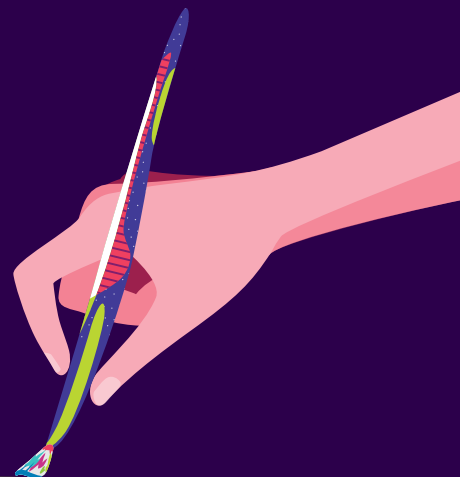


# CONVERSATIONS

with leading CxOs, startups, and academics

Towards Ethical AI



Capgemini   
RESEARCH INSTITUTE



# TOWARDS ETHICAL AI

---

# TABLE OF CONTENTS

**06**

## INTRODUCTION

*Infusing ethics and transparency into AI*

**12**

## INDUSTRY LEADER PERSPECTIVES



**12**

SASKIA STEINACKER,  
BAYER

*Designing ethical and transparent AI for healthcare*



**20**

MICHAEL NATUSCH,  
PRUDENTIAL PLC.

*Taking a human-centered approach to building ethical and transparent AI*



**26**

PAUL COBBAN,  
DBS

*Leveraging the power of ethical and transparent AI for business transformation*



**34**

NICOLAS ECONOMOU,  
H5

*Understanding the role of regulations and standards in creating digital ethics*



**46**

MARCIN DETYNIECKI,  
AXA

*Building robust AI through interpretable AI implementation*

**52**

ACADEMIA PERSPECTIVES



**52**

DANIELA RUS,  
MIT CSAIL

*Building trust in AI-based decision making by understanding the strengths and limitations of machines*



**64**

LUCIANO FLORIDI,  
UNIVERSITY OF OXFORD

*The virtuous circle of trusted AI: turning ethical and transparent AI into a competitive advantage*



**70**

RYAN BUDISH,  
HARVARD UNIVERSITY

*Making AI accountable: Creating openness and transparency in AI*

**78**

VIEWS FROM A TRADE ASSOCIATION



**78**

CECILIA BONEFELD-DAHL,  
DIGITALEUROPE

*Unlocking the true potential of AI through its ethical implementation*

**86**

CAPGEMINI PERSPECTIVE

*Why addressing ethical questions in AI will benefit organizations*

# INTRODUCTION BY THE CAPGEMINI RESEARCH INSTITUTE

---

## ETHICS IN AI – WHAT WE HEARD FROM OUR GUEST CONTRIBUTORS

Every wave of technology innovation raises its own set of ethical questions, from unintended and harmful consequences for users to concerns about technologies being weaponized. To compound matters, the pace of technological advancement tends to outstrip the pace of regulatory and ethical frameworks. This is the position that AI finds itself in today. AI has unleashed a range of ethical questions, from concerns over autonomous vehicles to what constitutes end-user consent. In this Review, we interview a range of experts and practitioners, providing varied perspective on tackling issues of ethics and transparency in AI and the role of guidelines and regulations in this space.

We spoke to leaders from various industries, including insurance, banking, pharmaceutical and life sciences; leading academic experts from Harvard, Oxford, and MIT, and the director of an industry association called DigitalEurope. A number of critical insights emerged, from the importance of shared responsibility to the leading role of diversity:

- **AI algorithms need to be both transparent and understandable.** Recently, a number of expert committees issued guidelines on ethical AI. Examples include the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and the European Union's High-Level Expert Group on AI. These guidelines emphasize how important it is to make AI algorithms transparent if they are to be trusted. And while transparency is critical, it is clear from our interviews that organizations and regulators also need to ensure that these algorithms are understandable and interpretable. *"If you bring in transparency, it will enable anybody to see the rules, but you will not necessarily understand anything – especially if you have millions of them,"* says Marcin Detyniecki, chief data scientist, AXA Group, one of the world's leading insurance and asset

management firms. Nicolas Economou, CEO of H5, an American company using technology to help clients find the right legal information, echoes this sentiment. *"An excessive focus on transparency might confine the discussion to the elites able to understand algorithms, thus deepening the digital divide,"* he explains. *"What we need are broadly understandable and accessible gauges of the fitness for purpose of AI systems."*

- **The ethics of AI needs to be a shared responsibility.** In a world where AI becomes so advanced it takes certain business decisions, it is important not to abdicate responsibility, especially if a decision raises ethical concerns. *"In the case of ethics, this is not something where responsibility lies with any particular individual in the company,"* says Michael Natusch, global head of AI at Prudential Plc., a British life insurance and financial services company. *"It is a shared responsibility for all of us."* Harvard University Professor Ryan Budish agrees that even when specific roles are created, responsibility is still shared in areas such as privacy. *"Everyone in an organization has an obligation to respect the privacy of customers or to protect their data,"* he points out. *"Certainly, organizations have created positions like chief privacy officer to help ensure that the right policies and systems are in place. But the responsibility itself lies with everyone."*

- There is a need to balance legislation with self-regulation to avoid stifling innovation.** Guidelines and regulations give society reassurance and increase consumer trust in new technologies. However, regulators must be cautious about over-regulation as it can stifle innovation. Nicolas Economou emphasizes that cultural context will play an important role in the balance between self-legislation and regulation, *“Like in so many other technological domains, a combination of industry-driven endeavors and regulation will prevail,”* he says. *“The balance between these is likely to depend on the societal context.”* Paul Cobban, chief data and transformation officer, a multinational banking and financial services group headquartered in Singapore, believes that uniformity in regulations will be key. *“The other challenge around regulation is that in an increasingly connected world, regulations in one part of the world differ from those in other parts,”* he says. *“Regulators have a duty to collaborate among themselves and have some kind of baseline approach to this.”*
- Team diversity is an important tool to tackle bias in AI.** A significant challenge in creating ethical AI is ensuring biases are removed from input data. A study from investigative newsroom ProPublica<sup>1</sup> outlined how a computer algorithm used for predicting future criminals was biased against people of African American descent. Another study found that a natural language processing algorithm tended to play up the association between professions and gender stereotypes, thanks to the training data used for it.<sup>2</sup> Our interviews emphasized how having a diverse team can be an effective tool in tackling bias and

creating ethical AI. *“Diversity in every way – ethnic, gender, sexual orientation – are all very important,”* says Michael Natusch. *“It is not just about modeling accuracy, but also about asking the right questions and doing things that are culturally sensitive.”* Ryan Budish, assistant director for research at the Berkman Klein Center for Internet and Society at Harvard University, emphasizes the importance of diversity across the board, from developers to datasets. *“We need greater diversity in terms of the people who are developing the technologies,”* he says. *“We need more diverse datasets to go into developing those technologies.”*

