Crowe

Risk based approaches to Artificial Intelligence

Neil Currie

“Companies should take risk-based approaches to Artificial Intelligence and Machine Learning to evaluate their impacts on the business, its reputation, its individual customers and employees”

As next-generation technologies continue to proliferate, one thing is increasingly clear. By companies adopting good, and progressive, governance and ethical standards that foster consumers’ trust they will unlock the ability to harness the full potential of these innovations, to fuel future growth. At the heart of every company, consumer trust is key: this trust is created and preserved through an honest, open, and recurring dialogue with those consumers.

Weighing up the opportunities from using new technologies needs to focus on assessing the risks to the company. This is in order to determine if the use of the technology itself is appropriate and how that technology should be implemented and controlled. Using a defined and consistent framework to make that risk assessment will provide clarity, greater understanding, time saving and ultimately should lead to a higher chance of success. In addition, while legislation may exist and be a factor, this is evolving rapidly; by adopting a proactive, risk-based approach, the company can remain in control of strategies, key decisions and commercial outcomes.

Assertions

1. The opportunity to improve customer acquisition, cross sell, and retention from the use of artificial intelligence (AI) and machine learning is driving significant investment by companies, but the associated governance is not keeping up with their adoption.

2. The 'black box' effect of many AI and machine learning initiatives makes it hard for companies to inform customers / employees about what exactly is being done with their data.

3. Companies' excitement and appetite for the opportunities presented by AI and machine learning override considerations about what is appropriate for them to do with customers' data they hold.

4. Risk-based approaches should be a part of assessing whether an AI or machine learning model should be used for making certain decisions, or for determining what additional controls need to be in place.

5. Regulatory authorities are currently looking at ways to audit companies AI and machine learning activities as a part of their regulatory frameworks. It makes sense for a company to have a risk-based framework in place to assess its own initiatives rather than relying on government regulation to define standards, which may inhibit its business model.

Assertion 1

The opportunity to improve customer acquisition, cross sell and retention from the use of AI and machine learning is driving significant investment by companies, but the associated governance is not keeping up with their adoption.

Companies are using AI and machine learning to improve or in order to aid customer acquisition, an example of this is Upstart. With more than USD 3.1 billion in loans originated, Upstart views millennials as a strategic market segment it is able to target successfully. The consumer lending platform, founded by ex-Googlers, leverages AI and machine learning to price credit and automate the borrowing process, successfully automating the process for over 60% of its loans. Alternatively, other companies are using these technologies at the back end to speed up processes that traditionally have been manual. For example, in media buying, the introduction of "programmatic media buying" is using machines to buy ads instead of using humans to go through a more manual process of setting up ads and negotiating individual buys for each site. Programmatic leverages AI technologies to bid across display, social media, and video channels, allowing advertisers to reach their target audiences more effectively. This approach works by using algorithms to analyse audience behaviours and likelihood to convert, based on thousands of factors, and then adjusting bids accordingly.

In a world where many of us increasingly want and expect content edited and tailored to us as an individual, personalisation is increasingly determining our interaction with websites, apps and search tools. Historically, personalised content started with developing a hypothesis, creating proposed content to meet that hypothesis, and then testing that content with a control group. AI flips this process on its head. While still setting parameters, the machine can adjust the experience in real time adding offers, images, and streams so that when consumers come to a website, AI is making decisions and recommendations for them based on many different attributes reducing the need for time spent 'researching' on the site. By AI facilitating these product recommendations and interactions consumers are moved through the buying cycle more quickly without their knowledge.

In this environment, the machine is making decisions based on both real and simulated data, directly affecting an individual in real time. Clearly, the traditional approaches to governance that companies have taken lack the necessary flexibility, speed and attributes to keep up and be relevant. Gartner talks of "adaptive governance" and how this enables data and analytics leaders to apply different governance styles to suit the context of the business scenarios they are faced with.

