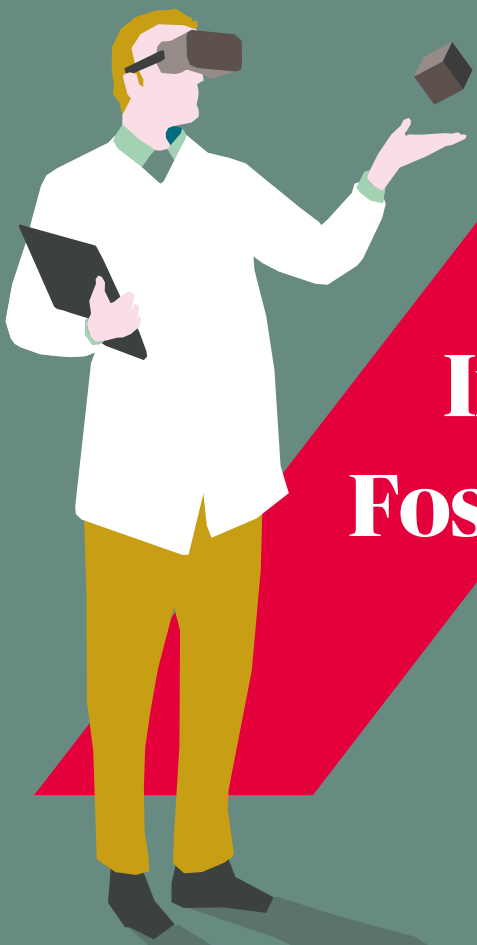
AXA Research Fund



Artificial Intelligence: Fostering Trust

AXA Research Guide

Artificial Intelligence: Fostering Trust

Table of contents

Artificial Intelligence: building trust through research

Artificial Intelligence (AI) is already part of our daily lives for example, behind what we see on our screens after an online search. Not a day passes by without hearing that AI has reached another milestone, be it detecting cancer as accurately as radiologists or beating human players at strategy games. The technology has many other applications and is bound to be ever more present in our lives. In insurance, AI offers many opportunities both for our business and our customers such as speedier claims processes or improved customer relations through tailor-made solutions.

Today, AI brings concerns on automation and fantasies on our ability to control technology. As it remains poorly understood, there is a polarized public debate between those with a “solutionist” approach, for whom a Golden Age is starting where "AI will solve all problems", and those with an apocalyptic vision, for whom AI marks the “end of humanity" by threatening human autonomy and free will.

This is where science comes in to foster an informed discussion and help us better prepare for the AI revolution in a responsible way.

At AXA, we are convinced that scientific research enables us to better anticipate and address the challenges we are facing. AI is a mighty one.

Fostering trust in AI

In this context, fostering trust in AI is crucial. Two approaches stand out:

- Abiding by a set of principles which define what one can and cannot do with AI, for instance, committing to train algorithms on diverse datasets to avoid biases or making sure that AI's decisions can be explained to users and consumers in a simple way and understood by them.

- Having rigorous corporate governance processes to control what is being done with these new technologies, such as setting up a panel of external experts to review AI's use.

Science can help us address the issues of transparency, accountability, security and privacy raised by AI.

Giving researchers a voice

This Research Guide has a two-fold ambition: to report on progress in research and to give researchers a voice by showing how AI projects interlock with each other. Multiform by nature, AI can only be fully grasped through a holistic approach.

At AXA, we are convinced that scientific research enables us to better anticipate and address the challenges the world is facing. AI is a mighty one.

This is where science comes in to foster an informed discussion and help us better prepare for the AI revolution in a responsible way.

This is why we are committed to supporting research projects and encouraging researchers to participate in the public debate. Their expertise and their voice carry weight and help light the way towards responsible AI empowering people to live a better life.

We hope this guide will serve as a basis for an open and fruitful debate.

Jad Ariss

AXA Group Head of Public Affairs and Corporate Responsibility
Member of the AXA Research Fund Scientific Board

“Artificial Intelligence has a fascination that makes us overestimate its capacities, and its dangers”

In 1956, experts met to discuss “thinking machines” at the Dartmouth College workshop. At this seminal event, John McCarthy coined the term “Artificial Intelligence”. Since then, AI has had an increasingly important role in society. Raja Chatila, Institute of Electrical and Electronics Engineers Fellow, Professor and Director of the Institute of Intelligent Systems and Robotics at Pierre and Marie Curie University in Paris and member of the European Commission High-Level Expert Group on Artificial Intelligence, answered our questions on the present and future of AI.

The term “Artificial Intelligence” has certain connotations. How do you define this type of “intelligence”?

The reference to artificial intelligence goes back to Alan Turing in 1950. I believe these concepts stem from an understanding of intelligence as an isolated process producing and manipulating abstractions, neglecting the fact that our brain and nervous system evolved to deal with the complexity and dynamics of the real world and to control our body.

How much AI is already present in our lives?

It is AI technology that delivers results in the blink of an eye when we make an Internet search. Many sectors ranging from medicine to defense use AI techniques for image recognition, while AI-based algorithms assist recruitment managers and inform judicial decisions. This trend will soon reach all sectors. What is called the 4th industrial revolution (or Industry 4.0) is transforming industrial plants and warehouses into cyber-physical systems in which robots are networked and interact with human workers.

Is this “simply” evolution, or is it revolution?

It will be both revolution and evolution. There will be radical and abrupt changes when some technologies are first deployed. But their inclusion into daily life could be gradual. In a few years, automated vehicles will be deployed, first for specific types of driving and in dedicated areas, and then in more challenging situations. As the adoption of automation spreads throughout industries, the impact on jobs and organizations will

increase, which might lead to radical changes in society.

Is AI a double-edged sword?

Every technology can have beneficial and negative aspects, and AI is no exception. But AI has a fascination that makes us overestimate its capacities, and its dangers. For example, the fear that AI might take over the world is not founded on any evidence, but this doesn't mean that the development of AI shouldn't include safeguards. One major issue is a lack of transparency, which raises questions about the use of algorithmic decisions in critical domains such as justice.

Beyond the regulatory aspect, does trust in AI require an ethical framework?

The absence of a regulatory framework is in part due to a lack of awareness about the impact and consequences of diffusion in society, and in part also due to the fear of taking precautionary measures too early, which would hinder innovation.

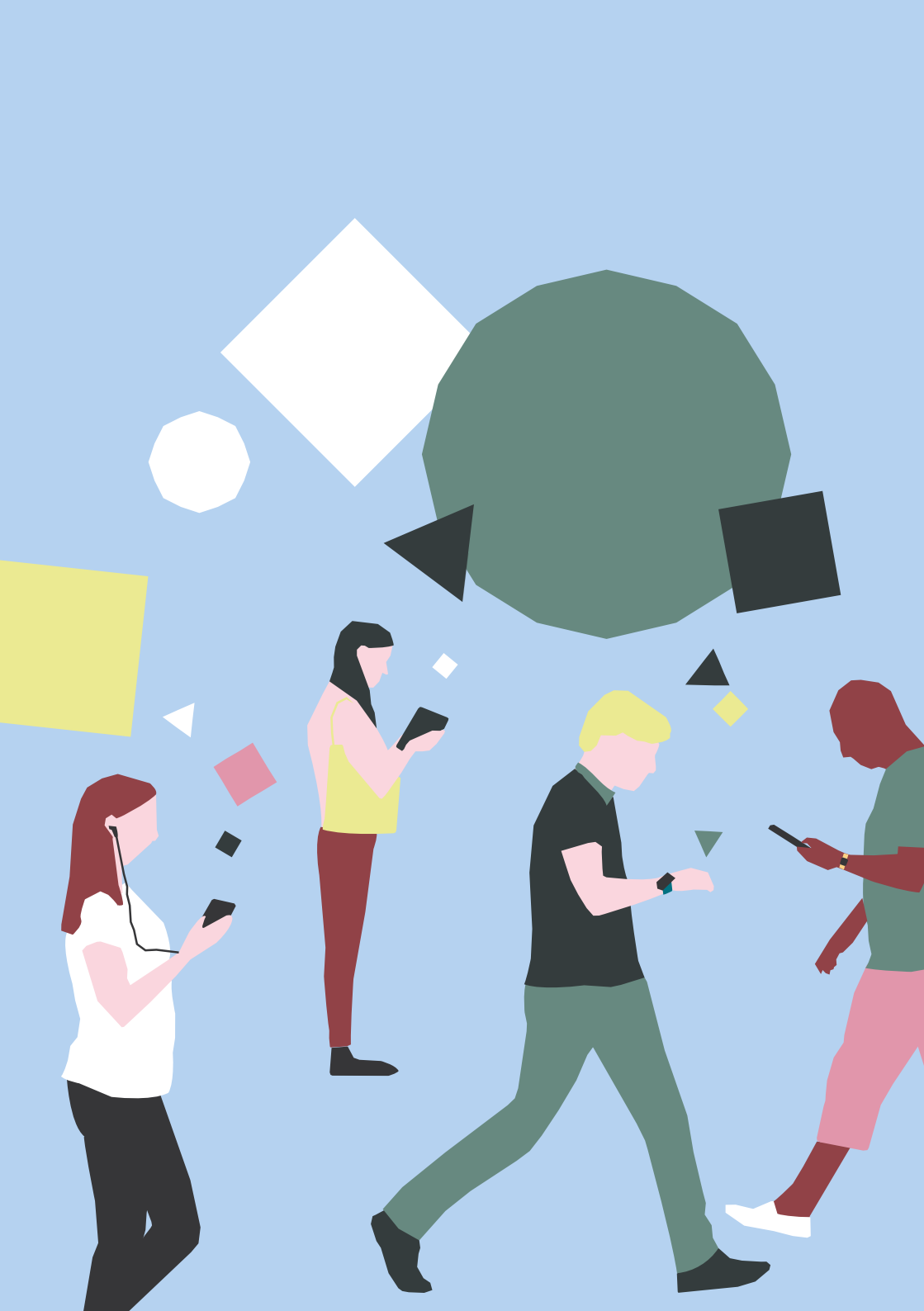
"No airline company would fly a plane that was not certified to be trustworthy. Nothing of the sort exists for AI today."