## ETHICS IN AI – THE CAPGEMINI RESEARCH INSTITUTE REPORT

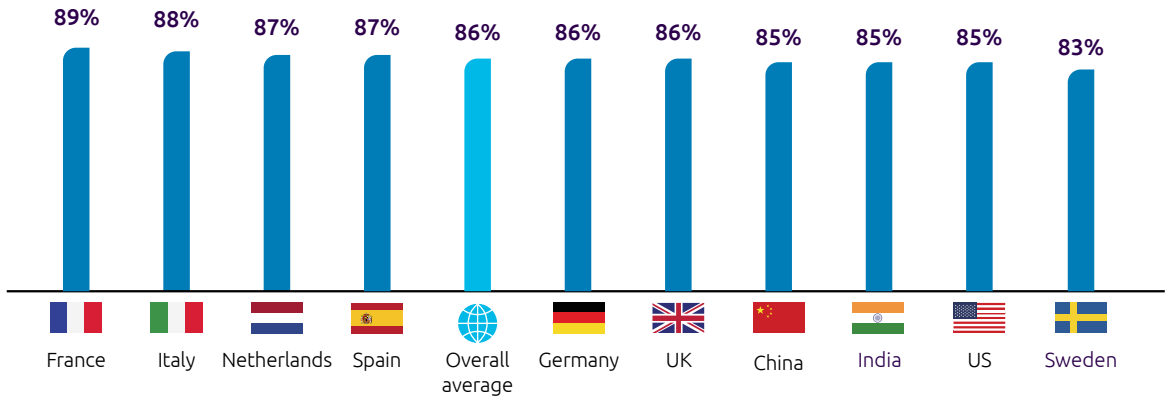
In this edition, we also present the key findings from our extensive global research into this subject, which gathered the views of over 1,500 executives and over 4,400 consumers to understand the case for addressing ethical questions in AI. We found that consumers and citizens are more trusting and loyal to those organizations where their AI-based interaction is seen as ethical. On the flip side, it is also clear that an interaction seen as not ethical will significantly dent a company’s reputation. Our research shows clearly that these issues are currently prevalent. As the chart below shows, the vast majority of organizations worldwide – 86% - have experienced ethical issues with AI.

1. ProPublica, “Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks,” May 2016.  
 2. Google Developers, “Text Embedding Models Contain Bias. Here’s Why That Matters,” April 2018.



**Figure 4.** Nearly nine in ten organizations across countries have encountered ethical issues resulting from the use of AI

**In the last 2–3 years, have ethical issues relating to use and implementation of AI systems, been brought to your attention?** (percentage of executives, by country)



We presented over 40 cases where ethical issues could arise from the use of AI, to executives across sectors. We asked them whether they encountered these issues in the last 2–3 years.

Source: Capgemini Research Institute, Ethics in AI executive survey, N = 1,580 executives, 510 organizations.

However, executives are starting to realize the importance of ethical AI and are taking action when ethical issues are raised. Fifty-one percent of executives believe that it is important to ensure that AI systems are ethical and transparent. We also found that 41% of senior executives report that they have “abandoned an AI system altogether when ethics concerns were raised.”

Our research examines these developments in more detail and offers perspective from experts and practitioners on the actionable steps that organizations can take in response.

We hope this edition offers both insight and pragmatic ideas for action. Please reach out to us if you would like to discuss any of these topics and the implications for your organization.

Please visit us at [www.capgemini.com/researchinstitute/](http://www.capgemini.com/researchinstitute/) or write to us at [research@capgemini.com](mailto:research@capgemini.com).

You can also subscribe to our research here: <https://www.capgemini.com/dti-report-subscription-2018-preview/>

# LEADERS AND EXPERTS INTERVIEWS

## United Kingdom

**Michael Natusch**, *Prudential Plc.*, London  
**Luciano Floridi**, *University of Oxford*, Oxford

## France

**Marcin Detyniecki**,  
*AXA*, Paris

## United States

**Nicolas Economou**, *H5*, New York  
**Saskia Steinacker**, *Bayer*, New York  
**Daniela Rus**, *MIT CSAIL*, Massachusetts  
**Ryan Budish**, *Harvard University*, Boston



Belgium

**Cecilia Bonefeld-Dahl**, *DigitalEurope*, Brussels

Singapore

**Paul Cobban**,  
*DBS*, Singapore





**SASKIA STEINACKER,**  
Global head of Digital  
Transformation



## DESIGNING ETHICAL AND TRANSPARENT AI FOR HEALTHCARE

**S**askia Steinacker is the global head of Digital Transformation at Bayer. She works closely with the Digital Transformation Board of the Group, which is composed of the three divisional heads and board members, the chief financial officer, the chief information officer, plus the digital leads. She has played a key role in developing the company's digital agenda with a focus on new business models to accelerate growth. She is also a member of the High Level Expert Group on Artificial Intelligence formed by the European Commission.

The Capgemini Research Institute spoke with Saskia to understand more about designing ethical and transparent AI in the context of the healthcare sector.



## DEFINING ETHICAL AND TRANSPARENT AI

### **What are your key responsibilities as they relate to AI at Bayer? Could you also talk about your role in the EU HLEG?**

I lead the internal board that is driving Bayer's digital transformation, which basically means transforming the value chain in all our divisions. Within Life Sciences, our focus areas are healthcare and nutrition and we see tremendous opportunities for AI in these areas. In particular, it is about developing digital health solutions as well as expanding the digital farming business. AI can help us to better fight diseases, such as cancer or strokes, and also feed a growing world population more sustainably.

Artificial intelligence is a key technology and its impact goes far beyond our business. Growth in computing power, availability of data, and the progress made in algorithms have turned AI into one of the most powerful technologies of our time. This power can be used for the good or for the bad. There are good reasons for concerns about self-determination and data privacy, as well as the impact on jobs and established business models. Risks and opportunities must be discussed in a broad social dialog and ethical questions must be taken into consideration. Trust in new technologies can only be gained by providing an ethical framework for their implementation. This is why I'm part of the EU Commission's High-Level Expert Group (HLEG) on AI: to contribute to the development of such an ethical framework. This is what I stand for and this is what Bayer stands for.

