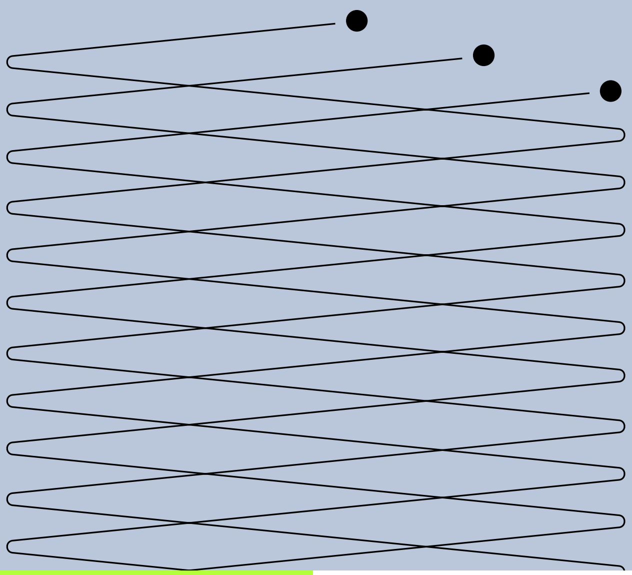
dreamquark

2021-03

State of Ethical AI in 2021: Challenges posed by ethics to companies developing AI



Review by
an ethicist and philosopher
& entrepreneur and scientist

Thomas Souverain & Nicolas Meric

Thomas Souverain, PhD student on AI (ENS Ulm)
Nicolas Meric, DreamQuark CEO

SUMMARY

Beyond the difficulty that many firms face while deploying artificial intelligence (AI), apprehensions and fears associated with AI pose a serious threat to its development. These fears are **a challenge that companies have to face** in order to use AI and allow society to benefit from it. That is where ethics comes into play: by incorporating the use of technology into a strong value system, companies can dispel customers' and employees' uneasiness.

In the long term, **including ethics into their business model** allows companies to **foster value creation** and reach their objectives, all while taking into account the **people affected by AI**. That is why ethics is a primary asset to be developed and a significant competitive advantage.

This paper seeks to identify the **fears** that need to be addressed, the **principles** that need to be promoted, and **the concrete applications** for an AI that is both trustworthy and ethical. **It delivers a vision of how companies can position themselves in order to handle the challenges posed by an ethical AI**.

INTRODUCTION

Overcoming AI development challenges

AI for good opens up markets

An increasing number of companies are choosing to implement AI: **in August 2019, it was projected to reach by 2021 a $2,9 trillion market value** according to Gartner. Accenture estimates that by 2035, it could reach a market value of $8.3 trillion in the USA, $2.1 trillion in Japan, $1.1 trillion in Germany and $814 billion in the UK. Despite Covid-19 having altered global activity, this positive dynamic is truer than ever for AI: it is currently being used to help predict the evolution of the virus by institutions such as by the Guangzhou Medical University in China.

2035

USA: **$8.3 TRILLION**
JAPAN: **$2.1 TRILLION**
GERMANY: **$1.1 TRILLION**
UK: **$814 BILLION**

> **“Economists predict that AI products will account for up to 15% of the total production of goods within a decade”**
> says Yoshua Bengio,
> recipient of the 2018 Turing Award,
> expert on deep-learning.

AI has the power to **learn and reproduce behaviour** that humans view as intelligent, by stocking and analysing data at high speeds; thanks to machine-learning and deep-learning technologies, it is becoming more and more efficient at performing a variety of useful tasks. It can assist judicial proceedings; help detect diseases; synchronise with drones to facilitate proper irrigation in agriculture; automate assembly lines for higher productivity...

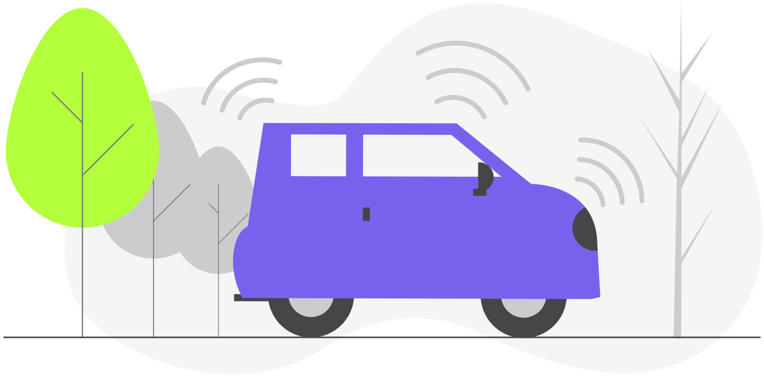
In that, AI is able to do what humans cannot. While humans do not always pay attention to details, AI offers a much higher **precision**. While our institutions are not always effective at protecting citizens, and human errors can make the law less fair, AI can compute data in a more **objective** manner. While society can have a hard time running smoothly, AI can make things **easier** for us: it is used for driverless cars which make accidents less likely, it can automate air conditioning and even our entire homes, it can be used by educators to find resources through search engines and visual interfaces...

In short, a great number of market applications for AI can serve as a tool to make daily life easier. We will consider them in this paper.

A number of countries, such as China, Japan, the UK and the US, as well as the EU, are developing strategies to profit from this revolution. So are NGOS and multinational companies, starting with the Tech Giants. Smaller companies have started to follow: in 2017, there were already 1393 start-ups specialised in AI in the USA and 383 in China. In fact, the UNESCO has presented this rush to master AI as a “**new space race**”.

The use of AI gives rise to concern

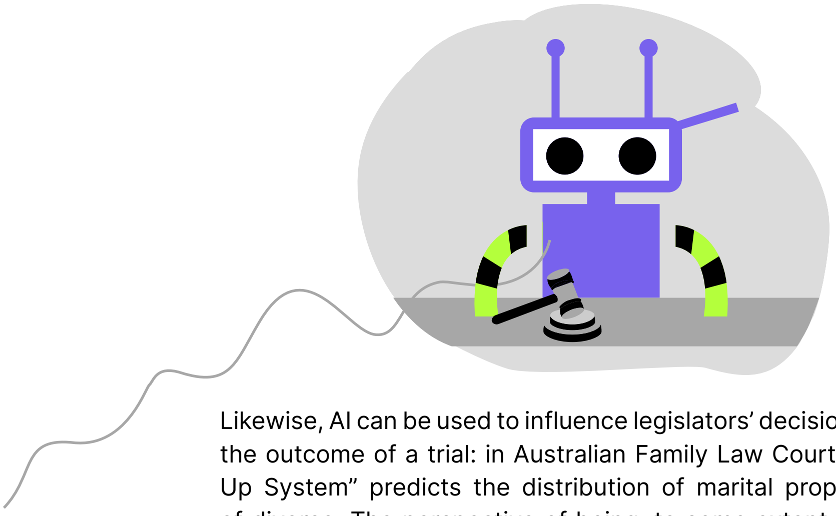
Nevertheless, before AI can be implemented on a wide scale, and its benefits can be felt all over the world, **DEVELOPERS NEED THE CONSENT OF CITIZENS AND EMPLOYEES**. The task is not an easy one.



In fact, the development of AI tends to be met with **doubts** and apprehensions coming **from consumers**. Take for example self-driving vehicles. They were put on the market in 2013, and only 6 fatalities have since been recorded: a very small number, compared to the 1.35 million deaths by car crash which happen every year worldwide, according to the World Health Organisation (WHO). But deaths caused by a machine are readily seen as suspicious and are widely discussed in the media, giving rise to a strong sense of mistrust in parts of the public.

In 2019, for example, an Uber self-driving car killed a pedestrian because of an error in its program: it was not able to recognise a jaywalking pedestrian, confused it with another vehicle, miscalculated its speed, and did not stop in time. Uber faced massive backlash and took nine months to **upgrade the safety** of its vehicles before relaunching its self-driving cars. Another example which worries the public is the prospect of "killer robots". An international campaign was launched in 2018 in the aim to make fully autonomous weapons – *i.e.*, machines with the power to kill humans – illegal.

But it is not always a matter of life and death. In other cases, people are simply worried about changes in their way of life. Take, for instance, the Social Credit System which the Chinese Government it currently developing. Its task is to rank each citizen: it gathers data on citizens' relationships, activities, opinions, and uses AI facial recognition in order to analyse social behaviours. While a higher score provides material advantages (for instance, being able to see a doctor without lining up to pay), a lower score might limit the ability to purchase some products (in 2018, the Chinese Global Times stated that **11 million people were prohibited from traveling by plane**). The system's purpose is to act as an incentive for the "right" behaviours, as in, conducts that comply with social and official norms. In the Western world, hearing about this system can create fears around invasive use of AI.



Likewise, AI can be used to influence legislators' decisions regarding the outcome of a trial: in Australian Family Law Courts, the "Split-Up System" predicts the distribution of marital property in case of divorce. The perspective of being, to some extent, judged by a machine can be disquieting to the public.

Job loss, loss of physical contact, diversity biases, discrimination, lack of transparency, control, oppression, misinformation, dehumanisation, digital burn-out... these words often come up in **debates surrounding AI**, which explains why many citizens, scientists, managers, employees, consumers, perceive the rise of this technology as a threat.

These fears are a potential obstacle to profitable AI applications

The way this technology is being put to use across the world legitimately raises questions. If it is not controlled to some extent, and if we do not question its implications, it could develop in a way that is erratic and counter-productive. Another risk is the mistrust it generates, which could make the public reluctant to have AI be a part of their daily lives, their professions or their leisure activities. This would close market prospects.

These growing worries indicate that AI is at a mature stage, and that we can realistically expect it to be deployed on a large scale. Yet, **as with nuclear energy in the 1950s**, techniques that are powerful invite us to reflect both on their potential benefits and on their potential downsides, leading to a debate on ethics.

Two options could be envisioned: either we simply forbid the use of AI, which means we deprive ourselves of all the positives it brings; or, we carefully experiment with AI, all while reducing its downsides. **This means it is time to think about which applications of AI are ethical and which ones are not. An ethical framework is necessary**, and it has to exist in a balance between free experimentation and cautious limitation.

It is absolutely crucial for us to discuss what we want to create, what our societies are ready to accept, and what goes against the liberty and moral compass of individuals. **THE DEVELOPMENT OF AI MUST BE IN AGREEMENT WITH THE PRINCIPLES WE VALUE.**

Consumers expect debate on AI

As a response to customers' apprehensions, a series of guidelines and laws have been put in place since 2017, by countries (eg. Canada, Japan), supra-national unions (*eg.* the EU), and a variety of organisations (*eg.* the UNESCO, the OECD, the World Bank, and the ILO).

Their aim is to establish safeguards against non-ethical design, deployment and use of AI, particularly regarding privacy and technical safety.

By helping build a standard for ethical AI, these legal frameworks provide an answer to the questions this technology generates and open up a debate.

We are at the very beginning of a long, collective reflection, which involves three aspects:

- from a market perspective: deciding for which tasks we need AI.
- from a technical perspective: understanding what AI can do and what it can not do.
- from an ethical perspective: determining what we want it to do, and what we prefer it not do.

It is only through a rational use of this technology that we will be able to benefit from the economic and social opportunities brought by AI. This means that, before acting, we need to consider what we want to achieve.

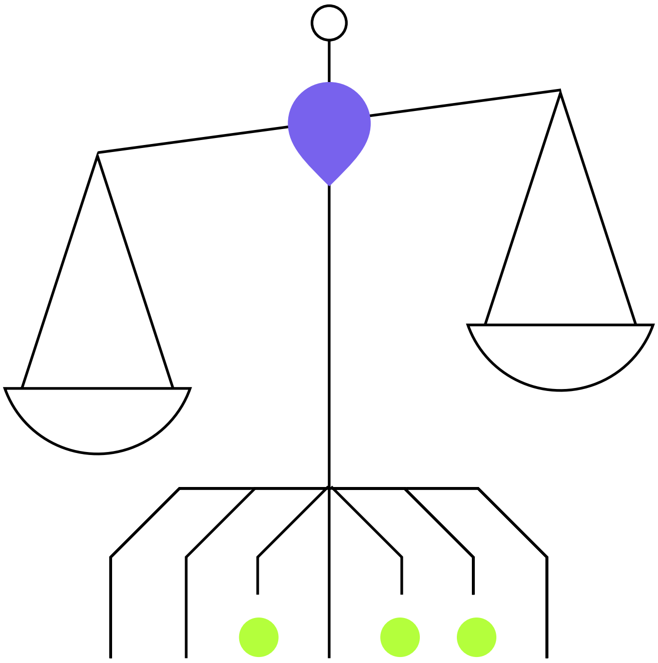
Depending on each company's specific sector – *e.g.* finance, mobility... – different ethical principles will come into play.

THIS PAPER SEEKS TO DISCUSS AND PUT INTO PERSPECTIVE THE 7 ETHICAL PRINCIPLES WHICH THE EUROPEAN UNION HAS IDENTIFIED AS THE MOST RELEVANT CONCERNING AI:

- › **RESPECT FOR HUMAN AUTONOMY,**
- › **RESPONSIBILITY,**
- › **JUSTIFIABILITY,**
- › **WELL-BEING,**
- › **EQUITY,**
- › **PRIVACY,**
- › **TECHNICAL SAFETY.**

These seven principles come from the seven key requirements presented by the European Commission in 2019. This paper also discusses their importance in different sectors, in order to show where the implementation of an ethical AI is most urgent and how it can be done: some crucial issues include selection algorithms for students wishing to enter an university, as well as privacy and technical robustness.

> **“The time is more than ripe to define the ethical principles [which will] ensure that AI serves collective choices, based on humanistic values.”**
> Audrey Azoulay,
> UNESCO



Laws are not enough: we need ethics

Law and ethics are often considered to be the same thing. This is a mistake; while ethics can act as laws do, by prohibiting some behaviours and encouraging others, there is an added complexity to it. Once a law is promulgated, it stands firm and cannot be questioned anymore. Ethical principles, on the other hand, invite constant reevaluation in order to remain relevant, and are always up for debate.

It is one thing to *be* compliant, as we are towards the law; it is quite another to think about *why* **we comply and in which way we wish to position our business and activities**. Those are ethical questions. While laws create rules, ethics invites us to debate whether those rules are fair; it incorporates points of view coming from other cultures, religions, educational systems...

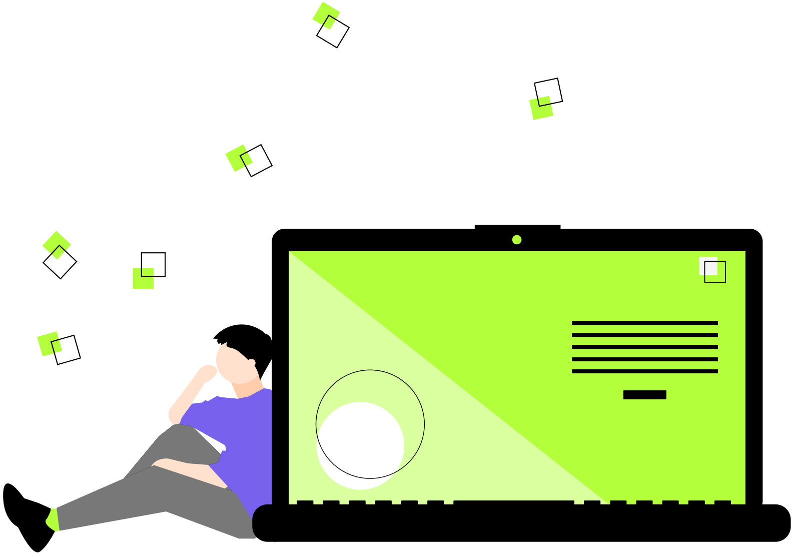
AN ETHICAL PERSPECTIVE ENCOURAGES US TO LOOK TOWARDS THE FUTURE, AND TO IMAGINE *WHAT WE WOULD DO IF WE WERE BETTER PEOPLE*, SO AS TO BE ABLE TO DO IT.

Ethics to anticipate market shifts

Therefore, ethics comes into play, not only as a safeguard against the potential excesses of AI – so to tell us what it should not do –, but also as a set of principles determining what it should do. If we discuss the way AI is to be deployed, then a great deal of apprehension around it will be put to rest.

> Lately, civil society has become particularly interested in the ethics of the products they buy, from their food to their medication. Consumers have begun to request labels which ensure that their production was "ethical", and this demand will continue to rise. AI will only accelerate this phenomenon, as users will want to make sure they can trust it not to get out of hand.

Leaders must take this trend into account. As with digitisation, those who join the movement too late risk being disadvantaged. **An ethical AI will anticipate market shifts**. That is why we need to go further and draw up an overview of the ethical questions AI poses; that is the aim of this paper.



The purpose of this paper: how can AI be associated with ethics rather than with fear?

The scope of this paper is to give an overview of our current situation: we are human beings faced with an artificial form of intelligence, which generates mistrust. Its scope is also to **present the ways in which an ethical AI can provide an answer**. To this end, we will begin by giving an overview of the current state of AI, including concerns which have come up (part 1), then link these concerns to our desire to be and remain human (part 2). As a solution, ethics appears to be a crucial part of designing and using AI as a tool for good (part 3). Far from substantiating the fears it may cause, a trustworthy AI can prove to be beneficial to society in many fields of activity (part 4). Lasty, we will present DreamQuark's vision for an ethical AI (part 5).

Developing a **trustworthy AI is highly beneficial**: it avoids the unwelcome consequences that an unchecked AI could bring, and it also prevents exaggerated worries that could unjustly curtail the development of AI.

Our society cannot afford to miss this boat. That is why the intended audience of this paper is broad: it goes from entrepreneurs that are already developing AI, and are specialists, to citizens who only have a vague idea of what it is. **THE IDEA WE STAND BY IS THAT AN ETHICAL AI COULD PLAY A SIGNIFICANT ROLE IN OUR PROGRESS AS A SPECIES.**

1

Emergence of AI and renewed interest: an overview

1.1. There is worry around AI

The newness of AI can be frightening

The development of AI has been picking up speed in the last ten or twenty years, and there is renewed public interest in it. This novelty and this quick development are disquieting to the public, as has often been the case with new technologies; in the past, initial apprehensions often turned out to be excessive, and this could be no different. According to skeptics in the 1990s, computers were bound to stop working after 1999; to those in the 1950s, the earth was bound to be destroyed before the year 2000 by progress in chemical technology.

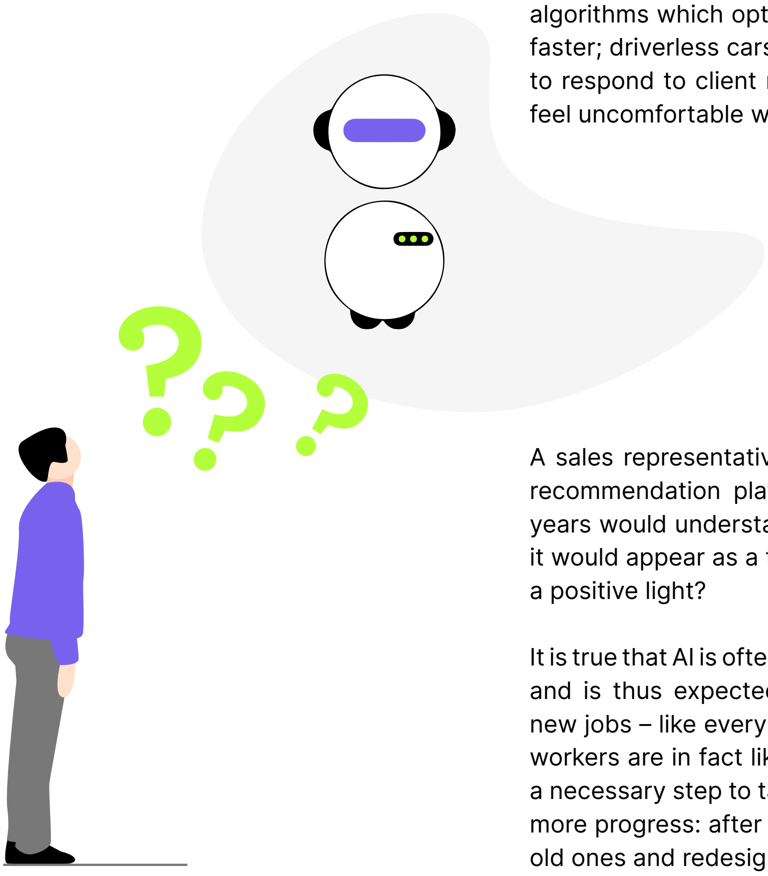
The fact that a risk may exist does not necessarily mean things will go wrong. But so-called "intelligent" technologies change the way we envision technological progress: they amplify the worries surrounding it. What fuels these fears?

Humans view machines as dangerous

Two points come to mind: first, we are afraid that a potentially over-performing technology may replace us; second, we fear it may overturn our established values and take decisions in our place. Delving into the root of those worries may help us understand them better.

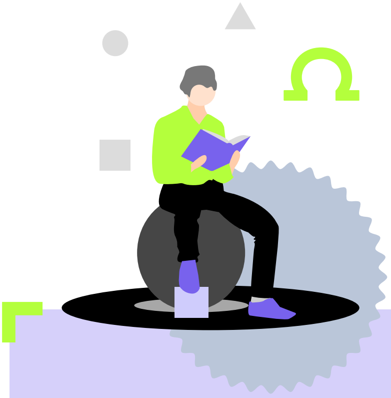
1. Humans are afraid of being substituted

Due to the progress of technology, in robotics above all, some machines may well be able to take our place at work. There are algorithms which optimise travel time, in order to deliver packages faster; driverless cars; chatbots which are more and more qualified to respond to client requests and complaints. The public tends to feel uncomfortable with these new intelligent technologies.



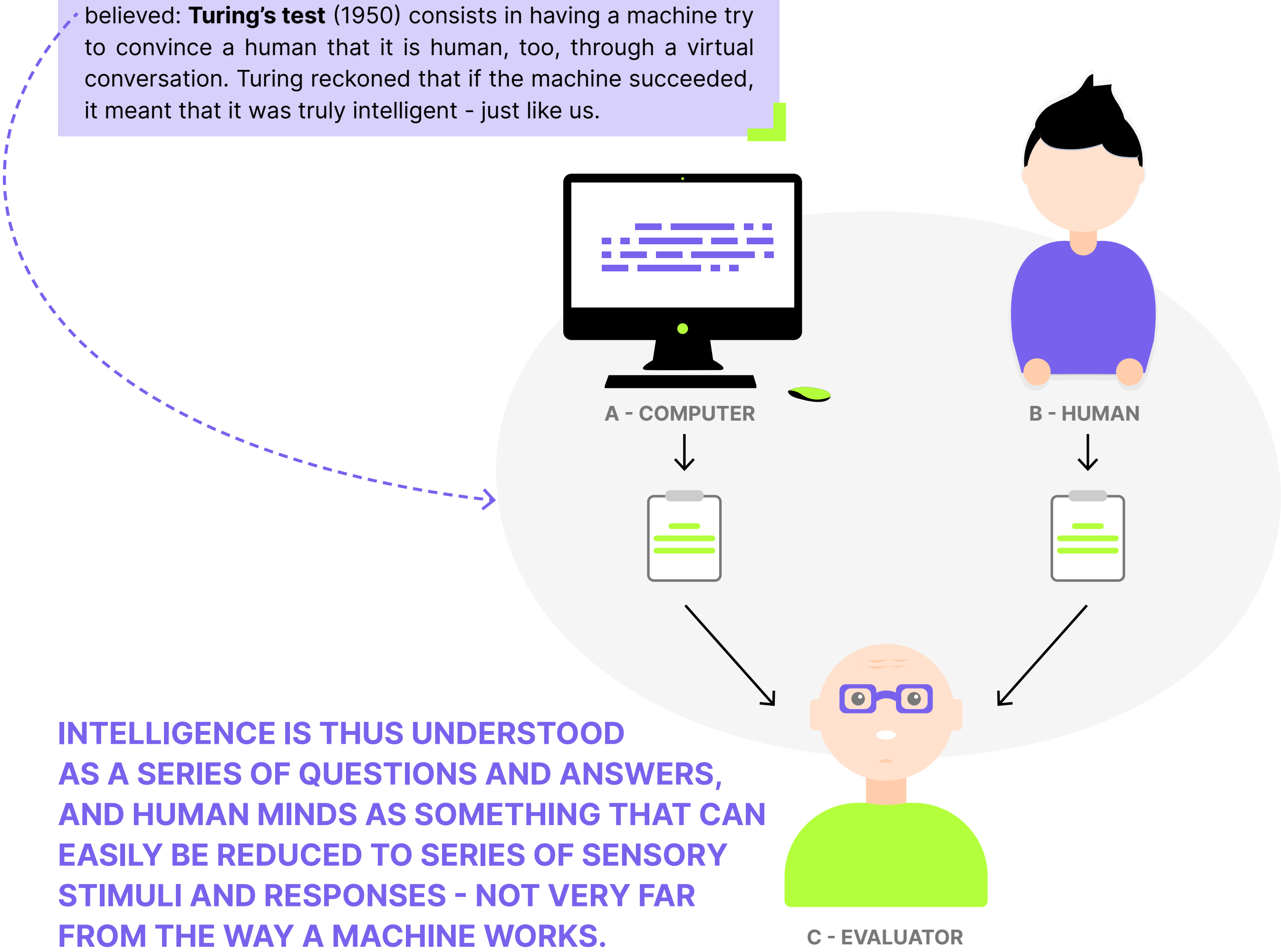
A sales representative bound to lose his job because an efficient recommendation platform is set to replace him in the next few years would understandably resent the development of AI. To him, it would appear as a threat to his well-being: how could he see it in a positive light?

It is true that AI is often considered to be a **new industrial revolution**, and is thus expected to come with job loss and the creation of new jobs – like every industrial revolution so far. A great number of workers are in fact likely to be affected. But many believe that it is a necessary step to take in order to keep moving towards more and more progress: after all, new technologies have always supplanted old ones and redesigned our way of doing things.



The difference here is that humans might no longer be simply assisted by machines, but rather be the ones to help machines, or even become useless, since their intelligence might be no match for the artificial one being developed. Because of this, there might be more job loss than job creation.

The apprehension we have of being replaced by machines stems from the way artificial intelligence is often considered in relation to our "natural" intelligence: that is, as something similar, but more advanced and more efficient. According to this idea, we are nothing more than less-sophisticated machines, with a mind and a way of reasoning that are similar to those of a computer. That is what Turing, the father of AI, believed: **Turing's test** (1950) consists in having a machine try to convince a human that it is human, too, through a virtual conversation. Turing reckoned that if the machine succeeded, it meant that it was truly intelligent - just like us.



INTELLIGENCE IS THUS UNDERSTOOD AS A SERIES OF QUESTIONS AND ANSWERS, AND HUMAN MINDS AS SOMETHING THAT CAN EASILY BE REDUCED TO SERIES OF SENSORY STIMULI AND RESPONSES - NOT VERY FAR FROM THE WAY A MACHINE WORKS.

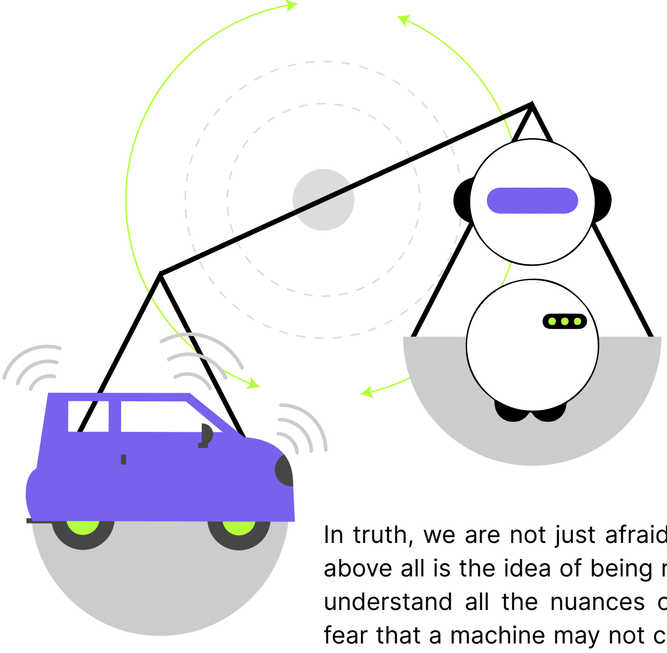
Turing's machines are built upon this idea and that is why they are at the basis of artificial intelligence. That is also why they are more efficient than us: while our ability to compute data is limited by our physical needs, such as eating and sleeping, and the relative slowness of our brains, AI is extremely fast and can store a great amount of data. In other words, artificial intelligence **is structured much like our "natural" intelligence**... but when given **a simple and precise task**, it is able to do it much, much faster and much, much more precisely. It lags behind when it comes to complex tasks, but it is quickly catching up, and that is worrying to some people who fear that AI might outperform humans in several fields.

2. Humans are afraid of being replaced by something that does not understand human complexity

Potential troubling AI applications

A possibility that some people find troubling is the development of judge-bots or police-bots, which might replace human judges and police officers. Automated judges would certainly be extremely skilled at remembering correlations between events which might help them judge an illegal action more accurately. They would also have the law memorised in all its detail, which could allow them to apply it without fail. The decisions taken by these judge-bots would then undoubtedly be fair: they would take every legal perspective and every element of the trial into consideration, and they would be unbiased by personal feeling.

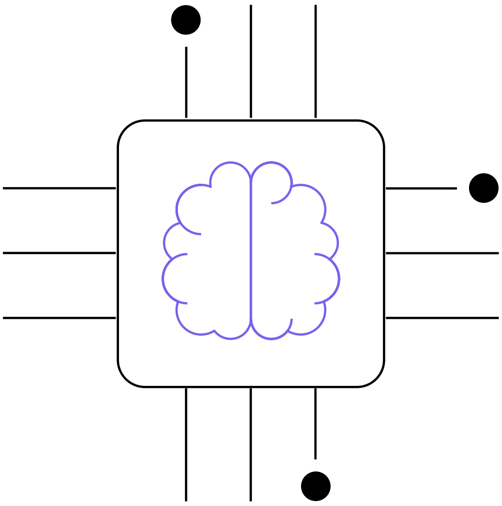
But a human being who values the principles of freedom and believes that, in some cases, mistakes can be forgiven – for example, if a poor individual stole some food – could find it troubling to be judged by something which cannot relate. The rigidity and objectivity of AI can seem cruel when it comes to evaluating human mistakes. The perspective of being replaced by something that does not think like us can be disquieting.



In truth, we are not just afraid of being replaced. What worries us above all is the idea of being replaced by something that does not understand all the nuances of our decision-making process. We fear that a machine may not consider the complexity of a situation before judging it. AI is programmed to act based on the way humans act in general: it is not trained to recognise exceptional situations and treat them differently, as we would. The 2019 fatality caused by a driverless car perfectly illustrates this flaw: the AI controlling the vehicle had been programmed to recognise a pedestrian as an object which crosses the road at a crosswalk, and, as soon as it encountered a pedestrian that went against the general rule – the pedestrian was jaywalking –, it was unable to recognise it and accidentally caused its death. **AI TENDS TO GO BY THE RULES AND DOES NOT HANDLE EXCEPTIONAL SITUATIONS WELL.**

Even if AI was trained to go on a case-by-case basis, and understood how to treat special circumstances, one thing it will always be unable to do is relate to humans on an empathic level. That is a quality when it comes to being neutral: a judge who cannot be biased by his or her own feelings, a selective algorithm which does not discriminate on the basis of race or gender can be valuable additions to our society. Neutrality is necessary in certain fields. It can also be useful for certain interactions to be automated, as empathy and human contact may not be essential when it comes to day-to-day activities such as shopping for food: some people may feel relieved to be able to avoid small talk and unnecessary discussions with the cashier.

But even though the neutrality of AI has a positive side to it, the fact that AI will never be able to understand empathy can be worrying. AI cannot be *humane*, as in, take the feelings of others into consideration. It cannot understand our irrational side. For example, AI might have no problem eliminating individuals who may be dangerous to society: murderers, thieves, and all sorts of criminals are undoubtedly harmful to social order and the safety of others. The rational decision might be to sentence them to death so that they may not cause any more harm. But it is a decision we, as humans, choose not to take – we decide to keep them alive and attempt to give them another chance, which is to some degree an irrational decision, based on our humane understanding of suffering and the importance of human life. A machine would be unable to grasp this decision.