Classically, a commercial product must follow certain standards and be certified – especially if used in critical applications. No airline company would fly a plane that was not certified to be trustworthy. Nothing of the sort exists for AI today, even though the technology is increasingly used in applications determining people's destinies. It is essential to devise industrial standards, to define certification processes and organize independent public certification authorities.

The IEEE has launched the "Global Initiative on Ethics of Autonomous and Intelligent Systems" (A/IS) which states that such systems should comply with clear general principles: respect of human rights, prioritizing well-being, system transparency, accountability of humans and organizations deploying A/IS, and misuse risk minimization. It's clear that ethical reflections and recommendations must be made at an international level, so that they are grounded on shared values while taking into account cultural diversity.

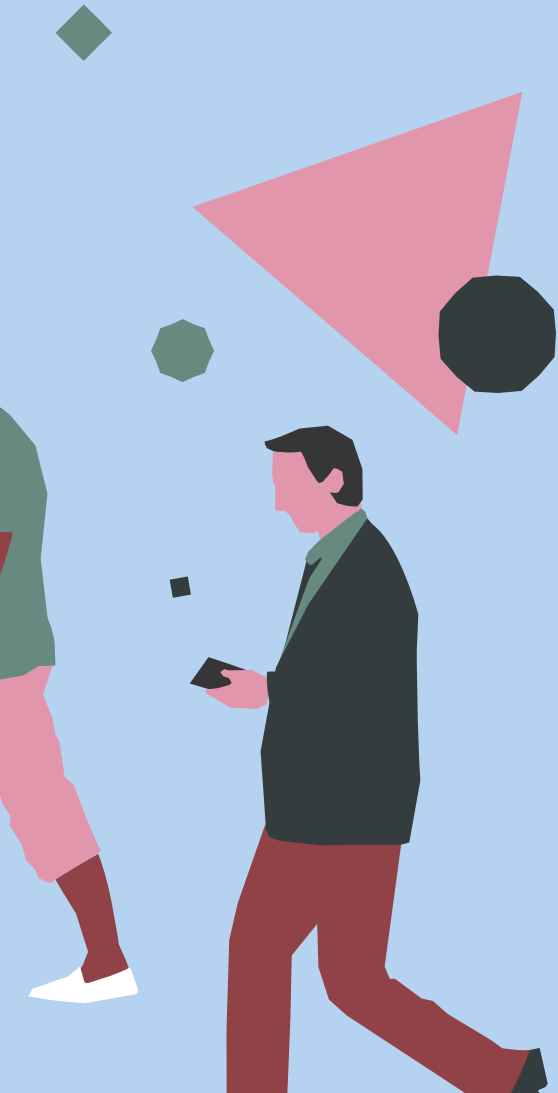
What will happen regarding accountability as AI becomes more autonomous?

Algorithms are designed by humans. Learning processes are designed by humans. Data are collated by humans. Systems are built and commercialized by humans. Therefore, accountability must remain with the humans and their organizations that design, deploy, operate or knowingly use autonomous and intelligent systems.



Chapter 01

Ensuring data security and privacy in the Age of Artificial Intelligence



Cultural recommendations, the quantified self, smart assistants, cities and houses, autonomous vehicles… Data is the driver of every smart system and is firmly established at the core of autonomous and personalized AI solutions. But while the amount of online (and personal) data grows day by day, many voices are raising the alarm about the security surrounding what many are calling the "fuel of the future."

This poses a simple question: how can we ensure a sufficiently secure environment to foster trust in a data driven future?

“We now live in a new quantification era of digital platforms, financial capitalism, digital traces and machine learning”

Data fuels Artificial Intelligence (AI) and stands at the core of autonomous and personalized solutions. Ensuring data protection is therefore a cornerstone in the adoption of AI in our daily lives. EPFL Digital Humanities Sociologist Dominique Boullier and Singapore Management University’s AXA Chair Professor of Cybersecurity Robert Deng discuss the prerequisites for long-lasting trust.

What makes data so valuable today?

Robert Deng: Data has been described as the new oil of the digital economy. Data represents information, which in turn represents knowledge and value. For example, personal data is specific about an individual; trade secrets are confidential information on which companies base their business decisions; and classified data are sensitive information used by governments to inform policies.

Leakage of such data may therefore result in the violation of personal privacy and can jeopardize businesses and even governments.

Has society’s relationship with data evolved throughout history?

Dominique Boullier: Since Mesopotamian times we can see that writing, accounting and the emergence of states are closely related. Identity documents and the traceability of goods and people have been at the foundation of modern states. Censuses, computing (Hollerith machines) and the welfare state were developed at the end of the 19th century, together with a social sciences model of society as a whole. More recently, opinion polls, sampling, mass media and brands have all emerged in the 1930s. And we now live in a new quantification era of digital platforms, financial capitalism, digital traces and machine learning, which I call third generation social sciences. The term “AI” is

frequently used in relation to the technologies emerging in this new age, but I prefer the term "machine learning", because "learning" is really the key feature. One of the defining characteristics of this new era is that two thirds of humans are now equipped with a mobile phone, i.e. approximately 4.5 billion individual users, according to 2013 figures. I've coined the term "habitele" to describe the anthropological transformation this represents. I've been carrying out empirical investigations into typical data usage, examining behaviors and mapping connected social worlds. Much of this research will be published in 2019, with many papers and presentations being made available on my website (boullier.bzh).

How do you think this technology will evolve in the coming years?

RD: We have entered into a new paradigm of big data. This is characterized by huge volumes of data, a wide variety of data, and its high velocity production from numerous sources, such as social network users, cameras, sensors, and mobile devices. With the introduction of 5G networks, the number of connected devices and services will increase in both number and types, accelerating the big data trend even more rapidly.

DB: Digital identities will be assigned to any entity, human or non-human, enabling their traceability to all companies with a high level of machine learning expertise. Since there are many security flaws in the network, we are likely to experience huge

Dominique Boullier

Digital Humanities Institute | EPFL

Dominique Boullier is a University Professor of Sociology, specializing in the uses of digital and cognitive technologies. After creating and acting as the scientific director of Sciences Po medialab, in 2015 he directed the Social Media Lab of the Swiss Federal Institute of Technology in Lausanne (EPFL) where he became a senior researcher at the Digital Humanities Institute.

He is the author of many books across various areas of expertise, with a particular emphasis on digital technology. His work in this field began in 1982 with an evaluation of the Minitel – a successful online service in France that predated the World Wide Web.

His **AXA Joint Research Initiative** conducted with Sciences Po Media Lab in Paris aims to better understand the role of sociocognitive barriers in the adoption of technologies and systems related to data sharing and financial transactions. Findings from this program will contribute to the evolution of a framework for the use of personal data.

data breaches in future, especially via connected objects that have already been the targets of cyber-attacks. Unfortunately, it might take a personal data disaster on the scale of Fukushima to trigger a reaction from governments and the public.

> **Some observers are convinced that large-scale data collection, along with data misuse and data security breaches, will mean an end to privacy. Do you agree?**

RD: Data security includes three aspects: confidentiality, integrity, and availability. The violation of data confidentiality results in the disclosure of sensitive data to unauthorized parties, which has been the subject of much attention and will continue to command the spotlight in future. Threats to privacy are normally the result of a violation of data confidentiality protection.
Going forward, data integrity will play a crucial part in the Internet of Things. For example, GPS signal spoofing is a form of attack on the integrity of location information, which will have serious implications for driverless cars.