"I DON'T BELIEVE THAT ETHICS SHOULD BE DEFINED BY A SINGLE COMPANY. TO DEFINE AN ETHICAL FRAMEWORK IS A TASK FOR AN ENTIRE SOCIETY AND SHOULD EVEN BE DISCUSSED AT A SUPRANATIONAL LEVEL, AS DIGITAL TECHNOLOGIES DON'T CARE MUCH FOR NATIONAL BORDERS."

**How does Bayer define ethics and transparency in AI? If there is a definition, how did you arrive at it? If not, are you working on building a definition/guideline to create a common understanding on the topic at your firm?**

I don't believe that ethics should be defined by a single company. To define an ethical framework is a task for an entire society and should even be discussed at a supranational level, as digital technologies don't care much for national borders. This is what makes the EU approach so compelling. We have different stakeholders with completely different backgrounds in the Expert group: from industry, society, and academia. This reflects the diversity of our society and gives us varied perspectives.

The "Guidelines for Trustworthy AI" we developed with the HLEG address a number of relevant key requirements. Trustworthy AI needs to be lawful, ethical, and robust with the aim of maximizing the benefits and minimizing the risk. The requirements AI systems need to meet are the following: human agency and oversight; technical robustness and safety, which includes accuracy, reliability and reproducibility; privacy and data governance; transparency; diversity, non-discrimination and fairness, which includes areas like avoidance of unfair bias; societal and environmental wellbeing; and, finally, accountability, which includes areas such as auditability.

The EU developed these guidelines and is currently in the piloting phase of the assessment list. This is a tool that will help companies to practically implement the guidelines. Once the pilot phase is completed at the end of 2019, the final assessment list will be published at the beginning of 2020, and companies should adapt their own guidelines accordingly.

**DESIGNING ETHICAL AND TRANSPARENT AI**

**When implementing AI, why do you think it is important to follow ethical and transparent practices in design? Do you receive questions from your clients about this?**

Acceptance of new technologies requires trust, and trust requires transparency. This holds especially true in critical areas such as healthcare and nutrition. With AI, we have the chance to shape a technology in a way that it is socially accepted and beneficial for the individual as well as for society as a whole. We have to admit that people have concerns with regards to self-determination, data privacy, as well as effects on the job market and current business models. These concerns have to be taken into consideration despite the excitement for new scientific opportunities through AI.

**How do we address the issue of ownership in AI? Who is responsible if an AI system makes a wrong diagnosis?**

Our goal in healthcare is not to let AI take decisions, but to help doctors make better decisions. AI has its strengths – analyzing huge amounts of data and generating insights that a human being wouldn't have thought of before. It is able to identify certain patterns, such as radiological images, and supports the diagnosis of a doctor. AI is meant to enhance or augment the capabilities of humans. How AI is actually leveraged within the healthcare system has to be defined by the different players and ultimately by the regulators.

“MOVING TOO QUICKLY AND CREATING OVERREGULATION WILL CERTAINLY MAKE MANY PLAYERS SHY AWAY FROM INNOVATION IN THE FUTURE, MEANING THE FUTURE WILL HAPPEN ELSEWHERE.”

## EXPERIENCE AT BAYER

**Do you have a defined governance framework for tackling ethical issues at Bayer? How do you ensure that relevant teams in your organization are aware and responsive to issues of ethics and transparency in AI?**

We do have our corporate values, and also well-established internal compliance systems, as is always the case in highly regulated industries such as ours. It is early days in the implementation of AI in our sector and we are one of the first companies to test the assessment list that supports the guidelines for trustworthy AI, focusing on a concrete lighthouse case in pharmaceuticals. It's a project where we try to help doctors identify patients whose cancer is likely the result of a special gene fusion in their tumor cells. It's important to know this if you are to choose the right treatment – this is about precision medicine.

**Bayer offers training programs to educate employees on topics such as AI or blockchain. Is there also a need for organizations to train employees on the topic of AI ethics too?**

Absolutely. We have regular global webcasts on AI topics and our ethics sessions had a full house. We don't have a full-fledged and dedicated AI ethics training program yet, but this could be developed once the assessment list on EU level is final and can be used. This will be helpful for employees who develop, implement, or use AI.

## REGULATIONS IN AI

**Do you see regulations as necessary for implementing ethical AI or is self-regulation the way to go?**

I think we need a common framework first, which is binding for all players. Then an area-specific form of self-regulation could make sense. But it's always about finding the right balance: It wouldn't make sense to have a regulation in place that would make it impossible to develop AI solutions here in Europe.

**Should there be GDPR-like regulation in this area? How can you build the right regulation practices that don't stifle innovation?**

This is exactly the balance that is discussed in the HLEG. If we figure that out right, Europe could be leading with ethical AI. Given the magnitude of the AI revolution ahead of us, there needs to be a certain degree of regulation with a special focus on ethical questions. This type of regulation needs to be binding for all players in a given market, and ideally, worldwide. At the same time, there is already an abundance of regulations that govern many aspects, such as data privacy. However, they are not always fit for the AI age, and AI brings a number of new ethical aspects to the table. We need a very broad discussions at all levels of society, including the industry that is expected to develop these new solutions. Moving too quickly and creating overregulation will certainly make many players shy away from innovation in the future, meaning the future will happen elsewhere.



## RECOMMENDATIONS AND OUTLOOK

### **What are the main risks facing an organization that does not take ethics in AI seriously?**

Apart from immediate consequences – for instance, not being able to sell a solution in an increasingly ethics-conscious world – image and reputation loss are probably the most apparent immediate results. However, what is probably most important are the consequences, which reach beyond a single solution and a single company. With the AI solutions we build today, we will shape the future we as a society and as individuals will have to live in. So, you could even say that there's the greater good at stake, and it starts with the people who build, use and deploy AI solutions today.

### **How can organizations make their AI systems transparent, ethical, and bias free? What concrete steps are necessary for this?**

To help companies develop ethical and bias-free AI systems is exactly the aim of the HLEG guidelines. If you follow the key requirements, which are translated for practical use in the assessment list, you are able to develop, deploy, and use a trustworthy application. Additionally, companies should sensitize and educate their employees and take ethical aspects into consideration right at the beginning of an AI project. As with other topics, it's always necessary to have diverse teams.

### **What is the one AI ethics policy that you would want every organization to adopt at the minimum?**

There are many essential topics to consider for ethics in AI, and it's challenging to prefer one over the other. Each of them has been long discussed in the HLEG. As a starting point, the most important aspect to me is the context of AI usage – it is certainly different if you use an application for a ticket machine versus a complex healthcare diagnosis system.

“WITH THE AI SOLUTIONS WE BUILD TODAY, WE WILL SHAPE THE FUTURE WE AS A SOCIETY AND AS INDIVIDUALS WILL HAVE TO LIVE IN.”

