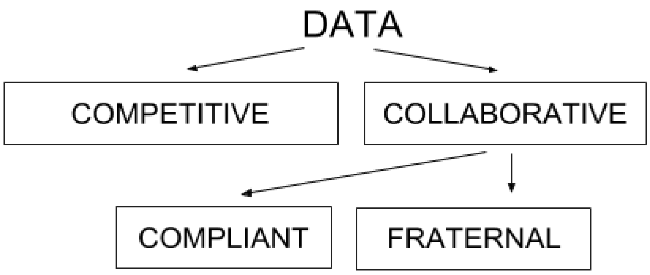
However, the largest companies have, for competitive reasons, made their ML libraries open to the world for use. Hence it has become increasingly cheap and popular to implement your own model based on their infrastructure and programming paradigms. And even if this democratisation effort doesn't stall third party service providers, providers still have discretion to use a combination of multiple models to inform a single prediction tas, all of which further discredit direct audits.

Models are also slowly developing a mind of their own, making it even harder for regulators to identify biases, errors, and contagion risks. Technologies like AutoKeras and Google's AdaNet can now automatically create context specific ML models. And to audit these models you would have to investigate each new generation of an automated model as opposed to the infrastructure of the automated system.

And even if regulators seek to audit these models, their lack of interpretability poses further issues. The complexity of AI models undermines traditional regulatory frameworks that rely on an expectation of transparency. For example, the Fair Credit Reporting Act requires that companies notify a consumer if consumer report information is used to deny credit. But with AI-enabled firms might find it difficult to provide the rationale for such decisions because the importance of each factor can't be reliably disentangled from other inputs. Lastly, there are thousands of models in action across industries and they differ greatly across and within industries. The sheer quantity of models should scare regulators away from this approach. As a result, auditing individual models would be not feasible, regulators would have to rely on an outcome-based approach, which is dealt with in the *Adaptive AI Regulation* section.

Data Competitively Defined

There is a distinction between *competitive* and *collaborative* data. *Competitive data* are generally used to generate more revenue. The sharing of *competitive* data can lead to increases in market failures as a result of its market concentration effects. An example would be the sharing of customer characteristics for the purpose of customer loan approval. *Collaborative data* are generally used to reduce unnecessary expenses. *Collaborative* data can be either *Compliant* or *Fraternal* data. *Compliant* data are mandated by the regulator with the purpose of decreasing market failures. An example would be a timestamped list of potentially fraudulent transactions. *Fraternal* data are not regulated or mandated but can, nonetheless lead to the fairer treatment of customers across different sectors. An example would be a dynamic list of clients who are in-fact, contrary to public credit scores, stand-up citizens. Collaborative data should be shared across sectors instead of within sectors to avoid obtaining a competitive flavour. It is therefore not beneficial to share all data within sectors.



Section 2: Regulatory Recommendations

AI implemented in a company is nothing more than a function replicator or automaton; to this extent whatever the corporate motive is, the AI motive becomes. Thus, **regulatory focus should be at the corporate decision level to ensure that AI enables a fairer, stable and more inclusive financial system as it risks doing the opposite without supervision**. Automation will improve processes, without necessarily changing the nature of undertaken activities. This "non-novelty" justifies the adoption of a technology-neutral approach to regulation.

Biases Mitigation Strategies

Correcting Biases

By spotting differences in activity outcomes among FIs, regulators could catch rogue perpetrators in the act before they induce permanently bad outcomes. FIs will be required to produce outcome logs about the customers served. These will prove that their models are free from bias and demonstrate their ability to detect and prevent models that discriminate against or exclude marginalised groups and individuals. **Instead of auditing models, regulators should focus on auditing standardised data outputs from collaborating institutions**. The standardisation can further help with the processes towards the automation of compliance supervision through big data.

However, this approach may lead to unintended consequences. Companies, who have stepped in a bias trap and are asked to change their client demographic, are unlikely to correct their original model's prediction. Instead they will create a new model to select the most profitable clients from the mandated quota. Thus, although the problem will be solved on one dimension, e.g. race or gender, it will still select for people with high incomes and socio-economic status within those groups. In other words, businesses will simply take on customers who are the closest to, but did not fully cross, the acceptance threshold. One solution is for **regulators to ask for granular model input data. Regulators could then themselves identify fine tuned subgroups that are least likely to be taken on-board because of companies' short-term profit motivated model.** Quotas set in a very unidimensional basis will ease the compliance burden on the regulator, but it will lead to substandard outcomes. It's unclear how to deal with this; the before mentioned step that involves regulators' self-identification of sub-groups would be very burdensome. Depending on the type of bias, correcting it could have a positive or a negative impact on a company's accuracy and profits. A *self-fulfilling bias*, whereby prospective customers with new circumstances are not accepted because they don't fit the pattern of past clients, if not corrected can persist indefinitely to the detriment of the company's profitability (i.e. will miss out potential customers). Therefore, correcting for self-fulfilling biases improves the company's performance.[27] *Mandated biases* are imposed by regulators onto companies, such as the inclusion of an undeveloped class. Conversely to self-fulfilling biases, mandated biases can lead to short term loss (i.e. undeveloped class leads to more losses) and has uncertain consequence in the long run: could either lead to a long-term loss (economically inefficient quota) or long-term gain (by developing a new consumer class).