One of the main focuses of my research is in improving data security. My team has designed and implemented an efficient Attributed-Based Encryption (ABE) scheme to enforce fine-grained access control of encrypted data in the cloud. In this system, a user encrypts his or her "plaintext" data and links it with an access policy. For example, an access policy could be: Access Policy = (Engineering AND Project Manager) OR Marketing. In this case, encrypted data could then be uploaded to a cloud storage system and only Project Managers in the Engineering Department, or anyone in the Marketing Department, would be able to access and decrypt this data. Other parties, including the cloud service provider, are not able to access the underlying plaintext data. In addition, my team is increasingly focusing more on data integrity protection, to make sure that outsourced data in the cloud are intact and retrievable, and that outsourced computations are performed correctly.

> **Are users aware of the risks?**

DB: People are aware, but services are often free of charge and highly addictive, which makes it hard to persuade people to change their behavior. For this reason, we need to take action at the level of nudging mechanisms on online platforms. Currently the main motto seems to be "Yes I am aware of the risks, however…", which is an outdated way of living with a cognitive conflict. Users need to regain control of their data, but this can only happen if states, regulatory bodies, and responsible companies take action on the basis that these issues are critical £for social trust in general, because individual behavior alone is not enough to change anything, given that whole populations are addicted to these free services.

"Users need to regain control of their data, but this can only happen if states, regulatory bodies, and responsible companies take action"

Recovering control of our data will require trust architects, designers, and educators to be trained to create design interfaces and means of controlling the security levels of "privacy-by-design" products and services. We must encourage communities that are building services for collective control and promote general encryption - as Professor Deng is doing - and support open source for all public services and companies in relation to transactions with the public.

What is the role of research in these matters?

DB: Researchers must contribute to legal creativity and to the design of technical user-centered services. To achieve this we need a real effort in theoretical thinking. My own work addresses two areas. Firstly, I am examining the ways that machine learning affects the social sciences. The second aspect of my work addresses the development of a meme-tracker using machine learning to account for propagation patterns in this new world where information flows and replicates at very high velocities.

My AXA Research Fund-backed project dealt with the high risk of data breaches generated by the flaws in Internet architecture – the issue is speed versus security. We produced a graphical representation of topics related to data privacy issues, conducted interviews in four countries with people who had experienced data breaches, carried out case studies of credit card trust systems and well known examples of data breaches, and produced a semantic map of terms of use.

Robert Deng

Cybersecurity & Information Systems | Singapore Management University

Robert Deng is a cybersecurity researcher at the Singapore Management University's School of Information Systems. He holds the **AXA Chair Professor of Cybersecurity**, in which role he conducts research into protecting data security and privacy to defend against cyber criminality.

Professor Deng has a Degree in electrical engineering from China's National University of Defense Technology, and a Masters and Doctorate in electrical engineering from the Illinois Institute of Technology. His interests include applied cryptography; multimedia security; mobile, wireless and sensor network security; and trusted computing and system security.

Professor Deng aims to develop new security models, algorithms, protocols, and analysis techniques to ensure data security in the cloud computing environment.

RD: In this age of big data and AI we urgently need technologies for scalable and efficient protection of data confidentiality, integrity, and availability. The IT landscape evolves rapidly, as do the threats, and therefore so must the technologies to counter these threats. In addition to new technologies, user awareness, government regulations and legislation must also be updated to catch up with the rapid progress in IT.

My AXA-funded research program is addressing this urgency for new technologies. I mentioned earlier that my team is investigating new systems for data confidentiality and integrity protection. For examples, creating new ways of enforcing access control of encrypted data by authorized users and ensuring computations on data are performed correctly when data is outsourced to a public cloud, while keeping the input and output of the computation confidential from the servers. This research effort will provide new security models, algorithms, security protocols, and security analysis techniques that will address some long standing data security and privacy problems.

> **As AI and data are strongly dependent on one another, could data breaches threaten the adoption of new technology?**

DB: No, because data breaches have already occurred and this has not led to a rejection of free services. On the contrary, AI innovators are promoting the ability of AI techniques to solve security issues and to deliver better data protection. This presentation of the benefits of AI is as potentially damaging as the misuse of AI technology itself. We must publicly criticize any AI techniques that are totally opaque and untraceable.

“We must publicly criticize any AI techniques that are totally opaque and untraceable”

> **To what extent is protecting data and privacy a way to establish trust in AI?**

RD: AI is a double-edged sword in terms of data security. It can of course be used for protecting data and information systems. For example, AI can be used to better detect spear phishing attacks, and to automatically respond to such attacks. However, cyber criminals can also use AI to launch more effective attacks, such as automatically sending spear phishing emails or to spread fake news on social networks

We need large amounts of data to train AI systems and it is vital that this data is protected from distortion at the various stages of data processing, collection and storage. This clearly requires data integrity and data availability protection.

DB: As I mentioned earlier, the public can be convinced that AI is an effective way to promote privacy, but that’s by no means the whole picture. The condition for sustainable data development, AI driven or not, is the need to radically reform the network itself and its architecture.

“I want to design quantum cryptographic protocols for the strongest security conceivable”

Are we on the cusp of a quantum revolution that will change our idea of data security? Researchers like Professor Antonio Acín are increasingly convinced that such a breakthrough is imminent. *“My main aim is to develop device-independent (DI) cryptography for quantum information applications,”* explained Antonio, holder of the AXA Chair for Quantum Cryptography for Enhanced Information Security. DI quantum protocols provide unprecedented security, as they do not rely on the trustworthiness of the physical devices involved.

Evolution at the core of data security

These new security protocols are highly resistant to hacking attacks. So, how do they work? Two honest users willing to exchange a secret perform measurements on entangled quantum particles. These measurements produce results between them that are perfectly correlated, but that cannot be predicted by an adversary. These correlated results therefore provide a secret key to the users that they can use to encrypt the secret in a secure way. Antonio’s work utilizes the unique rules that govern the quantum world, creating systems with no classical equivalent: *“I want to design quantum cryptographic protocols for the strongest security conceivable for current and near-future technology. Hacking these protocols would mean violating the laws of quantum physics, which is something that has never been done.”*

Advanced AI applications, and more…

Antonio’s research also focuses on AI. *“This involves the use of classical machine learning techniques to solve quantum physics problems, and also the use of quantum devices to design algorithms for machine learning and optimization.”* According to him, quantum computers may provide major advantages in terms of speed and performance for AI: *“Thanks to quantum superposition, they provide different and sometimes more efficient ways of dealing with large amounts of data. Understanding how quantum computers can boost AI problem is at the moment a very active research direction,”* said Antonio.

Antonio Acin
Institute of Photonic Sciences (ICFO)
AXA Chair since 2015

Chapter 02

Tailoring the ethical and regulatory framework of Artificial Intelligence

While Artificial Intelligence is increasingly permeating our daily lives, contacts and interactions between humans and machines are constantly growing. However, such new relationships often go beyond our current laws and ethical frameworks. For instance, who is accountable when an autonomous vehicle hits a pedestrian? What boundaries should be set for human-robot interactions?

All these questions boil down to a simple one: while innovation in intelligent systems accelerates, are laws and ethics able to keep pace?

“Artificial Intelligence is very much a legal and ethical issue because it can easily be used to exploit human vulnerabilities”

When well designed and carefully used, AI has great potential. However, its rapid adoption threatens to undermine these benefits. Examples of manipulation, discrimination and exploitation in machine learning have already been observed. Joanna Bryson and Philipp Hacker discuss some of the crucial regulatory and ethical questions raised by this technology.

> **As AI is increasingly adopted, is our relationship with this technology changing? And to what extent is AI becoming an ethical and legal issue?**

Joanna Bryson: The vast majority of our interactions with intelligent technology go unobserved. We don’t notice how much AI improves our spelling, photography, web searches, navigation, dishwashing… We use it without thought – in fact, without knowledge or informed consent.
So there is the pragmatic relationship where we are all being both empowered and monitored by intelligent systems in ways that are invisible to us. Then at the same time many people are forming beliefs –which may have little basis in reality – about the needs, desires, or intentions of the intelligent machines that look and somewhat interact like science fiction has led us to expect.

Philipp Hacker: AI is currently one of the most hotly debated topics in legal research. I see five main areas in which the discussion has become particularly intense: the rise of social media and the consequences for democracy; autonomous weapons and their implications for international humanitarian law; the potential collusion between self-learning algorithms and questions of algorithmic price discrimination (“AI cartels”); the impact of algorithmic decision making on meaningful individual choice; and finally, issues of accountability.

JB: AI is very much a legal and ethical issue because it can easily exploit human

vulnerabilities. For example, many people believe they have a new servant in their homes, an "intelligent speaker", forgetting it is really a microphone that uploads their information to the Internet.

> **What ethical or regulatory rules are currently in place to monitor the expanding relationship between humans and AI?**

JB: Within the EU, the General Data Protection Regulation (GDPR) has just come into effect, which requires that automated decision-making that affects human livelihoods should be explainable. Just as importantly, existing liability and tax laws apply to any product incorporating AI. Courts must treat intelligent technology like what it is, just another artifact, and uphold ordinary procedures for determining when a company is culpable. When corporations realize that they will be held accountable, then they will be much better motivated to create transparent systems.