---

## ETHICS GUIDELINES FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE

In April 2019, the High-Level Expert Group on Artificial Intelligence set up by the European Commission, released Ethics Guidelines for Trustworthy AI. According to the guidelines, “Trustworthy AI has three components, which should be met throughout the system’s entire life cycle: (1) it should be lawful, complying with all applicable laws and regulations (2) it should be ethical, ensuring adherence to ethical

principles and values and (3) it should be robust, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm.” The guidelines also list seven requirements that should be kept in mind while developing trustworthy AI. These requirements are: “(1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability.”<sup>3</sup>

3. European Commission, “Ethics guidelines for trustworthy AI”, April 2019.

## WHY DO WE NEED ETHICAL AI?



“In a system where a machine makes a decision, we want to make sure that the decision of that system is done in a way that ensures people’s confidence in that decision.”

.....

– Daniela Rus, MIT CSAIL



“We need to ensure that AI is acting in such a way that we can hold it accountable and also respond if we determine it is acting in a way that we don’t believe to be consistent with our values and/or laws.”

.....

– **Ryan Budish, Harvard University**



“Trust in new technologies can only be gained by providing an ethical framework for their implementation.”

.....

– **Saskia Steinacker, Bayer**



“Algorithmic systems are amoral... they do not have a moral compass. Yet, they can make decisions that have pervasive moral consequences.”

.....

– **Nicolas Economou, H5**



MICHAEL NATUSCH,  
Global head of AI

## TAKING A HUMAN-CENTERED APPROACH TO BUILDING ETHICAL AND TRANSPARENT AI

**M**ichael Natusch is the global head of AI at Prudential Plc. and also founder of the AI Center of Excellence in Prudential Corporation Asia.

With over 20 years of experience in data analytics and machine learning, he enjoys working with data and leading-edge statistical methods to tackle real-world problems, which today means applying machine learning and neural networks to large-scale, multi-structured data sets.

The Capgemini Research Institute spoke with Michael to understand more about creating an ethical and transparent AI and the technological challenges involved.



“WE ARE TRYING TO BUILD SOMETHING THAT EITHER REPLACES OR COMPLEMENTS AN EXISTING PROCESS WITH AN AI SOLUTION. SO, PEOPLE ARE ASKING – “HOW DO YOU ACTUALLY MAKE A DECISION?” ONE ASPECT OF THE “HOW” IS OBVIOUSLY AROUND ACCURACY.”

## ETHICS AND TRANSPARENCY IN AI AT PRUDENTIAL

### What is the model you have deployed to scale AI at Prudential?

At Prudential, we have both a centralized and a localized model. I am a big believer that a centralized-only or a localized-only model would be doomed to fail. In the former, you would find people who build amazingly clever things that nobody ever wants to implement. And in the latter, you would find people who would spend literally all their time on minute process improvement without ever being able to truly reinvent the business and move beyond sub-optimization.

So, we want to have some centralized capability as that brings efficiency in terms of being able to copy-paste approaches to different countries and the ability to hire AI experts. But, to supplement that, we need to have localized capability. If you only have one, then it is not going to work very well.

### How do you define ethics and transparency in AI at Prudential and what is driving action in the organization?

We do not have a working definition. Our position around AI and ethics is still evolving. We are still in the process of formulating as to what the position of the company is and what that means in practice. We have a program of action that, by the end of this year, we hope to have clearer views of where we stand as a company around ethics, AI, data, and all the associated aspects of transparency, privacy, and compliance.

### Why is it an important issue for Prudential?

There are three different strands that lead us to take this issue seriously. One is that there is an overarching conversation in society. For instance, our regulators are starting to look at it. The Monetary Authority of Singapore has published a paper called FEAT, which lays out some very basic principles. So, our vital stakeholders, our regulators, and even our board members, have thoughts and questions.

The second strand comes from our business. We are trying to build something that either replaces or complements an existing process with an AI solution. So, people are asking – “how do you actually make a decision?” One aspect of the “how” is obviously around accuracy. Are

you making the right decision? What is your false positive rate? What is your true positive rate? Those kinds of questions. The second aspect to that is transparency. Can I, as an employee, understand it? If challenged by a regulator or customer, can I justify the decision that has been made? The question that employees also need to ask themselves is, “am I making the right decision?” Even though the decision might be precise and transparent, it might still be the wrong decision. And that has a legal and ethical component to it. So, for instance, am I explicitly or implicitly discriminating against a particular demographic?

The third and final aspect is that we believe that ethical and transparent AI will be a competitive differentiator for us in the marketplace. We have a unique opportunity to seek consumer trust and a short window of time to realize this opportunity. We should demonstrate to people that they can trust us. And they can trust us not just in the world of the 1990s or the early 2000s, but they can also

“WE RECENTLY LOOKED AT FACIAL RECOGNITION, IN TERMS OF IDENTIFYING THE KIND OF DIAGNOSTIC ASPECTS THAT WE CAN READ FROM A SELFIE.[...] AND WHAT CAME OUT CLEARLY WAS THAT WHILE THE PRETRAINED MODEL WORKED ALMOST PERFECTLY ON SOME OF OUR TEAM MEMBERS, IT DID NOT WORK AT ALL ON OTHERS.”

trust us going forward because we will deal with their data in the right way. We will not take their privacy for granted, we will not misuse their personal data, we will not infer things about them from the data that we have that they would consider inappropriate. By being cautious and doing the right thing by our customers, we hope to differentiate ourselves in the marketplace.

### **Have you ever experienced any ethical issues in AI systems that you have deployed?**

We recently looked at facial recognition, in terms of identifying the kind of diagnostic aspects that we can read from a selfie. So, we started with some pre-trained models. And what came out clearly was that while the pre-trained model worked almost perfectly on some of our team members, it did not work at all on others. And it did not take a great genius to realize what was going on. For Caucasians, the model came out with the correct age, but people of South Asian origin tended to be estimated as being older than they were. People of East Asian ethnicity were estimated as being significantly younger than they were. So, even with this sort of five-minute playing around – and without doing anything really sophisticated – you realize that you cannot just bluntly apply pre-trained models using an off-the-shelf algorithm. There must be feedback in the middle. So, this is one simple and trivial example of that third aspect in our own work – where we became aware of ethical issues and what we need to do to attack these ethical issues head on.