Andrew Burt, Chief Privacy Officer at Immuta, states that "Governance, now, is actually beginning to impact what types of decisions can be made, and what types of rights the subjects of those decisions have. It is not a binary choice between letting machine learning models run amok or strengthening governance so much that there is no machine learning. Actually, there is a host of ways we can govern and actively control and monitor the process of creating machine learning models. There are really three buckets here. You have the data, the model and the decisions. There are ways to govern using each of those buckets."[1]

AI governance is the idea that there should be a legal framework for ensuring that machine learning technologies are well researched and developed, with the goal of helping companies and consumers navigate their adoption of AI systems fairly. Dealing with issues surrounding the right to be informed, and violations that may occur, AI governance aims to close the gap that exists between accountability and ethics in technological advancement.[2]

Because of this need for flexibility, adaptability and speed, we can ask ourselves do we need a machine to govern the machine? Does the opportunity lie in developing a machine learning model that is responsible for monitoring and governing our machine learning? In this environment, would the model verify the running of models against a set of risk-based parameters and control factors, and either turn off any model that has deviated from established boundaries until it can be updated, or alert a human moderator who can assess and take action? In today's world where adoption of AI and machine learning is still in relatively early stages for most sectors, we have not yet reached the point where many see this as a priority issue, but such considerations are likely to feature in the future solutions we arrive at for wider application.

Assertion 2

The 'black box' effect of many AI and machine learning initiatives makes it hard for companies to inform customers / employees about what exactly is being done with their data.

Historically, when companies responded to a subject access request from an individual, companies looked at the ways in which that individual had interacted with them. Those companies would then trace through a relatively simple set of electronic or manual processes to confirm what had been done with their data, invoking a basic set of audit and discovery protocols. Whilst time consuming, in most cases a satisfactory outcome could be reached using these processes

With the wider use of AI and machine learning, this scenario has introduced significant additional complexity. According to Dr Rob Walker, Vice President of Pegasystems, there are two types of AI. "When the company is able to explain the algorithm through which machine learning is making decisions and this algorithm can be audited, the machine learning system is classified as Transparent AI. Opaque AI consists of a machine learning model built on an algorithm that cannot be explained, and, therefore, audited – this phenomenon is known as black box effect. …… the requirement for a system to be able to explain its logic when making decisions can place a brake on its effectiveness and analytical capabilities; for this reason, Opaque AI tends to be more powerful than Transparent AI and, therefore, more popular among businesses."[3]

Often however, transparency is not enough. Transparency will enable a company to verify the mathematical side of machine learning, namely the calculation implemented to get to the result, but it will not explain the reason behind it.

For example, let us assume that a medical company is training a model to provide neural predictions. A transparent system will provide training parameters as well as final parameters that can be inspected. In this case, the outputs are validated, but there is no indication why the model behaves in a certain way. Verifying the outputs is not an option with respect to both time and effort, therefore negating the purpose of the model. We need to introduce the concept of justification. In addition to the identification of the algorithm and the parameters used by the machine learning model, justification explains also the reason behind it, defining the basis behind what the machine is 'thinking' during the decision-making process. The concept of justification also helps companies identify any systemic error that has been absorbed by the model and to correct it. This is pertinent when the decisions taken by the model are with regard to regulated industries such as financial services or healthcare, where the output could have a huge impact on the individuals.[4]

Professor Pedro Domingos comments: "When a new technology is as pervasive and game-changing as machine learning, it's not wise to let it remain a black box. Opacity opens the door to error and misuse …… If incapable of understanding and explaining the outcome of the machine learning process, companies will not be able to apply it properly, leading to mistakes and losses. Additionally, there is the need to take action regarding the leadership and governance of machine learning. In fact, leaders should make sure that the decision-making process implemented by the system is ethical and aligned with the values of the company. The governance of the machine learning includes systemic ways to formalise hidden assumptions and ensure the accountability and auditability of the internal process. This governance incorporates also the responsibility to verify that machine learning is not in charge of decisions that it does not have the intelligence and capability to make."[5]

This commentary clearly indicates a recognition that individuals as consumers and employees will want to know how their data is used. Furthermore, it supports the view that meeting this objective with an audit-based approach will be difficult because of the opacity within some models. In a 2017 article by Gurdeet Singh on "AI's Vulnerabilities Threaten to Limit its Progress", a justification-based approach was suggested to counteract the inability to create definitive validation.[6] While this approach seems reasonable and sensible, how would companies' define what justification might look like, and what should constitute reasonable justification?

In an article published by Morgan Meaker on "How Should Self-Driving Cars Choose Who Not to Kill?", she raises the ethical questions of choice. For example, should the self-driven car impact with the five middle–aged pedestrians on the crossing, or instead hit the barrier risking killing the three passengers, which include a young family and their child.[7] Clearly, one can expect that the answer to that question will vary depending on who you ask, and could be influenced by many considerations such as whether the risk of three deaths will always be better than the risk of five deaths, or whether the young family should be spared because they have more remaining life? If one reflects upon such ethical and cultural considerations, it becomes almost impossible to reach a consensus conclusion about either how the model's algorithms should be developed, or how to verify the model's outcomes.