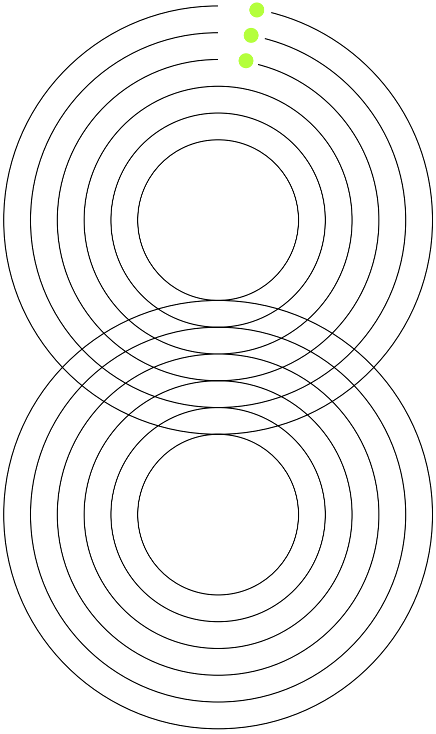


Oftentimes, two sets of values are presented as contradictory: on one hand, **efficiency**, and, on the other, human values such as **empathy**. Efficiency is the core value that AI is built on, and for that reason, AI is extremely useful to us – for productivity, to avoid wasting time, to find precise solutions which we would never find otherwise. **BUT BEING MORE EFFECTIVE AND BEING MORE USEFUL IS OFTEN SEEN AS BEING "BETTER", AND THIS MIGHT ALWAYS NOT BE TRUE. THERE IS A POINT WHERE EFFICIENCY CONTRADICTS HUMANITY.** Police-bots, who would be trained to be efficient, might not understand why we choose not to kill criminals on the spot – but that is where our humanity lies. Likewise, AI used for facial recognition might not be programmed to respect people's privacy – but it is a human right to not be watched all the time. These values are *ethical* values and could be said to stem from our "weak" side: the part of us that is emotional, imaginative, and compassionate, and therefore more fragile and less down-to-earth. That might be the reason dystopian films often imagine a future where machines are dangerous to humans because they do not understand feelings and suffering. These qualities are a fundamental part of being human and an artificial intelligence cannot relate to them, understand them, or act accordingly, which can be worrying to some.

1.2. These worries stem from centuries of myths around humans and machines

We have explored the logical arguments that stand at the base of worries around AI. It is also worth considering the way in which this narrative around AI goes hand in hand with ancestral stories, anchored in history and myths, which focus on the link between humans and machines.

1. Machines replacing humans?



The idea that humans could be replaced draws from 15th century perceptions of bodies as elaborated machines – or rather, of machines as less-elaborate versions of living beings. This comparison is particularly interesting in that it views the working of living bodies as automatic: it implies that a stimuli necessary brings a determined response, the same way a patient's knee jerks when a doctor knocks it with a rubber mallet. So to explore these similarities even further, and owing to a fascination for machines in the 15th century, scientists such as Leonardo da Vinci built automated animals: in 1499, da Vinci brought an automated lion to King Louis XII's court in Milan.

Stories such as *Frankenstein* by Mary Shelley also draw on this idea: **IF A BODY IS NOTHING MORE THAN A ASSEMBLY OF PARTS, THEN SURELY A FUNCTIONAL HUMAN COULD RESULT FROM SEWING BODY PARTS TOGETHER. AT LEAST, THAT IS WHAT THE MECHANISTIC VIEW OF HUMAN BODIES COULD IMPLY, AND THAT IS THE IDEA SHELLEY EXPLORES.**

AI goes further than this: it is built upon the idea that minds, not only bodies, can be reproduced by machines.

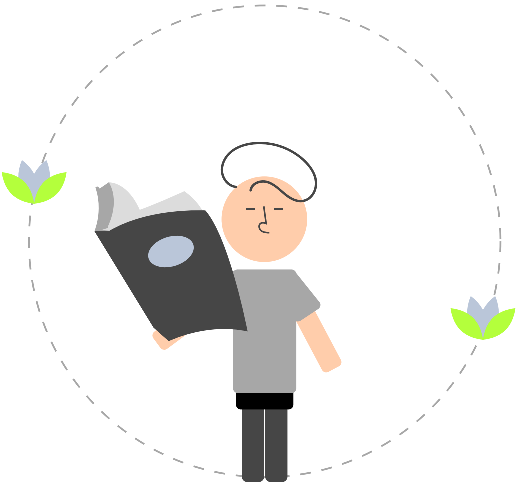
2. Copying humans without understanding them

The complex question of whether it would be desirable for humans to be replaced by machines can be explored through an ancient myth: the greek myth of Prometheus. According to this symbolic tale, the Gods created two titans whose task was to give the species of the earth the resources they would need to survive: claws to fight for food, speed to catch prey and run from predators. But the titan Epimetheus made a mistake: when he got to mankind, he had no resources left to give out. Quite fittingly, his name meant "the one who thinks after he acts" in greek. His brother Prometheus – "the one who thinks before acting" – decided to compensate by stealing the fire from the Gods and giving it to mankind.

The fire has been interpreted as a symbol for reason and technique, as in, the ability to think about one's actions and build things with one's hands. This gift symbolically allows mankind to survive while being physically weaker, slower, and smaller than many of its predators. **THE MYTH OF PROMETHEUS TELLS US THAT FROM THE MOMENT MANKIND BUILT TOOLS OUT OF WOOD AND ROCKS TO DEFEND ITSELF, TO NOW, WHERE IT IS DEVELOPING INCREASINGLY ELABORATE TECHNOLOGIES, REASON AND TECHNIQUE ARE AT THE BASIS OF OUR SURVIVAL AS A SPECIES.**

This tale allows us to put into perspective the difference between our reason, our intelligence, and that of AI. There are two interpretations of this myth which are worth discussing:

- The first interpretation – a post-humanistic perspective – views Prometheus's action as **a terrible mistake**. It argues that, because of the gift of reason, humans are too caught up in speculation, ask themselves too many questions instead of acting, and have lost contact with reality. While animals are entirely guided by their survival instinct which guarantees that they live in harmony with their surroundings, humans are able to question and resist this instinct; post-humanists argue that this leads them to waste time, to not be efficient enough, and to lose sight of their main function in the world. Imagination makes them too impulsive at times and too inactive at other times because they focus on daydreams rather than material reality. So post-humanists often view mankind as *incomplete*, faulty, and believe that **these inefficient beings deserve to be replaced by intelligent, more performing machines**.



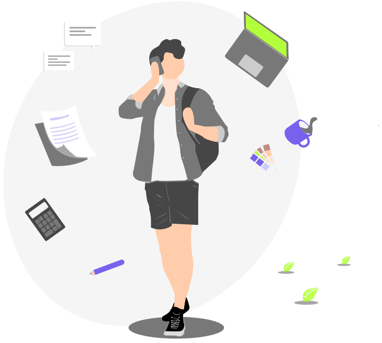
- The second interpretation – a classical humanist perspective – **views Prometheus's action as a gift**. By giving mankind the fire of the Gods, he gave it a spark of divinity. This means that humans are above animals in that their behaviour is not simply determined by instinct: since we have reason, we are able to reflect on incoming stimuli and can **choose not to respond automatically**. Our mind is much more complex, which makes it unpredictable, and humanists argue that this is **where liberty comes in**.

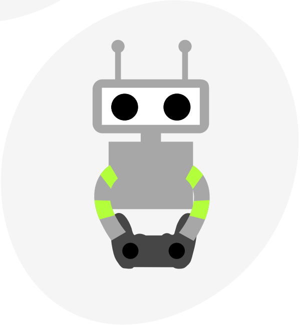
  The fire could then be a symbol of our ability to imagine, to create, to reflect on our actions, and to have subjective view on the world rather than following the crowd or our instinct. **This interpretation explains the potential threat posed by an intelligent machine**: if AI is only able to reproduce the computing part of our intelligence, which works much like a calculator, it means that no AI will ever be able to grasp what humanists hold most dear – the irrational, spiritual, contemplative part of the human mind. That part is what allows us to hold others accountable for their actions: if one is free, and one is able to reflect on one's actions, then they must also **bear out the consequences of their decisions**. That is the principle our entire law system is based on. But, since AI does not have free will, it can hardly be held accountable for its mistakes. At the same time, it might not be fair to blame the developer, seeing as AI is becoming more and more autonomous in its decision-making. That is why the question of responsibility is complex in the case of AI, and why its intelligence is so drastically different from ours.

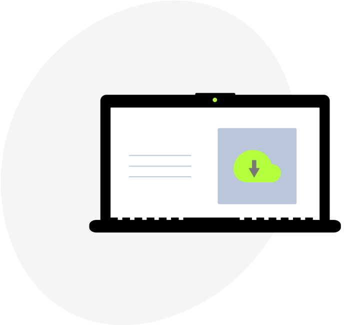
1.3. The current state of AI

Describing the current state of AI can help give us some perspective on **how and why these worries came to be, whether they are relevant, and what we can do about them**.

In truth, AI is not new to us, but it has been so largely covered by the media and has known such a **significant development in the last 5 years** that most of us are now aware of its presence in our lives. Developers have managed to create in such a short time AI-based solutions for tasks which no machine had ever been able to accomplish. As AI became more successful, fears associated with it grew as well, and politicians took hold of the topic to discuss legislation and voice concerns.







What made AI so successful?

But where does the success of AI come from? Some key elements:

Digitisation: the recent digitisation of our society was the fertile ground in which AI took root. Digitisation gave birth to cloud computing, video games, and large amounts of data. It allowed us to produce billions of pictures, video recordings, conversations on social networks, written text, customer data... all at a very low cost. These fuelled the algorithms that firms began deploying throughout the world, and the sum of all these pictures and words on the internet is what we call Big Data.

ImageNet: thanks to digitisation, ImageNet, a dataset of 13 million images split in 1000 different categories, was created. ImageNet caused the biggest leap since the beginning of AI research: in 2012, a research team using deep-learning technology managed to create an algorithm capable of labelling the images in this dataset with less than 25% error.

Graphical process units (GPUs): as the video games industry was met with a rising demand for high quality graphics, Nvidia created graphical process units, a computing technology initially used for graphics and video rendering. GPUs allow parallel processing, which means that machines which use GPUs are able to run several tasks at once. The researchers from ImageNet used this technology to classify the images in only a few hours, and it has been used for artificial intelligence ever since.

Cloud computing: cloud computing allowed researchers to find data to exploit without needing a datacenter. Thanks to cloud computing, early-stage startups were able to perform iterative experiments for little cost, which allowed them to better their machine-learning algorithms and lowered the capital required to build an AI startup. The number of startups therefore rose quickly, which also fuelled the demand for data-trained professionals and lead companies to share their progress through open-source software repositories so as to attract talents. In turn, this allowed the community to benefit from recent discoveries.

Thanks to these developments, we have made progress in several areas of AI: image recognition and synthesis, speech recognition and synthesis, language understanding, emotion labelling, predicting behaviours...

Two types of AI

Artificial intelligence can be split into two categories: **machine-learning and symbolic artificial intelligence**.

- **Machine-learning** is specialised in algorithms. Machine-learning algorithms learn to achieve tasks thanks to an existing corpus of data.

  For example, in the case of sites such as ImageNet, researchers can start out by labelling each image: “cat” for pictures of cats, “house” for pictures of houses, and so on.

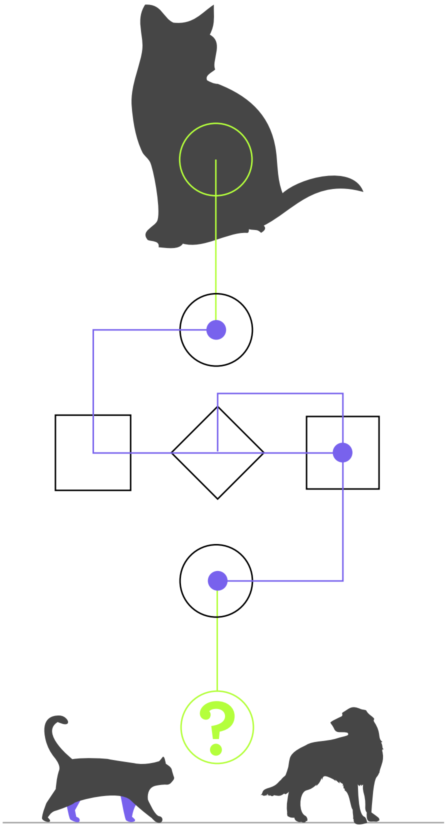
  The point of the algorithm is to predict the labels by analysing the pictures. To achieve this result, researchers take for example 80% of the images and build an algorithm able to make a guess for each image. The input is the image, and the output is a series of potential labels. If the algorithm is right, nothing happens, but if it is wrong, the parameters are updated thanks to a mathematical procedure. This process is repeated. Then, the algorithm is tested on the remaining 20%, and the testing ends when the error margin is below a certain threshold.

  Therefore, the **algorithm is designed to learn the right guess for new images by itself**.

- **Symbolic AI**: in symbolic AI, the algorithm is built out of the expertise that the developer has on a specific task. It does not learn by itself, but works as well as machine-learning for some problems. In the last ten years, most of the success has been attributed to machine-learning, and it has received much more attention than symbolic AI.

  Symbolic AI and machine-learning are generally not suited for the same problems and can be combined to build more effective models. Machine-learning is usually better suited for situations in which we lack a representation of the world, or in which the problem is to complex to be written as a program; symbolic AI is best when a formula or a specific set of rules exist, or when we lack data to learn from.

  To date, the most successful machine-learning models are deep neural-networks (also called “**deep-learning**”). Neural networks are algorithms made to mimic brain behaviour: **simple computation units are connected in order to achieve complex tasks such as recognising the content of an image**. These models are much more performing than other types of algorithms, but they have also raised red flags as to the potential unethical aspects of AI.



Fears and limitations

There are a few reasons why AI and in particular neural-networks can be worrying:

The Black Box effect: we do not generally understand how these algorithms make their decisions, and the process is not explained. This is referred to as the "black box effect".

Unfair decisions being replicated at large scale: our lack of understanding, and the great speed at which these algorithms work, make it difficult to stop unfair decisions from being made by the machines and to notice or fix unconscious biases that the developer may have had.

The algorithm's ability to **beat humans on some tasks**: playing board games, playing video games, quickly analysing large amounts of data. This is worrying to those who fear that humans may lose their place in the world.

Lack of understanding: public opinion tends to have a hard time grasping the way AI works, sees it as a topic for geeks or researchers, and is usually worried about what it does not understand well.

AI-models also have **limitations**:

- They do not work well when *data does not exist* or is available in small quantities.
- They are often unable to *explain* their decisions.
- They are unable to *extrapolate*: they do not adapt easily from tasks they know how to perform to new tasks which are similar.
- They do not have any *common sense, like humans do*.
- They are not *ecologically viable* in the long term as they consume a significant amount of energy.

Despite these challenges which remain to be overcome, AI is at a crucial stage: it has begun moving from the labs towards the first customer and enterprise solutions. Most innovations generate significant benefits when they move from the labs and start being used by the public, which also reduces the worries associated with these innovations.

The quick spread of AI

More and more startups are integrating AI into the solutions they offer their customers. As of 2020, 3975 for profit companies specialised in artificial intelligence, and this number may continue to grow.

Many recent products offered by large tech companies are based on AI: virtual marketplaces, vocal assistants such as Google Assistant, online meeting platforms such as Facebook portal.

AI is widely used by banks to assist customers – chatbots, such as the chatbot "Erika" created by Bank of America, are a good example of this. Banks also use AI to deploy use cases in order to reduce the cost of operations. AI is useful to wealth management firms as well: it allows them to create recommendation engines for their customers.

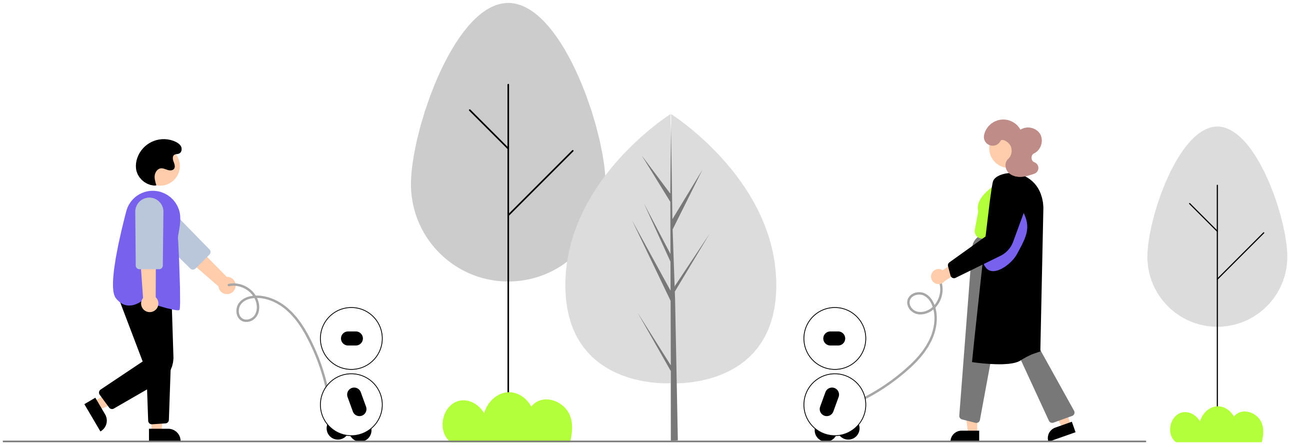
In 2018, the FDA (the US Food and Drug Administration) approved the use of AI to help diagnose health issues in the US. Insurance startups such as Oscar Health also release applications with AI to ease on-boarding, reduce the burden of compliance or automate repetitive tasks.

The 4 main challenges of deploying AI

While many use cases in production already work well, and while most large companies have already deployed at least one use case in production, there are four aspects to consider in order to successfully deploy AI on a large scale:

1. **Performance of the AI**: is my AI solution functional? Am I meeting the minimum requirements on critical metrics for the use case? Am I solving the customers' problems, and what is my success rate? Is there a return on investment?
2. **Compliance of the AI solution**: am I meeting the legal and regulatory requirements?
3. **Deployment of the AI**: am I successfully deploying my AI solution? Is it easy to maintain?
4. **Widespread adoption**: is it used by my customers? It is useful? Does it generate trust?

To overcome these four challenges, companies such as DreamQuark want to **democratise** artificial intelligence technologies. In order to do so, they are working to remove the technical aspects required to built an AI machine-learning algorithm, in particular the process of coding the solution. As a result, people with little to no technical skills in machine-learning will be able to develop their own AI applications, and the first challenge will be easier to overcome.



To overcome the second challenge, some AI solutions are beginning to integrate **explainability** as a core feature. Lack of transparency is the main issue which most of the recent AI solutions share. While they are valuable for solving use cases which other, more accessible algorithms are unable to solve, these non-explainable solutions cannot be largely adopted. They do not meet the regulatory criteria as they cannot explain how their algorithms make their decisions.

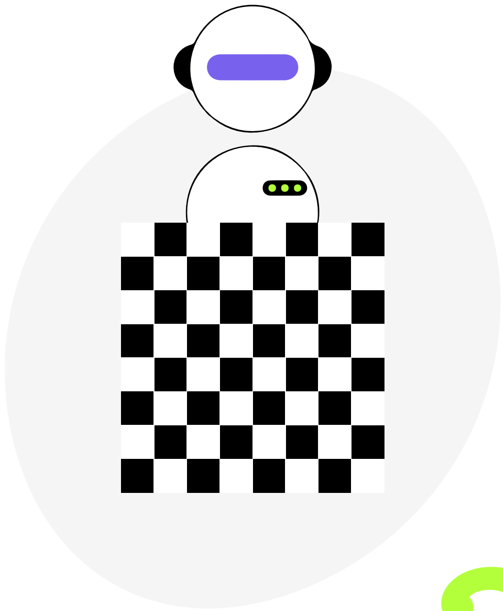
DreamQuark has been able to address this challenge successfully on several decision-making and recommendation-based applications, but we are working on ways to provide deeper insight into the logic of machine-learning algorithms.

The third challenge lies the **deployment and integration** of the algorithms, as well as in the constant maintenance of the machine-learning algorithms, since their performance evolves every time they receive new data. Pre-canned Software as a Service (SaaS) applications remove the burden of deploying a use case and make this step easier. Application Programming Interfaces (APIs) can also be used to further integrate the algorithm with existing applications.

The fourth challenge is **widespread adoption** by the customers of the company which has chosen to implement an AI-based solution. AI algorithms are all the more easily adopted if they are both valuable and trustful. Explainability, fairness and performance help create that trust. To achieve this result, solution providers aim to connect new AI solutions with pre-existing applications used by the clients. Solution providers also offer predetermined solutions which can be replicated in order to solve specific and recurring problems.

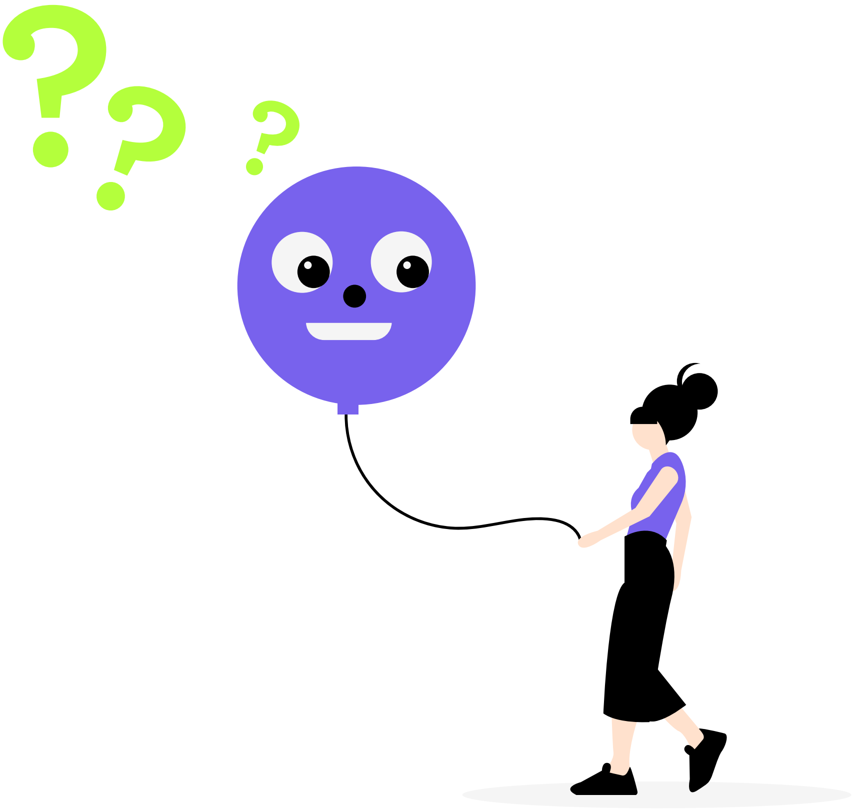
Further challenges to overcome on the long term

Some challenges remain to be solved in the next years. The main things an AI cannot yet do is use **common sense**, create **high-level** concepts by combining simple pieces of information, and learn to **generalise** off of a small number of examples. These three things are very simple for a human mind: common sense comes from having experienced a large number of different situations in a lifetime; creating high-level concepts is something we do all the time, such as combining age and gender – for example, "70" and "man" – to create the concept "an old man". Humans are also able to quickly see the bigger picture with the help of only a few details, while AI still needs quite a significant amount of data to be able to generalise.



For now, AI algorithms remain **highly specific** – but it may change in the near future –, and cannot accomplish complex tasks as well as humans. They cannot extrapolate and explain why they did what they did. AI's inability to work with general intelligence like a human does is what we call the "AI-complete" problem.

Algorithms are good at closed problems like board games, such as the Go Game, even when there are a great number of solutions to the game. Humans are good at open problems where there may be an infinite number of solutions and the problem is ill-defined or needs to be defined. As of now, we are far from bridging this gap between humans and machines.



1.4. Fears around AI: are they legitimate concerns?

In this part, we will attempt to provide some insight as to whether or not the current worries around AI are legitimate: do they hold up against the current state of artificial intelligence, are they likely to come true in the future?

We are far from an AI that could replace us

Records from earlier industrial revolutions indicate that workers are often intimidated by new technologies, as they are afraid that they will lose their jobs. During the First Industrial Revolution (18th century), the practice of machine-smashing spread across the English countryside for this reason: the textile workers known as Luddites objected to the use of mechanised looms and knitting frames as they were against the replacement of human labor by machines.

But in reality, the amount of people who lost their job was much lower than expected. New jobs associated with new techniques came to be. As a result of their ability to reflect on problems, to use context, and to come up with creative solutions, **humans still found a way to generally become employed**, even though machines were present. The jobs that were entirely taken over by machines were the most mechanical, repetitive jobs... jobs that were closer to the skillset of machines than to proper human qualities.

Artificial intelligence is likely to have a similar effect on the job market. In cigar fabrics, trials to replace human employees by machines have failed on the grounds that even our most sophisticated machines cannot rely on the intuition and dexterity that only humans have. In other areas where AI performs well, such as automatic translation, **we will still need high-level human translators** able to understand context, catch implicit nuances, and have an intuitive grasp of what an author or diplomat mean when they use certain words.

AI will do robot jobs, routine-jobs, and leave us operations which appeal to our human intelligence: creative, empathic jobs, complex and non-specific jobs which require general intelligence and life experience.

So what happens when machines take human jobs is that they challenge humans to be **adaptable**. When an intelligent program takes over a task previously assigned to a human employee, **the employee now has to perform a truly human function**. As long as an employee is adaptable and can be retrained to perform a non-automatised task, he is not threatened by substitution. Machines, however, are unlikely to be able to perform anything else that what their initial programming taught them to do.

Therefore, AI does not point towards a substitution of human employees by machines, but rather towards a metamorphosis of the job market. What is likely to happen is a form of **hybridisation between humans and machines**. Intelligent algorithms will challenge humans by helping them perform better. In journalism, for instance, AI can help journalists become faster at writing by correcting their mistakes, it can help with memory storage, and can thus simplify their routines. Journalists can then focus all their energy into writing an intelligent overview of the news and introducing creativity into their articles. Seen under this light, AI holds great benefits for the job market.

Many people have seen Charlie Chaplin's film *Modern Times*, in which the alienating aspect of repetitive tasks is highlighted. If an AI were to take charge of all these small, automatic, day-to-day tasks which exist in most jobs, employees could feel **liberated rather than threatened**. Artificial intelligence would help them be efficient at performing routine chores, effectively freeing up their time so they can accomplish high-level tasks which are intellectually or manually satisfying for the workers.

AI does not have general intelligence yet

While AI may transform the job market in a positive way, and be somewhat of a freeing force for workers whose job is alienating, these arguments are not enough to dispel all concern around AI replacing us. With good reason: **while previous industrial revolutions introduced technologies which replaced *human bodies***, **AI is set to replace *human minds* as well**. To some, there is something invasive to this new technological prowess, and it feels menacing that a machine may be able to think like them – or better than them.

As a matter of fact, **AI is currently unable to replicate the full workings of a human mind**. AI is good at computing data, at sorting it, at storing it, at analysing and taking decisions quickly. These are things that humans can also do, but they are only a fragment of our intelligence. Substitution by AI is therefore possible, but only **in areas where human intelligence is similar to the way AI reasons**. A deep-learning model learns by discovering its environment analytically: it processes data, finds correlations between variables thanks to what it observes… In short, AI is able to memorise and perform repetitive tasks in an effective manner. Even if it outperforms humans on those specific tasks, it cannot reach other human dimensions.

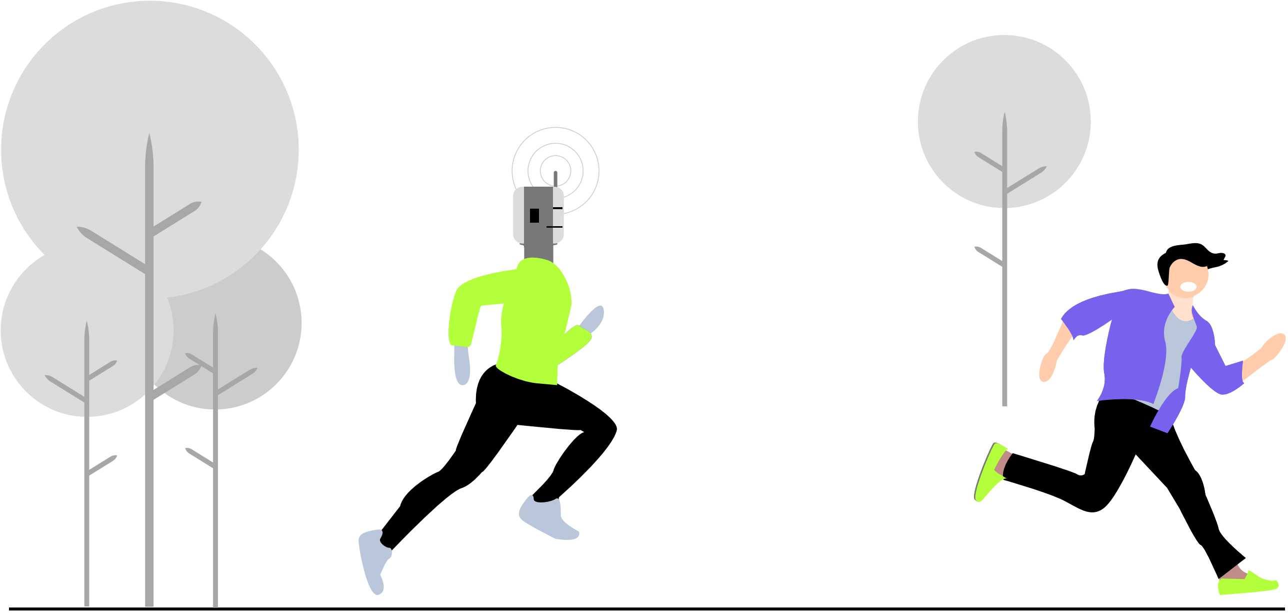
That is why we are more likely to work *with* machines than be replaced by them. When a human task is predictable, handles a great amount of data, and is specific to one problem, it can be delegated to AI. But when **flexibility of mind, imagination and creativity** are needed, humans will always perform better. Human minds are more complete. Therefore, distinct areas between AI and human expertise remain.

“Singularity”: the fear of being replaced by AI

The hypothetical situation in which AI could accomplish exactly the same tasks as humans is referred to as “singularity”. Thinkers such as John von Neumann believe than an exponential acceleration in technological progress could result in **a super-intelligence which would not be compatible with human existence**, or would put mankind’s place in the world at risk.

For now, although AI is very powerful in some specific areas, its is **far from having general intelligence**. A fundamental part of human intelligence which eludes machines is the ability **to think critically** and to **exchange with other humans**. Through dialogue and our ability to think collectively, we see ourselves reflected in the speech of others and gain consciousness of our own existence. We are also able to question our actions and choose them freely. AI is not able to think in the same way.

There is a way to combine the simplicity of the machine’s execution with the complexity of human thought. Since an AI algorithm does not think about morality before acting, it will perform a program if it is given one, no questions asked. **It will not wonder whether assigning a social score to citizens is a good thing**, or whether the variables it uses to make decisions are morally justifiable. Human principles and value systems can therefore easily be implemented into the program AI will follow. That is why we have a responsibility to think about the values that are important to us ahead of time.



2

AI
and mankind

In part one, we went over some recurring fears around AI and attempted to provide some answers. Now, we will analyse AI as a concrete power introduced into our world, a world made for humans, which might not be ready for another form of intelligence.

2.1. AI as a "pharmakon"

The greek concept of "pharmakon": the poison and the antidote

In ancient greek, the word "pharmakon" refers to a substance which can act both as a poison and as a remedy. Its effect comes to down how it is given to the patient: the right dose can heal, the wrong dose can kill.