### **ROLE OF DIVERSITY AND AN ETHICAL CODE OF CONDUCT**

#### **How important is the diversity of AI teams when identifying potential biases?**

Diversity in every way – ethnic, gender, sexual orientation – are all very important. It is not just about modeling accuracy, but also about asking the right questions and

“YOU HAVE TO HAVE HUMANS WHO EVERY NOW AND THEN LOOK AT WHAT IS COMING OUT, RE-THINK IF WE ARE DOING THE RIGHT THINGS, AND THEN ADJUST THE MODEL”

doing things that are culturally sensitive. I think diversity in every day interactions is extremely important for an AI team, because you are not going to ask yourself questions that somebody from a different background would come up with.

### **Does Prudential already have an ethical code of conduct and does AI feature in it?**

There is and it goes back quite a long time. What we are going through right now is translating it for the AI world. We are taking those principles, adapting them to AI, and extending them from an AI point of view. Hopefully, by the end of this year, we will get to a much more holistic, all-encompassing, ethical framework that is applicable across everything that we do.

In a low-scale, largely manual world, you can do things at a fairly slow, straightforward, manual manner. The ethical component is manageable because you can achieve that by training and very limited remedial actions. In a world that is dominated by AI, and where you work at scale, if you do something wrong, you do something wrong on a huge scale. And therefore, you need to be much more careful regarding ethics, transparency, privacy, and compliance. All these need to be incorporated by design right from the start. And that requires a very different way of working and thinking. Therefore, purely from an ethical point of view, the way we choose products and run processes in an AI-dominated world must be done in a very different way.

## ETHICS BY DESIGN

### **What does ethics by design mean in your business?**

Ethics by design has three different aspects. One, it is about mindset. As much as we want to move fast, we cannot afford to break things. And that is a mindset thing. The second is about automated and continuous, software-enabled checks. Are we doing the right things? Is there something that is coming up that that looks unusual? And that then leads to the third piece which is that, sooner or later, every model will misbehave. That is just a fact of life. So, based on the second step, you also need to have a level of human control. You have to have humans who every now and then look at what is coming out, re-think if we are doing the right things, and then adjust the model.

## ENSURING AWARENESS AND RESPONSIBILITY FOR ETHICS IN AI

### **How do you ensure that the relevant teams are aware and responsive of ethics and transparency in AI?**

We have some really smart and empathetic people in the AI Center of Excellence. So, we have an understanding of the kind of biases that we need to watch out for. But, what I am really hoping for, is two things. I am looking for validation and completeness, and additions from the overall process that I described earlier. And the other



thing that I am looking for is a checklist of things that need to be done less frequently, maybe just at the inception of a particular type of activity, and so on. Some of the checklists might be literal, whereas others might be more intangible. But those are the two kinds of things that I am hoping to get out of this effort, which will help us to supplement our own limited understanding around ethical issues and how to present them.

### **Where should the responsibility lie if some systems do not act the way they should?**

The seat of responsibility will not shift. Ultimately, the people who are accountable for what is happening in Prudential are the chairman and the CEO of Prudential. Our shareholders would ask, "Why did you not prevent this?" So, that will not change. In the case of ethics, this is not something where responsibility lies with any particular individual in the company. It is a shared responsibility for all of us. My team and I are cogs in the wider machinery. We are not the only ones. There are other people who have their part to play as well. It is a shared activity in every sense.

## **TECHNOLOGICAL CHALLENGES IN ACHIEVING ETHICAL AI**

### **What are the technological challenges with respect to achieving ethics in AI?**

It is essentially about applying the right type of technology in the right manner and for the right problem. For this, I actually have a framework in my mind which has two different axes. One axis is the volume axis, be it data points, the volume of transactions, or the volume of events. And the other axis is a value axis. And so, if you look at that space of volume versus cost of making the wrong decision, there are two extreme points that you can immediately identify. One is extremely high volume, extremely low cost.

A good example of that is doing a Google search. So, with 3.5 trillion Google searches a day, what is the cost of Google showing you the wrong ad on one of those searches? It is obviously virtually zero – the impact is minimal. And then, there is the other extreme. For instance, you are in a hospital, and you have a cancer patient, and you need to decide about the radiation dose for radiation therapy for that patient. Clearly, the volume is much lower, but the cost of making the wrong decision can be extremely high.

And then, you have kind of a gray area in the middle. And everything that we do in terms of the kind of algorithms and technology we use, and what kind of considerations we need to get to, depends on where you are on this chart. In the high-volume, low-impact scenario, there are no real ethical considerations there because the impact is so low. On the other extreme, you need to think very hard about what to do. Regulators need to look very hard at what is happening there so that they protect the consumers or whoever they are serving.

**“THE SEAT OF RESPONSIBILITY WILL NOT SHIFT. ULTIMATELY, THE PEOPLE WHO ARE ACCOUNTABLE FOR WHAT IS HAPPENING IN PRUDENTIAL ARE THE CHAIRMAN AND THE CEO OF PRUDENTIAL.”**

## FEAT PRINCIPLES<sup>4</sup>

The Monetary Authority of Singapore released guidelines “to promote Fairness, Ethics, Accountability, and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics (AIDA).” The objectives of the principles are:

1. “To provide firms providing financial products and services with a set of foundational principles to consider when using AIDA in decision making
2. To assist firms in contextualizing and operationalizing governance of use of AIDA in their own business models and structures
3. To promote public confidence and trust in the use of AIDA.”

Fairness encompasses the justifiability of decisions taken by AIDA systems and reviewing AIDA decisions for accuracy and removal of bias. The second pillar, ethics, involves ensuring that AIDA decisions are “aligned with the firm’s ethical standards, values, and codes of conduct” and “held to at least the same ethical standards as human-driven decisions.” Accountability refers to ensuring both internal as well as external accountability. Lastly, transparency involves data subjects being “provided, upon request, clear explanations on what data is used to make AIDA-driven decisions about the data subject and how the data affects the decision.”

4. Monetary Authority of Singapore, “Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore’s Financial Sector”, November 2018.

## WHAT ARE THE CHALLENGES THAT ORGANIZATIONS ARE FACING REGARDING ETHICAL AI?



“One big challenge is to transform a societal wish into a mathematical formulation. Defining fairness mathematically is quite a challenge.”

– Marcin Detyniecki, AXA



“It is essentially about applying the right type of technology in the right manner and for the right problem.”  
.....

– Michael Natusch, Prudential Plc.



“I think the biggest challenge right now is the information asymmetry that exists between the people who are creating these AI technologies and the people who need to decide whether to use them, and how to use them.”  
.....