PH: The international deployment of machine learning applications, and the size of the actors involved is posing significant regulatory challenges; however, private international law is already accustomed to establishing connections in highly international contexts, so I think these problems can be overcome. For example, for EU law to apply, a company need not necessarily have its headquarters in the EU.

It is only necessary for an offence to have occurred in the EU, or that websites or services can be accessed from the EU. Existing regulations are also capable of holding very wealthy companies to account, including The Big Four (Alphabet Inc., Apple Inc., Facebook and Amazon). For example, the GDPR now contains sanctions amounting to up to 4% of the global annual turnover. If these sanctions are effectively enforced, they are likely to be a game changer.

Joanna Bryson

Department of Computer Science | University of Bath

How does a robot's appearance affect our perception and our understanding of what it really is? In 2010, the UK published a set of five ethical rules for robotics – the first national level document on AI ethics. A delegate of the workshop that produced this set of rules, Dr Joanna Bryson of the Department of Computer Science at the University of Bath, is putting one particular ethical rule to the test.

Joanna and her team have experimented with non-humanoid robots in the past. The **AXA Award on Responsible Artificial Intelligence** is allowing them to do experiments with humanoid robots this time, and compare the results with their previous findings.

By investigating how a robot's appearance affects human/machine interactions, Dr. Bryson's experiments will greatly contribute to our understanding of what humanoid robots should look like to allow safe use in the future.

There are some specific laws geared towards reining in the pitfalls of AI. For example in Germany, an amendment to the general traffic law addresses liability for autonomous vehicles, and another law seeks to rein in hate speech on social media, while algorithmic high-frequency trading is addressed in EU capital market regulation. Nevertheless, the law is struggling to keep pace with this rapidly expanding and dynamic field.

"When corporations realize that they will be held accountable, then they will be much better motivated to create transparent systems."

What could bridge the gap between innovation and current laws?

JB: I recommend maintaining the present standards of accountability under the law. There must be substantial prosecutions to reduce sloppy and irresponsible software practices. We need to know where data comes from and where it goes. This involves cybersecurity and good accountability practices. Importantly, to ensure corporations are strongly motivated to get their houses in order, we must make it clear that AI by no means changes liability and accountability.

One special concern at the transnational level is that there is a great deal of value being created in what is essentially "information bartering" – we give data to get information, and the companies say this is a "free" service, but corporations obtain significant value from this data. In theory, these corporations are in one country and pay tax in that country, but they often claim to operate in tax havens. I believe we should tax corporations operating transnationally based on increases in their market valuation rather than on their annual income, and we need to find a way to redistribute a fair share of that income to support the infrastructure of the societies they benefit from.

PH: I see four main principles of regulation that would be helpful in this context:

- legislation needs to be future-proof, i.e, formulated broadly enough to cover new technologies that are not yet on the horizon;
- we need well-resourced supervisory authorities with the necessary prowess to deal with these complex phenomena;
- to avoid stifling innovation, "regulatory sandboxes" such as those established by the Financial Conduct Authority in the UK can be helpful for startup companies by enabling them to test their products in a relatively safe regulatory environment;

- and, finally, we should aim to implement legal and social values directly into code.

This would provoke a shift from ex post liability, which suffers from rampant enforcement problems, to ex ante mitigation of risks inherent in AI. Algorithmic fairness procedures, for example, offer a promising outlook in this context.

> **What is the role of research in these matters, and on which research projects are you currently working?**

JB: There are at least two roles for research: to solve problems we already know; and to understand things that we do not yet understand. Industrial and government research are often best for the first, universities and philanthropists are best for the second. Some countries have done very well by investing heavily in universities that not only educate the next generation, but also collect the brightest people together and give them a chance to pursue questions that experts have recognized as having potential significance.

My two biggest projects right now are my AXA project, and another concerning human cooperation. The AXA project involves determining how much of a problem it is to present AI as humanlike, and whether there's a way to get the advantages of such a presentation, while avoiding any moral harm.

PH: Researchers, particularly when supported by research grants, can take time to think matters through, which is a valuable resource. It enables researchers in several disciplines to understand the challenges posed by new technologies. Those designing and applying laws are often under much greater time pressure.

Philipp Hacker

Centre for Law, Economics and Society | Humboldt University of Berlin

Philipp Hacker is a Postdoctoral Fellow at the Law Department of Humboldt University of Berlin, an A.SK Fellow at WZB Berlin Social Sciences Center and a Research Fellow at the Centre for Blockchain Technologies and at the Centre for Law, Economics and Society, both at University College London. He was awarded an **AXA Post Doctoral Fellowship** in 2017.

Philipp's research interests include behavioral law and economics, regulation of AI and blockchain, contract law, securities regulation, and mathematical approaches to the legal arena.

In 2016 Philipp started a new large-scale project on principles of economic regulation in the digital age that deals with the opportunities and challenges Big Data and Blockchain hold for the law. He is the co-author, inter alia, of "FairEconomy – Crises, Culture, Competition and the Role of Law" on the post-financial crisis global economic regime, and an editor of a forthcoming book on blockchain and the law (OUP).

An important part of my research concerns the regulation of AI, and particularly its relationship to social and legal concepts of fairness. It seeks to uncover how existing laws regulate AI, and to what extent novel legal strategies are needed to cover regulatory risks. The second area of my research addresses regulation with AI. I ask to what extent AI can help fulfill regulatory goals and how legal and social norms can be directly infused into AI models. Both areas are financed by the AXA Research Fund.

> **Do we need to foster trust in AI? And if so, how is your research helping to achieve this?**

PH: Trust in the systems we interact with on a daily basis, be they political institutions or technological systems, is one of the most important resources of our societies. This trust is currently eroding at a dangerous pace. With respect to AI, it needs to be fostered not only for instrumental reasons, but also because it positively impacts on the well-being of those affected by AI. With jobs increasingly under pressure from automation, it is likely that trust in AI is also going to be important for social coherence.

“Trust in the systems we interact with on a daily basis is one of the most important resources of our societies. This trust is currently eroding at a dangerous pace.”

People will only trust AI if they can somehow, if only in particular instances, understand what a model is doing, and if they are sure that the workings of the system do not infringe on key social norms of fairness that are also inherent in market exchange. My research on algorithmic fairness seeks to make a difference in this area. I am developing fair algorithms and proposing feasible ways to legally implement such algorithms in practice.

JB: Actually, I think people trust AI too much! I think it is important to foster an understanding in AI, so we can know when to trust intelligent machines in our homes and on our person, and when to invest in cybersecurity, or when to delete an account or to engage in a class-action lawsuit.

My research project aims to give people direct access to the priorities of their AI system. I think the ideal would be that the code is fully open to inspection. Our transparency software and visualizations are an abstraction to make the complex more comprehensible, but the best is to have confidence that someone who wants to could check that the abstraction corresponds to reality. Even if only one user in a million made that check, if everyone knows they could, and that they could publicize the outcome if they found something suspicious, then we can all have more confidence.

“Data has been key to insurance for a century, but AI provides much more powerful tools”

With AI becoming increasingly important in our daily lives, shining a light on the underlying algorithms that power this new technology has become a major research focus. Alexandre d’Aspremont, from the École Normale Supérieure, and Guillaume Beraud-Sudreau, Head of R&D from AXA Global Direct, are carrying out critical work in this field through a collaboration with the French Institute for Research in Computer Science and Automation, and the Fondation du Risque.

Opening the black box

“Data has been key to insurance for a century, but AI provides much more powerful tools to analyze client risk and behavior. These new technologies create major changes in the way insurers work,” commented Guillaume. *“Conventional machine-learning techniques rarely provide explanations as to why one client is considered as lower risk than another, which is why these techniques are described as ‘black boxes’. Our partnership aims to create ‘transparent boxes’.” “This partnership also gave us the opportunity to test new methods using realistic data sets, which is best achieved through a collaboration between academia and industry,”* added Alexandre.

Developing advanced algorithms

The partnership was highly productive on all sides: Alexandre’s team gained clear feedback on how their methods behaved in real world situations, while AXA was able to work alongside high-level academics on state-of-the-art technology.
“A key component of the project was to test and implement algorithms in a core open source machine learning library called SCIKIT-LEARN. This is an important academic contribution and also has direct industrial applications as AXA uses this software”, said Alexandre. *“In addition to the SCIKIT-LEARN work, we conducted internal research to adapt these new algorithms to the needs of the insurance industry,”* said Guillaume. The next step for the partnership? To design robust algorithms capable of addressing the complex subject of insurance pricing.

Alexandre d'Aspremont
Ecole Normale Supérieure | AXA Joint Research Initiative since 2014

Guillaume Beraud-Sudreau
Head of research at Kamet Ventures

Artificial Intelligence must primarily serve to enhance our insurance value proposition

Artificial Intelligence is already part of my life, facilitating cultural recommendations, online shopping... But has it had a major impact on my way of living? Frankly, not really. The science fiction movies and TV shows that have fuelled our collective imagination still seem a distant fantasy.