[27] There is currently no good method to account for new users. The best is to test these users more theoretically or empirically using traditional models and then to progressively create a dataset of performance for this previously excluded group.

What is important to note is that, sans the persistence of inherent discrimination, these models perform very well in removing taste or preference bias as their error metrics are, to the most extent, profit-centric. This doesn't mean that the model removes statistical biases. It only means that these models optimise on nothing but profit, so it does not care about skin colour, gender or age apart from how it informs profit. Therefore, biases are financially justified (statistical discrimination). In saying this, FIs do have the tendency to ignore long-term economic thinking when building their financial models. To this point serving the underprivileged might in fact help to create a new future financial class and lead to better economic success. Therefore, policy advisors might seek to provide guidance at the company level where deemed economically appropriate. Some of the extra profits generated from improved performance and accuracy from AI can be used to subsides the disadvantaged.

Another important point for regulators to understand is that the company with the widest and largest quantity of data, *ceteris paribus*, would produce the least biased model. Therefore, the reduction of bias is also the result of the quantity and type of data. One potential method of regulation would be to, instead of inspecting all lender models, ask for granular lender data and to reproduce their results with the bias-reduction best practice. This approach might, however, place too much of the burden on regulators.

Alternatively, **regulators can use FIs' metadata (data about the FI themself as opposed to the FI's activity) to create bias-proclivity prediction machines and flag institutions that need an extensive audit.** For instance, in the context of lending companies, regulators can build databases that include among other things the ratios between application volume, approval and denial rates, average rates and even data on complaints. These databases can be fed into a ML model to assess the likelihood of bias without creating hard rules. This form of probabilistic regulation will, within time, become essential to regulatory efficiency and effectiveness.

Preventing Contagion

To tackle the new form of systemic risk created by AI, several options are available to regulators. First, **regulators should investigate past failures of AI to identify and mitigate downside risks.** Regulators can learn from case studies such as where platforms use AI to start self-regulating content. Regulators can also learn from each other.

Second, regulators can experiment with their own isolated machine learning simulations to exponentiate their understanding of the potential downside risks; within this process they should start small by drawing on a few FIs to help validate their system. Third, preventative measures have to be built into the regulatory framework, which could take the form of **regulators' support for small data silos for competitive data and large data silos for collaborative and compliance data.**

Comparables

A big issue in AI is the tailored experience. **Biases would only be identifiable using comparable data across different FIs, which would ultimately consist in comparing the least and most biased institutions**. The main challenge for regulators to use comparable data thus come from the data collection process and standardisation practices. **A regulator would have specific requirements for the data collected from FIs, which might cause some teething issues at the start.** It is important that the regulator does the necessary monitoring to ensure that this activity data is accurate.

It should be noted, that within this comparable approach the potential of collective collusion might render the comparable approach obsolete. By investigating the activity levels of multiple homogeneous firms, you would be able to obtain ‘optimal’ activity estimates. Some firms might differ from the average; this can be as a result of identifying a more competitive niche or can be the result of discriminatory behaviour. When all firms discriminate then the optimal level itself is discriminatory, i.e. if all lenders exclude a certain demographic, then you won’t notice the exclusion algorithmically.

To better further explain the comparable approach, we can use lending companies once more as an example. For lending companies, we can investigate the following activity outputs to assess risk and conduct: 1) look at the approval and denial rate by prohibited basis category, 2) identify the average price by prohibited basis category, 3) consider application volume and 4) give special attention to increases in the level of authentic complaints.

AI As Part of the Regulatory Machinery/AI Regulatory Toolkit

There is a growing consensus across the globe that regulators will be able to maintain a high-standard of compliance and prudential standards through effective AI, ML and big data solutions. It is time for regulatory organisations to proactively plan to migrate from legacy systems to the latest regulation technology and educate their staff to embrace the power of AI.

Data Approach

Regulators should focus on moving away from rules-based towards data centric approaches. In a recent speech, FCA[28] representatives indicated that they are moving away from a rules-based, prescriptive approach of regulation to a more data-driven, predictive place where regulators would use data to objectively assess the inherent risk posed by FIs.