– Ryan Budish, Harvard University

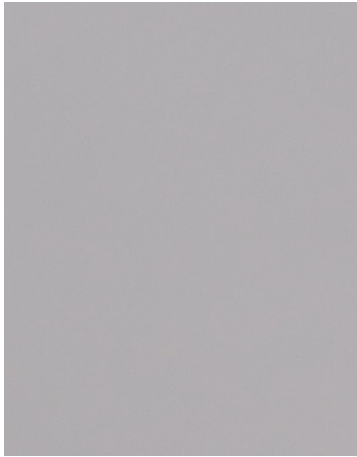


**PAUL COBBAN,**  
Chief data and  
transformation officer

## LEVERAGING THE POWER OF ETHICAL AND TRANSPARENT AI FOR BUSINESS TRANSFORMATION

**P**aul Cobban is the chief data and transformation officer at DBS, a multinational bank with total assets worth SGD 551 billion. The bank has won plaudits as the “World’s Best Digital Bank.” He chairs the Future Enabled Skills workgroup of the Institute of Banking and Finance and he is also a member of both the Institute of International Finance’s Fintech Advisory Council and the Technology Roadmap Steering Committee of the Infocomm Media Development Authority.

The Capgemini Research Institute spoke with Paul to understand more about the role of ethical and transparent AI in driving business transformation.





## AI AT DBS

**DBS has been recognized as one of the world's best digital banks. Did AI have a role to play in this transformation, and to what extent do you believe you have been able to leverage AI for business transformation?**

Our transformation has been 10 years in the making. In the early phases, AI was not part of the story, but it is definitely playing a critical role now. Going back five or six years, we partnered with A\*STAR, which is the government's research and development arm in Singapore. Through the partnership, we learned how to make use of our data in non-traditional ways. They taught us how to predict when ATMs are going to fail or which one of our branches is going to have the next operational error. Then we broadened those use cases and started using data to predict when our relationship managers are likely to quit, so that we can put in interventions.

Last year, we introduced an AI-chatbot to help our HR teams recruit and do a first round of interviews. We have seen a significant increase in productivity, mainly around augmenting people's jobs and making them easier.

## DEFINING ETHICAL AND TRANSPARENT AI

**How have you arrived at a definition of ethics and transparency in AI at DBS?**

The Monetary Authority of Singapore (MAS) issued a document on this called FEAT, which stands for "Fairness, Ethics, Accountability, and Transparency." We used that as a foundation for our own internal variant, PURE, which stands for "Purposeful, Unsurprising, Respectful, and Explainable." This was the foundation for the process we put in place to assess our data use cases. It is broader than just AI – it is about the use of data, and AI is a subset of that.

Talking about the PURE descriptors, the first idea about being purposeful implies that we should not collect data just for the sake of collecting data. Instead, we should have a very concrete purpose for doing so – with the intent of making the lives of our customers better. The way in which we use the data should not shock our customers, and it should be unsurprising to them. Respectful refers to how we should not invade the privacy of people without good reason. At the same time, we are also very mindful of the fact that there are certain use cases, such as fraud and criminal activity, where you have to have a balanced approach.

There are increasing expectations from customers that any decision that is made using an algorithm needs to be explainable, and the MAS guidelines are very clear that the explainability and accountability of a decision need to lie with a human being at some point.

We recognize this as a very nascent area, and we will need to continue to iterate as we learn.

## THE BUSINESS OWNER OF THE ALGORITHM IS ACCOUNTABLE

### Do you have a defined governance mechanism for tackling ethical issues in AI?

Yes, it is all based around the PURE concept. We have a process where everybody who is using data for a specific use case needs to do a self-assessment against the PURE principles. The vast majority of use cases are innocuous and do not need any formal assessment. Anything that triggers any of the PURE principles then goes to a PURE Committee, which I co-chair along with one of my colleagues from the business unit. Those use cases are then presented and discussed at the PURE Committee for ratification. They are then either approved or a mitigating control will be put in place to make sure that we do not trigger any of the PURE categories.

### When issues do arise with AI, where do you think accountability and responsibility should lie?

We don't have any issues yet, but we have plenty of questions. For example, what is surprising to you may not be surprising to me. And, what is surprising to me today may not be surprising to me tomorrow as things evolve and people get used to things. Nothing here is black and white. There is a lot of judgment at play, especially in these early days of AI. However, accountability needs to be very clear. So, we are in the process of compiling an algorithmic model inventory, which means we "inventorize" every model in the company and ensure there is an owner associated with that model – and, that owner is accountable for the decisions that model makes. It is therefore important for that individual to be conversant enough with advanced analytics depending on the model and know how it operates.

“THERE ARE INCREASING EXPECTATIONS FROM CUSTOMERS THAT ANY DECISION THAT IS MADE USING AN ALGORITHM NEEDS TO BE EXPLAINABLE, AND THE MAS<sup>5</sup> GUIDELINES ARE VERY CLEAR THAT THE EXPLAINABILITY AND ACCOUNTABILITY OF A DECISION NEED TO LIE WITH A HUMAN BEING AT SOME POINT.”

5. The Monetary Authority of Singapore

The other thing to note here involves the use cases of the model, as not all are sensitive. So, for example, we use algorithms to predict which one of our ATMs might have the next mechanical failure, but that is not very contentious. If the model gets it wrong, the worst that can happen is that the ATM can have an outage. However, if you are assessing people for credit, that is a different issue. You must have a judgment call around it and that is where some of the complexities are.

### **You mentioned ownership of these algorithmic models – could you tell us who the owner usually is?**

It depends on the model. Typically, it is the individual who is making decision before the algorithm. If I am responsible for the uptime of ATMs and I want to improve that, I will create an algorithm that helps me do it, and I will be accountable. The accountability and responsibility do not lie with the data scientist who develops the algorithm. The business owner in question needs to understand enough about the model to take on that accountability.

### **How do you ensure that all the relevant teams are aware of, and are responsible for, ethics and transparency issues in AI?**

We have a substantive training and awareness program called DataFirst. We also have various big data and data analytics training programs, and we have trained half the company on the basics of data in the past 18 months. Through these programs, we have equipped 1,000 of our employees to become data translators. Our senior leaders have also undergone specialized data courses.

## THE ROLE OF HUMANS IN ETHICAL AI

### **Current technology is not fully geared to deal with all issues – for example, explainability, bias, etc. So, how far can AI solve its own problems today?**

In the short term, we are seeing a remarkable acceleration in tools that can adjust bias and non-explainability. It is not simply a case of waiting for technology to solve all the problems – it comes down to human judgment to make the call. Going back to my previous example of ATMs, we found that ATMs in the west of Singapore break down more frequently than in the east. This is not of any concern, but if my credit algorithm was biased towards one gender, then that is a cause for concern. We always need that judgment overlay.