This concept can help us reflect on the place of AI in our world. While it holds great powers and great potential benefits for us, whether its outcome actually turns out to be positive depends entirely on **our way of implementing it**. This dilemma is already true for other aspects of modern technology: nuclear power, for instance, can serve both as a source of electricity – to help our societies lie comfortably – and as a weapon of mass destruction.

In the case of AI, smart robots can intervene to save lives in situations where humans cannot – for instance, by rushing into a burning building –, but they can also be used as weapons with an unfair advantage, or go against international laws. Either way, the ambivalence of AI is apparent. We must aim to turn AI into a remedy and not into a poison.

AI makes risky scenarios more likely

With AI, the risks for going astray are greater. AI can be perceived as a weapon more dangerous than the nuclear bomb because of **how many people it can reach in very little time**: if a social network such as Facebook changes their algorithm, 2,4 billion people are affected at once. New technologies linked to AI are therefore expanding quickly, without any physical barriers, mostly thanks to social networks which use AI.

2,4 BILLION PEOPLE
ARE AFFECTED AT ONCE

Therefore, AI amplifies the risk of technological repercussions. While old technologies could only be applied in a specific, limited domain, AI has a much wider reach, and the possibilities for it to be implemented seem to be limitless, all over the world.

Old technologies, which had fewer possibilities, had a lesser potential for causing damage; now that we are developing such a powerful tool, finding a way to use it for good has become a necessity. Our responsibility is to use this pharmakon as an antidote, and not as a poison.

2.2. Will AI make us less human?

Debates around artificial intelligence draw on fears dating back to Taylorism (end of the 19th century): machines, and technology at large, may dehumanise us. The words being used in debates around AI speak for themselves: people want a “trustworthy AI”, they want it to be “human-centred”... these debates imply that, **LEFT TO ITSELF AND DEVELOPED AS IS, AI WOULD NOT BE ALIGNED WITH HUMAN NEEDS AND EXPECTATIONS**.

We have identified **four areas where we need to take action quickly** to avoid a push-back and make artificial intelligence a driver of value creation: 1. employment; 2. management; 3. job definition; 4. quality of relationships. None of these points are linked to the fictional, dramatised idea that AI could take over the world; their aim is to highlight the very real, day-to-day problems that AI use could create.

Likely consequences and risks of AI

1. Massive job destruction. The first impact that AI may have on a large scale is **massive job destruction** as people are progressively replaced by digital interfaces, AI-based solutions, as well as actual robots. Job destruction appears to be inevitable, to some extent at least. AI-based agents may be able to do a better job and find unique solutions, all while being more productive: **THEY COULD WORK 24 HOURS A DAY**, every day, without being tired, and work laws and limitations would not apply to them...

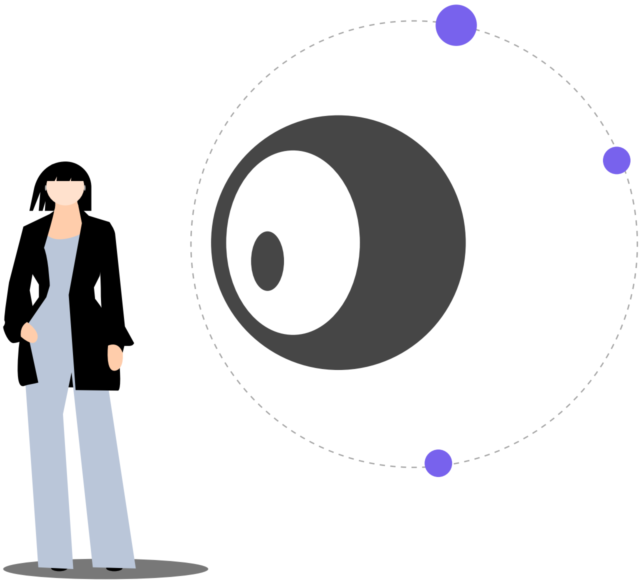
Losing our jobs might lead us to lose part of our humanity: we would socialise less, and the part of our identity that is strongly influenced by our work would disappear.

There are two reasons which make job destruction by AI possible: first, the fact that many tasks are robotic and could quite easily be automated. Second, the fact that we do not value the social part of our jobs very much: it seems natural for us that cashiers or delivery people should be replaced by robots, and in this, we might miss out on the casual, spontaneous social interactions that make our lives a little bit more interesting.

2. Inhumane management by AI. The second impact of AI could be a robotic managerial pressure, which is already starting to be apparent in some well-known US companies and the gig economy at large. Some human workers (generally with low qualifications) are **entirely managed by an artificial intelligence** which puts a great amount of pressure on them.

Everything they do is precisely measured: they have to follow the exact orders established by the AI, such as which path to take in a warehouse or between different locations in a city (to pick up and drop off passengers, for instance). They have little room for error and can not stray too much from what the AI asks them to do. Their objectives are very difficult to achieve, and they are rapidly fired if they fail too many times. The AI is in charge of recruiting and firing the workers, which means that some workers **never talk to a human superior**, and that the process can lack **transparency**. These people can truly say that they work for a robot.

These workers are in a precarious situation with weak social protection, in a competitive job market, with bills to pay at the end of the month, and a huge amount of pressure. Generally, these situations are made possible by work laws which do not protect employees, in a free-market and capital-driven economy. While they can be effective at producing profit, they can also be dehumanising and alienating for workers.



3. Significant transformation of skills required to success in the workplace. The third impact of AI on the job market is the automation of repetitive tasks which are necessary but do not drive value, such as paper work. AI can take charge of those activities and accomplish them better and faster than we can, effectively freeing us from them.

For the moment, workers in charge of those repetitive tasks spend a great amount of time accomplishing them and little time developing customer relationships. If these tasks are automated, they may lose their jobs.

However, in a competitive environment, winners are companies which favour their employees' creativity, put them in charge of identifying and solving problems in smart ways, and value the quality of customer relationships over more robot-like tasks. CEOs could choose to delegate these tedious tasks to an AI, while **reassigning their existing employees** to creative, innovative, or simply less repetitive functions.

This shift might take some time, and will certainly require retraining workers, but it is likely to lead to **greater satisfaction at work**, greater **fulfilment** in workers, and be a driver of market and societal value. The time spent on improving customer relationships will make companies more effective at targeting their desired audience and improving the value of their products.

4. Generalisation of automatised, digital and robotic interfaces. The last impact of widespread AI adoption may be the transformation of human interfaces into digital ones. Interfaces are made to be increasingly intelligent and effective: when phoning a service, it is frequent for the customer to be greeted by an automatised voice, telling them to press 1 for a particular query, 2 for another... and so on, leading to phone conversations where no human voice ever comes in.

This lack of human connection can have some negative effects: it becomes impossible for some customers to have their problems solved if their request is too specific or unique to have been programmed into the algorithm. Customers are also unable to call onto the other person's **empathy**, to convince them to find an original way to solve their problem, to draw some satisfaction from a conversation with another human being, to crack a joke... These drawbacks do not necessarily indicate that automatic interfaces are counter-productive, but they are worth considering, as some customers may be disappointed by excessive automation.

In the medical field, some patients are already complaining about robotic interactions, which they find too cold. Diagnoses delivered on a screen, through simple text, and without any doctor-to-patient interaction, are seen as inhumane and not tactful enough.

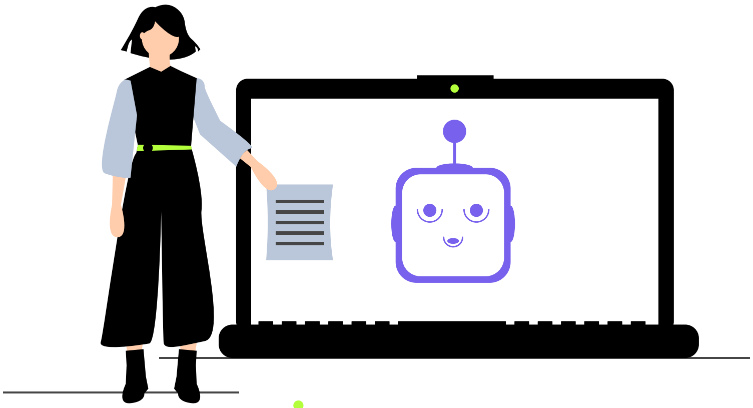
If a pre-constructed program has unfair biases, lack of true interaction with customers makes it impossible for customers to express their discomfort. When banks deny credits to clients on the basis of calculations made by an algorithm, there is no way of arguing, even though the algorithm used might have been unfair. In short, these interfaces tend to not be **adaptable**.

What many people expect from customer service is an advisor able to understand their specific situation, a discussion through which they may discover new ideas and open their minds, and an interaction with another human being where everything is not planned. While AI may answer customers' queries efficiently, the fact that everything is planned out in advance may remove the element of surprise and a form of spontaneity that helps make our lives rich and interesting.

In order to **remain in control** of the ways in which AI will affect the workplace, leaders must develop their algorithms carefully, work with the people who will be impacted by them, and train developers in good practices, so as to improve the transparency of their AI solutions. On the long term, we must teach the next generations to be wary of those outcomes: workers, entrepreneurs, managers in large private corporations and public entities must have the tools to apprehend the changes brought by AI.

In the present day, our task as citizens could be to **minimise any and all negative impacts of AI** through carefully thought-out regulation and work laws which protect workers if something were to go wrong.

In this way, **we can avoid humans feeling dehumanised** by AI and feeling as if their place in their workplace is compromised. Part of their humanity – as in, their ability to accomplish fulfilling tasks – might even be restored by AI taking over robotic tasks. Workers could have more time to think, to create relationships, to work on creativity and innovation, as well as improve the sustainability of our organisations and society.



2.3. Could we lose control over AI?

This overview of the four challenges posed by AI deployment lead us to a central issue: the question of whether we have true control over AI.

For now, the answer is is reassuring. Even cutting-edge AI technologies are unable to do much more than very specific, local tasks, which are precisely defined by the developer – for instance, telling an unattended bag from other types of luggage at an airport. This limited range of possibilities, coupled with AI's unawareness of their own existence, makes the creation of a dangerous super-intelligence unlikely.

However, we have to remain vigilant while deploying AI: it does happen for an AI to be diverted from its intended purpose. **Tay, a chatbot created in 2016**, which was meant to have conversations with American teenagers, was taught insults by malicious users. In this case, the users are responsible for the project going astray; but in other cases, our expertise is delegated to the machine, which can use it wrongly.

When an AI controls a car instead of a human driver, or when a smart program diagnoses a patient instead of a doctor, it means that we are allowing AI to take over in fields that are complex and **where small mistakes might have significant consequences**. This requires a great deal of trust, which must not be given lightly.

BEFORE TRUSTING AI TO ACT AUTONOMOUSLY, OR TO TAKE RELATIVELY INDEPENDENT DECISIONS, IT IS IMPORTANT TO IMPLEMENT OUR HUMAN VALUES INTO ITS DECISION PROCESS.

3

The rise of ethical concerns about AI

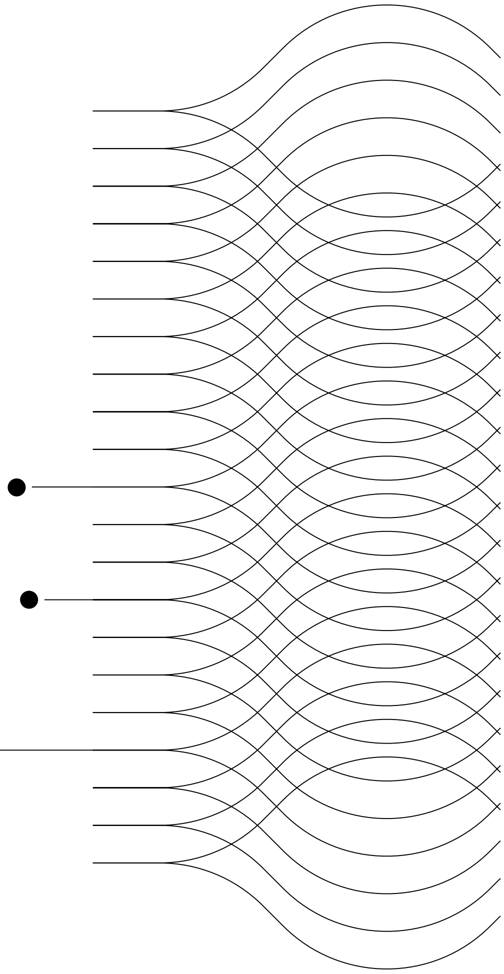
As we begin to apprehend the broad scope of what AI can do, we also begin to take stock of the very real possibility that it might stray and do more harm than good. This question invites us to critically reflect on the ethics of AI.

3.1. Ethics: looking for answers in a complex world

Ethics to help guide companies as they develop AI

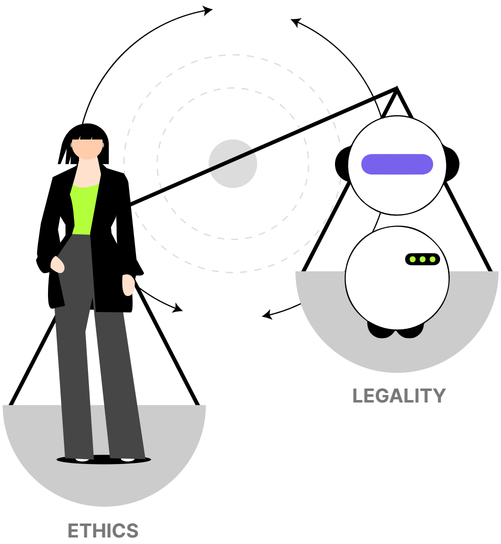
A reflection on ethics is threefold:

- First, it must consider **IMPERATIVES**, as in, base principles that one wishes to live by and which must remain unbroken. For instance, privacy is considered to be a basic human right.
- Second, it must try to anticipate and evaluate **CONSEQUENCES**, as in, the direct or indirect results of an action or a decision. For instance, the reason we authorise driverless cars is because their direct consequence is reducing fatalities.
- Third, it takes into account the idea of **VIRTUOUS BEHAVIOUR**: when we use robotic AI companies to take care of older persons, we do it because there is benevolence in doing so, and because society is made better by this action on a moral level.



Companies wishing to integrate AI into our society seamlessly should view it as **a new phenomenon which changes our way of inhabiting the world**: the fact that we are going to be increasingly surrounded by intelligent machines may alter our balance. Companies are likely to be in charge of leading and encouraging this debate, and the values they decide to integrate – be they implicit or explicit – will radiate towards other parts of our society. They should be careful of harmful dynamics they could introduce and which might be **harder to fix** further down the road.

What is relatively new for scholars in this area of thought is that their contributions to debate around AI are not destined to be quietly forgotten in a drawer, or hidden in a library, as research sometimes is. In this area, any and all reflections on values and principles are going to be essential for companies, employees, customers, and all those likely to be affected by AI.



Ethics to think more deeply about values, and not just about legality

A legal approach to considering social problems rests on the notion of *conformity*. If a certain behaviour is not *conform* to the rule of law, it is prohibited. Therefore, there are two distinct categories in the law: legal and illegal. For example, if a firm fails to comply with certain norms of technical safety or data protection, the firm is in the wrong **according to the law**. The advantage in approaching things this way, and the reason it is so effective at making our society work smoothly, is that it creates **rules that are directly applicable**: one can usually check rather easily whether or not something is legal.

But laws are not enough because they cannot define everything, and because not everything is either good or bad. Some abstract ideas are so complex that there is no right answer. Can we define the notion of human well-being or human accountability through the legal lens only? We might decide that some practices are unethical, and should be banned; but we cannot make a list of all the practices that would be ethical, and in what context, and why.

That is why deciding on their ethics is up to each *individual firm*. One might choose to focus on well-being, control, protection, or any other value which is important to the company. It must decide for itself which boundaries it is ready to cross for progress, and which ones it wishes to preserve at all costs. Ethics is not just about prohibiting or about dictating precise and absolute rules: **it is about determining which society the company wants to build**.

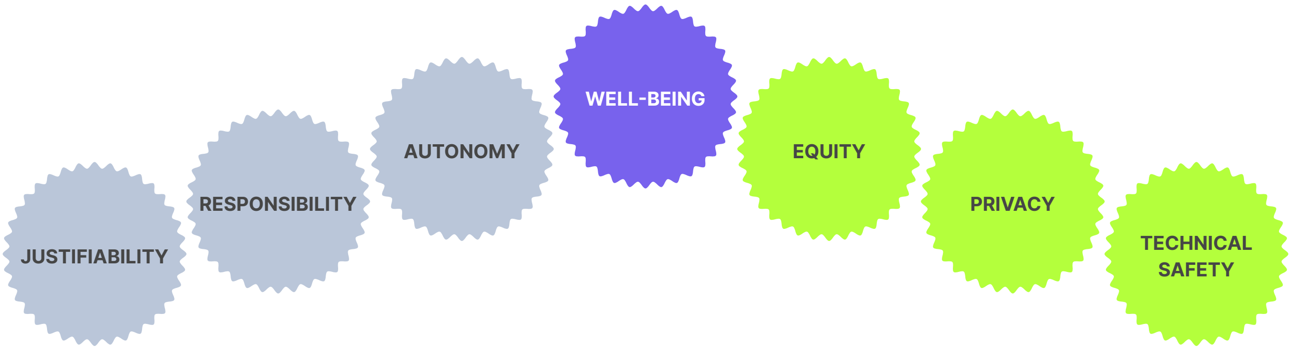
Therefore, ethics implies an open discussion, in order to address the problems that should be fixed, the rules that should not be broken, long-term principles that should be followed, and a vision for society.

3.2. An overview of ethical AI issues

WHAT ARE SOME CONCRETE PRINCIPLES WHICH COMPANIES CAN USE AS THEY DEVELOP AI? We have attempted to clearly formulate some of the fundamental questions and identify the critical challenges which must be faced by companies. What follows is an overview of the ethical AI principles being debated by firms, researchers and experts across the world.

We have grouped these principles into three categories, three essential aspects of human life which are affected by AI, three main points for companies to pay attention to:

1. **human control** (autonomy, responsibility, justifiability): drawing the line between humans and AI
2. **human life quality** (well-being): ensuring that firms develop AI for good
3. **human protection** (equity, privacy, technical safety): preventing AI from harming groups and individuals

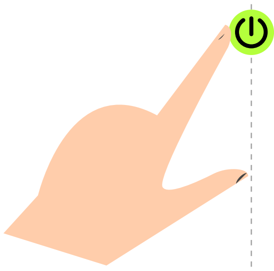


1. Autonomy: respecting human autonomy

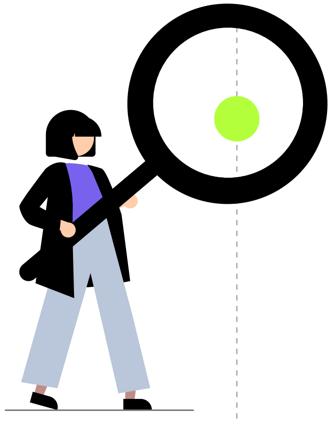
The word “autonomy” is derived from two ancient greek words: *autós*, meaning *self*, and *nomos*, meaning *the law, the rule*. Autonomy is living by self-imposed laws. An autonomous individual is someone who gives themselves their own code of conduct: they **rule over themselves**.

A philosophical tradition dating back to the Age of Enlightenment sees autonomy as a central part of **being human**, and more specifically, **human morality**: human beings become free by **determining which moral rules they wish to abide by**. In doing so, they gain initiative over their own actions and consciousness of the weight of their conduct. They also gain the ability to decide *for* themselves and *by* themselves, without needing someone to do things for them.

What does AI have to do with autonomy? The risk of AI is that we might depend too much on it and **lose control** over its actions, or be **unable to act** without it. From automatic spell-checkers, to checking out of a store, to AI assistance in judging cases in court, it accompanies us more and more in our daily lives. It thus changes our relationship to our actions: while we may act on our own initiative (initiative being a critical element of autonomy), several of our actions now tend to be supported by AI intervention, especially when they are on a screen.



If these new tools which heavily rely on AI are opaque to us, we risk becoming more dependent and losing part of our autonomy to AI. If, however, the working of those tools is **transparent**, and we use them while being aware of the ways they are helping us, they might increase our potential to act, **making us freer than before**.



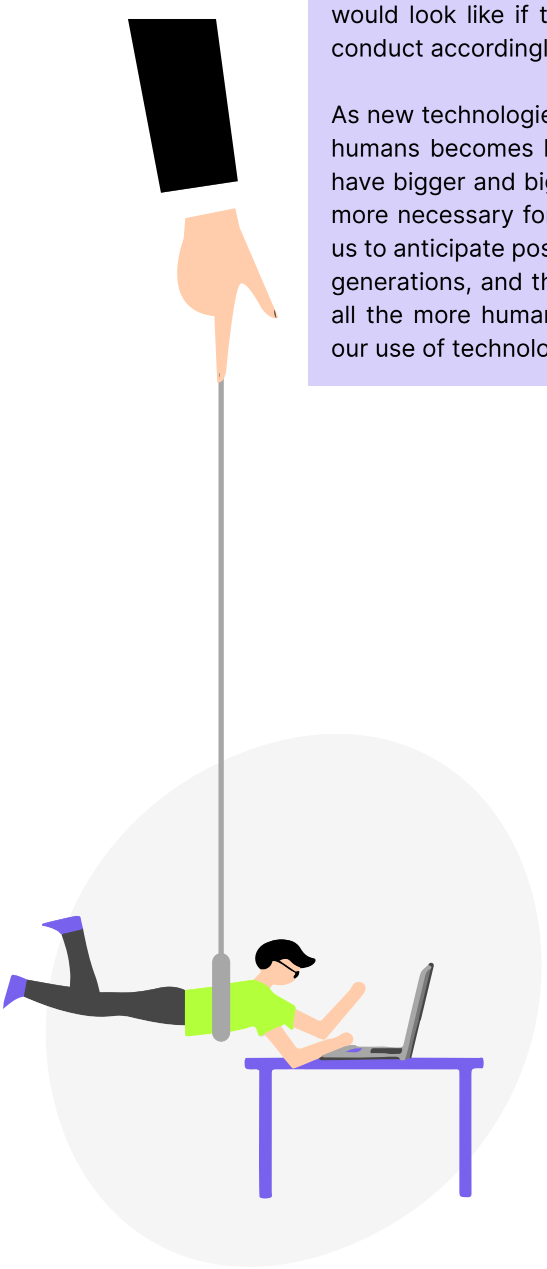
Operational principles for autonomy:
To help companies respect human autonomy when developing AI, some operational principles have been laid out. OPERATIONAL PRINCIPLES ARE CONTENTS THAT MINDFUL COMPANIES MUST CHECK IN ORDER TO PUT THEIR VALUES INTO PRACTICE. The more precisely the following issues are addressed, the more intelligently AI will be developed.

- **Human-in-the-loop**: every time AI makes a decision, a person has to give their approval before the AI can go through with it. Most of the time, this is difficult to put into place, and might not be desirable: though it might be reassuring, it inhibits AI’s power to accomplish mechanical tasks quickly and efficiently, making it less useful to us.
- **Human-on-the-loop**: humans can intervene when the system is being designed, they can monitor its evolution, they can update it, readjust it, or stop it.
- **Human-in-command**: humans can oversee the overall activity of the AI. Although this is difficult to set up, it appears to be necessary in the long term. It would mean that specific guidance and missions about ethical consequences as well as social expectations must be put into place inside the company. Firms must understand whether their particular AI product has an impact on social problems, and to which degree they can intervene while the system is being used: whether they understand what is happening, whether they can update it, or whether they can stop it.

2. Responsibility: determining who is responsible for what AI does

Human autonomy raises the question of responsibility: if people are not entirely in charge of what AI does, it is unclear who should be responsible in case something goes wrong.

The concept of autonomy makes us consider things in the present: who is in control of the decision being taken right now? But the concept of responsibility intertwines the **past** and the **future**, and makes us ask: who should be held accountable for the mistake AI **just made**? Who will be held accountable if AI makes a mistake **in the future**?



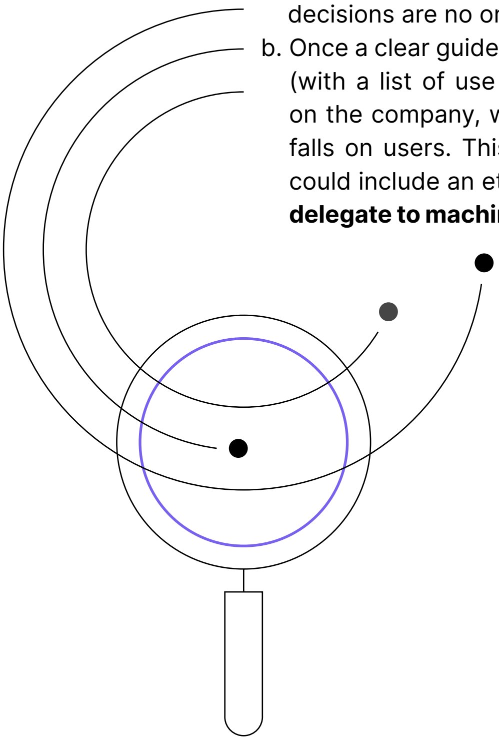
In short, responsibility is about **consequences**. In the second half of the 20th century, philosophers such as Hans Jonas defined responsibility as an imperative, a moral obligation: humans should think ahead before acting, and refrain from acting if they anticipate negative consequences, especially in regards to human life. They should **imagine** what the future would look like if they acted a certain way, and modify their conduct accordingly.

As new technologies are developed, the scope of action for us humans becomes larger and larger, and wrong decisions can have bigger and bigger consequences. That is why it is all the more necessary for us to think things through: it is critical for us to anticipate possible long-term effects of AI on us, the next generations, and the rest of the world around us. We become all the more human when we hold ourselves accountable for our use of technology and its potential downsides.

Operational principles for responsibility:
To help companies determine who is responsible when working with AI, some operational principles have been laid out. **Operational principles are contents that mindful companies must check in order to put their values into practice**. The more precisely the following issues are addressed, the more intelligently AI will be developed. On responsibility, we need a clear distinction between areas where machines are responsible, and areas where humans are.

- 1. Before making AI operational: what we want AI to do, and what we do not want it to do. AI might be able to have some latitude in some areas and not need a human to oversee everything. Companies should ask for **expert opinions** (data-scientists, but also law firms, social scientists and philosophers), identify **trade-offs**, and **include diverse points of view** to deal with these cases. For instance, when a driverless vehicle has to park, should a human get behind the wheel to perform this delicate task? At a higher degree, companies should adapt to different cultures and integrate different types of values: some countries might be more or less reluctant to give up control of their vehicles; some countries might value certain human rights, such as privacy, more than other rights.

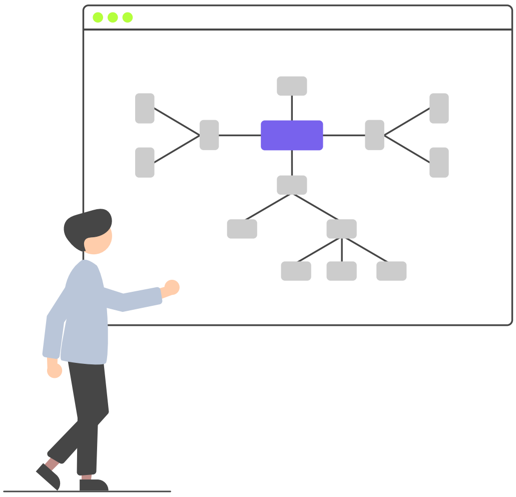
- 2. After AI has acted: in case something goes wrong, **who is responsible**?
    a. It is up to companies to **establish global guidelines** stating who is responsible in which case. For instance, some companies claim that humans (developers, or users, or others) are always responsible. Others believe that some or all AI decisions are no one's fault.
    b. Once a clear guideline is established, companies can create imputation systems (with a list of use cases) to have a clear legal distinction between what falls on the company, what falls on the factory, what falls on developers, and what falls on users. This document should be more than just a legal framework: it could include an ethical position, with an evaluation on **how much humans can delegate to machines**. It should therefore be justified by a specific set of values.



3. Justifiability: making AI's decision process transparent

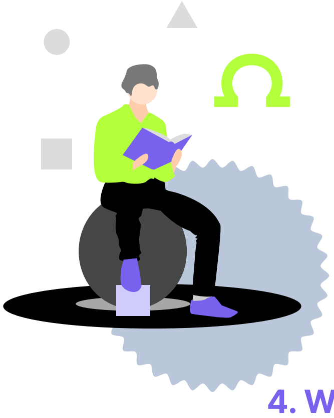
Figuring out the extent to which humans should take accountability for AI decisions implies **understanding the way the system works**. If users have no clue, they can hardly challenge decisions made by AI even when these decisions may seem unfair (*e.g.* not granting a loan). Justifiability means **the entire decision process must be decomposed and explained**.

In 2019, the philosopher Brent Mittelstadt determined two ways to make AI justifiable. The first is **transparency**: making clear to a human which steps the algorithm is taking, in such a way that the human can see the overall functioning of the system, and follow along. The second, **post-hoc interpretability**, goes a little deeper: faced with an explainable algorithm, a human can understand why and how the algorithm is taking its decisions. Whether a company chooses one or the other, attempting to make an algorithm justifiable poses ethical problems.



Operational principles for justifiability:
To help companies make their AI justifiable, some operational principles have been laid out. **Operational principles are contents that mindful companies must check in order to put their values into practice**. The more precisely the following issues are addressed, the more intelligently AI will be developed. On justifiability, guaranteeing technical transparency is one thing, ensuring humans concerned by AI have an ethical attitude is another. Hence, two criteria should be met for justifiability:

1. **Technical transparency**: the way the algorithms are built needs to be as clear as possible. Companies should aim for 3 degrees of transparency:
   a. The way the algorithm is **built** must be clear (*e.g.* how a neural network was structured)
   b. Individual components: the way important **features** are determined should be explicit
   c. The way the models are **trained** should be transparent: for example, how are the "better" models selected (*e.g.* balance between equity and efficiency) ?
2. **Human explainability**: people who use the AI should understand what they are doing.
   a. From the designer's point of view: the designer should give the user the keys to understanding the algorithm. Are there clear explanations provided, be it through the **graphic interface** or through **personal dialog**? For instance, when the algorithm gives an output, is it clear how this output was obtained?
   b. From the final user's point of view: based on a technical interpretation of the results, the user makes a decision. Is this explanation basis clear enough to be used as **a reference tool for humans**? Is the prediction delivered by the machine understandable enough to be seen as something that a human could have done (*e.g.* an AI gives investment advice which appears rational to a wealth manager)?



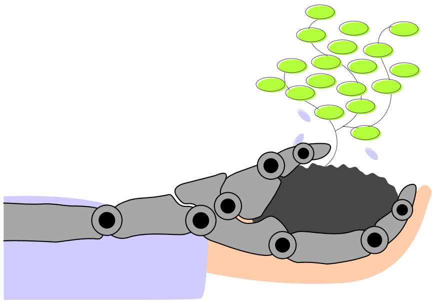
4. Well-being: ensuring that AI is beneficial to our lives

AI visibly allows us to be more efficient in many fields, increases productivity, and introduces more comfort in our daily lives. It is rapidly modifying some of our daily habits, concerning transportation, health, our work lives, and more.

But human well-being is complex and does not only depend on comfort or being more productive. While it might be tempting to be helped by machines all the time, it risks making our lives lose some of their **meaning**: what would there be left for us to do? Comfort is not always desirable, and effort is usually good for us to some extent.

Again, ancient greek philosophers can help us understand the concept of well-being better. The ancient greek word *eudaimonia* refers to a global state of **personal satisfaction** which is **sustainable** on the long term. It views happiness as a state of accomplishment, closely linked with taking action and revealing one's true potential. The philosopher Aristotle believes that it is when one reveals their *excellence* – by excelling at the task they have freely chosen, and in a way that is helpful to society – that they reach *eudaimonia*.

Since the criteria for well-being hugely vary from person to person, the development of AI should allow every individual to work towards their personal goals by providing them with the tools to be free – or, at the very least, it should not hinder their progress.