Currently, regulators are not paying attention to the vast amount of data available in the public space. Yet this might prove crucial in their endeavour to move towards a data-centric system. Regulators should invest in obtaining alternative datasets as a means of measuring compliance and policy effects. This can be done by purchasing datasets or leveraging easy to use technologies like web scraping to source data. For instance, real estate activity like volume and prices can be gathered from a range of real estate websites and used to guide interest rate and policies like property loan-to-value restrictions.

Machine Executable Format

Regulators should actively partake in converting their rules into unambiguous machine-executable format. This essentially means that rules should be so written that compliance with the rule can be easily manageable by computers. The most efficient way to do this, is a form of probabilistic regulation using AIs that have judgement and can understand intent. An example of such a rule follows,“In the event that a FI exceeds a probability threshold of 60% for doing X for more than a month while being aware of underlying change in Y, an audit will follow”. This rule is in a data-centric machine-executable format.

[28] https://www.fca.org.uk/news/speeches/ai-and-financial-crime-silver-bullet-or-red-herring

Another machine learning format is more binary in nature and would have been a decent automated solution before the age of AI. This system creates hard rules and minimise the use of judgement and requirements like intent; for example, “In the event that a FI has successively (more than twice) breached threshold X and audit will follow”. There is room for both of these automated approaches, depending on where intent is important. More human centered approaches should also considered; the context and importance of rule matters when considering different rule making approaches. However, regulators might want to focus on one form of rule writing to make these rules more easily interpretable.

Systems can be set up to flag FIs compliance directly by leveraging AI/ML solutions. One such example is machines to scanning regulatory filings for inconsistent submissions. Regulators can both make use of non-cognitive robotic process automation tools and cognitive activated systems. Doing this would facilitate the investigation of firm’s compliance as their actions become machine assessable. This automated can’t be done by hard rules, there is as there is a requirement of automated intelligence. You can use hard rules, but it would simply limit accuracy and performance and take much longer to develop. By necessity probabilistic approaches might be the only way forward.

Automated Monitoring

Regulators should, as one of their criteria in deciding among competing policies, identify how automatable the monitoring of a policy is.

Automated Policy Compliance

The two primary reasons for automated compliance is the benefit of timely regulatory responses and cost advantages. A lot of the systemic inefficiencies are highly correlated; the suboptimal process at one institution has knock-on effects on other institutions and across the ecosystem. If unchecked, there is a threat of a system-wide contagion. Timely response is critical to the resolution of these threats and automated compliance can help to significantly improve timely response.

Regulatory priorities stretch beyond leverage and capital-adequacy requirements and increasingly focus on collective issues, such as financial crime, privacy and data security. New regulatory requirements that emerge to address these areas will place pressure on operating budgets for institutions and regulators themselves. Automated compliance can help to create unbiased and data centric rules that can also help decrease FIs’ compliance cost. Cumulative financial penalties for non-compliance between 2009 and 2017 totalled $342 billion globally, with 89% of industry executives around the world expecting continued increases in compliance costs from 2017–2019. **To cut costs on the side of both enforcement and compliance, regulators should consider new technologies as a means of implementing their own systems to take care of regulatory slack.**

The recent (25/05/2018) GDPR regulation require firms to clearly explain how users’ data is collected and with whom it is shared. It further requires firms to explicitly consent to retain and process customers’ data. This type of policy and many others around the world can be more easily monitored using AI. For example, FIs’ privacy policies can be compared against EU GDPR policies using advanced NLP technologies to perform some of the following tasks:

1. Identifying 3rd parties with which user’s personal data is shared using entity recognition.
2. Assessing the readability and complexity of the policy.
3. Getting a sense for the sentiment of the policy document (giving special attention to coercion)

4. Identifying whether users' consent is implicit or indeed explicit (as mandated by GDPR)
5. Privacy policies can then be rated across various FIs and those who fall below a threshold can be flagged and sent to supervisors and agents for further investigation.
6. The actual threshold and rating metric would have to be set by humans, after which the system becomes fully automated.

Automated AI have already been implemented in the financial arena. Central banks in Russia[29] and China[30] have both invested in these technologies to improve their decision-making quality. The Russian Central Bank created a new ML economic indicator to assess national economic activity using online and other sources, while China has looked at ML as a means to identify and defend against market and sector financial risks. If this trend persists, central banks are soon to use predictive ML models to identify changes in short-, medium- and long-term economic trends.

Policy departments can create space for machine adjustable variables to advise decision-making. ML models can act quickly by scouring past data. The US Federal Reserve commented that ML lets the available data speak for itself, potentially revealing important relationships that have not yet been identified by theorists.[31] A little bit more speculative, but in the future we can imagine central banks that allow for a component of their interest rate to be managed and executed by automated AI systems. Real-time transaction level data can be fed in from central banks' and relevant governments' rich databases that reports on consumer behaviour, unemployment rates, government spending and production levels.