“WE HAVE A SUBSTANTIVE TRAINING AND AWARENESS PROGRAM CALLED DATAFIRST. WE ALSO HAVE VARIOUS BIG DATA AND DATA ANALYTICS TRAINING PROGRAMS, AND WE HAVE TRAINED HALF THE COMPANY ON THE BASICS OF DATA IN THE PAST 18 MONTHS.”



“CREATE A CROSS-FUNCTIONAL TEAM TO DO TWO THINGS – DO SOME EXTERNAL RESEARCH ABOUT WHAT IS RELEVANT WITHIN THE INDUSTRY AND BEYOND, AND, LOOK INTERNALLY TO FIND OUT WHAT IS BEING DONE WITH DATA AND DEFINE HOW QUICKLY YOU NEED TO ACT.”

**Do you believe that there has to be a human in the loop for all the AI systems before they make consequential decisions about people?**

If you look at autonomous cars, by definition, there is no human in the loop. So it is only a matter of time before AI will increasingly act on its own. But that is when you really have to pay attention to what is going on. For example, as we have seen with algorithmic trading, it can cause a massive shift in the market.

**THE NEED FOR A BALANCED APPROACH WHEN IT COMES TO REGULATION**

**Do you see regulations as necessary for implementing ethical AI or is self-regulation the way to go? In the latter case, can companies be trusted with building ethical AI practices in the absence of agreed standards?**

This is a challenging question. We have seen how unregulated big-tech companies, in the opinions of most people, have crossed the line. However, we have also seen where regulations with data have gone too far too quickly and have had negative, unintended consequences. The approach MAS is taking is sensible – it involves discussing the issues across the industry, putting in place some guidelines initially, and getting feedback to see how that operates before we cement any regulation.

You also have to think about the balance between the rights of the individual and the rights of business, and where you want to play. One analogy I often use is the measles vaccination. Should you make everyone take the vaccination for the greater protection of society? By doing so, you eliminate individual rights. These issues are difficult and regulating too much too soon can be an issue. But, on the other hand, leaving things completely unregulated is also very dangerous. The other challenge around regulation is that in an increasingly connected world, regulations in one part of the world differ from those in other parts. Regulators have a duty to collaborate among themselves and have some kind of baseline approach to this.

**ETHICAL AI – A COMPETITIVE ADVANTAGE**

**What would be your top suggestions for organizations across sectors that are just starting out on the journey of developing ethical AI?**

The approach we have taken is working quite well and we recognize that it is a competitive advantage and worth doing. Second, create a cross-functional team to do two things – do some external research about what is relevant within the industry and beyond, and, look internally to find out what is being done with data and define how quickly you need to act. My final recommendation would be to focus on the use cases rather than just the data collection.

H5

NICOLAS ECONOMOU,  
Chief executive

## UNDERSTANDING THE ROLE OF PRINCIPLES, STANDARDS, AND REGULATIONS IN CREATING DIGITAL ETHICS

**N**icolas Economou is the chief executive of H5, a legal automation technology and advisory firm. He was a pioneer in advancing the application of scientific methods, automation, and artificial intelligence in the law, and in advocating the development of sound norms for their trustworthy adoption. He chairs both The Future Society’s Law Initiative and the Law Committee of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. He served as chair of the Law Committee of the 2018 and 2019 Dubai Global Governance of AI Roundtable at the World Government Summit. He is also a member of the Council on Extended Intelligence (CXI), a joint initiative of the MIT Media Lab and IEEE-SA.

The Capgemini Research Institute spoke with Nicolas to understand more about the role of principles, standards, and regulations in the ethical and trustworthy design and deployment of AI.



## DEFINING ETHICAL AI

### How do you define ethics in AI?

Ethics is a long-standing academic discipline that provides ways to think about our behaviors. It offers pathways to rational debate, to critical evaluations of alternatives, and to decisions that have a moral foundation. Digital Ethics is the application of such methods to the challenges that AI presents. It is also important to recognize what ethics is not: it is not a universal law that just delivers the perfect answer, nor is it a simple “check-the-box” compliance exercise.

“IT IS ALSO IMPORTANT TO RECOGNIZE WHAT ETHICS IS NOT: IT IS NOT A UNIVERSAL LAW THAT JUST DELIVERS THE PERFECT ANSWER, NOR IS IT A SIMPLE ‘CHECK-THE-BOX’ COMPLIANCE EXERCISE.”

Another important consideration is that there are different types of ethics. Consider professional ethics: lawyers, for example, abide by certain rules of professional conduct. That is laudable and important, but the ethics of a corporation or a society may differ in certain ways from the ethics of the legal profession. Legal ethics may be consistent with the erosion of our privacy, as long as we have legally contracted our rights away. But, at the scale of society, the erosion of privacy takes on different ethical dimensions.

We should also remember that algorithmic systems are amoral. By that, I don’t mean that they are *immoral*. I mean that they do not have a moral compass. Yet, they can make decisions that have pervasive moral consequences. This places a particular responsibility on corporations not just to comply with their legal obligations, but also to develop and implement Digital Ethics, i.e. an institutional perspective on how to assess and address moral problems related to data, algorithms, and the practices that surround them.

Consider, for example, personal data: there may be uses of such data that are legal, but that may have certain business, brand, or societal consequences, that could cause a corporation to avoid certain legal, but ethically challenging uses.

## DEFINING TRANSPARENCY IN AI AND THE CURRENT STATE OF PLAY

### What role do you think transparency plays in AI?

The IEEE Global Initiative has done excellent work in promulgating principles for the ethical design and operation of AI. One of those key principles is

## “AN EXCESSIVE FOCUS ON TRANSPARENCY MIGHT CONFINE THE DISCUSSION TO THE ELITES’ ABILITY TO UNDERSTAND ALGORITHMS, THUS DEEPENING THE DIGITAL DIVIDE.”

The Institution of Electrical and Electronic Engineers (IEEE), established in 1884, describes itself as “the world’s largest technical professional organization dedicated to advancing technology for the benefit of humanity.” The organization has over 422,000 members across 160 countries and 39 technical societies and seven technical councils to represent the wide range of IEEE’s interests.<sup>6</sup> The IEEE launched

its Global Initiative on Ethics of Autonomous and Intelligent Systems with an aim “to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity.”<sup>7</sup>

transparency. Transparency is not only a topic of predominant focus in the international AI governance dialogue; it also has an intuitive appeal: “if I can see under the covers, I will be able to understand the system.” But I think that this predominant focus and this intuitive appeal can hide some dangers.

One of those dangers is that transparency can serve as an inadequate stand-in for what you really want to know. For example, in drug manufacturing, having transparency into the manufacturing process of a drug will not tell you if the drug is effective at treating an ailment – clinical trials do. Similarly, in car manufacturing, you will not know if the car is safe until you crash-test it. In both these examples, transparency – or transparency alone – cannot give you the answer you really want.