What is clear is that, however difficult it is, a consistent, well-defined measurement basis will be an essential part of any companies' governance around their use of AI and machine learning models. This is regardless of whether the measurement basis is tracked by the machine, or by an individual throughout the process, or if it is prepared up front in the assessment phase. Clear documentation and collection of supporting evidence around the rationale for what is being actioned will be critical as use moves more towards the mainstream.

Assertion 3

Companies' excitement and appetite for the opportunities presented by AI and machine learning may override considerations about what is appropriate for them to do with customers' data they hold.

Artificial intelligence technologies have already begun to transform financial services. At the end of 2017, 52% of banks reported making substantial investments in AI and 66% said they planned to do so by the end of 2020. The stakes are enormous — one study found that banks that invest in AI could see their revenues increase by 34% by 2022. As AI becomes more embedded in banks' most critical operations, particularly in ways that impact the financial stability both of institutions and their customers, this could expose new hazards. Two of the most dangerous and far-reaching areas of risk when it comes to AI in banking are the opacity of some of these technologies and the vast changes AI will inflict on workforces in those banks.[8]

Machine learning is beloved by ecommerce and marketing: Amazon, Netflix and hundreds of online shops built their recommendation engines on it. Hedge funds, such as Two Sigma or Binatix, have deployed machine learning algorithms which forecast stock prices. The medical company Medecision uses machine learning to predict avoidable hospitalisations in diabetes patients, Schneider Electric to prevent oil and gas pumps from failure and the Zoological Society of London to track endangered animals from photos taken in Africa. Have you ever seen a Facebook application posing the question "which celebrity do you look like"? This uses also machine learning to deliver the result.[9]

In December 2018, several adult videos appeared on Reddit "featuring" top international female celebrities. User "DeepFakes" employed generative adversarial networks to swap celebrities' faces with those of the original adult video stars. While face-swapping technology has been under development for years, DeepFakes' method showed that anyone with enough facial images could now produce their own highly convincing fake videos; these realistic-looking fake videos of well-known people flooded the Internet through 2018. While such use cases are technically not a "failure," their potential dangers are serious and far-reaching. If video evidence is no longer credible, this could further encourage the circulation of fake news.[10]

According to a new study from the University of California, Berkeley, advances in artificial intelligence have rendered the privacy standards set by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) obsolete. In fact, stripping healthcare data of identifying information does not guarantee HIPAA compliance. Current laws are simply insufficient to protect an individual's health data. In part, this is a problem because the same data is incredibly valuable for companies building an AI system. As AI in healthcare becomes more and more commonplace, data privacy experts are raising big red flags about the ethical implications.[11]

Facebook, the largest social media platform, uses AI to store and act on users' mental health data with no legal safeguards in place. HIPAA's healthcare privacy regulations in place do not cover tech companies. HIPAA only protects patient health data when it comes from organisations that provide healthcare services, such as insurance companies and

hospitals. In late 2017, Facebook rolled out a “suicide detection algorithm” in the US in an effort to promote suicide awareness and prevention. The system used AI to gather data from individuals’ posts and then also to predict their mental state and propensity to commit suicide. The Facebook suicide algorithm is outside the jurisdiction of HIPAA. Of course, it can be viewed as a positive use case for AI in healthcare. However, benevolent intent aside, the fact remains that Facebook is gathering and storing individuals’ mental health data without specific consent.[10] EU/UK readers should note that under GDPR and the UK DPA 2018, consent would have been required to collect such sensitive data and as such, Facebook have announced that they will not use this algorithm in the EU.[12]

Clearly, as the above examples help to illustrate, the definition of what is appropriate varies greatly and regulation is yet to catch up. As individuals become more aware of the ways their personal data is being used, it is likely they will become more concerned about how its use is determined and governed. With the increased adoption of these technologies, current attempts by hackers that centre on causing a data breach and some malicious damage will likely shift to a new focus. Hackers will seek to corrupt companies’ models and algorithms, hidden within a companies’ processes, in an attempt to cause even greater damage.

Assertion 4

Risk-based approaches should be a part of assessing whether an AI or machine learning model should be used for making certain decisions, or for determining what additional controls need to be in place.

Formal frameworks for risk management have generally been implemented in most large companies. Either consciously or sub-consciously, we make decisions taking account of the risks involved. At their most formalised, risk management frameworks include an explicit definition of risk appetite and tolerances and formalised practices to identify, measure, manage and report on levels of risk against that appetite.