This is especially true for central banks that do simple inflation rate targeting such as the USA, NZ, UK and Canada. The intelligent use of real time and and historical data points can lead to significant efficiency gains. Imagine a policy rate that fluctuates on a daily basis to assist economic growth and other purposes. Naturally there are disadvantages to automating and giving up direct control of these important economic levers, but this topic will be left for future discussions.

Automated Risk Assessment

Like any automated process, once the correct data pipeline has been established the rest follows quite easily. Once data is injected in the ML models they can, for example, speedily assess FIs' (especially banks) metrics and features to assess the level of risk and whether or not they comply with liquidity and capital requirements. Such systems can be used to study and flag institutions with anomalous looking indicators without much human intervention.

Automated Market Surveillance

It is relatively undemanding to use ML solutions to identify and monitor market misconducts. Automated market surveillance can be comprehensive and look at areas of market manipulation such as pump-and-dump schemes, insider trading and advisor misconduct, among other things. Datasets need to be large and diverse. As it stands there are a lot of data already available that might be fit for purpose. A good and simple start would be trading firms' and exchange data. A second option would be access more unstructured data such as

[29] https://financefeeds.com/bank-russia-embraces-machine-learning-heart-new-economic-indicator/
[30] https://www.bloomberg.com/news/articles/2017-12-18/central-banks-are-turning-to-big-data-to-help-them-craft-policy
[31] https://www.frbatlanta.org/-/media/documents/news/conferences/2017/1102-financial-regulation-fit-for-the-future/wall.pdf

conference calls, which can be easily converted into a machine-readable format. **Regulators should, however, remain aware of the quality of the data and try to follow strict standardisation protocols.**

It is also possible that some manipulative activities can span multiple types of known activities (like spoofing and layering) so as to, in aggregate, stay under the radar. By using ML models, regulators can go beyond the mere identification of market manipulation practices. As an example, FINRA[32] established a AI project to investigate cross-asset manipulation by investigating investors who holds underlying options positions in an attempt to move the market.

Instead of hardcoding rules, with ML regulators can perform cluster analyses to identify and flag anomalous behaviours. In the same way that marketing companies segment customers into different categories by performing cluster analyses over large datasets, regulators can cluster FIs based on different compliance behaviours. These tools can be developed to spot anomalies and outliers from structured and unstructured data. For example, the Division of Economic and Risk Analysis of the US SEC looked at patterns of regulatory filings of investments advisors and unsupervised topic modelling and tonality analysis machine learning techniques (basically intelligent textual analysis and clustering) to uncover strange behaviour and idiosyncratic risk.[33]

Automated AML/CFT

Investigating money laundering and fraud is extremely costly and time-consuming. There is an extremely large amount of suspicious activity reports filed by FIs. Institutions over-report ‘disclosures’ to remain on the safe side, which leads to a large number of false positives for the regulators to deal with. In such scenarios, **ML models can be developed to fairly confidently predict the likelihood of clean transactions, which in turn allows for shortlisting truly suspicious transactions.**

By simply focusing on individuals, regulators may lose touch of transaction flows. By taking a step back and looking at transaction networks, group complicity can be observed. **Techniques like network analysis can be used in combination with ML tools to identify and address fraud and AML concerns.** An example would be to compare bank transactions with regulatory sanction lists and unstructured data to uncover relationships and detect money laundering patterns. The AUSTRAC[34] agency have used such steps in identifying previously undiscovered money laundering networks.

Investigative Automation

ML models can also be used to investigate specific instances of regulatory concern. Ad hoc models can be created where they are shown to be economically efficient to investigate individual matters of concern. The ASIC[35] has for example used ML and web scraping technologies to unearth potentially misleading conduct on accountants websites for self-managed superannuation fund activities. Even though they might use it again in the future, the purpose is not for it to be a permanent monitoring tool, but rather a research tool.

[32] https://www.marketsmedia.com/finra-taps-ai-to-stop-mini-manipulation/

[33] https://www.sec.gov/news/speech/bauguess-big-data-ai

[34] https://www.arc.gov.au/news-publications/publications/making-difference-outcomes-arc-supported-research-2016-17/new-software-detect-money-laundering

[35] https://download.asic.gov.au/media/4064271/greg-medcraft-speech-corp-governance-discussion-group-published-3-november-2016.pdf

The Risk of Transparency

There are two main drawbacks of model transparency.

- A publicly available model opens up the potential for the use of GAN models to exploit regulations for financial gain. Hence companies would not be willing to do this.
- Making models openly available can also lead to customers bending the truth in their applications to ensure that they do indeed meet the minimum threshold of, for example, a loan.
- Both of these misuses would be noticeable in the clustering that occurs just above the minimum thresholds for loan acceptance.