Operational principles for well-being:
To help companies preserve well-being while developing AI, some operational principles have been laid out. **Operational principles are contents that mindful companies must check in order to put their values into practice**. The more precisely the following issues are addressed, the more intelligently AI will be developed. On well-being, companies should adopt a clear stance on how to develop "AI for good", by considering several dimensions of well-being:

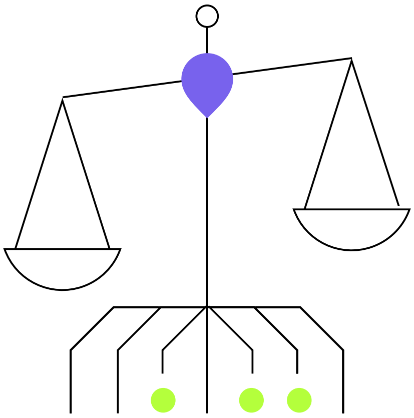
1. Providing people with services: companies can expose the ways in which its AI use cases could enhance quality of life (for instance, by automatising repetitive tasks which workers found alienating). They should also determine whether their AI solution has a positive or negative impact on important areas such as environment (for instance, if their hardware uses too much energy).
2. A **deeper insight** into these positive applications: are they good on the **long term**, or just on the short term? Are they merely comfortable, are they making users lazier or weaker, or are they improving mental health and **making society function better**? Do they line up with a strong **set of values** decided beforehand by the company – for instance, do they respect their users' privacy? Are they just useful, or are they also ethical?
3. Once the consequences of the AI solution have been clearly established by use cases, firms have a duty to **expose the potential risks** of their technology, as well as the positive impact it can have. They could use videos or other educational tools in order to let their customers know what to expect. This will create trust and allow users to make the best possible use of the product.

5. Equity: creating a fair AI

To understand the concept of equity, we must first link it back to equality.

Equality measures the extent to which **value is evenly divided** between people. If there are ten portions available, ten people in a room, and each person has one portion, then the situation is equal. According to French sociologist Bourdieu, value can be divided among people in very different ways, depending on which kinds of actions are recognised as valuable (is a janitor seen as valuable, is wealth management seen as valuable?) and how deep the gap between them can be (is an intern seen as more or less valuable than the CEO?). Equality would be attributing the same amount of value to everyone.

Equity is about making sure the attribution of value is **fair**. An example of equity would be giving an extra boost to those who start out with less. It is not about giving everyone the same amount, but about trying to even out the initial situation.



In short, the main idea behind equity is fairness and just, unbiased treatment. In AI, there is a risk that algorithms might **learn and reproduce unfair mechanisms**. Access to AI might also be a problem: how can we guarantee that this technology benefits everyone without discrimination?

Operational principles for equity:
To help companies develop a fair AI, some operational principles have been laid out. **Operational principles are contents that mindful companies must check in order to put their values into practice**. The more precisely the following issues are addressed, the more intelligently AI will be developed.

1. Documentation: the developer must make sure they are not reproducing **unfair social biases**. They must identify minorities which could be discriminated against by their AI model and aim to eliminate potentially discriminatory variables.
2. Inclusion: to what extent is the product accessible to different types of users? Designers should clearly distinguish **categories of users** in order to tailor the product to them.
3. **Redress**: make sure that the **identified biases have been corrected or avoided**. This is particularly important whenever these biases may cause harms, in facial recognition softwares, large modern natural language models, or decisions systems based on personal structured data.

6. Privacy: respect for the private lives of citizens

According to the *Stanford Encyclopaedia of Philosophy*, privacy acts as a sphere within which individuals can be **free from interference by others**. It is considered to be part of the fundamental rights of individuals: in fact, there is a legal right to privacy.

With AI, the main difficulty is knowing **where to place barriers to data sorting and synthesising**. Since the quality of AI comes from its training, and its training is based on the volume of collected data, privacy is a legal and ethical limit to its efficiency.

Operational principles for privacy:
To help companies respect citizens' privacy while developing AI, some operational principles have been laid out. **Operational principles are contents that mindful companies must check in order to put their values into practice**. The more precisely the following issues are addressed, the more intelligently AI will be developed. On privacy, there are both ethical and legal issues to be considered: our main objective should be to protect individuals and their personal data.

- Data exploitation: are there features making consumer products vulnerable (*e.g.* smart home appliances)?
- Data tracking: AI can *de-anonymise* data by cross-referencing multiple traces left by an IT user. Can this effect be avoided?
- Profiling: to what extent can and should data be used to deduct other types of private information, such as ethnic identity or emotional state, thanks to voice recognition and facial recognition? Are there concrete measures to prevent data from being diverted, for example to give citizens scores without their **consent**?

7. Technical robustness: ensuring that AI is robust, safe and secure

Technical robustness is at the base of all our other principles: if an AI is easy to manipulate from the outside, and data can be diverted, it cannot be trustworthy. When crucial decisions are delegated to AI, no error margin is allowed. Privacy cannot be guaranteed either if the system is easy to hack into.

Operational principles for technical robustness:
To help companies develop a robust AI, some operational principles have been laid out. **Operational principles are contents that mindful companies must check in order to put their values into practice**. The more precisely the following issues are addressed, the more intelligently AI will be developed.

- **Protection** of the system: are designers and developers taking rigorous steps to be resilient against attacks from the outside? Physically, how is the hardware protected? Are there systematic procedures to prevent data or models from being harmed (*e.g.* with data poisoning or model leakage), by adversarial attacks, for instance?
- **Fallback plans**: in case of emergency, a fallback plan should be elaborated (*e.g.* while blocking the data or referring to a human operator).
- Occasional **processing** errors: the system should immediately notify users of its dysfunction. It could also indicate how likely it is to malfunction.

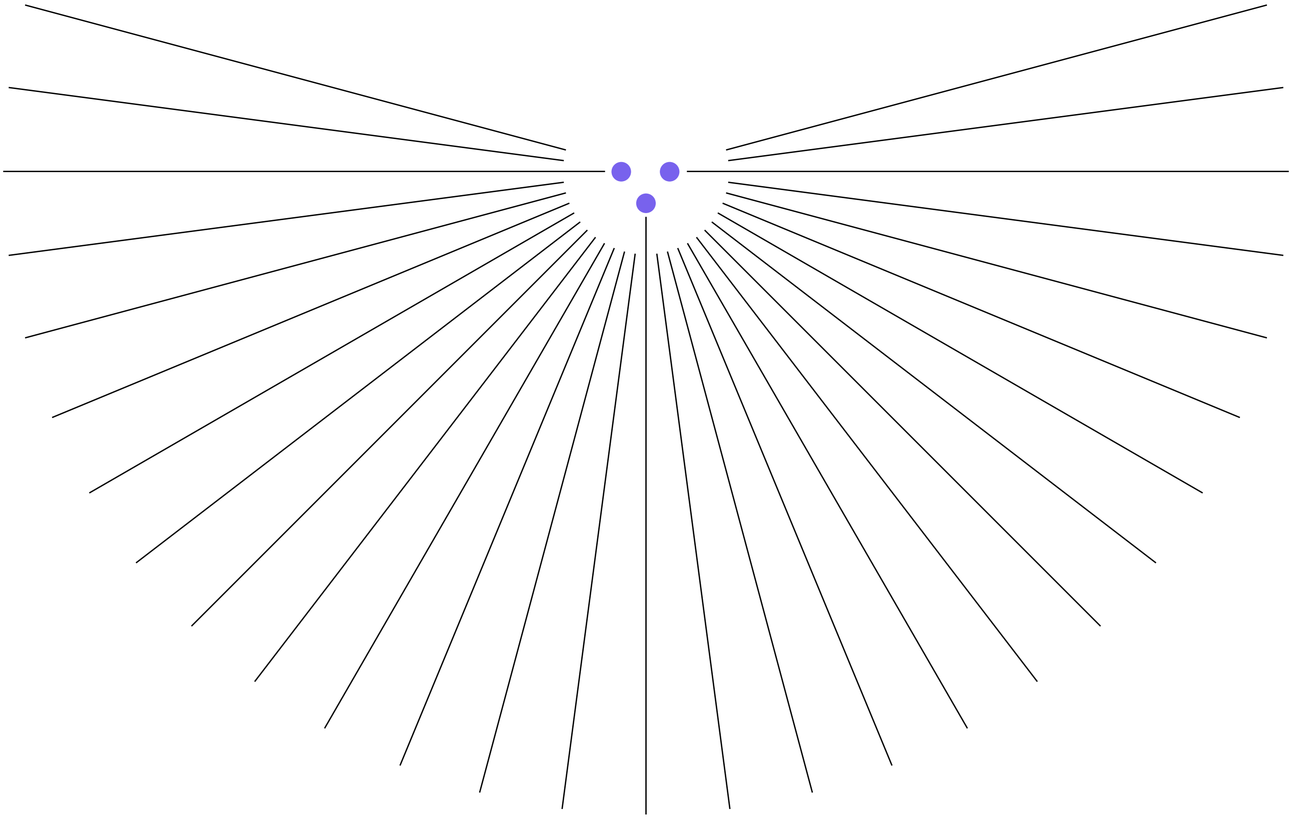
3.3. Ethics and AI around the world in 2021

A quick timeline

The idea that AI could be a threat to humans began to emerge **in the second half of the 20th century** with Isaac Asimov's books and films such as Stanley Kubrick's *2001: A Space Odyssey* (1968). The first debates around the ethics of AI date back to the early 2000s, and from 2010 onwards it has become a burning topic.

SINCE 2017, SEVERAL COUNTRIES HAVE RELEASED A STRATEGY AROUND AI, be it through executive orders (in the US), a declaration (in Canada), a high-level expert group (in Europe), a report (in France: the Villani report) and in China (the New Generation Artificial Intelligence Development Plan). They have all been followed by reports from the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), the International Labour Organisation (ILO), the United Nations and the Institute of Electrical and Electronics Engineers (IEEE).

Large companies such as Google and Microsoft have also created ethics committees on AI and released ethical principles for AI teams to follow as they develop AI-based solutions.



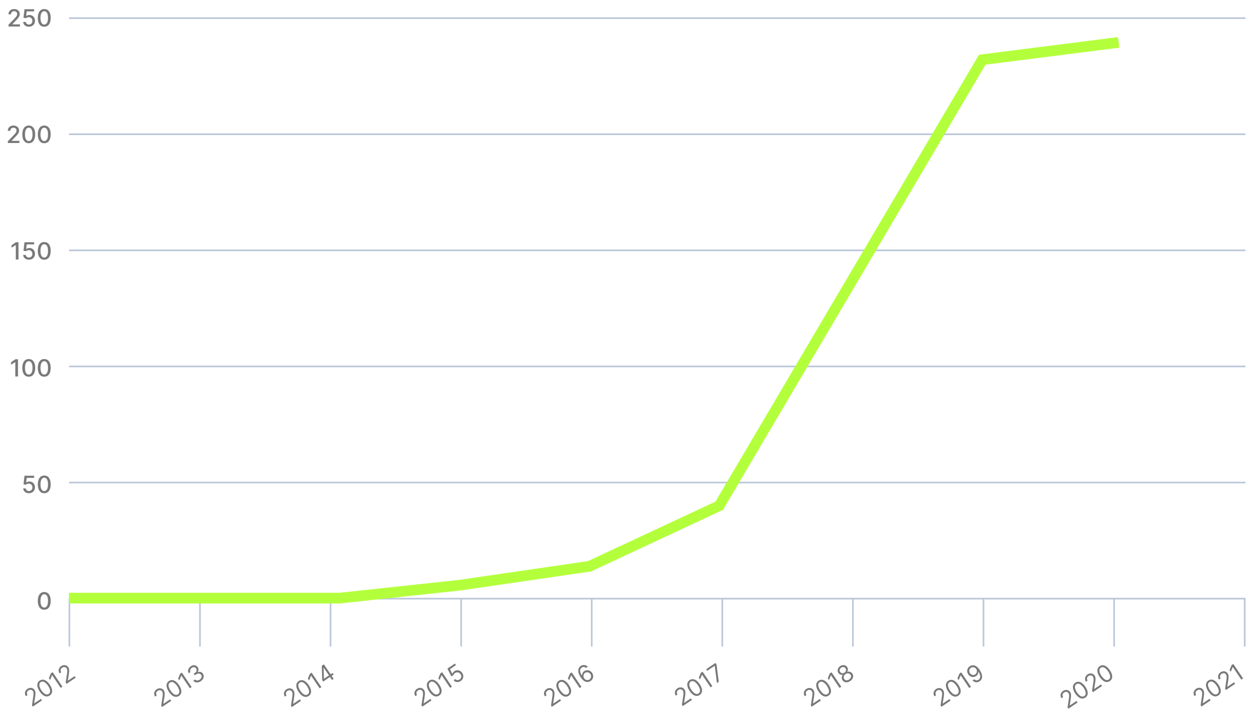
THE NUMBER OF PUBLICATIONS ON AI ETHICS INCREASES FROM YEAR TO YEAR

COMMENT:

Here we count the publications on ethical AI listed by Cathy Baxter, in “Ethics in AI resarch papers and articles” (a good proxy for the overall number of publications on AI ethics), using a small machine-learning algorithm (web-scraping).

INTERPRETATION:

Even if pioneers (companies and thinkers) already highlight the need for ethical AI (especially since 2016), there is a still increasing number of publications on ethical AI especially since 2018, when the first countries and organisations published their reflections and strategies on ethical AI.



Debates around ethics in AI

On one hand, some organisations and companies highlight the benefits of an ethical, human-centred and **trustworthy AI. On the other**, some individuals, companies and countries such as China and the US mostly focus on new profitable applications, even though they are open to the development of an ethical AI (aligned with their values).

They see AI as a **strategic technology** which will determine which countries or organisations have the upper hand in the years to come, and they do not want to miss out on potential profitable applications which might be in a moral grey area.

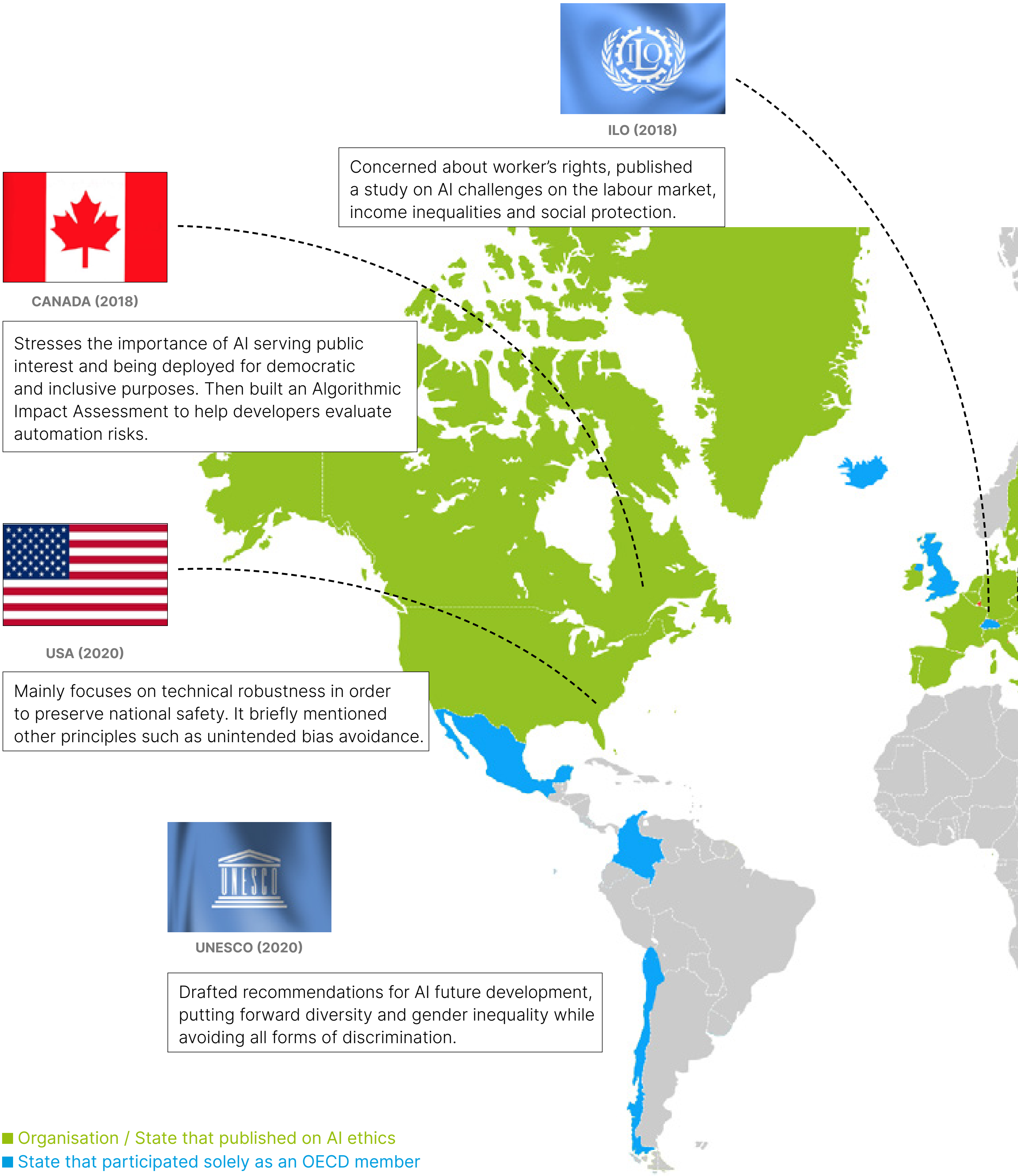
Some experts insist on the importance of explainability (as in, developing a transparent AI); but some companies argue that the most important thing is having an AI that works and is performing. Some associations, researchers, companies and politicians are also worried about AI incorporating discriminatory biases and argue for the need to create fair, robust AI algorithms.
Regulators tend to promote guidelines and safeguards around AI; however **most of them are not thinking about ethics, but rather about legality**: their concern is making AI comply with existing laws.

ULTIMATELY, IT MAY NOT TRULY BE ABOUT WHETHER OR NOT WE INCORPORATE ETHICS INTO AI BUT TO WHAT EXTENT. Some applications of AI may not require explainability, others may not need to be robust or fair. Some may influence the well-being of millions or billions of people, others may not have a wide-spread effect.

But more and more countries seem to be of the opinion that ethics is important in AI, and in 2020 the US Pentagon released 10 principles to ensure a controlled development of AI in the next years. There are strong discrepancies between countries and organisations on the principles that are thought to be meaningful for AI, but as of now, **what is important is that the issue is being debated** and that some answers are being formulated.

WHERE ETHICAL AI GOALS ARE PURSUED

THE MAIN ORGANISATIONS THAT PUBLISHED ON AI ETHICS WORLDWIDE



ILO (2018)

Concerned about worker’s rights, published a study on AI challenges on the labour market, income inequalities and social protection.

CANADA (2018)

Stresses the importance of AI serving public interest and being deployed for democratic and inclusive purposes. Then built an Algorithmic Impact Assessment to help developers evaluate automation risks.

USA (2020)

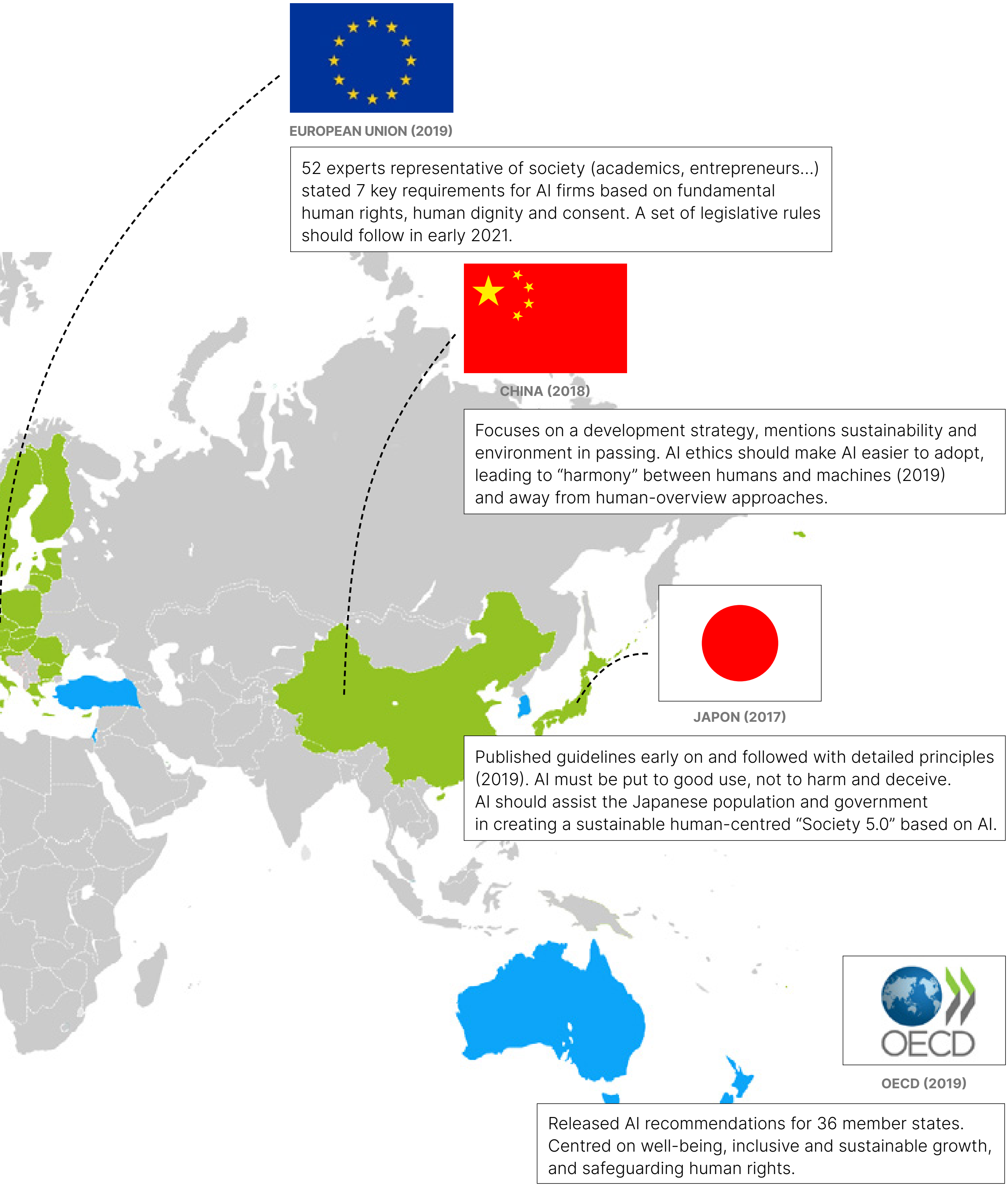
Mainly focuses on technical robustness in order to preserve national safety. It briefly mentioned other principles such as unintended bias avoidance.

UNESCO (2020)

Drafted recommendations for AI future development, putting forward diversity and gender inequality while avoiding all forms of discrimination.

🟩 Organisation / State that published on AI ethics

🟦 State that participated solely as an OECD member



EUROPEAN UNION (2019)

52 experts representative of society (academics, entrepreneurs...) stated 7 key requirements for AI firms based on fundamental human rights, human dignity and consent. A set of legislative rules should follow in early 2021.

CHINA (2018)

Focuses on a development strategy, mentions sustainability and environment in passing. AI ethics should make AI easier to adopt, leading to “harmony” between humans and machines (2019) and away from human-overview approaches.

JAPON (2017)

Published guidelines early on and followed with detailed principles (2019). AI must be put to good use, not to harm and deceive. AI should assist the Japanese population and government in creating a sustainable human-centred “Society 5.0” based on AI.

OECD (2019)

Released AI recommendations for 36 member states. Centred on well-being, inclusive and sustainable growth, and safeguarding human rights.

Text analysis: Applied AI ethics in 2021 – which ethics for which guideline?

COMMENT:
We tried to answer the following question: **what different ethical approaches are there across the world**?

To do so, we went through major papers detailing ethical strategies. We used a natural language processing algorithm, which outputs the most used words in a document (the top 250 words) depending on how long the document is – that is, by using the *term frequency-inverse document frequency* method (TF-IDF).

The more words are used (*e.g.* “well-being”, “control”) which are relevant to a given ethical principle (technical safety, responsibility...), **the more we consider that ethical principle to be significant in the paper’s guideline**.

Our analysis therefore led us to more precise keywords. For example, we split well-being into 21 key categories linked with the 17 Sustainable Development Goals adopted under United Nations Agenda 2030 (*e.g.* “no poverty”, “quality education”, “climate action”, “innovation”).

ethical
document
impact
mechanism
ai
trustworthy ai
fundamental right
principle
technology
requirement
ensure
human
may
must
data
right
decision
risk
ai system
society
assessment list
used
potential
individual
process
also
including
system
stakeholder
ensuring

HOW WE LED OUR ANALYSIS: THE PRINCIPLES, SUBPRINCIPLES AND KEYWORDS WE USED

well-being	**environmental footprint**	carbon footprint	environment, energy, ecologic, sustainable development, well-being, nature
		energy consumption	environment, energy, ecologic, sustainable development, well-being
		waste generation	environment, ecologic, sustainable development, well-being
		water consumption	environment, ecologic, sustainable development, well-being
		biodiversity preservation	environment, ecologic, sustainable development, well-being, preservation
	social-economic footprint	job production	future work, labour market, unemployment, employment, worker
		reduction of poverty	development, income inequality, social protection, human dignity
		education	education, knowledge, knowledge society, skills, competence, training, literacy numerics
		collaboration	collaboration
	cultural footprint	technical progress	public perception, research, ingenieering, innovation, technological change, new technology, human advantage
		liberty increase	freedom
		equality increase	inequality, developing country
		respect of human right	working conditions, fundamental rights, human dignity, human rights, fundational value
	health footprint	longer life expectancy	health
		fighting significant diseases	health
		day to day health	health
		access to better food	health
	economic footprint	value creation	development, market, application, productivity, value, benefit, economic, value
		wage increase	wage, income
		access to capital	financing, industry investment, capital, equity investment
		wealth distribution	tax
	discretion		

autonomy	**human in the loop**		life cycle, oversight, autonomy
	human on the loop		life cycle, oversight
	human in command		automation, governance, oversight, training
	human consent		oversight
	empowerment		oversight
human responsibility	**accountability**		accountability, responsible, responsibility
	control and capacity to audit		tracable, danger, concern, responsible, responsibility, control
	moral integrity		human dignity, moral, responsible, responsibility
	social responsibility		next generations, responsible, responsibility
justifiability	**transparency**		transparency
	explainability		
	intelligibility		
	auditability		auditable, monitoring
	unfalsifiability		
equity	**bias avoidance**		bias, equitable, justice
	fairness	equal opportunities	fairness
		equalized odds	fairness
		equalized outcomes	fairness
		demographic parity	fairness
	inclusion		human dignity, inclusive
	user centric design		design, end user
	diversity		diversity, gender
	open access		
privacy	**anonymity**		private, privacy
	personal data ownership		personal data, private
	data protection		personal data, private
	data use control and consent		personal data, private
	intimity		intimity, private
safety	**technical robustness**		life cycle, robustness, safety, security
	physical and psychological integrity		security and integrity, safety, security
	reliability of results		life cycle, safety, reliable, security
	prevent unattended harm	accuracy/precision	life cycle, harm, risk, safety, danger, avoid unintended harm, security, concern
		prevent model leakage	life cycle, harm, risk, safety, danger, avoid unintended harm, security
		prevent data poisoning	life cycle, harm, risk, safety, danger, avoid unintended harm, security
		prevent fallback	life cycle, harm, risk, safety, danger, avoid unintended harm, security
	sovereignity	national security protection	defence, war, security
		intellectual property protection	AI patent, security
		respect of the law	compliance, law, legal, regulation, security

INTERPRETATION:

Most of the organisations consider the 7 key principles as relevant for their ethical strategies, but **they attach greater importance to certain ethical issues** (*e.g.* technical safety or well-being) on which their texts focus.

For example, in the Canadian paper, or in the one by the EU, "human autonomy" is discussed more than in the Chinese ones or those from the US, which are more centred around their national strategy and their technical safety.

We propose below 30 subcategories to further precise the expectations for AI related to **technical robustness** and **governance**. This analysis could help create better documentation to precise what behaviour we want to favour for a specific AI product and how we ensured a solid governance and the technical safety of the product when we built it.

PRINCIPLES MENTIONED WITHOUT PRECISIONS	UNESCO	EUROPEAN UNION	ILO	MONTREAL	OECD	US DEFENSE
autonomy		🟪				
human responsibility						
justifiability						
equity						
privacy	🟪				🟪	🟪
safety					🟪	

	SUBPRINCIPLES MENTIONED WITH PRECISION	UNESCO	EUROPEAN UNION	ILO	CANADA	OECD	US DEFENSE	CHINA
autonomy	human in the loop	■						
	human on the loop			■				
	human in command	■		■				
	human consent							
	empowerment							
human responsibility	accountability	■	■			■	■	
	control and capacity to audit	■	■		■		■	
	moral integrity	■	■		■		■	
	social responsibility	■					■	■
justifiability	transparency	■	■			■	■	
	explainability							
	intelligibility							
	auditability	■	■				■	
	unfalsifiability							
equity	bias avoidance	■	■				■	
	fairness		■			■		
	inclusion	■	■					
	user centric design	■	■					
	diversity	■			■			
	open access							
privacy	anonymity							
	personal data ownership				■			
	data protection				■			
	data use control and consent				■			
	intimity				■			
safety	technical robustness	■	■				■	
	physical and psychological integrity	■	■		■		■	■
	reliability of results	■	■				■	■
	prevent unnattended harm	■	■		■		■	■
others	sovereignity		■				■	■
	respect of the law		■	■		■	■	■

The analysis below could **help create better documentation** to precise what behaviour we want to favour for a specific AI product and what potential effect it can have on **well-being**.

			UNESCO	EUROPEAN UNION	ILO	CANADA	OECD	US DEFENSE	CHINA
WELL-BEING	**environmental footprint**	Sustainable Goal 13 : Climate action	■	■		■	■		■
		Sustainable Goal 7 : Affordable and clean energy	■	■		■	■		■
		Sustainable Goal 6 : Clean water and sanitation	■	■		■	■		
		Sustainable Goal 14 : Life below water	■	■		■	■		
		Sustainable Goal 15 : Life on land	■	■		■	■		
		Sustainable Goal 11 : Sustainable cities and communities	■	■		■	■		
	social-economic footprint	Sustainable Goal 8 : Decent work and economic growth			■		■		■
		Sustainable Goal 1 : No poverty		■	■				■
		Sustainable Goal 4 : Quality education	■	■	■	■	■		■
		Sustainable Goal 17 : Partnership for the goals							■
	cultural footprint	Sustainable Goal 9 : Industry, innovation and infrastructure	■		■		■	■	■
		Sustainable Goal 16 : Peace, justice and strong institutions	■	■	■				■
		Sustainable Goal 5 : Gender equality							
	health footprint	Sustainable Goal 3 : Good health and well-being	■						
		Sustainable Goal 2 : Zero hunger	■						
	economic footprint	Sustainable Goal 12 : Responsible consumption and production			■	■	■		■
		Sustainable Goal 10 : Reduced inequalities			■				

SOME CONCLUSIONS:

We can draw some conclusions from our TF-IDF method. In particular, they enable us to identify that **some concepts are still relatively vague, as they are described in the AI ethical strategies** we analysed:

- Further research and strategies must define some concepts more precisely, such as autonomy or unattended harm (does it mean the AI models focus on preventing fallback, model leakage or data poisoning, or even on accuracy?). Concerning justifiability, research should also distinguish more clearly between justifiability, transparency, intelligibility, and explainability.
- Organisations will also have to adopt a clearer positioning on open-access, fairness (*e.g.* to choose between equal opportunities, equalized odds, equalized outcomes and democratic parity approaches) and privacy (besides the *individual* concept of privacy, intimacy *between people* talking or sending messages to each other is another challenge of AI that should be discussed).