Siri, Alexa and other intelligent assistants advance us ever closer towards this vision, with their gradual integration into our computers, smartphones, headphones, televisions...

"- What can I do for you?"
"- How can I help you?"

Weather, news, personal messages... everything becomes accessible simply through the power of speech. We know that such communication is not always straightforward. There's much scope for misunderstanding. Artificial Intelligence adapts to us as much as we intuitively try to adapt to it.

But who can deny that we are indeed at the beginning of a new revolution, and that the penetration of Artificial Intelligence into our everyday lives is still in its infancy?

AI-augmented insurance

Of course, this development is also perceptible in the professional sphere. As insurers, our role is to provide our customers with useful services, to be true partners to help them improve their lives. In this sense, there is no point in "doing" AI purely for its own sake. Utility comes first, whether improving service provision or customer experience. Otherwise, AI is merely a gadget.

"There is no point in 'doing' AI purely for its own sake."

One advance to be followed with interest is Natural Language Understanding, which concerns the design of algorithms to respond effectively (and understandably) to queries made using natural sentences.

This clearly requires, among other things, a large amount and diversity of data (big data), considerable computing power and, necessarily, upstream R&D capacity. Here, our intention is not to become like one of the Big Five tech companies (Google, Apple, Facebook, Amazon and Microsoft) – we are not and will never be one of these digital giants. And neither will we depend entirely on these companies to provide our services with improved customer experience, an option that, unfortunately, many actors across all sectors have chosen.

To avoid these twin pitfalls, it is essential in my view to focus our attention on our niches of expertise, namely risk estimation and insurance pricing. Artificial Intelligence that serves our core insurance business, specifically pricing and indemnification, will allow us to build a competitive advantage.

Predicting opportunities, anticipating risks

Indeed, our role is also to project ourselves into the future to envisage potential scenarios, to predict opportunities and anticipate risks for our customers. Parametric insurance, connected health... while Artificial Intelligence enhances insurance, it should not lead us to neglect our responsibilities.

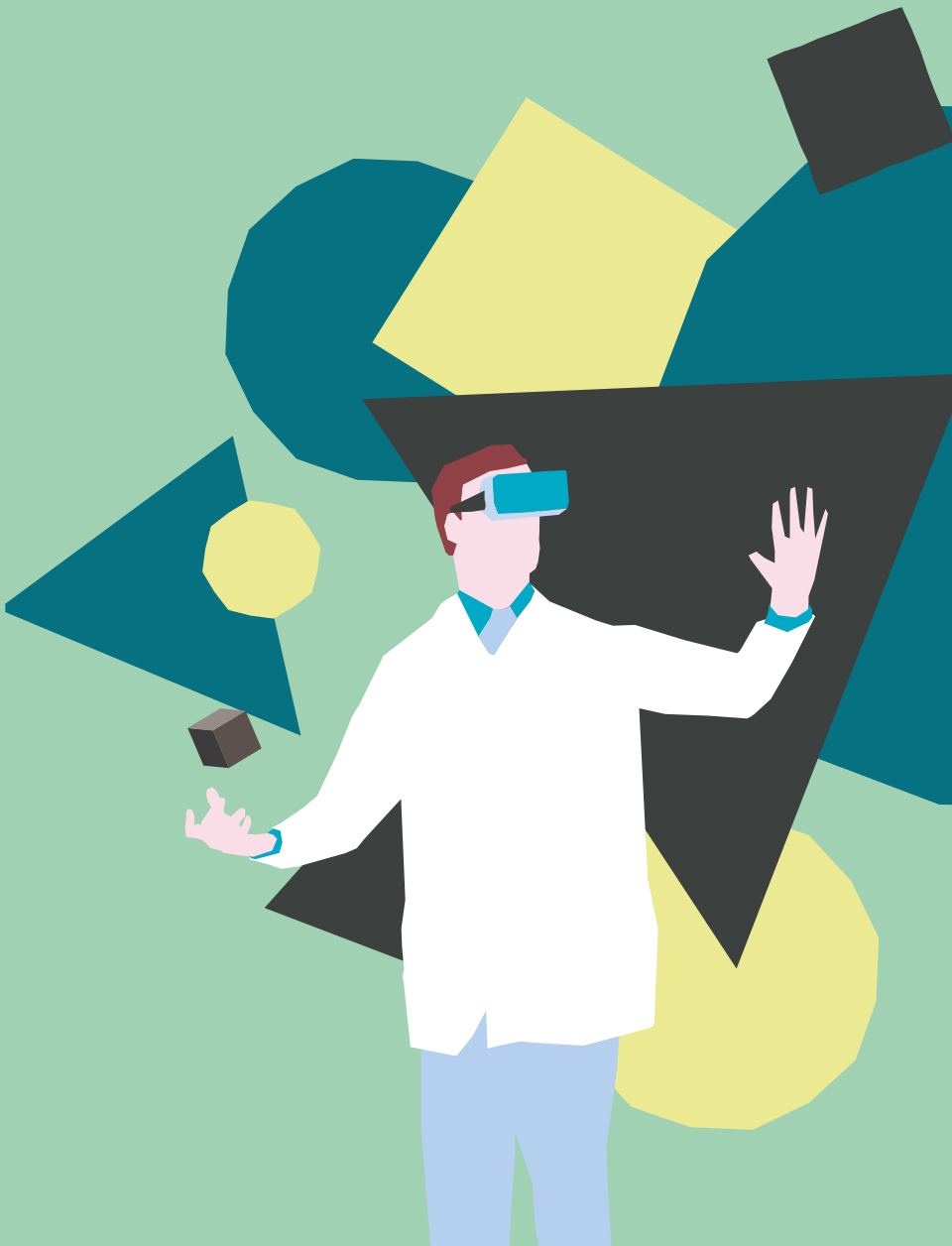
Such as our obligation regarding data security, for example. Customers have very high expectations on this issue. Because without assurances on data security, who would agree to share their data?

At AXA, we have always strived to develop and maintain the most secure IT environment possible. In the context of data collection and processing, this is an essential prerequisite. To fully grasp its significance we only need consider the sensitivity of banking and health data, or data generated in the context of smart cities. And the same goes for compliance with regulations, which are evolving very rapidly in this field.

> **“Without assurances on data security, who would agree to share their data?”**

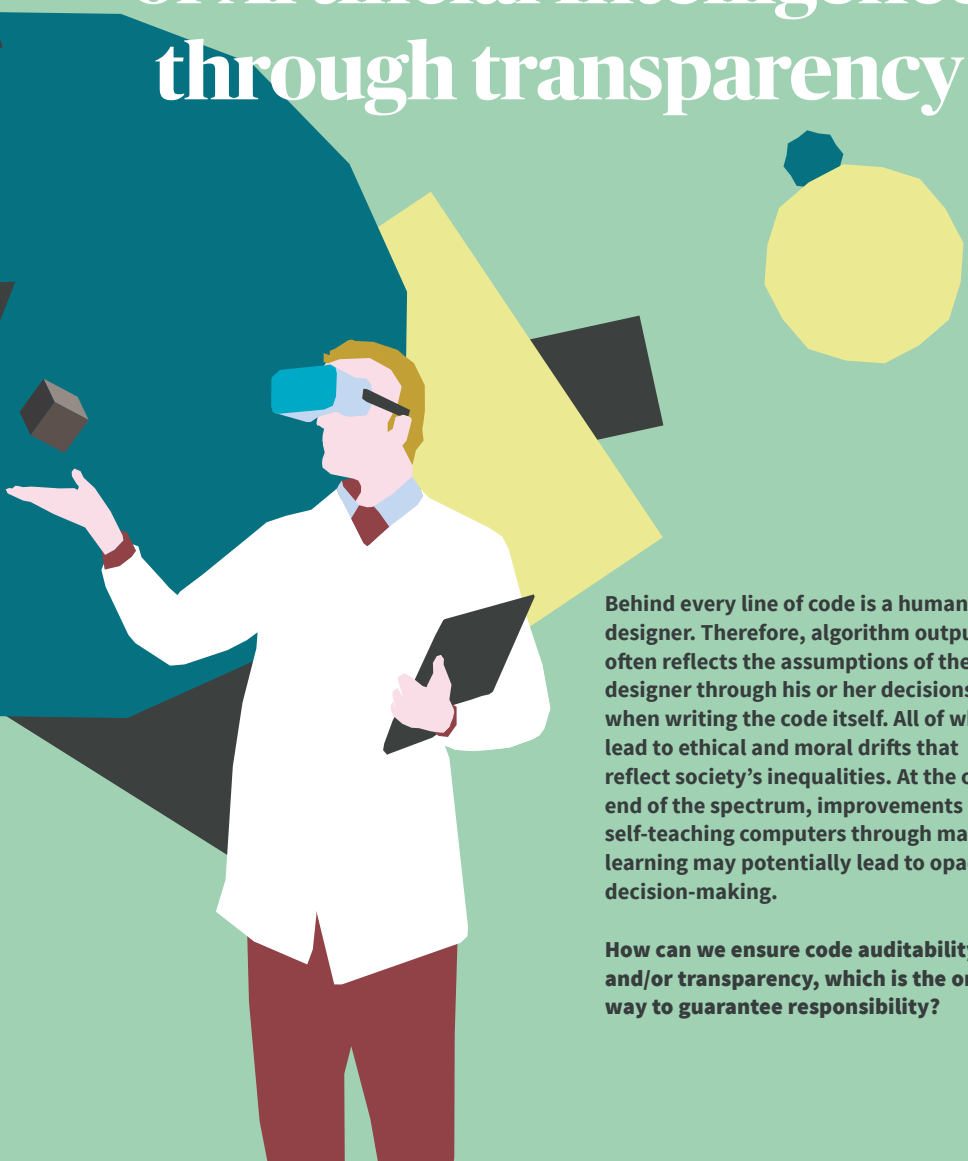
There is also the question of privacy. At the international level, current debate on the issue is intense and questions abound. Research is helping us to better understand this issue, while a multitude of criteria, many of which we are unaware of, define our digital identities. This debate goes hand in hand with discussions on the transparency of calculations and algorithmic decisions. Transparency is essential to maintain and grow the trust of our customers.