Further, even if the model is not publicly available but allows for user query-feedback process (input output), then the ML model can effectively be reverse-engineered. This has become very in vogue and is in fact not that hard to do[36]. For example, filling out characteristic for multiple online loan application while recording the outcomes. The recorded labels and features can then be used to reconstruct an ML model. In essence, the evidence points away from model visibility and open source development strategies.

Ease of Use for Regulated Entities

ML and good user experience (UX) design can be used to significantly improve overall compliance and user satisfaction. **Regulators should by all means possible, foster compliance by facilitating firm's and users' access to regulatory information.**The FCA in the UK collaborated with Corlytics to create an intelligent, easily searchable regulatory handbook.

Further improvements can be made with their client and/or user complaints interface. AI technologies like chatbots will be well-suited for this purpose. For instance, FINRA[37] has noted that chatbots has as a potential use case in broker-dealer relationships. These chatbots can further be utilised to conduct sentiment analysis of customer complaints and topic modelling to flag complaints according to specific categories, e.g. unjust discrimination by a FI.

Regulatory Response to Adversarial Agents

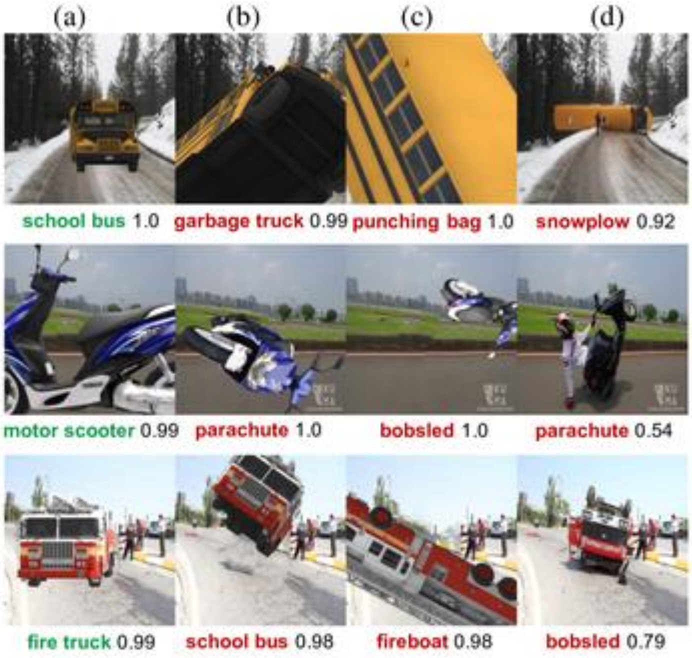
Generative adversarial networks (GANs) are two ML models pitted against each other (hence the word “adversarial”). In simple terms GANs attempt to simulate reality. An analogy can be drawn between a hacking defence system that has to stay one step in front of hackers that are simulating honest intent. A notable and recent use of GANs is Christie’s auction of an AI generated painting to simulate art between the 14th and 20th century[38]. GANs are also the technology behind voice, video and image replication.

As noted by the US treasury, while ML is currently being used to enhance fraud protection, it could potentially be used to circumvent these same fraud detection capabilities. Researchers have shown how GANs can be used to fool a Google AI model into believing that a

[36] https://arxiv.org/abs/1609.02943
[37] http://www.finra.org/industry/special-notice-073018
[38] https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx

helicopter is a rifle (see other examples, **Figure 1**).[39] Similarly, in the future a publicly known regulatory AI model can be fooled into believing that non-compliance is compliance from perfectly machine massaged data in an adversarial attack. This would lead to 'battles' between automated systems, where one system would update to fill a loophole as the other systems seeks another way to represent the data so as to identify another loophole. Adversarial agents are going to attempt this in two ways: first, they will try and simulate automated regulatory ML systems using data available to them. When data is not available to them, they can seek to reverse engineer the automated regulatory system by creating 'fraudulently' flagged instances and receiving feedback. In this way they can discover model decision boundaries and probability thresholds.

Figure 1: How GANs can fool an AI model



Link

It is important to understand that the agent does not need regulators exact data. Instead the agent simply needs the public record of all companies that have and have not been flagged for a certain regulatory issue and then retrofit *any* correlative data they might have for these companies. If the model achieves great success on an out of sample dataset then this model can essentially be used to decipher correlated features with the 'proprietary' regulatory data.

Such strategies have been observed in other areas. For example, Moody's and S&P rating agencies can be reasonably successfully reverse engineered using publicly available data and ML models. ML models have been successful at replicating bank analysts and credit rating agencies using publicly sourced data.[40]

Here is an explanation of a possible future scenario. In a standard ML model, managerial or financial accounting data can be fed into the model with data labelled as fraudulent and non-

[39] https://www.wired.com/story/researcher-fooled-a-google-ai-into-thinking-a-rifle-was-a-helicopter/
[40] See the example of random forest by Moody's https://www.moodysanalytics.com/risk-perspectives-magazine/managing-disruption/spotlight/machine-learning-challenges-lessons-and-opportunties-in-credit-risk-modeling

fraudulent. Future observations can then be fed into this system to predict the likelihood of fraud. With a GAN model, instead of company accountants playing around with financial data to ensure that the predicted level of fraud remains below regulated thresholds of materiality and/or probability, the model does this by constantly probing the regulatory model.