Our text analysis led us to another conclusion. The notion of well-being associated with AI is conceptually vague for the moment, but the way different guidelines describe it enabled us to build a taxonomy of different well-being aspects. Yet, we noticed afterwards that our 21 categories are extremely close to the 17 UN Sustainable Development Goals. For example, the 4 categories inside the "social-economic footprint" correspond exactly to 4 goals of the United Nations for 2030: "reduction of poverty" and Sustainable Goal 1, "education" and Sustainable Goal 4, "collaboration" and Sustainable Goal 17…

Noteworthy, the *17 Sustainable Goals* of the United Nations could become an analytical framework for the *well-being* principle applied to AI.

SUSTAINABLE DEVELOPMENT GOALS
1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
SUSTAINABLE DEVELOPMENT GOALS

4

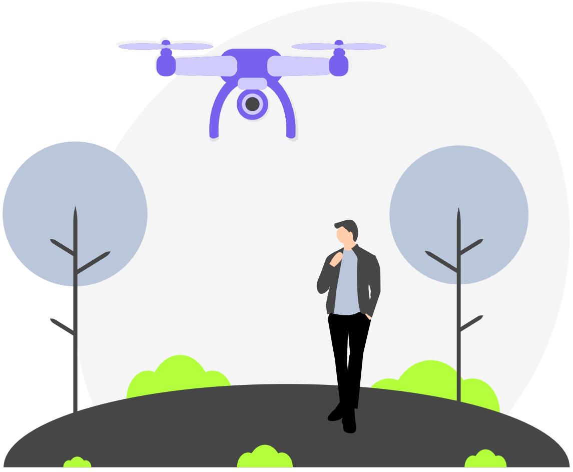
The benefits of an ethical and trustworthy AI : concrete applications

4.1. Selected applications of AI and their ethical dimension

Here, we present some fields where AI is being or has been deployed. We also highlight the **most important ethical principles in each field** and give some clues as to **how companies can handle ethical challenges** posed by AI. Then we put into perspective these ethical issues in each sector, estimating their level of urgency.

Agriculture

Due to the modernisation waves which have taken place since the 1940s in the West, workers in agriculture are accustomed to rapid changes in their tasks and organisation habits. This makes agriculture a fertile ground for implementing AI.



Agriculture has to supply the entire world population with food without using up the Earth's resources. AI can help by making agriculture both more efficient and more sustainable:

- In livestock, farmers have to supervise their cows night and day to guarantee the quality of produced milk. To help them do so, a company named **Cowlar** produced a collar capable of measuring the cows' temperatures, detecting when they are in heat or when they have a health issue, and monitoring their behaviour. Farmers are instantly notified on their phones if something is wrong. AI therefore helps them care for their cows better and maintain the quality of their milk.
- **Drones using AI can help flag diseases affecting plants**, and identify areas which lack irrigation or nutrients. A spectral sensor named Sequoia (developed by the French firm Parrot) accomplishes this by comparing sun activity with the status of the crops, and deduces crop vitality by capturing the amount of light the plants absorb and reflect.

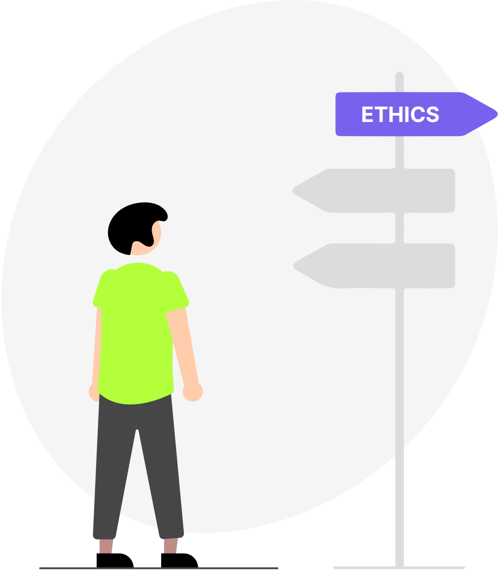
As AI promotes a resource-efficient management of the land, it can be broadly developed in an ecological way, even outside of agriculture:

- Since it is able to collect and analyse a great amount of data, it can help researchers identify zones that are critical for protecting the **environment**. For instance, drones are used in Guinea to monitor changes in vegetation cover and to improve the life of chimpanzees in the Mont Nimba biosphere.
- Drones equipped with AI can also be used to combat **deforestation** and **illegal hunting**. The World Wildlife Fund has launched a 5 million dollar project which counts of AI embedded systems to track poachers in the Maasai Mara, Kenya.
- AI can also help predict **natural disasters** such as hurricanes. Private firms in Sendai, Japan are currently testing AI devices to alert the population in case there is a tsunami.

Most relevant ethical principles in agriculture: justifiability, equity, well-being

Justifiability is important to understand the sensors and what they have detected, enabling human to be in control; **inclusion** of developing countries is essential. Technical safety is also important and humans must be responsible when failures have a strong impact (*e.g.* AI must be robust enough to avoid spraying of pesticides or to detect irrigation problems). Preserving **well-being** must be taken into consideration above all.

Well-being appears to be the most important issue here: AI in agriculture can allow us to use resources in a careful manner, so as to feed everyone on Earth in a sustainable way while respecting the environment. We should be wary when AI is used to pollute or make agriculture more efficient without respecting any other ethical norms aimed at preserving global well-being.



Military

In a military context, the main application for AI is automatising weapons. In theory, fully automated weapons could replace humans to fight humans. These are sometimes referred to as "killer robots". There are two **opposing points of view on this issue**:

- One side believes that making **quick** and **effective** weapons is the most important thing. Countries such as China, Israel, Russia, South Korea, the UK and the US are developing weapons which are increasingly autonomous.

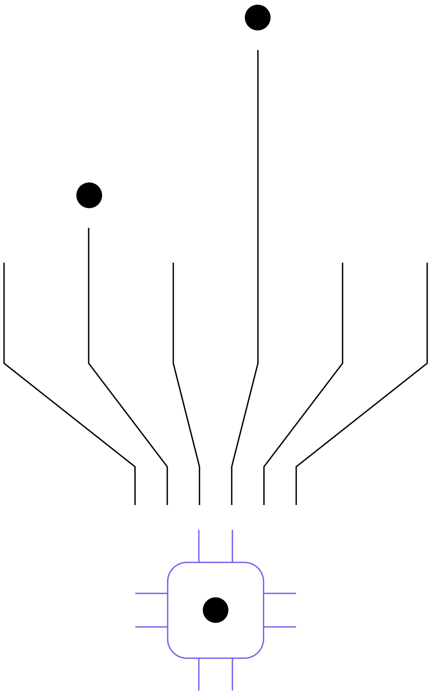
- The other side calls to **ban autonomous weapons**: its main representative is the "Campaign to stop killer robots". This side is supported by AI experts and organisations, as well as 28 countries and rapporteurs appointed by the United Nations Human Rights Council. **In 2018, the European Parliament adopted a resolution** to make autonomous weapons illegal. In March 2020, a global parliamentary campaign was launched on this matter: their aim is to obtain an international treaty which would ban "fully autonomous" weapons.

"Autonomous" weapons are weapons which use machine-learning to **recognise perilous obstacles** and weapons to neutralise: for instance, **RAPIDFire**, an autonomous weapon turret manufactured by Thales, is able to neutralise rockets in the sky. These weapons would be a great asset to soldiers on the battlefield thanks to their accuracy and their ability to remain **alert 24/7**. AI could therefore be used to prevent conflict, to dismantle danger, and thus to resolve tense situations – but these weapons could also become more dangerous than beneficial.

Other uses for AI on the battlefield are less controversial and less discussed, but also worth noting. AI can help predict social unrest and plan for battle. In cybersecurity, machine-learning can neutralise malware before it breaches military networks. In logistics, AI can ensure maintenance and develop autonomous vehicles.

Most relevant ethical principles in the military: autonomy, responsibility, justifiability and technical safety

Autonomy and human *responsibility* obviously pose a problem in this field: to be autonomous and free, we humans must have a way to defend ourselves which we can **control**. Here, legal regulations are important: AI must not go against the **law of war**. The central question is: **when is automatised violence legitimate**? For instance, we could say that automatic weapons can intervene only if the target is not human.



Therefore, *justifiability* comes into play, as we must be absolutely certain that **AI never confuses a human with a rocket**. The system must therefore report on its accuracy and be explainable. The way the machine analyses its surroundings and learns what its targets are must be very explicit.

Technical safety is also crucial as no exterior agents should be able to infiltrate our means of defence. Here more than anywhere, a safe and robust AI is a matter of life and death.

There are more distant issues related to AI in the military, such as well-being (in the long term, fights via the intermediary of robots could reduce battlefield deaths) and equity (*e.g.* reducing the gender imbalance in the army through military recruitment and equipment).

Justice

Justice stands at the basis of our modern democracies and the freedom of citizens. In countries with a fair and functional judiciary system, citizens feel protected, they feel **respected**, and they are quicker to **trust** their government. Because of this, justice is the key to a stable and functional society.

Using algorithms in criminal justice is not new, and scientific methods and technologies are constantly being tested in order to assist proceedings. For instance, databases of **DNA** samples are already used to help identify suspects. AI could give an extra boost to these techniques and bring many benefits to law enforcement, which could help make our society a little more secure and a little more fair.

Solving complex cases, cold cases, and identifying fraud

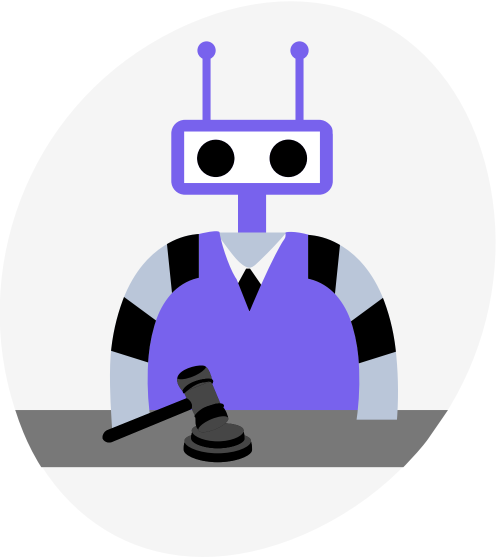
It sometimes takes several years or decades to solve complex cases or to uncover obscure murders, massive frauds, trafficking rings. AI can be an ally in these long, difficult investigations, and help accelerate them. It could **spot subtle clues** that may otherwise have been missed, or anomalies that can contradict alibis. It could help identify suspects by analysing hours of camera recordings in little time, or identify a phone call sequence in the database of a telephone company. It can also detect networks – such as fraud networks – in a sum of seemingly unrelated cases: in France, a software named **Anacrim** helped reopen multiple cases and was able to relate several cases together.

As this technology becomes more complex, it may be able to uncover more and more subtle relations, using different types of databases, such as **text, videos, audio recordings**. But law enforcement professionals are increasingly afraid that these algorithms may have **biases**, and they are opposed to the idea of being entirely replaced by these technologies over time. This is a key issue to address if we want to draw only positives from AI in justice.

Automated legal services

Many individuals or corporations need legal assistance at some point in their lives; as a result, many law firms provide hotlines to answer their customers' questions. The most frequent queries are where to get the documentation to create a company, how to get a divorce or sign a wedding contract, how to solve an issue with an existing contract, how to solve an issue with their boss, their neighbour, their landlord... Answering such questions may not be complicated, but it often requires going through legal documents, knowing codes of law, finding the right document for a given situation.

These frequently asked questions could be automated into databases ready to respond to all these queries; natural language processing technologies, question answering technologies and language generation technologies could be used to create **bots** able to solve all sorts of legal questions, even if they are a little more complex. This could reduce the cost of these services and **democratise** their use: people who do not have the means to pay for legal help could therefore receive guidance, and our world could be a little more fair and equal.



Emotion analysis and lie detection

Lie detection technologies have been heavily portrayed in films, but have rarely proven to be effective in real life, and many countries see them as entirely unreliable. However, AI could improve the accuracy of lie detection technologies. It is already able to identify emotional states by analysing voices, videos, or subtle body language clues. So far, it is considered to be about **80% ACCURATE**.

But artificial intelligence can be highly biased: the data collected to train it may not be good quality, depending on whether the people seen as guilty did actually lie or not, or vice versa. The algorithm may also end up working better on types of people who are more often found in courts: men and some ethnic groups. So we should be wary when using AI in this field.

One last application of AI would be to build psychological profiles of suspects more accurately than with only the help of a psychologist.

Most relevant ethical principles in justice: responsibility, justifiability, equity, and privacy

As algorithms and robots become more and more present in our society, and are more and more involved in our economy, it becomes crucial to know who is *responsible* if something involving AI goes wrong. If two autonomous cars collide, and human passengers are hurt, who is at fault? Who must pay back whom?

As algorithms are not conscious – and as long as they remain that way –, **RESPONSIBILITY WOULD PROBABLY GO TO THE PEOPLE WHO CREATED THE MACHINE OR TO THOSE WHO USED IT**. In fact, they were the last ones to symbolically approve of the algorithm's actions, either by declaring that it was a finished product, or by putting it on the market, or by buying and using it. But this definition clearly is not foolproof and will evolve over time as we embed systems of accountability into our development of AI.

AI should also be *justifiable* in the legal field because knowing how and why AI made their decisions is crucial. **Why did AI conclude from this emotional tension that the suspect was lying?** Why did that unusual phone call count as a decisive clue in favour of conviction? How is it treating statistics, what is it getting from them, and why? AI's decisions must absolutely be transparent for judges, lawyers, and all law enforcement professionals to be able to trust them, and for citizens to still see their law system as reliable.

Equity and privacy: law professionals may wish to make sure that companies and individuals who are using AI algorithms are in accordance with international laws and regulations on AI. They may also wish to ensure that the way companies use AI does not undermine the rights of their customers or other citizens.

The General Data Protection Regulation (**GDPR**) is an example of a law which focuses on the right to privacy, on the **consent of consumers when automated decisions are made**, and on rejecting discriminatory data from automated decisions. New similar laws may emerge in order to make sure that consumers are treated equally and that algorithms are more and more explainable, especially for those affected by them. Canada, Europe and Japan are at the forefront of this global dynamic. The main issue is making sure that AI treats individuals in a **fair** manner and respects their **rights** as well as their **freedom**.

There are other important issues dealing with justice, including autonomy (if an algorithm assists humans in court decisions the judge must be in command, and even in-the-loop when it is not an easy case to rule) and technical safety (the impartiality of the models must be guaranteed).

Well-being is also relevant in 3 cases where responsibility and equity issues harm individual lives: to prevent innocents from being unfairly incarcerated; and to keep intolerable events to the victim from happening, while reducing legal delays and making sure that a guilty person is not found innocent.

Healthcare

AI in healthcare is a **very delicate subject**. For a long time, healthcare professionals have believed that their expertise cannot be replaced by that of a machine when it comes **to something as critical as health and sickness, life and death**. While robotics are already sometimes used to help surgeons be as precise as possible, fully automatising surgeries may pose significant ethical problems. Patients want to be treated on a **human-to-human** basis. It may be hard for a patient or a patient's family to accept that a medical intervention went wrong because of a machine.

But there are other ways to use AI in healthcare, in **areas that are less critical**. AccuVein, for instance, specialises in vein visualisation tools: their technology allows healthcare professionals to visualise their patients' veins, which is useful for many procedures such as drawing blood or inserting IVs. Robots can also help researchers make lab work more efficient and less repetitive: companies such as Opentrons and Andrew Alliance have released tools to automatise lab work, such as pipetting robots.

Despite some resistance by healthcare professionals, AI-based tools to detect disease are also being released. For instance, Face2Gene uses AI to analyse photos of patients and detect signs of genetic diseases. Grail, a company started in 2016, is developing software able to detect cancer in its early stages by analysing their patients' blood, and has already raised $1,6 billion in funding from Sequoia Capital China. For now, **machine-learning detection of diseases through facial recognition, skin, urine or blood** are mostly used for research purposes; but they may also have a clinical application soon and open up a large market.

Most relevant ethical principles in healthcare: autonomy, justifiability, responsibility, technical safety and privacy

Since the medical field has to do with the most important parts of individuals' lives – illness, death –, it is one of the fields where human autonomy is most important to maintain: humans should always be in charge when it comes to health. AI can be used as a tool which is always supervised by humans ("human-in-the-loop") and which helps professionals make their decisions. As of now, responsibility for mistakes always falls on humans in the medical field, as machines are not entitled to make any important decisions in the place of health professionals.

It is also critical for AI in healthcare to be justifiable. In disease detection, for instance, professionals should be able to know why AI believes a patient has a certain disease. Otherwise, its point of view can only be accepted or rejected, and professionals cannot understand the reasoning behind its decison or use parts of it to draw a better conclusion.

Responsibility is a key issue: as doctors' decisions can have a great impact on individual lives, they must be accountable for it. *E.g.* using AI as a diagnosis support, **doctors have to challenge machine recommandations**: they are ultimately responsible for prescribing a treatment or diagnosing a disease, that can have a great impact on individual lives.

There are other issues in healthcare, such as technical safety: data have to be handled very carefully, and the algorithms with caution before establishing a diagnosis, because providing inappropriate treatment can endanger humans' lives.

Accurate diagnosis (*e.g.* based on DNA sequencing) could improve and speed up access to care. But health professionnals would have then to protect the millions of *private* genetic and medical data handled by AI. As we see progress in this field, it is also important to consider equity in our approach to healthcare: is our society moving towards an hyper-personalisation of medical treatments using AI, or towards an equal treatment? The solutions to be built to ensure fairness will be different in each of these cases.

Consumer goods

There are many uses for AI in the consumer goods sector, from smart gadgets to industrial applications. Final consumption is also going to be greatly impacted. It is worth noting that **THE RETAIL INDUSTRY, WHICH IS BOUND TO BE HEAVILY AFFECTED BY AI, IS AN EXPANDING MARKET WHOSE OPERATING PROFITS COULD REACH $2.95 TRILLION BY 2025**. Some examples of AI in the consumer goods industry include:

- Using AI to **reduce energy consumption** by supermarkets and restaurants. To that aim, Wattman Lite has developed a smart app which automatically adjusts the temperature of air conditioning when needed by recommending the appropriate room temperature to the system.

- Using computer vision to get data on shopping habits or assist shoppers. The Singapore-based firm **Trax** released a software which films the shelves in supermarkets 24/7. It collects data on consumer behaviour, which it **links to product placement and pricing**. It can then tell manufacturers and vendors which products perform badly or which products perform well, and give clues as to why.

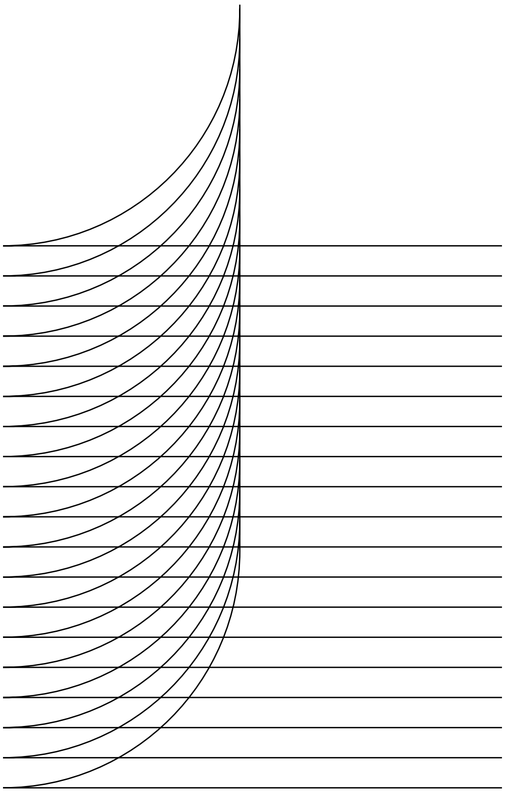
- Services such as the Chinese software Face++, which raised 750 million dollars in funding in 2019, also help authenticate payments through facial recognition.

- Companies like Amazon, Microsoft or IBM are also using tools such as the Watson Tone Analyser for emotion analysis, which allow them to better tailor their products to their audience and analyse customers' response to them.

AI can also have significant industrial applications, such as automatising factories partially or even in full. In the food and beverage industry, **smart sorting devices already exist**: for example, Tomra uses AI computer vision to analyse the aspect of French fries and eliminate those whose size and colour are imperfect. In 2014, it helped Agristo, a company located **IN BELGIUM AND THE NETHERLANDS, SUSTAIN ITS 140,000-TON PRODUCTION**. This technology can easily be transposed to other products made in **assembly lines**, such as ready-to-wear clothing or small technological devices.

Some companies are also working on hand gesture recognition by AI: thanks to cameras and sensors which transform hand and finger motion into data, they allow users to control some devices from afar by using their hands. This technology is already being used in home automation, shopping, consumer electronics, in medicine to help patients regain their range of motion through virtual reality, and in navigation.

For now, AI is used above all by recommendation engines, which analyse our search histories and personal networks in order to recommend products we might enjoy.



Most relevant ethical principles in consumer goods: well-being, privacy and equity

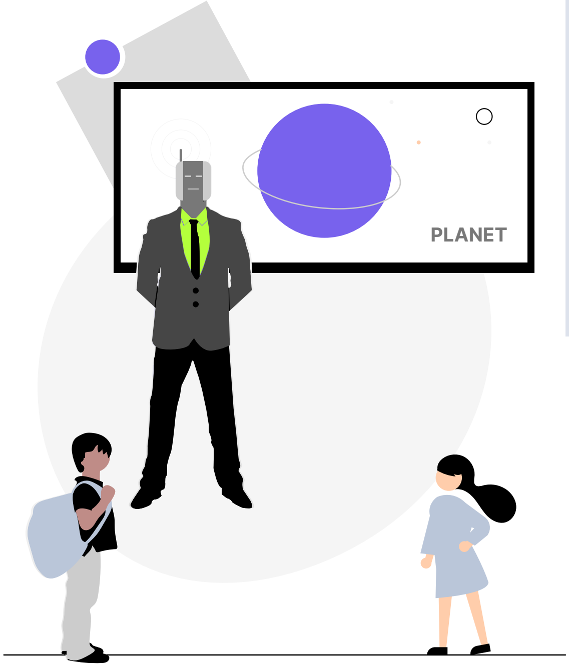
Well-being: having access to quicker and **more effective services does not necessarily equal well-being**, and that might be the biggest issue we will have to consider in the next years. In our opinion, AI use in the consumer goods industry is not a problem as long as **consumers give their consent**. But AI might not make citizens more happy, it might just make them better consumers.

Due to **different cultural traditions**, or different religions, or different values, all countries will not react to AI use the same way: companies or governments should not attempt to erase these differences but rather include them and make their products adaptable.

Hence AI has to face the challenge of autonomy, due to the Western tradition. By deploying AI in purchasing suggestions or to organise the shelves of supermarkets, will our future society enable humans to keep control of their consumption choices, even the most commonplace?

Privacy and equity: if AI is used to identify customers – for example, through facial recognition used to authenticate payments –, it can lead to client-profiling. Therefore, AI will know about each client's preferences, and be able to recommend more suitable products. The question is then: what use is made of these data?

The potential issue with profiling is that it could discriminate and recommend different food to certain ethnic groups, effectively making one group healthier than another. It could also lock people into their set preferences, encouraging individuals not to stray from their usual choices, and encourage a form of determinism: some people could remain stuck in fixed preference groups, which take into account factors such as their social background, and feel unable to escape from them. AI might also not take into consideration the health of the customers and keep recommending unhealthy products to someone who is trying to change their eating habits, because it knows the client will statistically «crack» and choose to buy those products. All of these issues should be kept in mind before putting AI products on the market.



Education

AI can be used for educational purposes. **Interactive toys which use speech recognition software are already on the market**. These toys use machine-learning to converse with children, to tell them stories, to ask questions and memorise answers. An example is Wonder Workshop, a company which makes age-appropriate robots and apps to help children develop their **conversational skills** and their interests.

Most relevant ethical principles in education: well-being, autonomy, responsibility, privacy and equity

Well-being: use of AI in education makes us question the **true role of an educator**. If machines teach children, the children's' sense of reality could be biased, or human-to-human interaction could become unnatural to them. Do we want educators to always be human? What is best for young minds? These questions fall under the principle of well-being because they determine the kind of society we wish to build.

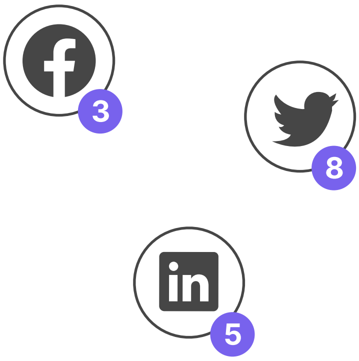
Autonomy: in philosophy, **autonomy is often considered to be the main purpose of education**. From Aristotle to Kant, philosophers believe that educators should teach children the basis of how humans ought to act with one another; as a result, children become autonomous because they know what knowledge and which actions they should pursue to become better.

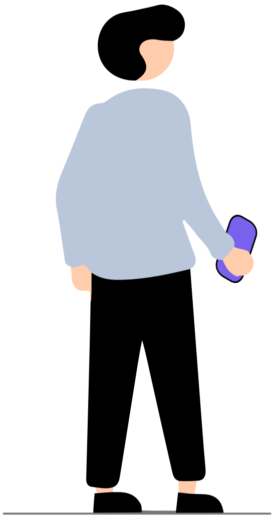
Therefore, an AI which acts as a teacher should still be able to create the same results, and adults should always be in control (following the "*human-in-command*" principle) to check that the information given to the children is appropriate for young minds.

Responsibility: If AI tools do not reach their educative goals (*e.g.* if the conversational robot is not robust enough to avoid recording insults or discriminatory remarks, as in the case of the chatbot Tay), the consequences are as damaging to a young person's education as they could have been positive, entertaining and instructive. Designers must therefore anticipate and bear the unintended consequences of educative AI for young people (involving they should know how to justify their models' functioning).

Privacy: contact with children can provide AI with plenty of data. Since young people are unable to give consent to data collection, and might regret having communicated that data in the future, companies developing such products should be very careful while accumulating this information and always check with the adults in charge of the children.

Equity: if educational AI is only available to wealthy families, it may reinforce pre-existing social biases. But if it is widely accessible, it may have the opposite effect and **provide high-quality education to parts of the population which would otherwise never have obtained it**. The deployment of AI in education should therefore take equity of education into account.





Communication

When we communicate, two criteria are important to us: *speed* and *clarity*. If the tool we use to communicate is **quick**, our message will be delivered at the right time, which is crucial; if the message is delivered **clearly**, and not distorted, we will have been able to make ourselves heard. These two points determine whether our attempt at communication is successful or not.

AI will certainly enhance the speed and quality of communication networks, in several ways. AI is sure to enhance the *speed* of communication networks: recommendation algorithms allow us to find relevant contacts, word correction softwares enable us to write our thoughts out quicker. By making the work of communication professionals such as journalists easier, AI hence improves *quality* of communication. It can help with memory storage and word analysis, so that journalists can focus rather on accuracy of content and creative delivery of news.

Natural language processing technologies can also be beneficial in other sectors, such as search engines, to match employers and job seekers for instance.

Above all, AI can be used to communicate on natural disasters or imminent danger, in order to alert people of what's coming. **In Sendai, Japan, an AI-based system is being tested to alert people on their phones in case of a tsunami**. Other applications could include alerting people of terrorist attacks or giving instructions to people entering dangerous zones so they can get out of harm's way.

Most relevant ethical principles in communication: equity, privacy, responsibility, technical robustness

Equity is important to consider, as data analysis will create user profiles on search engines which might include discriminatory biases. **When searching for content online, people might not find certain subjects or certain ethnic minorities, because a biased algorithm might have labeled them as less good or less relevant**. Youtube is currently dealing with this issue, as its algorithm has labeled several LBGT videos as “adult” content even when no inappropriate subjects are broached.

We should not forget that communication often involves **power dynamics**: those whose messages are heard loud and clear have more power than those who are not listened to, and the way messages are communicated affects their understanding by others. We must therefore always make sure that AI in communication treats users **fairly** and does not stigmatise categories of people. This is all the more complicated to ensure as profiling is already used and is sometimes beneficial: **recommending cooking videos to profiles which have shown an interest in cooking**, or science videos to those with an interest in science, is certainly relevant. So would AI be right in **recommending male profiles over females to employers** who have shown to favour males? Probably not. We should think about where we wish to draw the line.

A broader issue is *privacy*. People using products with AI are vulnerable to unexpected use of their personal data. Critical information could be used and sold against the will of their owners, and private text conversations, videos or audio recordings could be viewed by people who were never meant to see that content.

It appears to us that laws framing what can be done and what cannot be done with user data are necessary, now more than ever. They could draw on the General Data Protection Regulation. In this, as for other issues regarding AI, ethics and laws can work together to define privacy safeguards for individuals.

Top players in the field of communication (states, large companies or celebrities) using AI are also *responsible* for its use. “Fake news” broadcasted by new technologies, or contents suggestions reinforcing people prior beliefs (either artistic or political), can have a great impact on reputations and even on political stability. Thus major communication actors must be aware of the great responsibility they have in the AI and new technologies era.

Mobility

AI can be of great assistance to **increase mobility and make transportation safer**. It can help develop ride-sharing services, delivery systems by drone, and autonomous cars.

THE BIGGEST CHANGE AI IS BRINGING TO MOBILITY IS AUTONOMOUS CARS, WHICH COMBINE COMPUTER VISION, AUDIO DETECTION, AND MACHINE LEARNING TO BE AS SAFE AND PERFORMING AS POSSIBLE. The Google-owned company **Waymo**, for instance, wants to use this technology to make roads safer by **reducing accidents caused by human negligence** (*e.g.* drunk driving, taking unnecessary risks, being distracted, texting at the wheel...).

The first technology self-driving cars use is **visual: they have a 360-degree view of their surroundings**, which allows their cameras to scan and analyse the road in real time and with a very large scope. After having been trained with millions of different images, the database then identifies surrounding objects and acts accordingly: for example, it slows down automatically when it sees someone crossing the street.

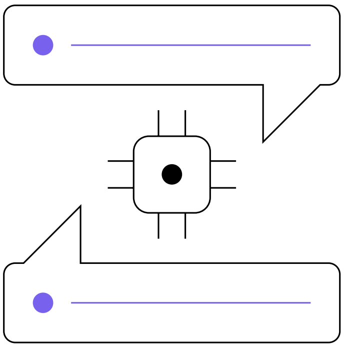
The second technology self-driving cars use is auditive. Waymo is developing sensors to recognise the sirens of fire trucks, or ambulances, or police cars, so autonomous cars can let them pass. Other audio sensors can be used to recognise the sounds made by the brakes or by the wheels, which give indications as to the state of the roads or problems in the car. All these elements could be combined with machine-learning to let the driver know when check-ups are required. In a few years, autonomous cars might even be able to drive themselves to be repaired...

Others applications are being developed to enhance the performance of autonomous cars: Jaguar is developing their own self-driving vehicles, and Microsoft is creating virtual environments to train and develop secure AI models.

Most relevant ethical principles in mobility: autonomy, human responsibility, justifiability, technical safety, and well-being

Autonomy: when it comes to autonomous cars, debates will centre around two very important principles, and it will be difficult to choose which one should be favoured. **THESE TWO PRINCIPLES ARE LIBERTY ON ONE HAND, AND SAFETY ON THE OTHER.**

Autonomous cars are undeniably safer than cars driven by humans. But is this security more valuable than the freedom humans get from being able to drive the way they wish? If a machine controlled our steps in order to stop us from running into others or veering into the street accidentally – would we experience this control as beneficial or as a prison? Or maybe both? **The question is whether we should be free to put our own lives, or others' lives, in danger.**



On the road, human autonomy might not be desirable. It might be more beneficial to give up our initiative to AI and to allow it to manœuvre cars in our place.

That is why varying degrees of autonomy could be explored. Semi-autonomous vehicles could allow us to be safe all while enjoying the freedom to drive as we wish, as long as no lives are in danger. Should humans then be "in-the-loop" (as in, approve or deny every action AI wishes to take), or "on-the-loop" (by keeping a general eye on what is going on and being able to regain control when they wish), or rather "in-command" (by being able to change the way the car behaves)?

Technical safety: it is crucial for designers to ensure AI in mobility is reliable enough to be in control of an entire car ride without fail. If AI is a hundred percent secure, the question of liberty versus security becomes easier to settle. Legislators are currently reflecting on norms for these types of vehicles.