Antoine Denoix
Chief Marketing, Digital, Data & Customer Officer
AXA France



Chapter 03

Strengthening the accountability of Artificial Intelligence through transparency



Behind every line of code is a human designer. Therefore, algorithm output often reflects the assumptions of the designer through his or her decisions when writing the code itself. All of which lead to ethical and moral drifts that reflect society's inequalities. At the other end of the spectrum, improvements in self-teaching computers through machine learning may potentially lead to opaque decision-making.

How can we ensure code auditability and/or transparency, which is the only way to guarantee responsibility?

“I believe we have a fundamental right to understand the decisions algorithms make”

Interpretability of machine learning systems is vital, especially when the decisions made by such systems impact on human lives, which is becoming increasingly routine. Indeed, machine learning systems have permeated into multiple fields, including medicine, justice and policy-making. But what exactly do we mean by interpretability in a machine learning context? Paul Ohm, Christophe Marsala and Marcin Detyniecki answer us.

How would you define data science and what are its current and potential applications?

Christophe Marsala: Data science involves the analysis and extraction of information from data to improve the way humans understand or use these data. It is based on approaches and techniques from several domains, including AI, computer science, statistics and mathematics.

There’s a myriad of potential applications for this technology, in a wide diversity of fields, for example helping physicians to diagnose and monitor patient symptoms, or for predictive maintenance to minimize the risk of equipment failure.

Marcin Detyniecki: It’s about utilizing statistics, machine learning and computer science to interpret situations or make predictions based on data from actual phenomena. A good example of data science at AXA is the creation of predictive models based on past customer data, which can automatically flag up the possibility that a certain behavior might be fraudulent.

How is data science linked to AI?

Paul Ohm: The power of AI technology comes from large data sets. To function properly, AI needs large “generic” databases as well as personal data to tailor the service to each individual. To enhance its decision-making, AI needs increasingly accurate databases, or at least more consistent ones, which means more data gathering in a seemingly endless process.

CM: AI benefits from human-based or heuristic-based approaches to solve difficult problems that could not be modeled or solved by classical mathematical approaches.

How does algorithmic bias occur? What can we do about it?

PO: No technology is neutral. All technology is infected with the preferences and biases of its designers. With machine learning, this bias is further compounded by the bias inherent in the datasets used to train the systems. In the AXA-supported project, Playing with the Data, my co-author and I showed how bias can surreptitiously creep into every stage of the machine learning process.

Unfortunately, we are at the very early stages of understanding how to eliminate algorithmic bias. We can certainly scrutinize the output of any machine learning system to look for examples of discrimination and other forms of invidious bias, but it is difficult to spot. Many researchers are exploring the possibility of creating AI systems that are explainable or interpretable. For example, if a computer program determines that I do not qualify for credit, at the very least it should be able to identify the critical factors that led to this conclusion.

"No technology is neutral. All technology is infected with the preferences and biases of its designers."

How would you define a "black box" and what can you tell us about interpretability?

CM: Let us consider a basic medical model, that when given a set of symptoms can specify a set of corresponding diseases that are a possible match for the symptoms, or a model that predicts the likelihood of malign tumors from a mammography. A black box describes a model like this: it provides a valuable

Paul Ohm

Paul Ohm - Law Center | Georgetown University

Professor of Law at the Georgetown University Law Center, Paul specializes in the areas between law and computer science, often focusing on how new technologies affect privacy. His **AXA Award project** involves designing tools for regulators and ordinary citizens to mitigate the power of AI systems. He argues for stricter legal standards for larger software platforms, and is proposing new tools to help determine when AI decision-making should replace human decision-makers.

AXA also funds Paul's work on "desirable inefficiency", which is the addition of tailored inefficiencies to software as a means of protecting important human values such as privacy, trust, dignity, and autonomy.

output but is unable to give additional information to elucidate the basis on which this decision was made. For example, a model can analyze a mammography and predict the presence of a malign tumor with very high accuracy, but it cannot explain its reasoning.

MD: In the past, interpretability has been regarded as a necessary compromise: the price to pay for performance optimization. Let's consider that a predictive model is based on a machine learning algorithm, which is essentially a set of rules. To be effective, any such algorithm has to handle real world complexity with all types of exceptions and special cases. It therefore needs millions of rules. This makes it difficult to design a model that can show the reasoning behind its decisions. To better understand what the model is doing, we would need simpler models, which means less rules and thus less accurate predictions.

PO: Some great advances on interpretability could be on the horizon. For example, some European scholars have high hopes for counterfactual explanations, i.e. machine learning algorithms with the ability to indicate the characteristic a person would need to change to reverse a model's negative result. For example, a system might say: you would have qualified for this loan if your income was 10% higher and you closed one line of credit.

An important point is that some machine learning techniques are more amenable to interpretation than others. For example, deep learning neural nets may be especially difficult to render legible. We may need to consider regulations that make AI researchers focus on the most interpretable approaches.

Christophe Marsala

Christophe Marsala - Computer Lab | University Pierre et Marie Curie

Based at the University Pierre et Marie Curie's Computer Lab, and AXA's Data Innovation Lab, Christophe is involved in a **AXA Joint Research Initiative** to improve understanding on the different facets of interpretability in a data science context. This work focuses on algorithms intended for classification tasks, with the objective to provide the basis for building a new generation of big data and machine learning systems with human-friendly features.

Through his research, Christophe aims to establish new approaches to define and study interpretability in data science at several levels, including global coherence and readability, local validity of components and output consistency.

Do we all need to understand the decisions made by these machines?

MD: When a decision has an affect on human interactions, particularly when it may influence a person's life, such as a medical intervention or a decision about justice, then a machine should be able to explain its decisions. But if the decision is not at that level of importance, then interpretability is not such an issue.

In the insurance context there are several actors that need to understand our AI models. First, the customer should know how the price is set. Second, the regulator may need to understand potential dangers that could lead to discrimination or other problems. Next, the business manager who uses these technologies may want to understand why customer A is made one offer, but not customer B. Finally, the data scientist who designs a model needs to understand if a false positive makes sense or if there's a bug in the algorithm.

CM: When a human is the target of a decision made by a computer program, which is the case in several domains, then it may be unacceptable to blindly follow the decisions of a machine. Explicability and interpretability are therefore crucial properties for these applications.

What has the GDPR changed?

PO: The GDPR is a welcome advance, particularly its provisions around automated decision-making. But we need much more. In this globally interconnected world of data and computation, the GDPR demonstrates how every nation has a role to play in protecting the rights and interests of all people. We are tracked, no matter what we do on the web, but it does not have to be this way.

"I believe we should have the opportunity to react to algorithm decisions."

How do you define interpretability? Why is it important?

MD: In a machine learning context, interpretability is used in relation to whether a model or its predictions can be understood by a human. However, this informal definition is not specific about what is interpretable and to whom. For instance, a mathematical linear model is interpretable to a data scientist, but possibly not to anyone else.

Marcin Detyniecki

Data Innovation Lab | AXA

As the Head of Research at AXA's Data Innovation Lab, Marcin leads innovation projects with a strong focus on advanced analytics. His research, in conjunction with AXA's operational business entities, provides strong technical insights. He plays a key role in AXA's interactions with the academic community, as well as defining AXA's research strategy.

Marcin is working with researchers at the Sorbonne University on a project to equip complex black box models with the ability to provide explanations for each individual prediction. This work has many applications, including in the insurance sector where it will allow consumers to be informed about the key factors driving risk prediction.