What can regulators do to fight off competing models?

1) **To fight off these attacks it's first of all essential for regulators to keep their models hidden**, as it is extremely easy to just copy a complete model (intelligent system) on a pen drive and distribute it to the highest bidder. What makes this more pressing is that these models are not exclusionary, the leaked model can be widely shared amongst FIs. To rub further salt in the wound, it's very hard to prove whether or not FIs are using the model.
2) Regulators would however be able to measure the intensity of adversarial attacks. Similar to the use of a funnel plot used in research to identify publication bias, regulators can investigate the clustering of companies who pass regulatory inspection just above the selection threshold. Even though these companies could have legitimate reasons for being that close to the thresholds, companies close to the threshold can still be sent warnings.
3) **Regulators should prioritise the use of privately available data to fight off some competing models**; as a corollary, they should also refrain from making this data public.
4) **Regulators can act in unsystematic ways to 'fool' these agents: regulators can simply dupe GAN's** by simulating and distributing false information about non-existent companies and regulatory breaches.

5) GAN's are the perfect automatable agent for spotting regulatory loopholes. Hence, regulators themselves can also use this technology to test the strength of their regulatory defence.

6) Lastly, regulators should be aware that the biggest pressure will come from third party services who would have accumulated enough data to execute these types of attacks. These companies can act under the guise of cheap ML service providers while they are in fact harvesting data that can be used to undermine regulators and select FIs in favour of others.

In the future it will not be enough to simply use ML models to flag suspicious transactions. Although this approach provides a good line of first defence, regulators would also have to investigate entities that try to bypass regulation by duping the system. **Regulators can have both a proactive strategy by trying to detect adversarial attacks, or they can investigate systematic biases in submissions and then investigate those clustering around the selection thresholds.**

Scott Bauguess, Chief Economist of the US SEC, gave a speech on the risk of AI and ML[41]. He stated that this technology will no doubt make the risk assessment process more efficient and effective, but is not likely to replace human judgment in regulation of financial markets. The same conclusion can also be drawn from the above analysis of GANS. The existence of GANs means that the future of regulation is indeed one where models will be fooled to the extent where humans would not. In its current form GANs are, however, still reliant on regulators adopting automated AI models and the collection of data that follows from AI adoption.

[41] https://www.sec.gov/news/speech/bauguess-big-data-ai

A Participative Regulatory Process

As a first regulatory response to liability, privacy, and competition issues posed by third-party data providers, this paper argues that third-parties should not be allowed to store any non-encrypted and non-anonymised data for the several reasons, including the following:

- Third-party data providers draw a safety-blanket across multiple companies, effectively creating a group of rent seekers;
- The bigger the alliance, the more homogonised the models, the less competitive the market and the more likely it becomes for discriminated parties to be fully neglected by the market;
- A single breach can have disastrous implications for third-party data providers who accumulate a large amount of data;
- Data can be used to dupe regulators with adversarial attacks; and
- FIs will likely shift the blame on third-parties' data/models to absolve themselves from any biases or errors.

A less stringent option would be to force third-party data providers to delete the data once transferred to the improvement of data-reliant models.[42] Another alternative would be to allow third-party providers to use the data only improve models and not to host data, and, as such, avoid being a harbinger for privacy issues. Thus, there should be a big push away from data-alliances towards model-alliances. Model-alliances do not eliminate the monopolistic and blame-shifting issues, but they are a first step towards addressing privacy concerns.

