Data challenges associated with AI

Kostas Botopoulos

Advisor to the Governor,
DPO, Bank of Greece

Artificial Intelligence and data: a steep two-way road

The link between artificial intelligence (AI) and the use of data is obvious: AI is about empowering machines to "think" as humans on the basis of huge amounts of data "fed" to those machines. The matter is not theoretical.

AI is already used or tested in the financial sector, especially by banks, which may see, according to studies, up to 50% improvement of their capabilities through the use of AI. Such use concerns most frequently trading (algorithmic trading), combatting anti-money-laundering (AML) and financial fraud (by identifying and preventing complex criminal behavior), enhancing consumer service (by mimicking human interaction and giving tailor-made answers), facilitating risk-management (by improved analysis and prediction) and even helping in compliance (by sorting out and giving solutions to the great complexities of the regulatory requests).

> *"The existing regulatory framework is not adapted and has not taken into consideration the emergence of AI."*
>
> - KOSTAS BOTOPOULOS

The basic attribute of "machine learning" is that it is able to detect patterns -of action, behavior, prediction- and thus make data "talk" –be more salient-and even "decide"– make choices and give advice- in a way that enhances financial capabilities. One could thus say that AI is both an opportunity and a danger, especially when viewed from a regulatory perspective: the main problem is that the existing regulatory framework is not adapted and has not taken into consideration the emergence of AI.

On the data front, the regulatory cornerstone in the EU, the GDPR, in principle prohibits, in the name of data protection, automated processing, which is the >>>

>>> basis of machine learning. In other words, although AI is not exempted from the regulatory perimeter of the GDPR, it does not fit into such perimeter either. The issues of supervision of data (who feeds what into the machines), organization (who decides what is or becomes relevant) and protection (who knows and who can oppose use of data) are both relevant and difficult to solve. New regulatory approaches are needed in order not to "tame" AI but to profit from its potential in the financial sector. ●

Olaf Sleijpen

Director Insurance Supervision, De Nederlandsche Bank

Explainable AI for regulators

Artificial intelligence (AI) is becoming more commonplace in the financial sector. Consequently, various legislators and supervisory authorities, including the European Commission and the European Supervisory Authorities, are considering the development of AI-specific policies and/or regulations.

Explainability seems to be at the heart of the current debate on responsible AI. When financial firms start to deploy advanced data analytics, such as deep neural networks, in their business processes, the need we feel to understand what is going on "under the hood", naturally, becomes stronger. After all, while deep learning is able to achieve extremely high levels of accuracy, it is also at the root of the "black box" problem. A systemically important financial institution that deploys an unexplainable black box to determine its capital requirements is clearly undesirable.

The obvious answer to this problem? Simply demand that AI applications used by financial firms should at all times be completely explainable. Indeed, many reports advocate a regulatory response along these lines. Taking our responsibility of safeguarding a sound financial system seriously, however, I would like to caution against such a regulatory reflex.

When considering the application of AI to their business processes, financial firms inevitably are faced with fundamental trade-offs. Should the application be as accurate as possible, or does our need to understand how the model works warrant a few extra wrong outcomes here and there? Should type I errors ('false positives') be minimized, or are type II errors ('false negatives') more detrimental? Is process fairness (to prevent disparate treatment) more important for a certain customer-oriented application, or is outcome fairness (to prevent disparate impact) of primary concern? There simply is no single right answer to these questions. What the most sensible trade-off is in any given situation depends, amongst others, on the activity for which a firm wants to deploy AI, the role the application plays in the decision-making process, and the materiality of the application for the firm's business continuity and/or its customers.

"The challenge we face is to assess what level of explainability is warranted in different situations."

- OLAF SLEIJPEN

To give an example: money laundering transactions are notoriously hard to identify using traditional (often rule-based) approaches. Globally, we do not even intercept 1% of all suspicious transactions. If someone were to develop a deep neural network that somehow increases performance by a factor 10, is rejecting this system for not being fully explainable really the right regulatory response? And how about the use of AI to automatically analyse legal contracts? Would it be sensible to favour a simple model that only gets it right 70% of the time over a – perhaps not fully understood – model that manages an accuracy of 99.99%? Probably not.