Moreover, the goal of interpretability can be very different. For example, for the data scientist it can be to understand if the predictive models she is trying to train is performing correctly; while for the user it can be to understand why the algorithm predicted an insurance policy price increase.

Interpretability is important because the predictions, decisions and actions of algorithms are increasingly impacting on humans. I believe we have a fundamental right to understand these decisions, and also the opportunity to react to them. Furthermore, there is always a risk of a flaw, which a model's interpretability will allow us to correct or counteract.

To what extent is your research fostering more trust in AI?

MD: It is not only about trust. It is also about making AI more responsible, or at least using it in a more responsible way. Research allows us to better understand the risks and even address them before it is too late. My work is looking at providing the means to interpret the predictions of machine learning algorithms, without any trade-offs, which will lead to more human friendly products, whether intended for the end customer or internal business teams.

CM: Explanation, explicability, and understanding are highly important when people are affected by the decisions of computational systems. Beyond our theoretical contributions, our research aims to provide tools, such as algorithms and programing code, to measure interpretability, to improve existing big data processing methods, and to propose new machine learning approaches.

"Explanation, explicability, and understanding are highly important when people are affected by the decisions of computational systems."

Moreover, we aim to conceive and build a new generation of more human-friendly big data and machine learning systems. Such models could facilitate the shift from black boxes to interpretable machine learning, opening up the possibility of understanding and changing the actions of decision-makers and providing new ways to enhance ethical behaviors. This should raise our trust in the output of AI-based models.

PO: My approach is to bridge the gap between technology and law, engaging both sides of the divide with depth and rigor. When it comes to AI, I am imploring scholars to think about the biases and other problems that creep into machine learning at every single phase of the development lifecycle. I am also focusing on what I've termed as "desirable inefficiency", which may point the way to novel regulatory approaches for limiting AI's potential for harm.

“Our advanced modeling could have a huge impact on public policies”

The term “complex phenomena” describes systems that cannot be easily modeled due to the level of detail required to predict their exact behavior in unforeseen conditions. Professor Maurizio Filippone, holder of the AXA Chair on New Computational Approaches to Risk Modeling, is addressing this by applying optimization and inference techniques to fine-tune computational models. The climate is an increasingly pertinent example of such a system: *“We can easily measure climatic factors, such as temperature and pressure, but accurately modeling their behavior on a planetary scale is a huge computational challenge,”* said Maurizio.

Informing public policy

Improving the way we quantify uncertainty with modern statistical models is key to enabling accurate risk prediction. *“The computational statistics and machine learning techniques developed through this AXA Chair will be applicable to numerous fields, such as predicting the location of the next high-magnitude earthquake, examining cancer treatment outcomes, or solving traffic management problems,”* said Maurizio. This work could have major consequences on decision-making by governments and authorities. “*Policy makers already use data analysis to inform their decisions. In the future, our advanced modeling could have a huge impact on public policies,”* he added.

Avoiding black box opacity

With the increased use of computational modeling and machine learning tools by public agencies, legitimate concerns are being raised about the accountability of “black box” systems. *“The last decade has been dominated by models that achieve impressive predictive performance but are not easy to interpret and it is hard to decipher their predicting mechanism,”* explained Maurizio. However, *“The computational statistics and machine learning community has realized the urgent need to improve on this aspect, and there is a growing literature on the topic. I believe that effective solutions to radically improve the interpretability of modern machine learning models will soon be available.”*

Maurizio Filippone

EURECOM
AXA Chair since 2016

“Once you embrace the idea that code is law, then you embrace the idea that code has to be public”

Do we have an outdated vision of privacy in today’s world of big data and Artificial Intelligence? How should we collectively deal with algorithms? Should we allow innovation in data and AI to continue unchecked, or should we try to limit its growth to mitigate risks? Lawrence Lessig, Professor of Law and Leadership at Harvard Law School and AXA Research Fund Scientific Board member, is a world renown expert in deciphering the digital era in which we are living, and also the one into which we are heading.

AI needs copious amounts of data to function properly. At the same time, most of its applications are aimed at gathering more data. Does this never-ending process mean the end of privacy as we know it?

The problem is that we have a very 20th century idea about what is meant by “invasion of privacy”. We imagine an intelligent actor penetrating into a space that we regard as protected, and using what he learns against us. When we carry this conception over into the 21st century world of AI and big data, it terrifies people because all sorts of private activities are now subject to surveillance. Surveillance is not the exception, surveillance is now the norm.

So, what can we do in a world of perpetual persistent surveillance? There are two extreme solutions: one is to say that any surveillance is fine, so long as people agree to it; and the opposite is to prohibit all

surveillance, so companies like Google and Facebook have to stop operating in the way they currently do. I think both extremes are a mistake. The notion of setting policy based on opt-in data agreements is completely crazy: people have no time to even think about what it is they are agreeing to. On the other hand, we are too far down the road to ban companies like Google...

What do you think about efforts to protect data such as the EU's General Data Protection Regulation (GDPR)?

It is very optimistic to think that programs like the EU's GDPR will give people control over their data. Let me be clear: we can certainly empower people to control whether data gets produced, how it gets produced, and how it gets shared. And if this type of control was effective, I might be sanguine about the future. But such barriers will be wiped away by the incentives created by companies to persuade us to waive all of our privacy rights.

So then what? Companies will say that people were given the choice, and these choices should be respected by allowing companies to do whatever they want with this data. That's what I'm rejecting. Choice is fine, but in addition I think it is vital that we begin a conversation about which uses are appropriate, and which are not.

Because in addition to any protection derived from people saying no, we also need governments to say that society considers certain kinds of uses as appropriate, and others as inappropriate. And we need to develop that description of values through a democratic process, to guide what these companies are doing.

"We need governments to say that society considers certain kinds of data uses as appropriate, and others as inappropriate."

I don't say this with an inherent belief in the current capacity of government to make these kinds of decisions, but I believe we have to build the recognition that this is a central part of architecting the right kind of future for life in this digital space. Rational, collective, decision-making is essential. We don't have it now, but we need it or we will be taken over by these machines.

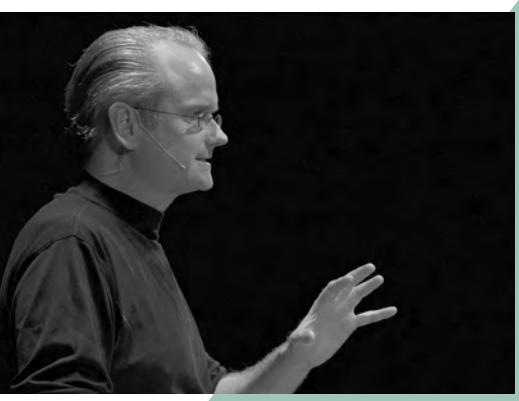
Algorithms are now making decisions on credit applications, health insurance, recruitment, and even justice and police decisions. If "code is law", are autonomous algorithms becoming the new lawmakers?

In the US right now, we are having battles about AI machines fed with data that decide whether or not prisoners get parole. Thus,

autonomous algorithms are indeed becoming new lawmakers.

This development raises important questions. Defense lawyers want to know what information these machines are taking into account to make such decisions. While manufacturers are either claiming not to know because the AI operates by machine learning, so they can't explain how it will develop, or they say that this is proprietary information which they can't be forced to give up in the context of selling products.

Both of these arguments are invalid. We at least need to know what are the values of the machine, so that we can decide if we agree with the way it regulates us. Do we approve or do we want to change it? Do we need interventions to force it to change? At the very minimal level, once you embrace the idea that code is law, then you embrace the idea that code has to be public, because the law that's not public is not law. To regard this information as private defeats a fundamental value of a democratic society: we get to decide on the rules that will constrain, control or enable us.



"We at least need to know what are the values of the machine, so that we can decide if we agree with the way it regulates us."

The first step will be incredibly hard to enforce, because the intellectual property resistance argument is strong, and the basic fact is that developers do not know exactly what their machines are doing. But we can't accept this as the end of the argument. It is simply the beginning of the conversation. The end has got to be that we understand the values embedded in this regulation, and that we approve of it, because if we don't approve of it, then it should not be regulating us.

> **You once said that we're not yet smart enough to understand how to live with AI. Do you think we should carry on innovating in the field of data and AI or apply a "precautionary principle"? In other words, to limit innovation because we don't know the consequences.**

When I say we're not smart enough, I am talking collectively. Smartness is the capacity to make a collective judgment about what makes sense for this AI technology in our society. Now, recognizing that we are not smart enough, we can either try to stop development, or we can become smart – we can develop the capacity to make collective judgments about this technology. I certainly believe in the second option. I don't think trying to stop progress is the solution. The solution should be building the capacity to become smart.

This technology is capable of bringing about a dystopian or utopian future. In the former, the technology takes over and multinational corporations own all of our personal information; while in the latter scenario the technology gives us a much more balanced existence on the planet, where we spend more time doing creative things. We will still work, but work will not define our lives because the machines will work for us.