By nature, machine learning systems are based on algorithms that are complex and unpredictable, which introduces additional risks to be managed and adds some complexity to those already being managed. As an example, there is a risk that customers and competitors who know about and who are impacted by the implementation of a machine learning model may change their behavior, and in doing so distort the inputs to the model. This often happens in models that aim to predict fraud, multi-party competitive scenarios and in cybersecurity, whether it is through deliberate behavior change in order to circumvent the prediction or an unintended resulting impact. Consequently, because the nature of risks changes in such a way, the integrity of the models being used will degrade faster than it would in an uncompromised environment.[13]

Machine learning algorithms are typically developed to improve functional performance, and to provide the ‘best’ possible response to a question that humans would either not be able to answer, or at least not answer quickly. This involves a machine, built with complex functions that often will not provide visibility or insight into the logic followed, or the structure of the decision process. Moreover, considering that the machine learning algorithms are trained with input data generated by people, the algorithm’s decision-making process is characterised by the same bias that applies to human decisions, and influenced by the culture, assumptions, points of view and stereotypes of people.[14]

Typically, traditional approaches to risk management have not been structured to accommodate the kinds of variables that are introduced by AI and machine learning. They do not take account of the 'black box' effect that exists in Opaque AI and machine learning models. Accordingly, companies will need to update their risk management frameworks to consider factors such as data ethics and to align that with their corporate values, or a similar demonstrable measure, in order to be able to justify and explain to individuals the intent behind the companies' use of data.

Google, Microsoft and many other companies have defined principles around fairness, safety, benefits to humanity, and other similar principles in an effort to define up-front whether a project should be undertaken. As is inevitable in using those terms, the measurement of those benefits will vary depending upon each company and who within them applies the principle. While such principles can be viewed in similar terms to 'data ethics', when applied to AI and machine learning projects they need to be embodied within a broader risk framework that considers a broad range of criteria including business risk, model risk, reputational risk , data accuracy to properly determine the risks associated with such initiatives.

Another key aspect is business continuity in the context of adoption of AI and machine learning, in that a company must still be able to carry on business whenever the model fails to operate.

A key element of a company's overall approach to risk management needs to be the development of a governance framework that allows it to evaluate and evidence a range of considerations: suitable corporate responsibility; the application of appropriate behaviour towards its customers and employees; and compliance with any relevant current and future regulatory requirements.

Assertion 5

Regulatory authorities are currently looking at ways to audit companies" AI and machine learning activities as a part of their regulatory frameworks. It makes sense for a company to have a risk-based framework in place to assess its own initiatives rather than relying on government regulation to define standards, which may inhibit its business model.

While the regulatory agencies and legislation are still trying to catch up with the pace of change and adoption of AI and machine learning, the UK Government has made very clear its intent to update its regulatory frameworks. The EU's General Data Protection Regulation (GDPR) included specific provisions regarding the requirement for consent in the areas of data profiling and automated processing. The UK Government embraced the spirit of GDPR and reflected it fully within new 2018 UK data protection legislation. This legislation included the right of review by data subjects regarding data profiling and automated processing, and the requirement for companies to explain to them the logic used in applying these technologies. As explained above, because of the nature of opaque models, this is a virtually impossible.

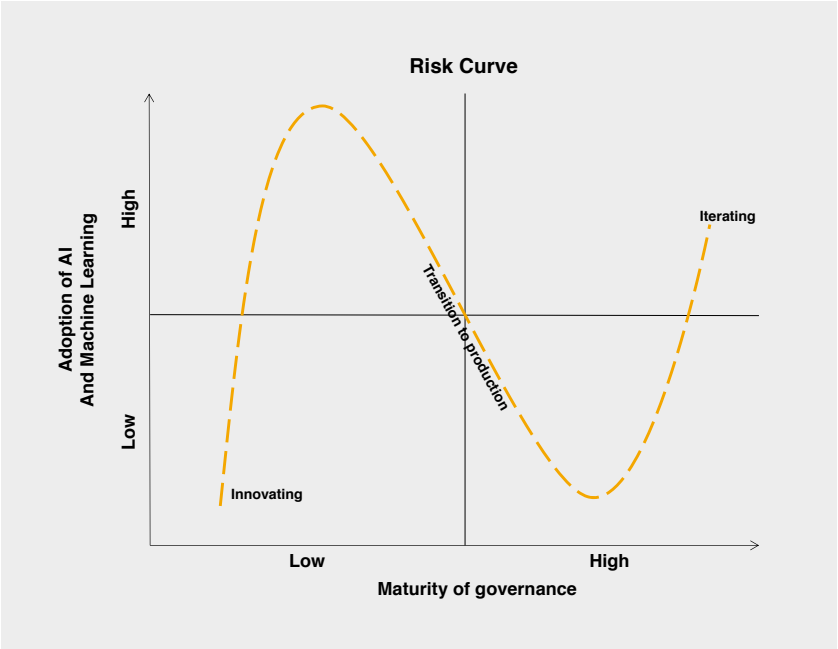
In addition to GDPR, The Council of Europe Treaty ("Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data", Convention 108) has been given an overhaul to bring it in line with the GDPR. Signed in October 2018 by 47 countries, including the UK, this treaty provides a robust set of principles and rules to protect personal data in an AI environment.