Responsibility: if we imagine that, at some point, AI-based vehicles are largely deployed, there will certainly be accidents now and then. A crucial ethical issue we must therefore solve is: who is responsible? Two stances exist:

1. **We can decide that humans are never responsible**. On the long term, machines could become legal persons, which would mean that mistakes are their fault. It would simply be difficult to know how we should punish them, or how they could make things right, since machines *do not own property (so they cannot be sued*), and they do not have emotions or physical sensations, so they cannot suffer. **The entire concept of punishing crimes would have to be rethought**. For now, legislators in Europe reject the idea of machines as responsible individuals.

   In fact, responsibility can only go to free individuals, who are able to reflect on their decisions. If we build machines to prevent our mistakes, and a machine makes a mistake... should we consider that machines have a **right to be wrong**?

2. **We can decide that humans are responsible for the mistakes made by machines.** In that case, we must decide in which ways they are responsible.

   We must therefore begin by trying to minimise damage and take into account problems which are in a moral grey area. These can be thought about in terms of collective well-being. For instance, the **trolley problem** imagined by Philippa Foot in 1967 makes us question the extent to which killing someone to save someone else can be legitimate: **should AI-driven cars choose to kill one person to save two**? What if more criteria are introduced, such as age, gender, ethnic group? Would that change our answer?

   We must also make a list of possible cases and decide who is responsible in which case. For instance, who is responsible for a crash? The driver? *What if it is a technical problem? Is the designer responsible? What if thousands of copies of the same car were sold, and only one industrial piece had an imperfection?* Is the factory responsible? What if they had no way of planning for this specific situation?

   In the near future, companies should therefore reflect on the products they sell and determine who is responsible in which case.

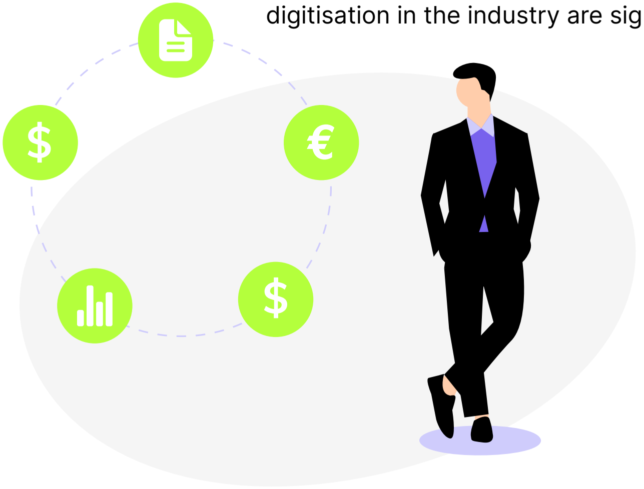
Justifiability: all these issues mean that the inner workings of a self-driving car must be as transparent as possible for us to make informed decisions about their actions. How are they trained, what data are they exposed to? **Some self-driving car crashes were due to road shapes that the cars had not been previously exposed to**, or situations where objects such as barriers were broken and the car had only been trained in a "perfect" environment. To determine who is responsible in case of failure, it is crucial for AI to explain why it made the decisions it made.

Besides these critical issues, AI must face other challenges in mobility, such as privacy (*e.g.* should we keep users' journey data safe?). Equity is also concerned: by deploying the driverless car, AI could for example give to some disabled people more freedom to travel and to move around alone where they want.

Finance and banking

The financial services industry was one of the first industries to significantly invest in the development of **modern AI-based applications, with increasing investments after 2014**. The insurance industry **was quick to invest** but hedge funds and investment banks rapidly followed.

Artificial intelligence is a great asset in the finance and banking industry, in several ways. **IT CAN HELP CONCEIVE NEW PRODUCTS; IT CAN IMPROVE CUSTOMER SERVICE; IT CAN REDUCE THE COST OF RISK AND COMPLIANCE; IT CAN HELP INVEST IN FINANCIAL MARKETS**. The massive amounts of data as well as the significant level of digitisation in the industry are significant enablers for drawing benefits from AI.



Conception of new products
Insurance, banks, hedge funds and asset managers all generate revenue through a diverse portfolio of products: insurance products to cover risk, financial products to manage the wealth or the daily financial needs of customers, providing access to corporate and government bonds, mortgages, and more.

Excelling at developing appealing and profitable products is an essential skill for banks and insurance companies: it allows them to extend the reach of their capabilities or succeed in new regions, and it is more important than ever in an industry which has recently become more competitive, less profitable, and needs to be transformed under the digital and environmental transition.

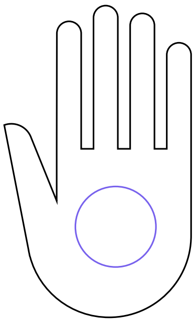
Since banks and insurance companies cannot protect their products with patents, they need to release new products more frequently, sell them for less, or make the products particularly useful to retain their customers.

Creating a new product requires a careful analysis of the risk it covers: specific insurance risk, risk of default, risk of liquidity, interest rate risk, risk of loss, prudential risk, etc. **AI CAN BE A STRONG ASSET TO MODEL THE RISKS INSURANCE COMPANIES NEED TO MANAGE FOR THEIR CLIENTS**, which means they can also manage new risks more easily, release new products quickly, reduce the cost of risk or take market shares by selling more attractive products. Increased access to data and its effective treatment by AI are powerful enablers for banks and insurance companies to become more competitive.

Multiple trends are affecting the conception of new products in the financial services industry:

The **use of new emerging data** has enabled banks and insurance companies to better model existing risks and provide discounts to certain types of customers based on the data they shared. Some ways to model risk are: using telematics to assess the risks associated with driver behaviour; using satellite data to better estimate risks associated with house insurance or farming; using mobile data to create financial services products for African or Indian citizens; using merchant website data to create credit products for small businesses; and more.

Artificial intelligence has allowed financial services to release products with more attractive digital experiences. These products use intelligent applications, capable of recording a large amount of data, and enable new services, such as an effortless on-boarding experience, faster treatment of customer requests, first instalments of a credit by sharing data from other applications… As a result, customer service is much more effective. For instance, if a customer is unhappy, AI enables “fast claims”: under certain conditions, such as when the cost threshold is low or when a significant climate event occurs, customers can be paid back in less than two hours.



The quick development of AI has enabled a more careful analysis of new risks, making them insurable. By using data from social networks, open data, and data from connected devices, insurance companies have been able to create differentiated prices to attract more customers with lower risk, thus increasing their market shares and pricing.

This is particularly true for specialty insurance companies: since they cover very specific risks, such as droughts, owning a Tesla car, having a specific disease, missing a plane, and so on, their pool of customers is relatively small, and risk needs to be managed more carefully. The introduction of advanced algorithms, coupled with the gathering and analysis of a larger and more specialised set of data, can help these companies create more competitive products. AI can therefore give **specialised insurance companies** an edge in comparison to more generalised companies.

Tesla is a striking example of this. The company has gathered a great amount of data about its customers and its cars: this data allows them to model their drivers’ risk in order to propose embedded car mechanisms to reduce that risk and the cost of insurance. This is very beneficial for customers who have a hard time finding competitive insurance products in generalised insurance companies. It is also beneficial for Tesla, because it can generate extra revenue more easily and create customer loyalty.

This principle applies to several financial products which are not yet profitable because assessing the risk behind them is too difficult. More advanced AI and new sets of data help solve this issue: they create sizeable opportunities for the financial services sector, either by helping them release entirely new products, or by allowing them to remove exclusion causes from existing contracts.

AI can also allow new business models to emerge. For example, car insurance providers have released new offerings labeled “pay as you drive”, “pay how you drive”, or “pay when you drive”. These offers are appealing to good drivers or to younger profiles who often need to pay a higher premium, and they are beneficial to insurance companies which need to adapt to new forms of mobility.

Customer relationships

Artificial intelligence allows companies to find new ways of interacting with customers, measure their satisfaction, and improve their brand loyalty.

The most significant example of how AI can improve customer relationships is the development of chatbots. **Chatbots allow customers to access their bank or insurance company whenever they need to**: customers can open a bank account, transfer money, invest into exchange traded funds (ETF), manage an insurance policy or make a claim all without speaking to an advisor, and at any hour of the day. Users do not need to go through complex menus and only have to write down what they are looking for in the chat box.

Some banks and insurance companies rely on popular assistants like Alexa to help their customers through vocal interactions, but the downside is that distribution and contact with customers are outsourced.

Another way AI can help is by analysing phone calls to assess customer satisfaction through emotion analysis in real time. It can thus help predict the departure of a customer or redirect the call to another customer success professional.

IN A POST-COVID ERA WHERE CONSUMPTION MAY DROP DUE TO INCREASED UNEMPLOYMENT AND REDUCED CONFIDENCE IN THE FUTURE, BUILDING SMOOTH CUSTOMER RELATIONSHIPS WILL BE CRITICAL, AND AI CAN BE AN INVALUABLE ASSET.

Operations

Banking operations account for approx. 20% of bank budget, 40% of a P&C insurance budget, and 20% of a life insurance budget. AI can reduce the cost of banking and insuring operations, which are a significant percentage of the total cost in delivering associated financial services, and is therefore an opportunity to improve the margin of these companies significantly.

Automating form analysis

Since banks still rely on paper documentation for many customer requests, the paperwork associated with processing these forms is very heavy on employees and longer for clients. Optical character recognition technologies (OCR) can help reduce the cost of operations and increase the speed of delivery, making for a better customer experience.

Predicting absenteeism and employee departure

Human resources costs can be optimised with the help of AI. In recruitment, AI can be used to find the best candidates for a position by screening CVs and motivation letters to reduce recruitment bias or by identifying top candidates on social networks. It can also be used for talent management: it can help identify employees who need training, build personalised training programs, increase their success and reduce the cost of making them. Lastly, it can help identify employees who are preparing to leave, understand how to limit absenteeism and **create employees retention plans to lower turnover**.

AI can also be used to reduce the bias that exists when choosing candidates for a position, to reduce the mismatch between candidates and their new position, and to bring more women to executive positions.

Improving data processing activities

AI can help improve data processing activities, such as data reconciliation, verification, data cataloguing, automatically processing logs to ensure that the IT is up and running, and automatic recording.

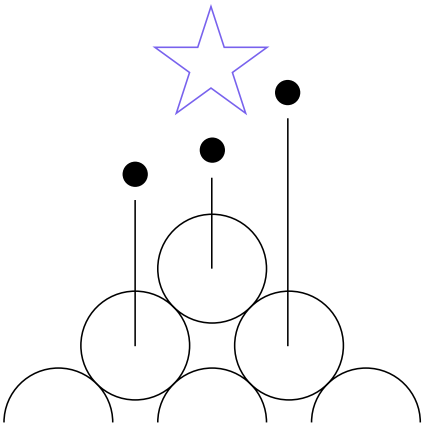
Data processing and analytics have an increasingly important role: since banks and insurance companies are digitising their process more and more, they can use data to train algorithms or take better strategic and operational decisions.

The data they need to analyse is often scattered across many systems and **diverse formats**: text, forms, pictures, tabular data, logs, etc. To be able to exploit this data, companies need to make regular inventories of the available data. They also need to identify privacy issues, security issues, outdated data, missing data, and set up processes to update their internal data dictionaries regularly as new systems are rolled out or removed. This allows them to keep up with new standards in data regulation.

Banks and insurance companies also need to comply with tightening risks and compliance regulations which force them to improve the quality of their data and data-related processes. Lastly, they need to open their data to comply to open banking and data porting regulations.

AI-based technologies can significantly decrease the cost of these activities and improve their overall quality. AI can analyse large amounts of data in different formats and uncover invisible relationships between data sets which can help internal activities. **IT CAN ALSO OPERATE IN REAL TIME ACROSS MANY COUNTRIES AND RAISE ALERTS WHEN DATA IS MISSING, OUTDATED OR INCORRECT**. It is therefore particularly useful to accelerate the integration of external subsidiaries during **M&A operations**, for instance.

Lastly, AI can be very useful to identify forbidden IT operations, fraud, security breaches, or failing systems that may disturb banking or insurance operations.



Underwriting, assessing the risk of a customer to provide a quote
Insurance underwriting is a key activity for insurance companies. It requires careful analysis in order to identify what risk the insurance company will actually bear, and issue a quote that the individual or company need to accept for the transaction to be finalised.

As speed and competitiveness are two factors that will play a role in wether or not policyholders accept a given proposition, insurance companies need to optimise their process. AI can be used to automatise this process, with several benefits: **speed**, as AI can provide a quote almost immediately; **improved risk modelling** and **competitiveness**, as a larger number of variables are taken into consideration; better **compliance**; reduced **fraud**.

This improved risk assessment can also help insurance companies provide more options to fit the exact needs of their customers, rather than providing standardised packages. This increases customer satisfaction and lifetime value.

Customer service
Improving customer service is key to improve customer retention. Artificial intelligence-based agents can handle an ever-increasing number of customer queries, and are getting better and better at responding to complex requests.

Operating a customer service centre is generally costly, and companies tend to outsource this service in low wage countries where working conditions are harsh. In these customer service centres, there is background noise, break time is limited, work days are strenuous, and, as a result, workers are unhappy and customers are not satisfied with the experience.

AI can help answer many low-level queries quickly or redirect customers to the best agent to solve their problem, which significantly reduces the cost of operations and improves flexibility. The number of virtual agents can vary depending on the number of customers and be reduced in down hours. Lastly, customer satisfaction is greatly improved, as customers get **better and faster service at any hour of the day**.

Finance and accounting

Risk management

Banks and insurance companies are inherently risk management companies: to invest into the market, lend credit and insure goods or persons against loss or injuries, one needs to carefully assess the risk those activities pose in order to draw a profit from them.

Therefore, risk management is the core activity of banks and insurance companies, and **billions are invested into risk management software** as well as conceiving actuarial models and risk models. Banks and insurance companies manage several risks for their clients in order to protect their assets and generate a financial return:

- Credit risk associated with lending activities
- Counter-party risk associated with market activities
- Foreign exchange risk
- Liquidity risk associated with non liquid assets
- Rate risk associated with international trade activities
- Operational risk due to internal errors
- Actuarial risk (accident, climate risk, customer default...)

Since risk management is such an important part of the profit and loss (P&L) of financial services companies, improving risk models and the speed of decision-making – even just by a small percentage – can improve revenue significantly. It can also help manage a more diversified portfolio of risks and therefore propose new appealing products to customers.

Since risk management heavily relies on mathematical models – such as statistical or probabilistic models or weather forecast models – and algorithms, improving these models is a good way to improve profitability.

The production of **fine-grained data** in large amounts and the emergence of machine learning **have changed the game** for risk management. Machine learning can help capture behaviours that would be missed otherwise and increase control over risks to manage them better.

Credit scoring is a widely known example of how machine learning can improve the precision of risk management algorithms. Credit scoring helps assess how likely it is for banking clients to default, by giving these clients a rating. This rating indicates how fragile customers are and helps determine whether or not those customers should be granted a credit.

For example, **in the case of a household credit, banks will gather information on potential customers** such as the type of home and its price, the customer's job, their expense patterns through their credit card activities, their ability to repay past consumption loans; then, the **AI will calculate the probability for the credit to be repaid**.

Compliance and fraud management
AML, contract compliance (leakage), fraud detection, analysing models with regulation (IDD, Solvency), ensuring validity with contract clauses…

NEW REGULATIONS ARE CONTINUOUSLY EMERGING ACROSS THE WORLD IN THE FINANCIAL SECTOR around customer communication, reporting, establishing capital ratio, identifying money laundering activities, distributing financial services products, assessing customer risk, fighting market abuse, data privacy…

It is very costly for banks and insurance companies to ensure that their operations follow regulations, to detect breaches to existing regulations, to determine how impactful forthcoming regulations will be, to follow emerging regulations, and to see how they evolve. **The cost of non-compliance can be significant, so banks and insurance companies heavily invest in technologies and workforce to ensure that they are compliant**.

In many ways, AI can reduce cost in these areas. Natural language processing technologies can be used to go through the many law articles, question-and-answer sessions, reports of analysts, in order to identify the impact of these new regulations and what needs to be done or changed.

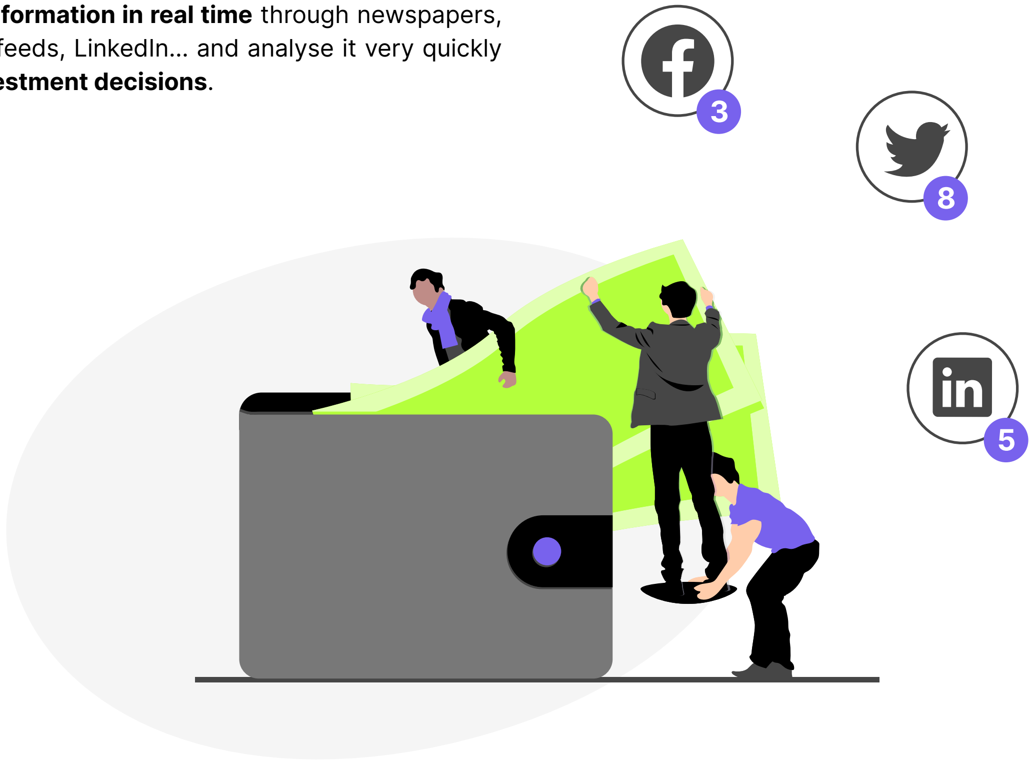
Artificial intelligence can then be used to model that impact for bank and insurance companies and to update existing models more easily. It can also be used to identify compliance issues, breaches to data privacy, missed money laundering as well as internal fraud patterns or abuse patterns. Lastly, it can be used to automate controls and internal processes in order to analyse inputs in bank systems and **avoid compliance issues from the beginning**.

For instance, AI can analyse calls between traders to identify potential frauds or insider trading. It can thus raise alerts when certain phrases are said. It can also **analyse transactions and warn the bank or insurance company if suspicious activity is detected** on a given bank account, such as if a series of payments are made on a dormant account from a third party country.

Financial markets

Analysing emotions and adapting to new information
Artificial intelligence can help banks and insurance companies take into account alternative data which can help them perform better. It can also remove the burden of going through lengthy company, analyst or notation agency reports to determine what drives the value of the company, how it is performing financially, what its strategy should be, its risk, and to what extent it depends on stranded assets.

Moreover, markets generally react quickly to news such as announcements by BCE commissioners, the release of numbers on unemployment, reports from country or company representatives, quarterly results presentations or disclosure of M&A operations. AI can gather this **new information in real time** through newspapers, presentations, twitter feeds, LinkedIn... and analyse it very quickly to take automated **investment decisions**.



Using new data to analyse financial health
Some new types of data which can help analyse the financial health of a company include: the turnover of the company, changes to top management positions, news published on social networks, open source data on flights or transportation boats, or data from satellites or cameras – to analyse the number of cars in the company's parking lot, the number of new buildings in a country, the effects of global warming or natural disasters, and so on.

AI can quickly link this data to real-time financial fluctuations in the market, and can be used to better assess the growth of a country or company. It can therefore select undervalued or overvalued companies to invest in with short or long-term strategies, or can hedge or insure against situations which could impact investment portfolios.

Stock picking

There are several strategies that can be successful to invest in markets: investing in a given instrument, in a given sector, in specific geographic areas, building a diverse portfolio of variously-sized companies from different sectors or geographic areas... Both short-term and long-term speculative strategies are possible: one can invest in growing companies, or bet on the failure of companies by owning short positions.

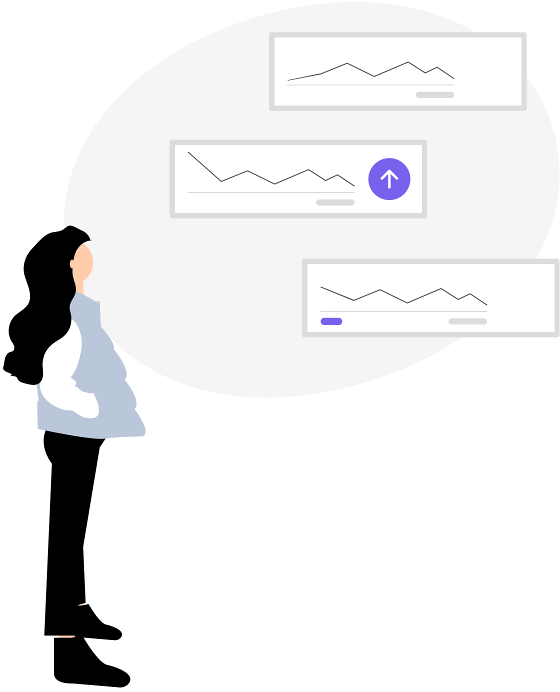
Several instruments exist: equities, fixed income, commodities, real estate, ETFs, derivatives, and so on.

IT IS VERY DIFFICULT TO EXTRACT VALUE FROM INVESTING IN THE MARKET, BUT AI CAN HELP PREDICT THE FUTURE VALUE OF A SET OF ASSETS. It is able to analyse a varied set of data to determine which sectors should be invested in, which instruments should be traded, and what amount should be traded. It can therefore help build powerful investment strategies.

Since an AI-based algorithms is right more often than a random predictor, there is much profit to be drawn from investing sufficiently large amounts of money and using several different strategies across a large number of instruments, sectors and geographical areas.

Because it may be complicated for smaller investment companies to generate performance by investing in financial markets, these smaller firms may adopt stock picking strategies. AI may be used to analyse the communication of companies on a given sector, their strategies and their financials; this can help identify companies that are undervalued and have strong growth potential in order to generate long-term performance.

Stock picking is particularly beneficial when a market is in a downturn – the **Covid-19 crisis is a good example** – and may help companies come out stronger. It helps them have stronger fundamentals and appropriate solutions for the challenges raised by the downturn, which sometimes helps them weather the crisis better than other companies which were in better shape when markets were rising.



Algorithmic trading

Algorithmic trading was popularised during the 2009 financial crisis, when high frequency trading crashed the stock market and companies lost hundreds of millions of dollars in minutes. In algorithmic trading, algorithms are entirely autonomous: they make decisions by themselves and act on them right away. What makes those algorithms successful is their speed and their accuracy. Oftentimes, these algorithms are simple and based on rules defined by their developers, such as “if this company’s share goes above 100 euros, then sell”.

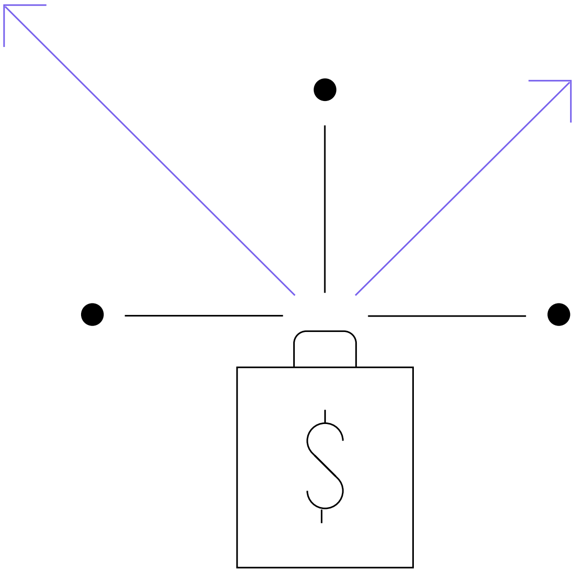
AI and machine learning algorithms are different: by integrating the data they get from analysing financial markets, macroeconomic data, twitter feeds, and quarterly reports, **they are able to create more complex rules to go by, and are more successful**. For now, this approach has not proven itself completely, but some funds get good results by using it. **Several companies (funds and startups) have released funds which are entirely managed by AI for some prospective investors**.

Market simulation

Under certain economic conditions, market simulation may be useful to make long-term investment decisions or define monetary and economic stimulation policies on a national scale.

Artificial intelligence and machine learning can be used to learn from the market or model the behaviour of a set of companies or individuals under certain past economic conditions. They use an agent-based modelling technique: each actor on a market is presented as an agent that will react to what other agents do. For example, if a customer asks a bank for credit to buy goods from a company, his or her consumption will be altered depending on whether he or she gets the credit they asked for. This will have consequences on whether or not the companies he or she was going to buy from go bankrupt, and in turn, it might signify credit loss for the bank.

By combining agent-based modelling with AI, the behaviour of each agent can be predicted more accurately and hypothetical future scenarios can be tested. Thus, the evolution of financial markets can be predicted more accurately, and efficient market simulation policies can be issued. These can be particularly beneficial after a downturn such as the one generated by the Covid-19 crisis.

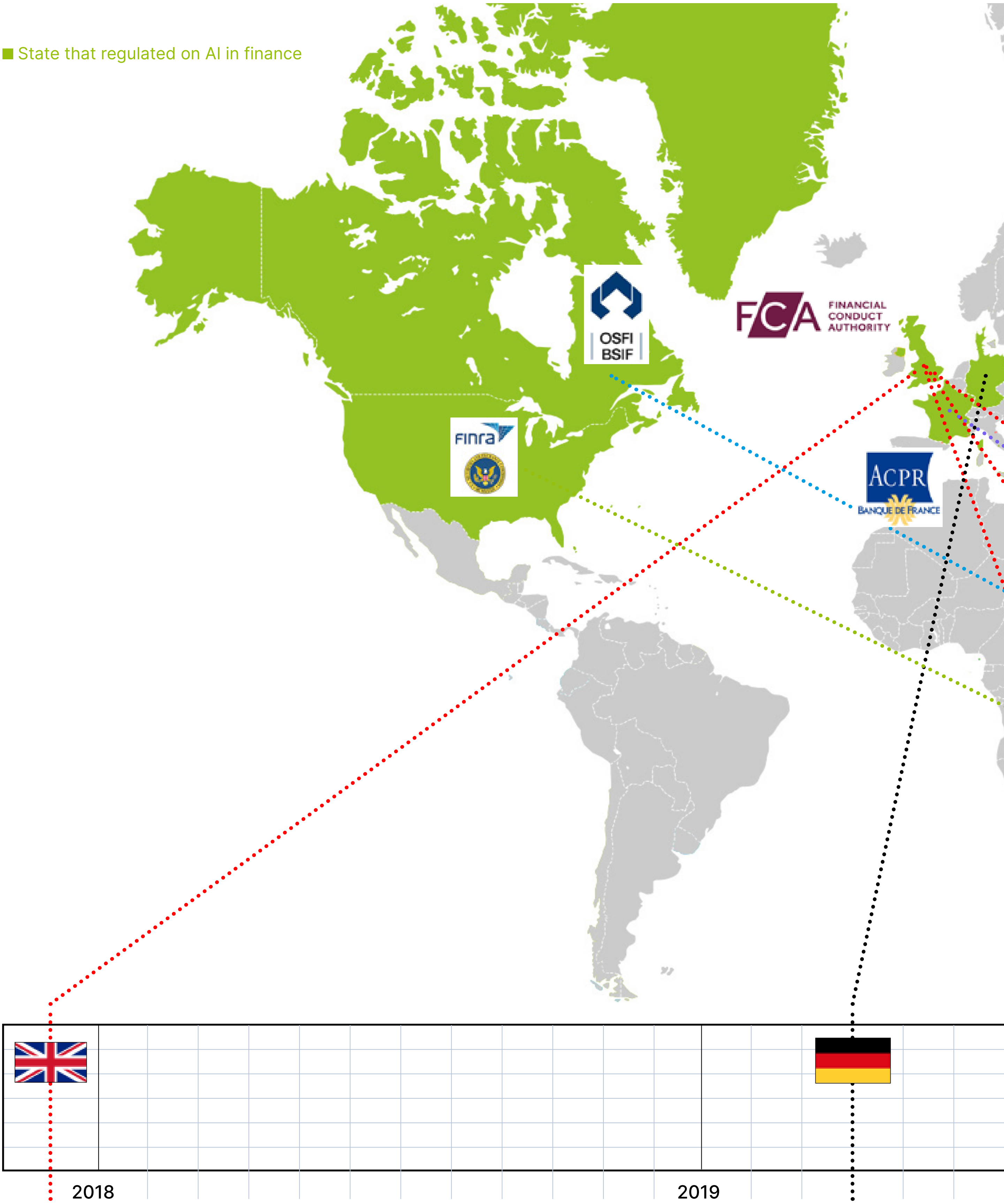


Most relevant ethical principles for finance: equity, justifiability, well-being, privacy

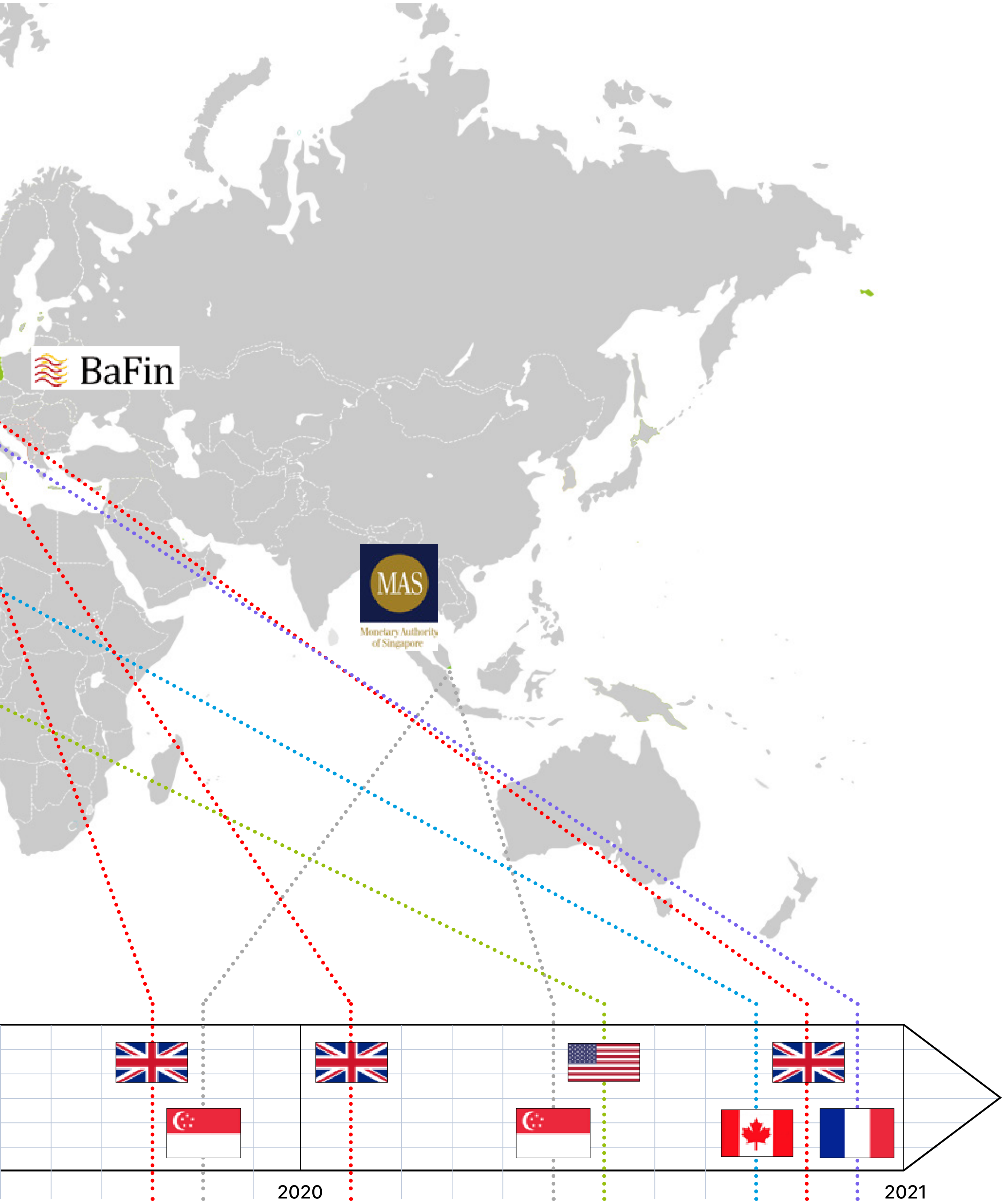
As these applications of artificial intelligence are being investigated, and some of them already used on a daily basis at least at the scale of a bank or several banks. It becomes urgent to question what is right, what is expected from these applications and how we can create trust across the different stakeholders – in particular employees, clients and regulators. Regulators in Singapore, France and Germany already released ethical guidelines for several of these applications and are even starting to test ways to turn these guidelines in an operational way.