The challenge we face, both as regulators and as financial firms, is to understand the risks we are dealing with and to conscientiously assess what level (and what kind) of explainability is warranted in different situations. While this certainly is no easy task, it is part and parcel of what constitutes responsible AI in the financial sector. ●

Garrett O'Neill

Assistant Commissioner, Data Protection Commission, Ireland

Key GDPR essentials for AI processing of personal data

Imagine that you just become the proud owner of a fully automated driverless car. You would expect the following before you took a journey in it:

That the car:

- had been fully programmed with the rules of the road, such as which side of the road to drive on;
- was able to distinguish the relevant speed limits;
- was able to adjust to bad weather conditions.

If the car has, none of these primary programmable features then your first journey in it will probably be your last.

Similarly, for Digitisation and Artificial Intelligence (AI), you require that certain features are running in the background to ensure that regulatory rules are being complied with.

From a data protection perspective this means that any collection and processing of personal data of an individual, needs to be done in a transparent and accountable manner that is consistent with the General Data Protection Regulation (GDPR).

How can this be done? Answer: - Data Protection Impact Assessment (DPIA).

The Article 35 DPIA, is a means to collate, analyse and implement measures to ensure that the eventual outcome of a project can have appropriate safeguards applied.

A DPIA should not be a "paint by the numbers", exercise. It must demonstrate the proportionality and necessity of the data being processed and how it could affect individuals and their rights. There is no point to a DPIA if it just highlights in bright colours the different level of risk as being either red, yellow or green. Instead, the DPIA should be exploring the whole purpose of the project and what the overall outcome of a project will be. It is therefore an objective, living document that will change, as the project progresses with the ultimate aim to give an evidential analysis of all the potential issues and problems that could affect the outcome of the project and what can be done to mitigate or reduce these problems by introducing appropriate and relevant safeguards.

What are relevant safeguards? Answer: Data Protection by Design & Default.

The safeguards can be a number of things but when it comes to A.I., it should incorporate Privacy by Design and Default features as set out in Article 25. This could be done by implementing and embedding regulatory EU/ National legislation, into algorithms, which have to be designed to be transparent and accountable. A core safeguard feature is that the processing of personal data is viewed from the perspective of a law-abiding citizen who expects that his /her personal data be treated with respect and confidentiality.

As A.I. is primarily a bunch of algorithms then there should be a built-in protocol algorithm that monitors and reports on non-compliant issues. Similarly, other financial and insurance laws could be built into a "regulatory algorithm", to ensure that the A.I. is within the regulatory obligations of AML rules or other requirements.

In conclusion, there must be protocols and limits, as to how any A.I. can operate and what is proportionate and necessary to achieve a beneficial outcome, for everyone. Potentially the best way to do this is to purpose build, a regulatory algorithm into the A.I. to ensure that the process is done in accordance with existing regulatory rules. ●

Ermir Qeli

Head Stargate Services, Swiss Re

Are there regulatory obstacles to innovation in insurance?

The rapid growth of new technologies is bringing a significant change to the (re)insurance value chain. Technologies such as Big Data and AI are expected to improve significantly underwriting, risk assessment, costing and pricing and claims management. The ability to predict risks and accurately quantify losses allows a better understanding of risk, enabling thus more attractive insurance products for existing risks as well as for risks that were previously excluded.

Wearables and lifestyle tracking technologies may lead to more rational and bespoke insurance solutions covering actual needs rather than subjectively perceived risks. At the same time, completely new types of insurable risks, such as cyber, are starting to emerge for which Big Data and AI might provide the means of assessing these risks properly.

"Some examples reveal that changes to regulation are needed to support financial innovation."

- ERMIR QELI

Technology also opens up new opportunities for "parametric" insurance products where claims pay-outs are not determined based on manual assessments

of resulting damage, but on the occurrence of predefined triggers, usually based on data, such as for example drought >>>

>>> insurance where the pay-out is linked to measures of lack of rainfall.