The fundamental decision about whether we will head towards the dystopian or the utopian version of the future depends on whether we have a government that is capable of making sure that the future benefits of this technology are shared by all of us, as opposed to being owned by the tiniest fraction of people. That choice is fundamental and can only be made sensibly if we have a governing structure that is capable of making the right decision. In Silicon Valley right now people are talking about ways to help bring about this utopia: so, say in fifty years, your children or grandchildren may have all of their basic needs taken care of, such as healthcare, food, and living places. And perhaps they will spend five hours a week working in a coffee shop, or thirty hours a week composing music. What is wrong with that world?

It is a better world than the one in which we currently live. But the only way we can get to this world, where machines work for us and generate enormous wealth, which we can make widely available, is if we can deal rationally with these machines. And by rationally, I mean collectively. So I don't think we should stop progress. I think we should fix what is stopping us from making these collective judgments sensibly.

What place should researchers have in the ongoing discussions?

Ever since I started writing on this subject 20 years ago I've been saying that researchers need to become responsible for the consequences of their technology. Which means they've got to develop the capacity to understand the values their technologies are enabling or disabling, and make that part of the way that the technology is accounted for.

"I don't think trying to stop progress in AI is the solution. The solution should be building the capacity to become smart."

For example, say I develop a technology that filters Internet content. That's not just a technical statement, that's also a political statement, and also a values statement. I think technologists need to understand that they at least have the obligation to make those value choices transparent. If they don't then they are compromising their obligation as citizens. Aerospace engineer Wernher von Braun said that it was not his job to worry about where the bombs came down, but whether they take off. That is sometimes the attitude of technologists today, and that has to change.

“Technologists need to understand that they at least have the obligation to make their value choices transparent. If they don’t, then they are compromising their obligation as citizens.”

> **Do you think the right spaces exist for trans-disciplinary dialogue, to support a creative exchange between technologists, developers, engineers and researchers?**

That’s an important point, and I think that we don’t have enough of these places. We don’t feel sufficiently obligated to speak more than one language. For example, 40 or 50 years ago there was an emergence in the US of something called law and economics, which put economics at the center of understanding how law works. When it first emerged, many lawyers rejected the concept on the grounds that they were experts in law, not economics.

But today, among academic lawyers, if you don’t have a sense of the economic aspect of legal rules, or how such rules interact with economic incentives, then you don’t understand the law. Law and economics have become so intertwined, that you need to have an understanding of the language for both.

It’s the same for technology. Recognizing that technology is itself a law, we have to develop among lawyers an awareness that they need to understand the choices and options, and the range of alternative worlds that are made possible by this technology. They need to understand these things and have the capacity to criticize the fact that one choice was made over another.

For this, we need a cultural change in education. We need law schools and technology universities to train students in a way that they understand the need to speak the different languages of governance, technology and law. If we can become multi-lingual in this sense, then there’s a hope that this sensibility can become a central part of how we understand our roles in the public space. We have to put collective self-governing into the mix and that’s the hardest thing to do.

About
The AXA Research Fund

The AXA Research Fund: Supporting research on the most important issues facing our planet today

Since our inception in 2008, our mission has been to support outstanding researchers who are committed to contributing to some of the most important issues facing our planet. Today, supporting scientific research is more important than ever due to the acceleration of both economic and physical phenomena affecting our societies.

"Today, supporting scientific research is more important than ever due to the acceleration of both economic and physical phenomena affecting our societies."

Going beyond AXA's mission as an insurer to cover and manage risk, through a one-of-a-kind private sector grant program based on independent and rigorous academic criteria, the AXA Research Fund strives to better understand and mitigate the climate, health, economic, and technological risks that mark our lives today and will affect them tomorrow. Through our grantees' work, we seek to work towards providing solutions to develop resiliency and reduce risk in these critical areas.

We also have a firm objective of ensuring that science plays a role in contributing to public debate. The AXA Research Fund provides our partner researchers the tools and network to help disseminate research program findings to a larger audience so as to enlighten decision making for a better future.

Marie Bogataj
Head of the AXA Research Fund

€ 179

millions committed

563

Research projects supported

Researchers of

58

nationalities

in

35

countries

Find out more

axa-research.org

@AXAResearchFund

AXA Research Fund

AXAResearchFundLive

Want to learn more about one specific project? Need to reach one of the researchers supported by the AXA Research Fund?

Please contact us at:
community.research@axa.com

Antonio Acin

ICREA Professor at ICFO | AXA Chair since 2015

Doctor in theoretical physics in the University of Barcelona, Antonio Acin is leading the Quantum Information Theory group at ICFO-The Institute of Photonic Sciences where he holds an AXA Chair on the coming quantum revolution in data security since 2015.

Alexandre d'Aspremont

Ecole Normale Supérieure | AXA Joint Research Initiative since 2014

Alexandre D'Aspremont is working at CNRS and is attached to Ecole Normale Supérieure. His research focuses on convex optimization and applications to machine learning, statistics and finance. In 2014, he leads a joint research initiative with AXA focusing on Machine learning for Large Scale Data Insurance.

Dominique Boullier

Ecole Polytechnique Fédérale de Lausanne (EPFL) | AXA Joint Research Initiative in 2014

Since 2015, Dominique Boullier is senior researcher at the Digital Humanities Institute at the Ecole Polytechnique Fédérale de Lausanne (EPFL). In 2014, he led a joint research initiative between Sciences Po médialab and AXA.

Joanna Bryson

Department of Computer Science, University of Bath | AXA Award in 2017

Dr Joanna Bryson of the Department of Computer Science at the University of Bath was one of those who formulated the set of five Principles of Robotics – ethical rules published by the United Kingdom in 2010. Her research investigating how humans behave around humanoid robots was funded by an AXA Award in 2017.

Robert Deng

Singapore Management University (SMU) | AXA Chair since 2017

Since 2017, Professor Deng is directing the AXA Chair of Cybersecurity at the Singapore Management University (SMU) focusing on insuring data security and privacy protection in the cloud computing environment. He is also professor of Information Systems and director of the Secure Mobile Center.

Maurizio Filippone

Eurecom, Sophia Antipolis | AXA Chair since 2016

Associate Professor at EURECOM, Sophia Antipolis, Maurizio Filippone is also leading the AXA Chair of Computational Statistics. His work aims to better qualify uncertainty when studying complex phenomena in risk modelling through new computational approaches.

Philipp Hacker

Humboldt-Universität zu Berlin | AXA Post-Doctoral Fellowship in 2017

Specialized in law and new technologies, Philipp Hacker was awarded a post-doctoral grant in 2017 for his researches about algorithmic discrimination and exploitation as a challenge for European law. Alongside his work regarding fairness in machine learning, his domains range from the legal implications of Blockchain to those of Behavioral Economics.

Christophe Marsala

University Pierre et Marie Curie | AXA Joint Research Initiative since 2016

Informatic professor at University Pierre et Marie Curie, Prof. Marsala is working within the Laboratoire d'Informatique de l'Université Paris 6 (LIP6). Since 2016, he is directing a joint research initiative with the Data Innovation Lab at AXA that aims to provide the basis to conceive a new generation of big data and machine-learning systems offering a human-friendly feature.

Paul Ohm

Georgetown University Law Center | AXA Award in 2013

Paul Ohm is a Professor of Law at the Georgetown University Law Center. He specializes in information privacy, computer crime law, intellectual property, and criminal procedure. He serves as a faculty director for the Center on Privacy and Technology at Georgetown. His researches regarding big data, privacy, and discrimination won an AXA Award in 2013.

Contributions from AXA experts

Guillaume Beraud- Sudreau

Head of research at Kamet Ventures

Guillaume Beraud-Sudreau is leading an AI-driven venture at Kamet. He graduated from the Télécom Paris, holds a master degree in cognitive science and is member of the French Institutes of Actuaries. Guillaume Beraud-Sudreau has been working for AXA since 2010, and was previously in charge of the R&D topics at AXA Global Direct.

Marcin Detyniecki

AXA Data Innovation Lab

As the Head of Research of AXA's Data Innovation Lab, Marcin Detyniecki leads innovation projects - with a strong focus on advanced analytics - in conjunction with AXA's operational business entities while providing strong technical insights. He plays a key role in animating and leveraging the academic community as well as defining the research strategy for AXA.

AXA Research Guide –
Artificial Intelligence: Fostering Trust

September 2018

Published by The AXA Research Fund
25, Avenue Matignon,
75008 Paris, France
community.research@axa.com

Content, design & artwork
Spintank <</>

This AXA Research Guide is printed in Munken Polar Rough (300 g/m^2) and Magno matt volume (135g/m^2). It is set in Source Sans and Publico typefaces.

This guide was printed by Manufacture d'Histoires des Deux-Ponts (Bresson, France) in 600 copies.

AXA Research Fund

axa-research.org
@AXAResearchFund