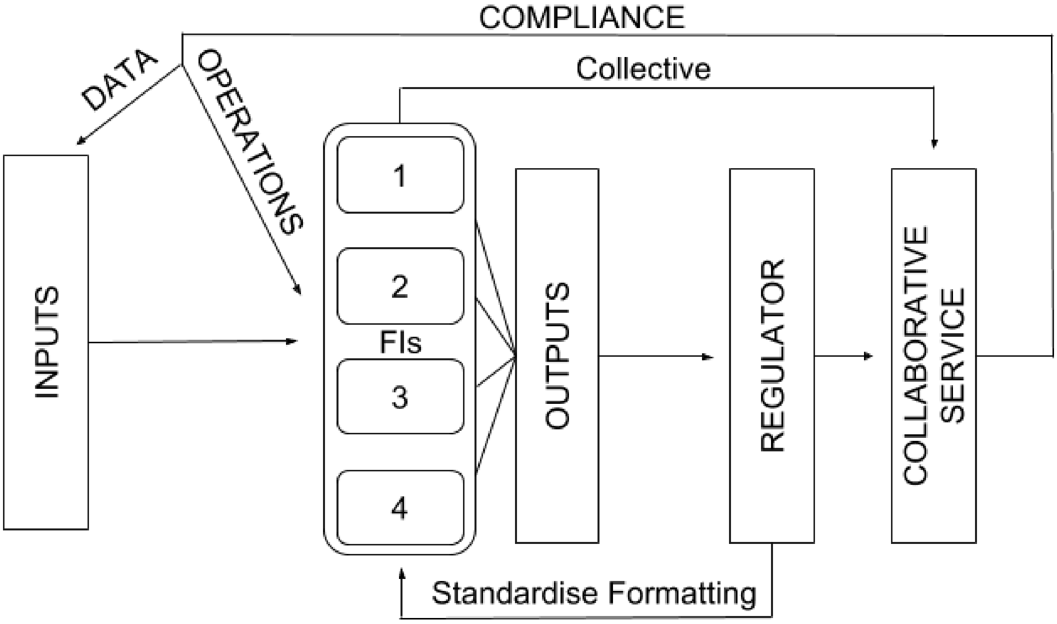
Institutions might become hesitant to use non-transparent third-party data for AI systems because regulators will likely hold the primary institutions responsible if there are damages to recover. **Regulators would have to find a balance between holding the primary and the third party responsible because third parties are responsible for a lot of innovation and progress in the field.** Although the marginal benefit for a FI to share their data to a third-party provider is high in the short run, the benefit to the joined FIs dilutes with each additional FI contributing and gaining access. Hence the *give some data in exchange for a service* model will likely succeed game theoretically, but it is to the disadvantage of the market in a long run. And in the long run the third-party provider obtains the competitive advantage as they have access to an increasing data set.

Regulator as Data Platform Intermediary

Unlike private institutions, regulators’ objectives are not to understand the inner workings of AI models, instead regulators should focus on redeveloping the way they assess compliance in the age of AI. **Regulators would have to become platform and data support administrators, which would allow them to automate compliance using big data. Together with institutions they have to develop ways for companies to share all their non-competitive (collaborative-compliant) information**, which will involve many intermediary steps, such as data standardisation protocols, data collection and transformation steps like data anonymisation, and also help with internal data auditing best practice . Regulators have a unique chance to take on this intermediary role and take advantage of institutions that are inclined to combine their data to reduce the compliance

[42] Processes such as encryption and deletion of data can, to a degree, be audited to ensure compliance.

costs of individualised regulatory interactions. The FI collective and the regulatory should work together to establish the collaborative service.



Corporations are already motivated to create strategic data alliances by sharing collaborative data. Companies want to form strategic alliances to collaborate on issues where the current inhouse processes are rarely a key differentiating capability, which in turn helps them to relieve costs through economies of scale and additionally allow them to improve the function due to both shared data and shared expertise. A few early-stage collaborative utilities are emerging, backed by key service providers. Collective institutions, such as SWIFT and EarlyWarning, have started developing service that leverage AI to address money laundering and other fraudulent threats. SWIFT for example launched a new intelligent in-network solution for fraud control that combines real-time monitoring, and the alerting and blocking of sent payments, with daily reporting; and EarlyWarning has developed AI fraud and risk-management technologies which were started and is currently being used by a collective of large US banks.

Other companies include ComplyAdvantage and Shift Technology who have demonstrated significant benefits in using AI-based algorithms to monitor transactions. ComplyAdvantage, for example, claims to have achieved an 84% reduction in false positive alerts for AML risk data, while Shift Technology is helping insurers fight claims fraud using AI."[43] More of these type of services can be kickstarted with help from regulators. Opportunities exist in situations where FIs can successfully balance their competitive impulses against collaborative opportunities. Some untapped collaboration can also come from sector agnostic processes, such as disclosure and financial statements submissions as well as annual audit requirements. Regulators are perfectly positioned to take advantage of FIs push towards coordinated efforts to minimise compliance costs. This engagement should preferably take place with bodies that are somewhat removed and independent from the main regulatory body. Regulators can help to establish semi-autonomous fee-servicing companies to significantly lessen their supervisory burden. As institutions seek to create common utilities,

[43] http://www3.weforum.org/docs/WEF_New_Physics_of_Financial_Services.pdf

new frameworks that address talent, governance and technology standards should emerge and regulators can help play an active role in this transition.

The safety of the financial system will be radically improved if collective solutions succeed: real-time scanning using aggregated market data has the potential to dramatically increase institutions' ability to react proactively to threats and catch malicious activities. Two crucial elements to the success of collective utilities are: (a) the development of a ownership and responsibility framework to ensure that collective utilities' interests are aligned with their stakeholders, and (b) the design of a liability framework to deal with errors and compliance failures. A liability framework would determine whether liabilities are shared collectively by participating FIs, or whether the wrongdoer should be identified individually and liability isolated. A shared liability framework would have to make sure that collective accountability is fostered rather than the offloading of risk to central utility established by the regulator.