WHERE AND WHEN REGULATIONS ON AI IN FINANCE APPEARED

More and more States are regulating on AI applied to finance: some regulation texts are identified here.
Click on State flags to see the texts.



🟩 State that regulated on AI in finance
OSFI
BSIF
FINANCIAL CONDUCT AUTHORITY
FCA
FINRA
ACPR
BANQUE DE FRANCE
2018
2019


BaFin
MAS
Monetary Authority of Singapore
2020
2021

Equity: a central question when AI is deployed is, **am I being treated fairly**? This is especially true in finance and banking. In many financial applications of AI, treating people in a fair way is crucial, especially when customers are not financial experts and have to rely on the company's or the bank's expertise. There are several stakes at hand: for example, it is important to prove to people that the reason they were denied a loan does not lie in discriminatory criteria, such as "sex" or "ethnic group".

Justifiability is important to make sure that individuals affected by finance and banking decisions are fairly treated. One must therefore consider the way the algorithm is designed: it must deliver means to follow what the AI is doing and to interpret it in the most transparent way possible. If borrowers wish to understand why they were denied a loan, and if they object, the decision must be discussed with rational arguments. This can only be done if we know why the AI made the decision it made.

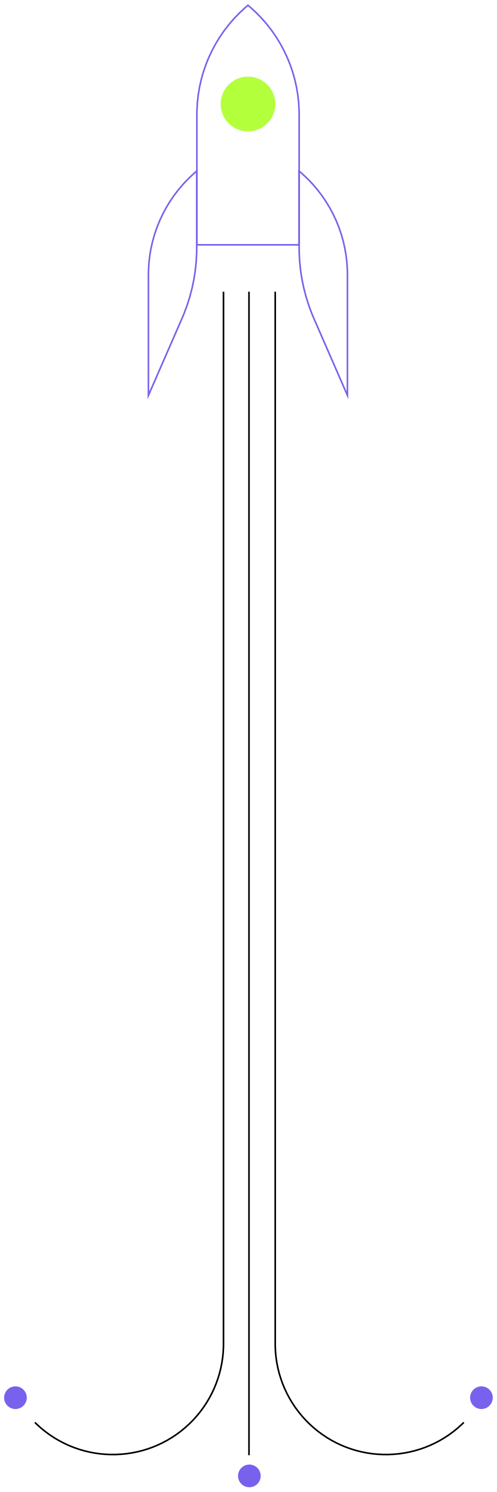
Justifiability is crucial for humans – investors, clients, regulators – to understand the process by which their data was handled, especially when it comes to decisions around emotion analysis, departure predictions, credit scoring, and so on.

It is worth noting that we do not systematically consider technical safety to be one of the guiding ethical principles in finance. In fact, mistakes are inevitable in this sector, as AI can only attempt to predict what may occur without having the absolute certainty that what has been predicted will indeed occur. A margin of error is to be expected when evaluating risks (or opportunities), on which most of the decisions focus. While it is crucial for AI's decisions to be a hundred percent trustworthy when it comes to driving a plane, that is not the case when it predicts the future hypothetical value of bonds. Nonetheless technical safety is crucial when a decision has a strong impact on the market (*e.g.* when AI is in charge of applying compliance and regulation can not be breached, or of avoiding systemic risks and bankrupcy), involving we have to know which person will be accountable in case of failures.

We consider autonomy as a less critical principle: indeed the financial sector is strongly regulated to avoid bankers and assurers too considerable leeway in interpreting the results of internal models (as established in BASEL III and Solvency 2 regulations). Thus regulations frame individual autonomy, controlling the market to avoid systemic effects.

The financial industry must notably follow the rules on *privacy*. Banks and financial agents have to keep safe health, family or income data gathered by their systems to train or use the AI.

The final important principle to deploy AI in finance is that it must improve general *well-being* (there are already regulations protecting consumers in Europe). It can help develop economic value, which can be shared among stakeholders and bring benefits to the general population. For instance, it can **enhance trust in newly designed green bonds**. In that regard, well-being is a core principle to keep in mind.



Where is it urgent to implement AI ethical principles?

Urgent vs important

In the body of this paper, we identified the most important ethical principles in different sectors. **In the following graph, we used examples from current events (articles, reglementations, technical solutions) to estimate which ethical principles are most *urgent***. We also estimated how far existing solutions have come for now. We have used the Eisenhower Matrix to help distinguish important problems and urgent ones, and what we recommend to do as **we need to focus collectively on developing and deploying the solutions to the most important and urgent challenges hindering AI development now**:

	URGENCY	NOT URGENT
IMPORTANT	ACT	PLAN
NOT IMPORTANT	RECONSIDER	DO NOTHING

Moreover if some ethical issues related to AI are considered urgent but not important, we need to reconsider whether they are really that urgent. Maybe the current existing solutions are already satisfactory or it can be easily solved with a good enough solution, meaning that investment need to be redirected to find and deploy solutions for the most important principles.

For instance, **responsibility in communications is both important and urgent**: it is important because an AI spreading fake news or potential-life changing news such as natural disasters can have an instant effect on millions of people. It is urgent because these technologies are getting more and more performing by the day, so solutions must be proposed *now*.

Other challenges, such as justifiability in agriculture, are important as well (to better understand how drones which detect and irrigate dry patches of land work, for instance), but they **do not appear to be as urgent as technical security** (for example, avoiding massive dissemination of pesticides in fields).

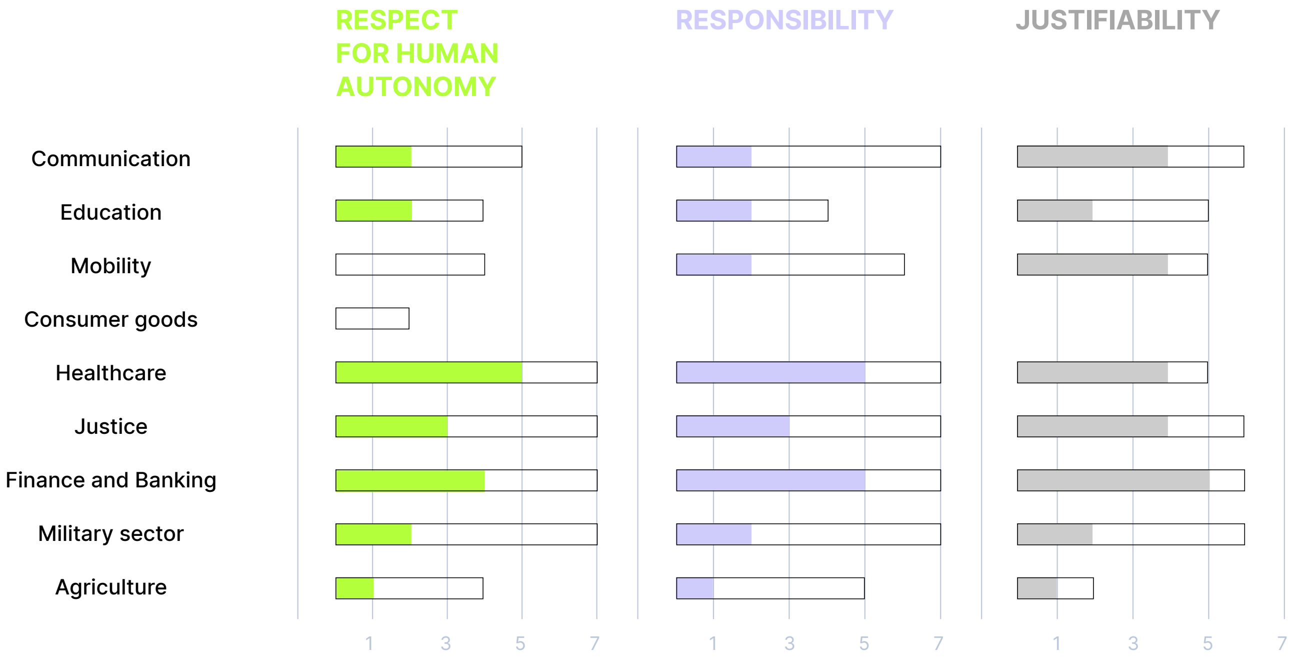
Similarly, **privacy is an issue that has been raised a few years ago**, and some laws exist around it, thanks to the GDPR (General Data Protection Regulation). In many areas, **finding solutions to protect privacy therefore becomes an urgent problem**.

> **We therefore publish these analyses to invite a discussion on ethical AI.** While our analyses on what might be urgent and what is important are founded on facts and arguments, they remain our own.
>
> They are thus meant to be challenged, and the authors of this paper are open **to debate and discussion, which are at the root of philosophical thought**.

IN WHICH AREAS IS ETHICAL AI MORE ADVANCED?
IN WHICH AREAS DOES IT LAG BEHIND?

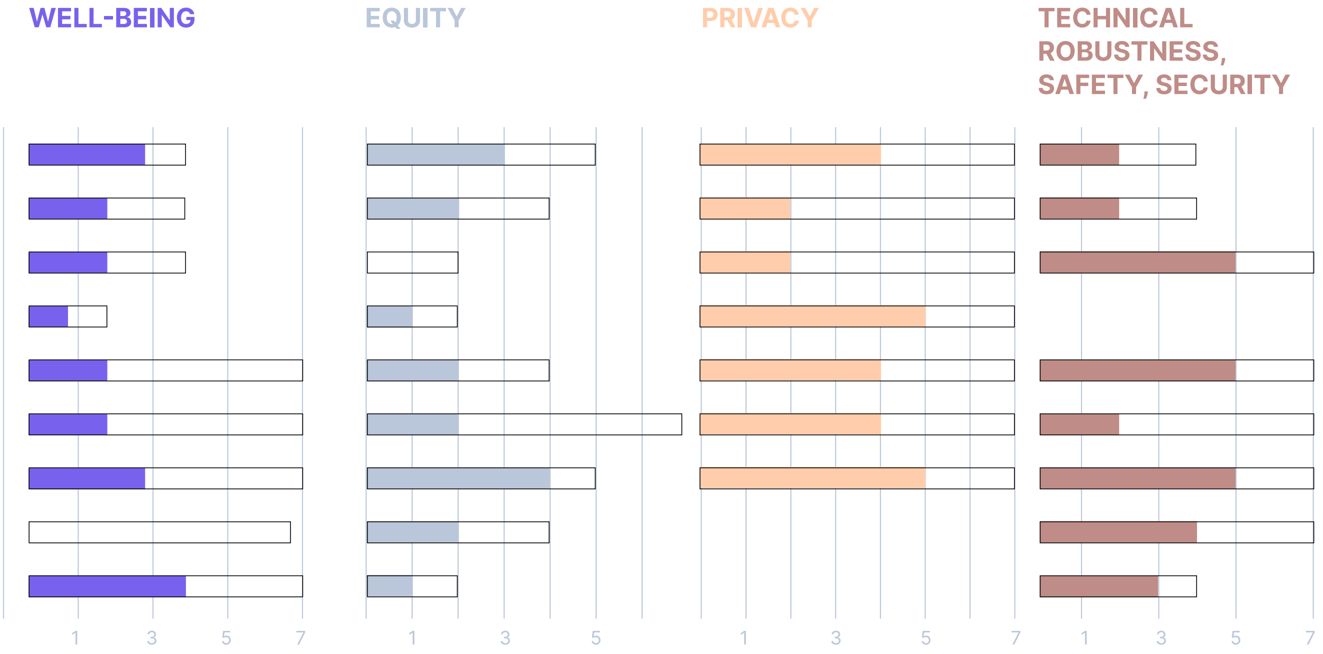
COMMENT:
We have measured the level of progress of ethical AI in a certain field on a scale from 1 to 7. Both the urgency of finding a solution and the current progress of these solutions are taken into consideration.

PROGRESS IN FINDING SOLUTIONS: WHAT IS BEING DONE	URGENCY: WHAT SHOULD BE DONE
(0) there is no discussion around the issue	(0) the issue is not very urgent for now
(1) the issue is coming up	(1) the issue should start to come up
(2) the issue is starting to be discussed	(2) the issue should be discussed
(3) some solutions are beginning to be looked into	(3) some solutions should begin to be looked into
(4) some solutions have been proposed and are being tested	(4) some solutions should be proposed and should be tested
(5) some solutions are being put into practice by companies, but only on a small scale	(5) companies should be putting solutions into practice, even just on a small scale
(6) solutions are being put into practice by companies on a significant scale	(6) companies should start putting solutions into practice on a large scale
(7) the problem is solved	(7) the problem must urgently be solved



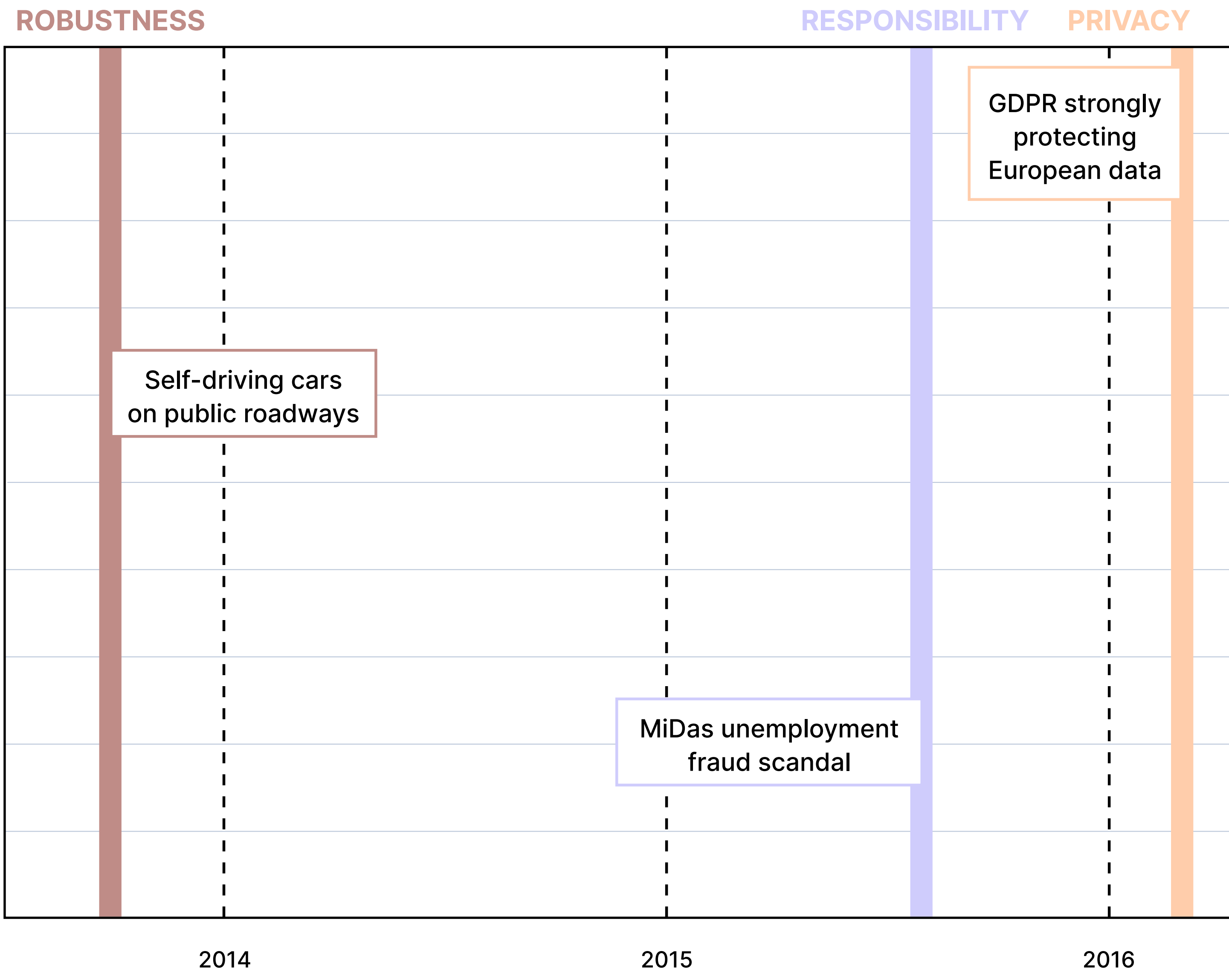
INTERPRETATION:
In agriculture, there is no need to reflect on privacy for now (level 0), and the existing problems (confidentiality of the growers' data) are either irrelevant for now or do not solely concern AI (grade 0).

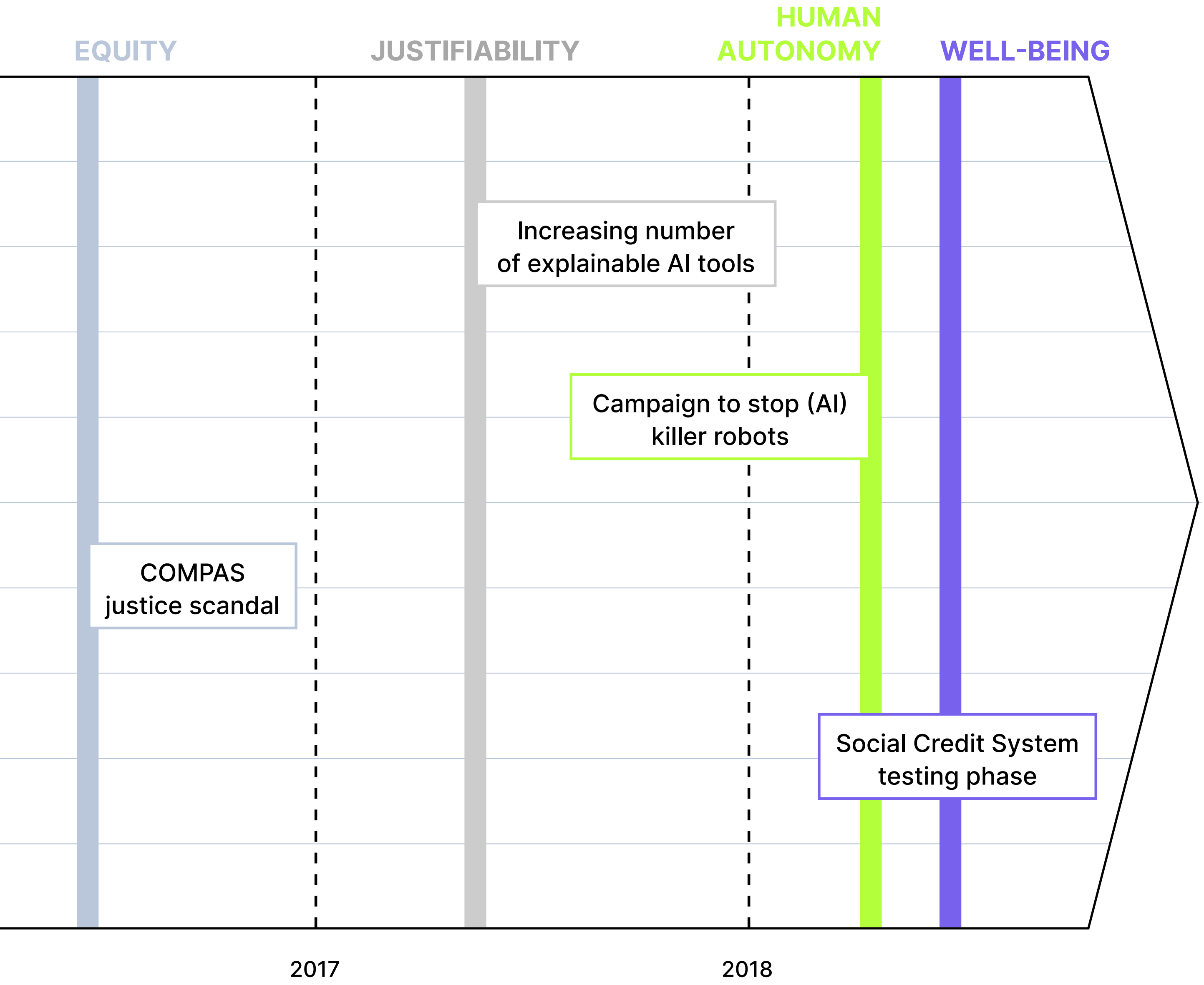
In other cases, such as security in the army, some solutions are being tested (level 4), but they must be approved and deployed as soon as possible, because human lives and human safety may be at risk (grade 7).



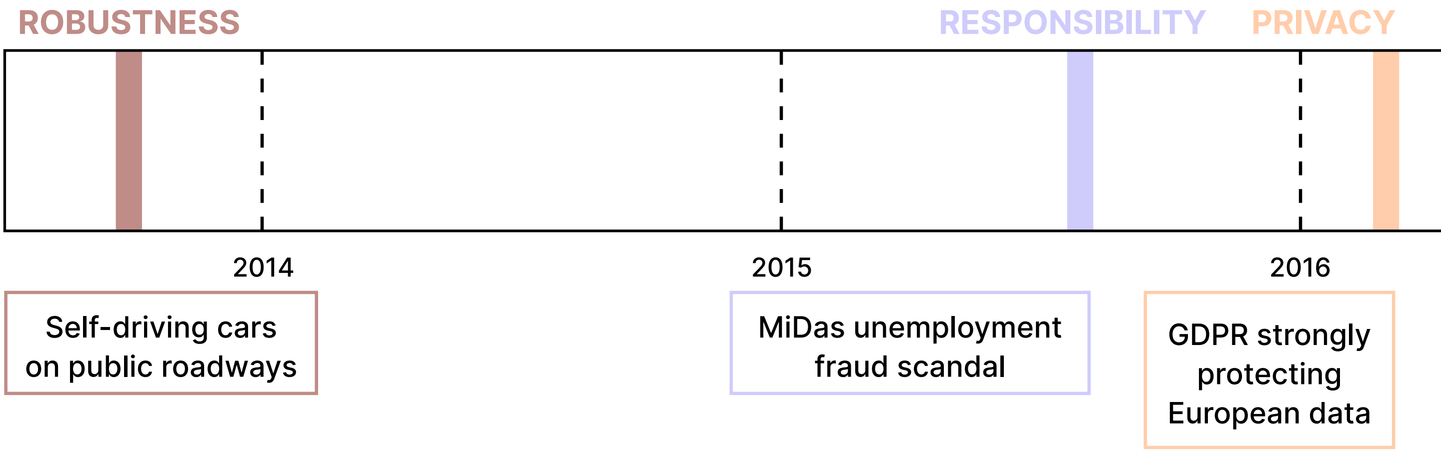
WHEN DID THE 7 ETHICAL PRINCIPLES ON AI EMERGE?

Ethical concerns regarding AI emerged gradually. Here, we list the moments they began to come up, due to events (*e.g.* self-driving cars on public roadways) showing society that an ethical principle (*e.g.* robustness) must now be taken seriously. See more details on the events below.




EQUITY
JUSTIFIABILITY
HUMAN AUTONOMY
WELL-BEING
Increasing number of explainable AI tools
Campaign to stop (AI) killer robots
COMPAS justice scandal
Social Credit System testing phase
2017
2018

WHEN DID THE 7 ETHICAL PRINCIPLES ON AI EMERGE?



ROBUSTNESS

RESPONSIBILITY

PRIVACY

2014

2015

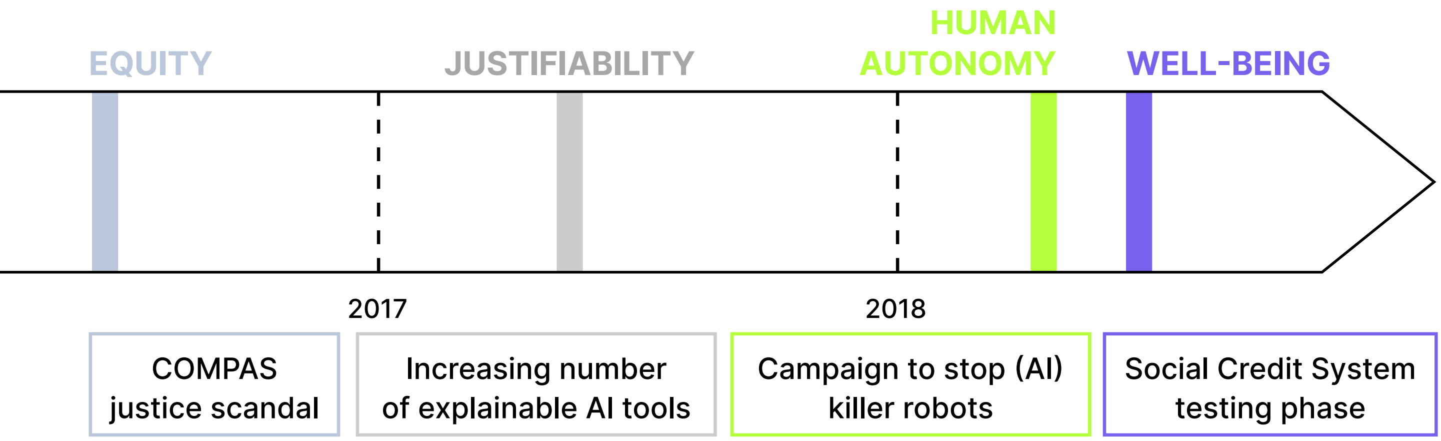
2016

Self-driving cars on public roadways

MiDas unemployment fraud scandal

GDPR strongly protecting European data

ROBUSTNESS	As self-driving cars were appearing in large numbers (already 4 American State is safer than humans'. This is representative of an obvious AI challenge: it shoul
RESPONSIBILITY	34 000 individuals were wrongfully accused of unemployment fraud in Michiga Indeed the Michigan Unemployment Insurance Agency (UIA) focused on reduci It was UIA's responsibility to program MiDas not only to gain cost efficiencies, b
PRIVACY	The GDPR (General Data Protection Regulation), adopted by the EU, strongly p mandating consumer's consent to gather it.
EQUITY	An analysis reveals a racist bias in the COMPAS Recidivism Algorithm, deploye whereas white defendants which were *real* "high-risk" recidivists were twice m
JUSTIFIABILITY	Following the EU regulation (GDPR) and other national programs (USA, China), and interpretability (heatmaps, decision trees) about what AI systems do.
HUMAN AUTONOMY	The "campaign to stop killer robots" launched in 2018 all over the world high-li
WELL-BEING	The Social Credit System is being tested extensively in different Chinese provi In 2018 also, for the first time an AI system supporting medical diagnosis is vali



COMPAS justice scandal

Increasing number of explainable AI tools

Campaign to stop (AI) killer robots

Social Credit System testing phase

es in 2013) on public roadways, a major purpose of manufacturers has been to make AI cars secure so that their driving d be used only if unpredictable behaviours and unexpected risks are low.

n form 2013 to 2015: some saw their credit and reputations ruined, or even filed for bankruptcy... ing efficiency costs by introducing the unsupervised algorithm MiDas. out also to check carefull that cases of fraud were real ones.

rotects European's personal data (IP addresses, cookie IDs, customer contact details...),

d in Florida. COMPAS was twice more likely to incorrectly judge black defendants as "high-risk" recidivists, ore likely not to be detected.

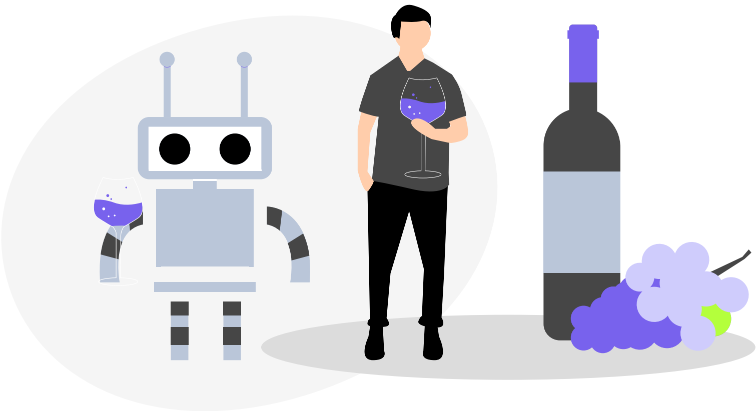
tools are largely deployed with the aim to provide users with more transparency (simplifying neural-networks models)

ghts the need to keep control over AI in case humans life and autonomy are endangered.

nces (in 2018, the Chinese Global Times stated that 11 million people were prohibited from traveling by plane). dated. For our society it raises the question: in which cases is AI deployed for good?

4.2. Ethical AI development in other domains

DEVELOPMENT OF AI COULD BE EXTENDED TO FIELDS WHICH WERE PREVIOUSLY THOUGHT OF AS SPECIFICALLY HUMAN. THIS COULD POSE NEW ETHICAL PROBLEMS which are very interesting to consider. Three striking examples are art, democracy and religion.



Art and good taste

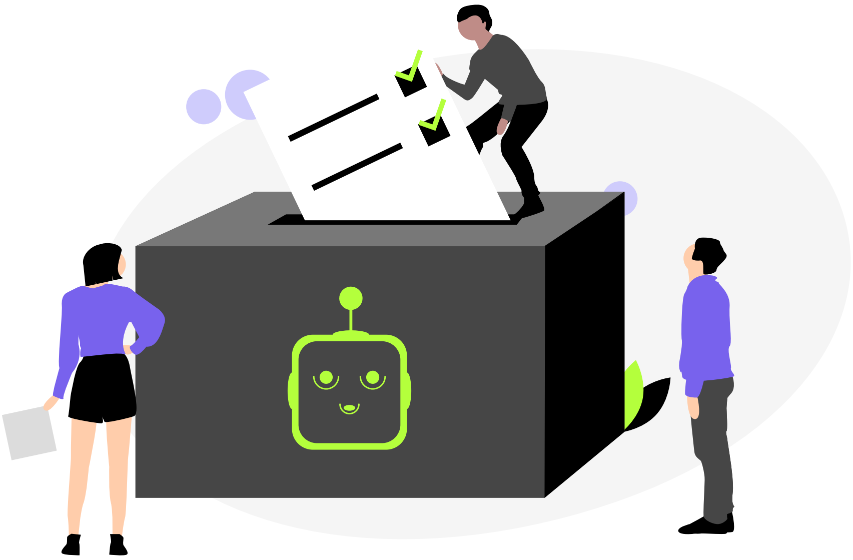
We tend to see art as a specifically human skill that AI could never truly take part in or understand. **As a matter of fact, while it may be unlikely for AI to be moved by an artwork or to produce a masterpiece, it appears that it is able to determine what moves us, recognise our tastes, and give us relevant recommendations**.