With new technologies come new risks, and the (re)insurance industry will be more relevant than ever as a financial shock absorber for unforeseen losses for individuals and institutions alike. (Re) insurers are adjusting their products and services to address new risks created by technologies in the most efficient way and to narrow the huge protection gaps worldwide. At the same time, regulators and policymakers will have a huge influence over whether the industry is able to develop new products and services that are relevant to customers' evolving needs.

Some concrete examples are already emerging where regulators and policymakers could make changes to the regulatory framework to support technological innovation in (re)insurance. For instance, under Solvency II Directive there are significant limitations on the types of products that (re)insurers can offer – it is not clear whether a reinsurer could receive a fee-based remuneration for the service it provides to the cedant where no prior reinsurance contractual relationship exists. Restrictions around business activities which are not directly linked to (re)insurance should be interpreted more liberally to allow incumbents to experiment with new business models and technology. Another example where we need regulatory action is on parametric solutions. Currently, in many markets across the EU, parametric products are not recognised by the national Codes of Insurance, and therefore, such solutions cannot be offered to consumers in these markets. Such products could eliminate all complexity of a loss investigation process and can give customers the confidence when it comes to liquidity and speed of payout in emergence situations. We need an EU-wide action by regulators and policymakers to ensure that such innovation solutions could be offered to consumers in Europe.

In the age where many companies in the (re)insurance industry are embracing new technologies, it is more important than ever for regulators and (re)insurers to engage in an open dialogue to ensure that the industry can harness technological development in the interest of consumers and society as a whole. ●

Diana Paredes

Chief Executive Officer, Suade

How do you solve a problem like AI?

If data is the new oil, artificial intelligence (AI) is the engine that turns that oil into something that can power a car. The comparison seems so obvious in terms of the benefits we extracted from moving from horses to motored engines, but the financial industry did not get the memo. Given how history repeats itself it is shocking to observe how little attention and investment is done for AI in our sector. AI is still seen as a pet project rather than a solution to real complex problems. It is often drowned in budgets for analytics rather than have a legitimate cost centre. AI has the potential to transform KYC, catch fraudulent transactions, catch a rogue trader, clean data and optimise margins for our industry. Then why is it not at the top of the agenda beyond the buzzword? There are two main reasons for this and they both come down to data.

Firstly, it is a matter of how the financial industry views its relationship with the data of its clients. There is no doubt that AI could allow the financial sector to leverage the vast amount of data it has. A bank could become as efficient as Amazon at selling products, advertisement, services to a client. But that is not something that the financial sector is used to do and would it be a popular decision with clients? Are clients ready for the invasion of their privacy in their bank accounts in the same way Google and Facebook do? The solution here is to change the relationship the industry has to client data. Let's have a good deep creative think about products and services that would truly be valued by clients in exchange of their data. Better mortgage rates, relevant savings products, a dating app, market comparisons are the first things that come to mind. The only way to commercialise client data using AI is by doing it more transparently and better than what tech giants have been doing for years instead of refusing to enter the game at all.

The second blocker for AI is that the results for the experiments conducted so far in finance are poor. This is mainly due to poor data and the misuse of data scientists. AI models are simply not delivering interesting insights proportionate to the cost of the investment. The lack of data standards in finance has perpetrated bad habits from enrichments and adjustments to proprietary, unsuitable ontologies that make data impossible to be easily reused.

"If data is the new oil, AI is the engine that turns that oil into something that can power a car."

- DIANA PAREDES

Data is locked within legacy systems from one vendor to another which makes it hard to clean and make sense of. The importance of starting an AI project with clean data is misunderstood. 99% of the efficiency of an algorithm is dependent on that. Due to this lack of focus on clean data, the spend goes to hiring an unnecessary amount of data scientists who consequently will not be able to do much. They will cost money but given they do not come from finance they usually cannot make sense of the data without support. If our industry is serious about AI it should:

1. Clean and harmonise its data using the regulation as a way to do it;
2. Hire fewer data scientists to pair up with existing subject matter experts within their organisation. This is how you will get to the best models and insights. This is also an opportunity to upskill the financial services workforce, it is always cheaper and more efficient to leverage your existing staff. ●