Furthermore, the UK's national data protection regulator, the Information Commissioner's Office (ICO), has engaged an AI specialist, in partnership with the Turing Institute, to assist in drafting a market consultation around approaches to audit AI in the future. In parallel with this, the UK government has established the "Centre for Data Ethics and Innovation", and published its own data ethics framework for use by government organisations. More broadly, within Europe the European Data Protection Supervisor has published its "Declaration on Ethics and Data Protection in Artificial Intelligence".

The UK's Financial Conduct Authority (FCA) regulatory agency recently summarised that "Algorithmic trading is a thoroughly embedded part of how markets function now and it is continuing to evolve …. Currently, artificial intelligence and machine learning techniques, whilst proving highly successful in certain fields, tend to lack the ability to explain how they derive their results. There should be caution in blindly accepting answers just because the computer says so …. this technology is only as good as the data that feeds it and the requirements and constraints to which it is put …. the FCA cannot prosecute a computer, but we can seek to prosecute the people who provided the governance over that computer". (15)

This is all indicative of an appetite for and trend towards further regulation in this area. With this level of activity, it would be well advised that companies start considering how best to evidence the process adopted, ethical considerations around, and decisions made regarding their AI and machine learning initiatives. The FCA's specific reference to governance suggests that the process should be structured and visible and which enables a demonstrable evaluation of risks and controls.

Conclusion



The rapid adoption and attempts by companies to scale AI and machine learning has created a nervousness about its risks and impacts. The FCA report that 1800 algorithms are running against the FTSE100 on a weekly basis, with 75% of proprietary trading decisions on equities now being executed by algorithms. In this environment, and this level of deployment, how will companies preserve customer and employee trust, maintain evidence to meet the regulators' changing expectations, and manage risks around reputation, business performance, business continuity, and corporate responsibility?

Historically, when making decisions, companies assess risk to determine whether or not it is acceptable – a risk appetite consideration. Through the adoption of AI and machine learning the risk profile has changed. The introduction of opacity in models which blurs the decision making processes means that different approaches to the assessment and justification of that risk need to take place. For example, to reflect data ethics principles within the decision making process.

The diagram considers Crowe's view of how companies see risk and how it changes within the lifecycle of an AI or machine learning model.

As the concept of AI and machine learning starts to be considered the risk is typically low, because nothing is entered into or is scheduled for a live 'production' environment. Governance models are based on traditional measures, but equally they do not yet need to evolve. As adoption increases, and current processes within the business begin to pilot or utilise AI and machine learning models, the risk increases accordingly. This challenges the governance and operating models to have to 'keep up', because factors including operational delivery, business continuity, culture change and potentially transformational change are introduced. For example, if in production a company's new machine learning model were to fail, would that company still have people to perform those tasks or will the skill base of its employees have already changed so that this now presents a risk to the company's continued existence?

As models get transitioned to production, it is likely that governance arrangements will have matured to accompany adoption of AI and machine learning within the business. The models being run will likely be relatively current, tested and being deployed in a controlled manner lowering our risk. The challenges from early adoption will likely have been recognised and processes, policies and operating models will have been updated to reflect the current environment. Formal risk frameworks will have been updated to include factors such as data ethics and to recognise the specific risks linked to AI and machine learning environments.

However, as those AI and machine learning models become embedded they can be expected to mature, learn and develop. This can mean that the results from the models will be less understood, the data within the model becomes less pure and the model will need to be formally validated, refined and updated. Despite relatively formalised governance arrangements being in place to reflect this operating environment, the speed of market, operational and ecosystem change facilitated by computing power will tend to introduce new risks to the company. This will mean that the company's governance arrangements, and maybe the model itself, will need to be refreshed and iterated to mitigate those new risks.