Take for example musical taste. Because they gather a lot of data, platforms such as Spotify can have an accurate 'idea' of what the tastes of each user are. Thus, they can understand what "good music" is for different types of users, and recommend music to them that they might like. **In fact, they influence consumer preferences on a large scale: 70% of the $7 billion earned by the music industry in the USA come from subscriptions to music streaming platforms.**

70% OF THE $7 BILLION
EARNED BY THE MUSIC INDUSTRY
IN THE USA COME
FROM SUBSCRIPTIONS
TO MUSIC STREAMING PLATFORMS

It can therefore understand what leads us to appreciate certain things, even though it cannot feel the emotions or the flavour we get out of them. However, it would be more difficult for an AI to say why we like certain things on which we do not have data while data capture is difficult. Take flavours, for instance: **it seems unlikely for AI to be able to translate things such as wine appreciation into data**. While it may be able to isolate parameters such as chemical components, colour, consistency, local origin... it would be difficult for AI to combine those norms into an objective evaluation of how the wine tastes and why we like this taste.

It seems even less likely for an AI to be able to say why humans like a certain artwork by isolating parameters such as the thickness of the paint layers, the way colours and shapes combine, the geometrical distance between patterns... **For now, these dimensions are too complex, and an AI is unable to truly relate to what we like and why** (but yet that could change quickly in the future, as artificial intelligence is taking major leaps forward).



Democracy

There is a lot of talk about how AI could change the workplace, how it could change our approach of wars, how it could change our law system. But how is AI going to impact democracy? Specialists in democracy issues such as Alexander Görlach have used technological metaphors to refer to the current crisis our **democracies** are going through, saying that they **would need "an update"**.

AI could intervene as an answer to doubts on democracy. The Canadian government, for instance, is already working on strategies to make AI a part of governance processes. Aside from the indirect positive effects that smart applications of AI could have on the lives of citizens, AI could seriously improve democratic processes. **ITS POWER TO COMMUNICATE AND EDUCATE EFFICIENTLY COULD BE USEFUL TO ACCELERATE REFERENDUMS, POLLS, OR OTHER WAYS TO CONSULT THE PEOPLE, ALL WHILE REDUCING ADMINISTRATIVE INERTIA AND BUREAUCRACY.**

Other applications of AI can have an indirect positive effect on democracy, by making society more inclusive, for instance. AI can predict optimal ways of allocating budget for governments to act in an effective and cost-effective way. This could help governments achieve their goals, such as **more precisely locating the spaces in a city or in a country in which poverty or crime rates are high** and the subtle set of reasons behind these discrepancies, in order to find better solutions.

At the same time, recent events in the United States have shown that the recourse to artificial intelligence could have **side effects that could significantly damage democracy with a more effective spread of fake news, the automatic suppression of contradictory contents, the creation of large scale filter effects** (due to algorithms recommending quickly the same content to a wide population eager to read about it without any critical analysis of this content's trustworthiness).

Algorithms were moreover able to facilitate connections between people from different regions based on their joint interests, which facilitated the creation of large communities of people sharing the same views without anyone to debate against them, as the potential debaters were assigned to disjoints community. It reinforced an existing separatism that weakened our democracies. A more skilled leader could have taken advantage of the situation to obtain a different outcome than the one we know. For these reasons artificial intelligence needs oversight.

Religion

For a long time, people usually believed that humans where set apart from all other living beings by their spiritual side, their soul. **The ability to bond with divinity, to act rationally, to have a moral compass were all seen as being part of the human soul**. Now, the idea that only humans have a spiritual side, and the idea that only humans can have a soul – or that souls even exist – are **put into question**.

In 2019, a group of Japanese Buddhists created a robot deity: it is named Mindar and they honour it as a divine creature. This robot priest uses AI and is able to provide services such as funeral ceremonies or religious chanting.

Two rather prosaic benefits come from these sorts of robots: first, they make it possible to hold expensive rites at a lower cost; second, they attract young people to Buddhism. A Silicon Valley engineer even founded a church named the Way of the Future where several forms of AI are seen as divinities. Since they appear to be omniscient, and they have the knowledge of multiple consciousnesses, these beings are seen as omnipotent, god-like.

But one thing remains to be questioned: is our spiritual life a sacred part of us, in which machines should have no part? On one hand, integrating technology into religious ceremonies could mean **adapting rites to our times** and giving a new meaning to them. On the other hand, smart technologies **should not erase religious beliefs** which are often deeply anchored in different cultures.

For example, a company which has the ability to build priest-robots could choose not to put them on the market. Why not? Because the company could decide that human-to-human interaction is the only desirable type of interaction in a religious context, if we wish religion to remain sacred. **Situations felt to be crucial for humans such as weddings or funerals could understandably require a human presence**, which would guarantee human empathy and understanding.

New ways to integrate ethics into AI development

Two ways of integrating ethics into the development of AI come to mind:

1. Regulations:

IN SECTORS SUCH AS HEALTHCARE, JUSTICE OR FINANCE, THERE ARE ALREADY REGULATIONS THAT CONSTRAINTS THE USE OF ALGORITHMS. BUT WE CAN GO FURTHER.

In these areas, ethical considerations can be applied by turning them into laws and supervising their application. Compliance with those rules can then come from control by external authorities or from an internal adaptation by the firm itself. Regulation requires experts, so that applicable principles can be determined and their implementation can be monitored. In several aspects, we need new regulations to give further clarity and facilitate the development of a more human centered artificial intelligence.

2. Initiative, self-regulation:

The principle of **Enterprise Social Responsibility, born in 1999**, invites companies to respect values which are not directly linked with profit. It was initially developed to promote a change of actions in the financial sector, which was destabilised by speculative bubbles, then went on to expand to other sectors. The goal of these measures is to take care of all stakeholders linked with the firm, to state internal rules (from recycling to respecting hygiene) and to organise activities which show that the company takes care of its employees, minorities, and the environment. Initially, these actions were strongly encouraged by tax reductions (for instance, approx. 5% of benefits).

It is worth noting that the Social Responsibility principle went on to generate benefits by itself, especially in its ESG form (Environmental, Societal, Governance). Companies noticed that, in the long run, non-ESG companies (*e.g.* companies which are not environmentally friendly) bring a lower return on investment. On the contrary, ESG-firms tend to gain a positive image, are favoured by investors, and reap long-term benefits.

But integrating ethics *inside* the company, and not just for image, implies an open dialog: companies must reflect on the concrete ways in which they can incorporate their values into their AI products and into their business model. That is broader than just following the Social Responsibility principle.

Unfortunately, self regulation is often not sufficient and we have many recent exemples showing that artificial intelligence needs to be better regulated to create trust and prevent dramatic side effects. **By anticipating future national and international laws** on the relationship between AI and the liberty of citizens, that will certainly be formulated in the long run and become more and more binding, companies can stay abreast of changes in the way of doing business with AI.

In all of these cases, **well-being** and the defense of our **fundamental principles** are the main principles around which our efforts to control AI must be centred. We must understand whether or not the changes AI brings are for the greater good before deploying AI.

The emergence of AI, as it becomes more and more present in our society, should encourage us to think about what makes us human deep down. We must use AI as a tool to serve human intelligence rather than let its power subsume or intimidate us.

From farming to driving, AI appears to be a solution to free us from robotic tasks. But in literature, in art, in justice, it should be seen as an assistant and not as a substitute.

Therefore, AI as a tool to guide our actions can help us determine what makes us specifically human and open our eyes to the many ways in which machines can never replace us:
HUMAN-TO-HUMAN CONTACT, CONVERSATIONS TO SOLVE PROBLEMS,
CREATIVE ENDEAVOURS, EXPRESSING EMOTIONS,
FEELING ENTERTAINED, SHARING OUR INNERMOST SELVES WITH OTHERS.
WHERE IS THE LIMIT BETWEEN HUMAN QUALITIES AND MACHINES?
How much will that limit evole over time? Those remain open questions.

5

Dreamquark,
ethics and AI

5.1. Our story

DreamQuark was **founded in March 2014 by Nicolas Meric** to leverage artificial intelligence in order to optimise the delivery of healthcare.

Even though DreamQuark was one of the first new generation artificial intelligence (*deep-learning*) companies to focus on healthcare, it was too early in 2014-2016 to enter into long-term engagements with pharmaceutical companies and the public sector. Data was not easily accessible, most companies did not understand the potential of data yet and it was hard to get access to sufficient amounts of data to train machine-learning algorithms.

In its early days, DreamQuark developed solutions to improve the maintenance of in vitro diagnosis devices as well as diagnosis software for eye diseases.

Because of these operational difficulties, **DreamQuark quickly pivoted to focus on the insurance sector** and worked with major names in the industry in France. We have worked with private and public entities that led us to develop **the premises of a machine-learning and deep-learning explanation framework in 2015** to facilitate the justification of our algorithms decisions. This framework was initially applied to the diagnosis of eye diseases and retirement insurance fraud detection.

In 2016 we closed a seed round of investment through 3 business angels and started to work with Ag2r la Mondiale, a large French health and life insurance company which became a DreamQuark shareholder in 2019.

In 2017 we initiated our relationship with BNP Paribas through a Plug and Play (the most active early stage venture capital fund in the world) accelerator program which helped us get a foothold in Fintech and finalise our transformation into a **'tech for fin' company** (that is, the sector of tech companies which provide technology solutions to the financial sector).

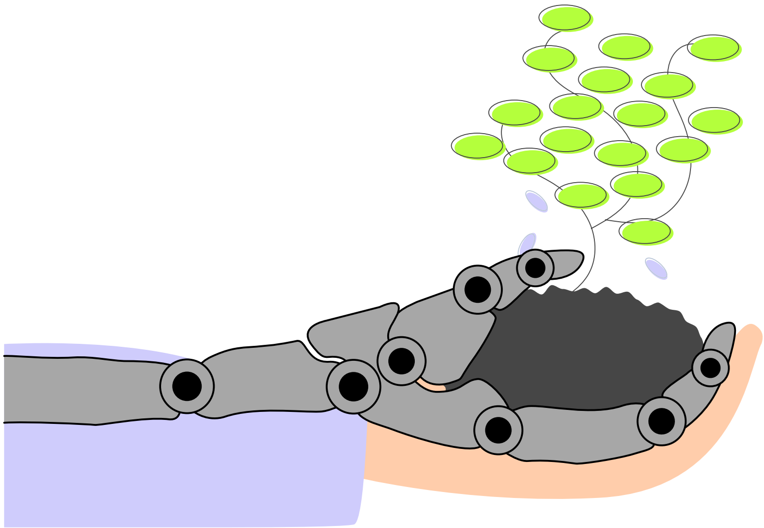
The same year, we also closed our series A round with Caphorn Invest, a French Venture capital firm, as well as Plug And Play.

In 2017, DreamQuark started developing Brain, a software meant to accelerate the deployment of business applications based on artificial intelligence in the financial sector. Brain provides intuitive, business-oriented interfaces for business teams or analysts which help them leverage DreamQuark's proprietary algorithms and open-source algorithms to solve their most important use cases (**upsells, cross-sells, targeted product recommendations, churn prevention**)... Through automation, we make using AI and deploying robust applications in the financial sector significantly easier. The emphasis of Brain is on generating value through artificial intelligence, on making it **faster to adopt for business teams**, on rendering AI decisions **explainable**, and on being **compliant** with existing regulations.

Responsible AI is at the heart of what we have been doing since the beginning.

IN 2021, FOUR YEARS LATER, DREAMQUARK HAS RAISED ALMOST 20M$ OF INVESTMENT FROM A SET OF TECH AND FINTECH FOCUSED ON VENTURE CAPITAL FIRMS AND INDIVIDUAL INVESTORS, ORIGINATING FROM FRANCE, SPAIN, THE NETHERLANDS, HONG-KONG.

DreamQuark operates in **France, UK, Germany, Singapore and Japan to build an European leader of artificial intelligence for the financial services sector**.



5.2. Our vision

We see AI as a transformative and turnaround technology for absolutely all sectors. It can potentially transform many of the business models we know today in a massive way.

Artificial intelligence can further economic growth: it can bring new revenue to companies by helping them identify which customers are the most profitable, it can reduce the cost of operations through smart automation, it can make us more efficient by determining the optimal way for industrial systems to function, it can help companies deliver new products or an improved customer experience. It can help us find new solutions to challenging problems we are facing. In particular, as we need to accelerate our ecological transition, **AI can help us design systems which are ecologically viable** and operate them with a reduced ecological footprint.

However, artificial intelligence in its current form is still only accessible to experts. Even though significant efforts have been made to ease the development of new algorithms and make them more open and accessible through open-source licence and software, there are only few people who really understand the underlying mechanisms behind the functioning of the algorithms.

The solutions that have been proposed tend to focus too much on the technical and algorithmic parts of AI, all while missing many ingredients that are essential to make sure that is can be used at scale by companies across sectors.

If we want to truly witness the promised benefits of artificial intelligence, the field needs to address several shortcomings and disconnects that currently hinder its development and the wider acceptability of the technology.

At DreamQuark, we have identified **7 AREAS IN WHICH WE THINK IMPROVEMENTS ARE NECESSARY TO FOSTER THE SUSTAINABLE DEVELOPMENT OF THIS TECHNOLOGY** and ensure that it will become our tool of choice to address the greatest challenges that lie ahead.

1. Democratising AI

AI providers currently speak mostly to data experts and developers. You generally need to be able to code to use artificial intelligence or have an advanced degree to understand what is done behind the scenes. If you are not a researcher, a data-scientist or a developer, artificial intelligence is hardly accessible.

Technologies become widespread when they speak to a broader audience. **Think about the internet, computers or iPhones. We need to make the same transition**: artificial intelligence must be used for meaningful applications which make sense for end-users and create value for them.

Great AI needs to be accessible to everybody

2. Adapting AI to each sector

Artificial intelligence is quite specific as the underlying algorithms are generic. Algorithms start to become meaningful when they are fed and trained with data to be able to become 'intelligent'.

But feeding them with data is not enough: we must combine that with expertise on the challenges addressed by the AI model being developed. This is necessary if we wish to create AI solutions which truly solve problems and do not add further hindrance in the process.

Having an algorithm ready is often the beginning of a longer process. AI requires a full-stack vertical approach with specific data, business expertise, algorithms, regulation and a way to support a specific integration.

Tesla ships great autonomous cars because they do not only have cars, but also millions of hours of **qualitative data** about how the cars are being driven, specific chips, specific devices to gather the data and specific algorithms.

Great AI companies are vertical

3. Demonstrating the economic benefits of AI

For now, we are only just beginning the transition to a world where artificial intelligence is widely used. Many leading consulting firms and technology providers claim that artificial intelligence can bring massive economic value. We have not yet seen proof of this except for a few select tech companies. In more traditional sectors, most companies are waiting to be shown that AI can be beneficial to their specific use case and be convinced.

Great AI creates sustainable economic value

4. Lowering the ecological impact of AI

As artificial intelligence becomes more widely used, we understand that it comes with an ecological cost. It needs significant computing power to learn to replicate different tasks and to become predictive. It also needs increasingly large amounts of computer storage, and to be updated frequently.

We need to prepare for this challenge ahead of time. At the same time, artificial intelligence can help our economies become more green and more sustainable, in many ways: it can optimise power grids, increase the longevity of many devices and industrial systems, **optimise the consumption of data-centres and energy providers**, help us design more efficient systems, assist us in identifying which investors are most likely to invest into sustainable products...

Great AI should ease the ecological transition

5. Promoting trust in AI

Even though artificial intelligence can bring many benefits, it is not without imperfections or biases. Moreover, the fact that algorithms are built on complex concepts, coupled with the potential job loss that could result from increased automation and the debates around a series of unsuccessful experiments all create a climate of mistrust around AI in the public.

AI is seen as a blackbox which needs to be opened, and that is a significant barrier to its widespread adoption. There are several questions around AI which need to be addressed: they touch on the quality of the output of AI, the fairness of algorithm decisions, how these decisions can be explained in an intelligible way, the transparency of the process of training an AI, the negative impact that automated decisions can have on human beings who are unable to object to those decisions.

Great AI is a white-box

6. Accelerating the integration of AI into an existing ecosystem

AI produces value when it leaves the labs and starts working with existing applications or industrial systems. In a sense, AI produces value when it goes live. **Creating an algorithm is the easiest part of the process and the integration part is generally where things get complicated**. In fact, **85% OF AI EXPERIMENTS DON'T GO INTO PRODUCTION**.

From the ground up, artificial intelligence needs to be designed in the aim of being integrated with existing software ecosystems. The following questions must be asked throughout the process: how can AI work with the constraints of production? Which data is available in production? How can I process my data and execute my algorithm so that the experience is not disrupted? How do I integrate it with my existing systems? How can the results be displayed in an easy way for my frontline officers? How can I do all that and ship in time? What governance process do I have in place? These are the questions that modern organisations need to ask.

Great AI goes in production quickly and is compatible with your systems

7. Maintaining AI over the years

Once artificial intelligence is deployed and connected with other systems, it starts to live on. At the same time, data keeps changing quickly: **new data is constantly produced, new signals are produced, customer habits change**, customers leave and are replaced by new customers, internal systems stop working, new regulations emerge or new products are launched.

Artificial intelligence (as it is now) is not aware of all these changes and needs to be maintained to reflect the changing nature of data. This generally means identifying when something changes, updating an algorithm if a new product or regulation is launched or an issue needs to be corrected, and redeploying the algorithms without affecting their environment and user experience. Facilitating all these steps is important to accelerate the deployment of AI and the costs associated with it. But only a few companies are already at this stage.

Great AI can be maintained

5.3. Our mission

DreamQuark offers solutions based on artificial intelligence that **empower our customers** to solve their most important challenges, create a positive impact and drive economic value creation.

We are a fin-tech company that sells its solutions to banks and insurance companies.

Banks and insurance companies face multiple challenges that are currently exacerbated by the Covid-19 crisis: they need to rethink their business models in light of the current low and negative rates, they face **increased regulatory pressure**, they are in a context of intense digitisation, with **extreme competition from new digital entrants**. They also need to accelerate their low carbon transition and find a way to transfer their high risk and stranded assets to more sustainable ones.

We are creating technological solutions to help them transform their business model accordingly.

DreamQuark has developed Brain, an artificial intelligence-based platform which allows financial services firms to deploy faster business applications in production in order to solve some of their key use cases in sales, marketing, customer engagement, risk or compliance.

Through Brain, we are democratising artificial intelligence. With it, business users can leverage the historical data inside their bank or insurance company which is necessary to solve their use cases.

Brain provides our clients with a **smooth user experience**, with user interfaces that guide them across all the necessary steps to prepare data, choose the right data, and use it to train powerful, robust and explainable artificial intelligence algorithms capable of delivering the insights they need to increase sales, reduce churn, improve risk modelling or reduce the cost of compliance.

Once they have seen for themselves that the Brain smart apps can **generate a significant return on investment for them**, businesses can autonomously deploy these applications, generate an API (Application Programming Interface) and connect this API with their existing systems, a CRM (Customer Relationship Management), a campaign manager or a anti money-laundering alert system.

The models that are built with Brain follow **responsible AI best practices**. Explainability is at the core of all we do and all the algorithms that we provide, from decision trees to the most advanced deep-learning algorithms. **With Brain, our customers do not have to choose between accuracy and transparency**. Users can know which predictive factors play a role in a given decision and how they affect decisions. We also develop technologies to assess various biases that can alter the accuracy and the robustness of the algorithms trained by Brain, or biases that might lead users to create unfair algorithms.

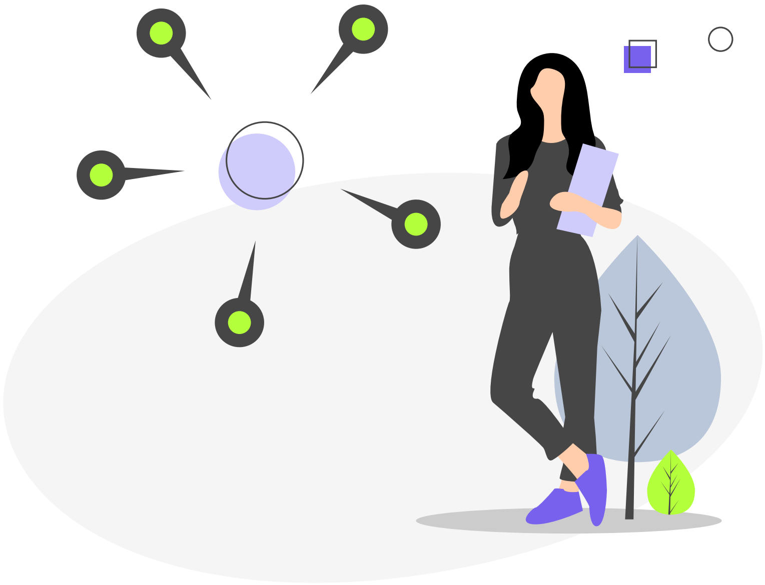
Brain is built from the ground up with production in mind, which means **models can be easily deployed but can also be easily maintained**. Maintaining AI models is a new paradigm but is critical as data and its patterns change quickly over time. Maintaining AI means regularly assessing how the AI models behave with the stream of data they analyse and how they to produce decision scores. Brain makes it extremely easy to know when a model needs to be updated and redeployed.

Lastly, as we need to accelerate our transition to a more sustainable way of producing, DreamQuark enables its investment management customers to use AI and ESG data to **help identify private and corporate customers likely to invest in sustainable assets**. This helps accelerate the transfer of funds from high carbon intensity investments to low carbon intensity investments.

Brain already powers several banks and insurance initiatives in France, UK, Switzerland, Singapore and Luxembourg and we are currently developing our presence in Germany and Japan.

We are deeply committed to furthering the development of several key points which are crucial to intelligent development of AI:

- **Ethical and responsible AI**, with an emphasis on fairness, robustness, security
- Artificial intelligence **deployment** (a field known as MLops), integration and maintenance
- **User experience for smart business applications**
- Artificial intelligence for **sustainability** (energetic efficiency, AI recommendation of green bonds...)



5.4. Our lines of research

Research is a key component of our activity. Since 2014 DreamQuark has been built around a strong research component in order to further the field of artificial intelligence and build products that accelerate the deployment of meaningful solutions.

We believe that we need to continue developing new algorithms to improve our way of tackling the most important use cases for our customers and increase the economic value they generate over time.

We see explainability as such as fundamental topic that it is at the heart of all we do and all the algorithms that are added to our platform. Therefore we are continuously researching to make sure that we can explain the decisions of an increasing number of algorithms such as:

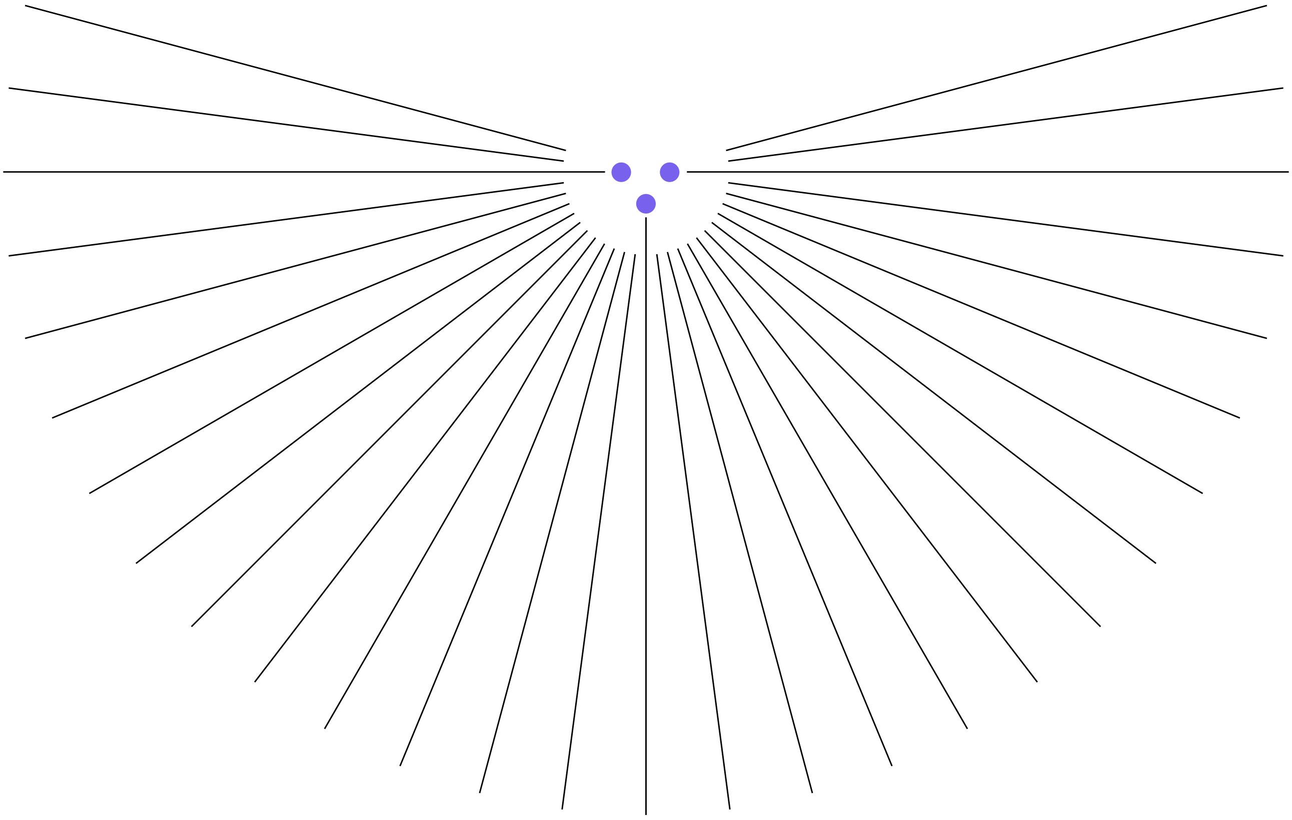
- Recommender systems
- Pricing algorithms
- Unsupervised algorithms for segmentation
- Algorithms for time-series based challenges (like stock-picking algorithms)
- Algorithms for natural language understanding and processing to extract and structure data from a wide range of documents with free text.

We believe that algorithms should be fair and robust and we develop technologies to assess the robustness and fairness of the algorithms we propose.

We believe that algorithms should be **easily deployed** and maintained and that we need to manage the life cycle of artificial intelligence models.

Lastly, we believe that artificial intelligence should be **accessible** and we invest in the development of a simple and engaging user experience, for individuals and companies, in order to democratise AI.

Customers will need state-of-the-art technology to succeed in their transformation journey. Through proprietary research, the development of open source and proprietary software, as well as thought-leadership, we are pursuing our ambition of removing every obstacle in this journey that is only just beginning.



BIBLIOGRAPHY

The main guidelines and ethical strategies we analysed:

- Canada
  www.declarationmontreal-iaresponsable.com/la-declaration

- Europe
  https://ec.europa.eu/futurium/en/ai-alliance-consultation

- UNESCO
  https://unesdoc.unesco.org/ark:/48223/pf0000373199/PDF/373199eng.pdf.multi

- OECD
  https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

- World Bank
  https://blogs.worldbank.org/impactevaluations/how-can-machine-learning-and-artificial-intelligence-be-used-development-interventions-and-impact

- ILO
  www.ilo.org/wcmsp5/groups/public/---dgreports/---cabinet/documents/publication/wcms_647306.pdf

- Japan
  www.soumu.go.jp/main_content/000507517.pdf

- China
  www.sppm.tsinghua.edu.cn/eWebEditor/UploadFile/China_AI_development_report_2018.pdf

- USA
  www.defense.gov/Explore/News/Article/Article/2006646/defense-innovation-board-recommends-ai-ethical-guidelines/

Some of the publications we reviewed about ethical and technological advances:

“Fairness Metrics” to Aid Responsible AI Adoption in Financial Services (mas.gov.sg)

- Gouvernance des algorithmes d’intelligence artificielle dans le secteur financier | Banque de France
  banque-france.fr
  ethical AI in communication - Bing
  zmz022.pdf (silverchair.com)

- AI Explainability and Communication | Reinventing Communication
  commission-white-paper-artificial-intelligence-feb2020_fr.pdf (europa.eu)
  Health care needs ethics-based governance of artificial intelligence - STAT (statnews.com)
  (PDF) Ethics of Using AI and Big Data in Agriculture: The Case of a Large Agriculture Multinational (researchgate.net)
  Agri-food to balance AI ethics with supply chain transparency opportunities (foodnavigator.com)
- The term ‘ethical AI’ is finally starting to mean something | VentureBeat
  www.un.org/development/desa/disabilities/envision2030.html
  www.researchgate.net/publication/228600407_Ethical_guidelines_for_AI_in_education_Starting_a_conversation
  https://fas.org/sgp/crs/natsec/R45178.pdf

On ethical principles:

- François Dubet (dir.), *Inégalités et justice sociale*, Paris, La Découverte, coll. “Recherches”, 2014
- Pierre Bourdieu & Loïc Wacquant, *An Invitation to Reflexive Sociology*, Chicago University Press, 1992
- Hans Jonas, *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*, Chicago University Press, 1979
- Brent Mittelstadt & Sandra Wachter (Oxford), Chris Russell (Surrey), “Explaining Explanations in AI”, the Alan Turing Institute, 2019
- Aristotle, *Nicomachean Ethics*, R. Crisp (ed.), Cambridge: Cambridge University Press, 2000.
- Crisp, Roger, “Well-Being”, *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.)
- DeCew, Judith, “Privacy”, *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.)

To stimulate further thinking about AI:

- *https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd*
- *https://www8.cao.go.jp/cstp/tyousakai/ai/summary/index.html*
- *https://blogs.worldbank.org/eastasiapacific/artificial-intelligence-big-data-opportunities-enhancing-human-development-thailand-and-beyond*
- Francis Wolff, Notre Humanité. *D'Aristote aux neurosciences*, éd. Fayard, 2010
- Antoine Garapon & Jean Lassègue, *Justice digitale*, éd. PUF, 2018
- Platon, *Protagoras* (320c-321d)
- *https://medium.com/center-for-effective-global-action/machines-fighting-poverty-ded0cbae6ae9*
- Dominique Lestel, *À quoi sert l'homme ?*, éd. Fayard, 2015
- AI and the future of religion : *www.youtube.com/watch?v=QLafg_mpHRA*
- *www.cnbc.com/2018/05/11/how-artificial-intelligence-is-shaping-religion-in-the-21st-century.html*
- Kant, *Critique of Practical Reason*, 1788
- *https://emerj.com/ai-sector-overviews/artificial-intelligence-for-video-marketing-emotion-recognition-video-generation-and-more/*
- MIT Technology Review report
- Hursthouse, Rosalind and Pettigrove, Glen, “Virtue Ethics”, *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.)
- *http://stephanus.tlg.uci.edu/lsj/#eid=48064*
- *https://medium.com/legal-design-and-innovation/ai-goes-to-court-the-growing-landscape-of-ai-for-access-to-justice-3f58aca4306f*
- *www.un.org/development/desa/disabilities/envision2030.html*
- Alexander Görlach, *Homo Empathicus. Von Sündenböcken, Populisten und der Rettung der Demokratie*, Herder Verlag 2019
- *www.nextnature.net/2019/04/the-religion-named-artificial-intelligence/*

State of ethical AI in 2021: challenges posed by ethics to companies developing AI

a white paper by DreamQuark

email : info@dreamquark.com
www.dreamquark.com

dreamquark