Sandbox and Preventative Measures

AI calls for scientific reactive as well as preventative measures; adaptive sandbox strategies might be a good solution; there is a certain beauty in small isolated failures. Many small AI trials are needed to investigate the outcome and downfalls of automating intelligence before we can subscribe reliable preventative measures.

Other preventative measures include compliant data submissions (by FIs), the application of models to compare activity outcome datasets across FIs (on the regulator side) and the development of prediction machines (bias-proclivity, monopoly formation, etc.).

Interagency and International Cooperation

Along with the potential development for special use cases, it is also important to create awareness and establish strategic partnerships across government agencies. In supporting the development of AI in the financial services sector, the US Treasury recommended that agencies pursue interagency efforts to advance AI and enable research and development. The treasury further urges agencies to run real world experiments (with appropriate limits) to better understand the benefits and risks of AI and how such technology should be appropriately regulated.

Despite recognising potential issues and challenges in fully implementing AI into the financial services industry, the US Treasury has indicated that the increased use of AI would provide significant benefits to the US economy by "improving the quality of financial services for households and businesses and supplying a source of competitive strength for U.S. firms" and, therefore, recommended that regulators "should not impose unnecessary burdens or obstacles to the use of [AI] and should provide greater regulatory clarity that would enable further testing and responsible deployment of these technologies by regulated financial services companies as the technologies develop."[44]

Differences in data standards and regulation are apparent across regions. This challenges institutions that operate globally and necessitates international frameworks to manage common systemic issues that would have an impact across jurisdictions. Risks such as rogue trading, capital adequacy and cybersecurity require cross-border collaboration to reach a resolution.

[44] https://home.treasury.gov/news/press-releases/sm447

The increasing breadth and depth of data flows within and across organisations increases the risk of improper use and data breaches. This new development requires further regulatory oversight in data management and sharing.

However, **regulators should be aware that AI models can easily be translocated, i.e. developed in a permissive data regimes but operating in less permissive ones.** For example, Google does not need all the data from EU citizens to drive certain campaigns, they just need a few bread crumbs to expand models that are trained in proxy countries like Canada, Australia and the USA.

The revised Payment Services Directive (PSD2) of the European Union came into force in January 2018, with the aim of enabling more innovative payments across Europe. In conjunction with the General Data Protection Regulation (GDPR), this means institutions have to carefully balance requirements to share data with third parties against the risk of substantial penalties in cases where data is mishandled. International regulatory frameworks have the disadvantage of allowing competing countries to strip your country of their competitive ‘moat’.

Data regulations formulated in the coming years will have long lasting effects on financial markets. What norms will develop regarding international data flows, and how will divergent domestic rules affect cross border data flows?[45]

Futureproofing and Flexibility

Regulations should be broadly written; otherwise FIs will simply not comply in the new do-first-apologise-later era. In this modern age, companies have the option to establish themselves where regulation or the threat of regulation is marginal. The aggregate effect of narrow regulations could constrain innovation and national competitiveness. However, FIs are not always inclined to play regulatory arbitrage for reasons relating to trustworthiness and reputation. Some companies want to show that they can satisfy a higher standard of regulation. Regulators should not see AI as complex; it should simply see it as a much better way to make predictions and act on decisions. More automation almost always leads to better aggregate accuracy, so special attention should rather be given to systematic inaccuracies.

Future Analysis

In the future it might be worth investigating different FI business models that includes the use of AI systems under the following regulatory classifications.

- Prudential regulations -- created for the purpose of addressing excessive risk taking. This can be further classified into micro and macro prudential regulation. Micro is to address issues relating to a single institution and macro to the whole market.
    - Get a log output of customers default rates.
- Market structure regulation -- addresses issues that are embedded into markets like asymmetric information.
    - Ensure that customers fairly report their financial position.
- Conduct regulation -- seeks to protect customers from not being able to properly assess the risks and rewards of financial products.

[45] http://www3.weforum.org/docs/WEF_New_Physics_of_Financial_Services.pdf

    - Ensure that both borrowers and lenders are aware of the financial consequences of defaults and interest payments.
- Public interest regulation -- are in place to guard against the use of the financial system for illicit purposes like money laundering.
    - Make sure that lenders are not receiving money from known terrorists or money launders.

Conclusion:

Data tends to corrupt and absolute data corrupts absolutely: a slight modification of Lord Acton's famous adage. If you would like to know where the power lies, follow the data.

I want to thank **Apolline Blandin** for her thorough review, insight and countless suggestions and **Leon Szeli** for reading and editing the first draft.