As AI and machine learning become more mainstream, updated risk management frameworks that reflect specific factors in this new environment will be essential to preserve customer trust, manage business performance, help ensure business continuity and to deliver regulatory compliance.

Frameworks and operating models for managing data-related risks in today's and future environments need to be far more capable of being able to adapt to cope with the speed of adoption and change that is now taking place and more adept at managing attitudes, relative to traditional approaches and ways of thinking.

Crowe

About the Author

Neil Currie is a Head of Data Management at Crowe's consulting team in London. He focuses on data strategy, governance, culture, protection and value.

As the current generation of consumers becomes increasingly mobile and borderless, leveraging and managing technology effectively through machine learning and artificial intelligence initiatives becomes a real and practical challenge for an increasing number of firms. Neil works with client stakeholders to help them think through and consider issues such as how the 'black box' effects of these technologies impact their data protection risk exposures and governance models. Working in partnership with Crowe's Head of Data Analytics, he looks at the opportunities and challenges brought by data analytics, machine learning and artificial intelligence and how these are impacted by data ownership, trust and ethics.

Contact Neil: +44 (0)203 752 3503, neil.currie@crowe.com

Crowe – a highly effective and trusted consulting partner

Crowe's risk heritage and broader capabilities help us to reinforce that risk management should be as much about leveraging opportunities as 'protecting the downside'. This is an important mindset when it comes to helping our clients to succeed in today's heavily data-driven and rapidly changing business environment.

For each of our clients, whether large or small, we keep their business objectives, priorities and specific cultural characteristics at the heart of what we do. They are fundamental to delivering outcomes with our clients that will be sustainable well beyond 'the project'. Our progressive approach and the way in which we consult, collaborate and partner with our clients, and being trusted by them to work with us again and again, is a key factor in our ongoing success and growth.

www.crowe.com/uk-risk

Reference Sources

1. (2017, Sep) Machine learning Algorithms meet data governance, Jack Vaughan, Retrieved from https://searchdatamanagement.techtarget.com/feature/Machine-learning-algorithms-meet-data-governance
2. (2018, Nov) AI Governance, Margaret Rouse, whatis .com , Retrieved from https://searchenterpriseai.techtarget.com/definition/AI-governance
3. (2017, July 27). Customer Experience, Opaque AI And The Risk Of Unintended Consequences, A Swinscoe. Retrieved from Forbes: https://www.forbes.com/sites/adrianswinscoe/2017/07/27/customer-experience-opaque-ai-and-the-risk-of-unintended-consequences/#5881df769630
4. (2017, December 12). AI's Vulnerabilities Threaten to Limit its Progress, G Singh. Retrieved from Corporate Compliance Insights: https://www.corporatecomplianceinsights.com/beyond-transparency-ai-justification-essential-risk-management/
5. (2017, April). AI & Machine learning Black Boxes: The Need for Transparency and Accountability, Colin Lewis. Retrieved from KDNuggets: https://www.kdnuggets.com/2017/04/ai-machine-learning-black-boxes-transparency-accountability.htmachine learning
6. https://www.corporatecomplianceinsights.com/beyond-transparency-ai-justification-essential-risk-management/
7. https://medium.com/s/story/how-should-self-driving-cars-choose-who-not-to-kill-442f2a5a1b59
8. https://www.americanbanker.com/opinion/adopting-ai-isnt-without-risk
9. https://www.netguru.com/blog/when-to-use-machine-learning-does-your-app-really-need-ml
10. https://medium.com/syncedreview/2018-in-review-10-ai-failures-c18faadf5983
11. https://www.lexalytics.com/lexablog/ai-healthcare-data-privacy-issuesGeneral Data protection regulation ("GDPR") and Data Protection Act 2018 ("DPA 2018")
12. https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12?r=US&IR=T
13. (2018, June 5). Lessons learned turning machine learning models into real products and services, Talby D. Retrieved from OReilly: https://www.oreilly.com/ideas/lessons-learned-turning-machine-learning-models-into-real-products-and-services
14. (2018, January 26). On bias, black-boxes and the quest for transparency in Artificial Intelligence. Dignum, V. Retrieved from Medium: https://medium.com/@virginiadignum/on-bias-black-boxes-and-the-quest-for-transparency-in-artificial-intelligence-bcde64f59f5b
15. (2019, February) Market abuse requires a dynamic response to a changing risk profile, Julia Hoggett, Director of Market Oversight at the FCA. https://www.fca.org.uk/news/speeches/market-abuse-requires-dynamic-response-changing-risk-profile