Whereas transparency is very important, I worry that, considered as a panacea, as it sometimes tends to be, it entails certain challenges.

### **Could you elaborate on concerns you have with an excessive focus on transparency?**

Beyond the issue I just raised, I think there are two other concerns:

One has to do with the cost of achieving transparency at the scale of society. For example, could courts handle extensive reviews of each socio-technical system and algorithm in matters before them, if transparency were the sole instrument available to them? Surely, in some cases such examinations will be indispensable. But it

6. IEEE, “IEEE at a glance”

7. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, “Ethically Aligned Design, First edition.”

is important to pause and think whether and when, to use my earlier analogy, a simple crash-test may offer a better answer than a complete review of a manufacturing process.

I also worry that an excessive focus on transparency might confine the discussion to the elites' ability to understand algorithms, thus deepening the digital divide. What we need are broadly understandable and accessible gauges of the fitness for purpose of AI systems, akin – if you will allow me to rely on the same analogy anew – to car crash-test ratings. Such gauges can empower citizens. Achieving those gauges requires complementary thinking to that which underpins transparency.

### **How do you view the current state of these ethics and transparency issues in organizations?**

Companies are struggling with this, because it is such a complex challenge. In addition, there are at present no consensus standards that companies can choose and certify adherence against. I think that, over time, corporate Digital Ethics will involve at least three elements: first, a Digital Ethics Charter published by companies; second, a set of standards that companies will be able to affirm adherence against (the IEEE is developing such standards); and third, auditing mechanisms. A good analogy in this last respect is financial audits: we trust companies to be able to produce sound financial statements, but it is auditors who attest the extent to which such statements meet the representations companies make.

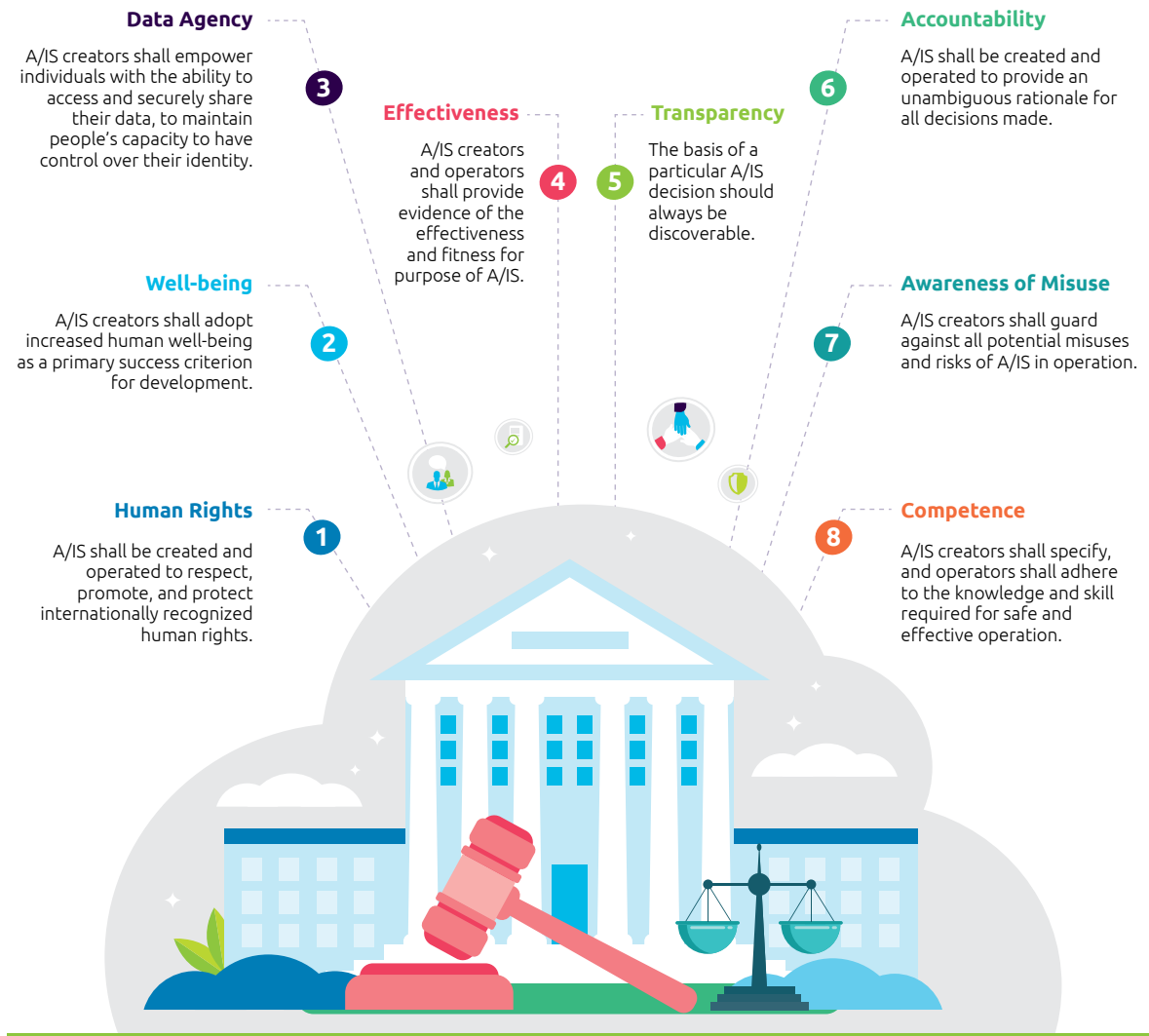
**“ AUDITING WILL BE KEY:  
WE TRUST COMPANIES TO BE ABLE  
TO PRODUCE SOUND FINANCIAL  
STATEMENTS, BUT IT IS AUDITORS  
WHO ATTEST THE EXTENT TO  
WHICH SUCH STATEMENTS MEET  
THE REPRESENTATIONS  
COMPANIES MAKE.”**

---

## REGULATIONS AND STANDARDS IN AI<sup>8</sup>

The IEEE released its General Principles of Ethically Aligned Design for creating and operating autonomous and intelligent systems (A/IS) that further human rights, human well-being, and ensure dependability.

**These principles are:**



8. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design, First edition"

### Could you tell us more about the IEEE's principles of effectiveness, competence, accountability, and transparency, and how these relate to trustworthiness?

In my personal view, these four principles are individually necessary and collectively sufficient in determining the extent to which AI-enabled processes should be trusted. They are also globally applicable but culturally flexible, as they are all evidence-based, rather than normative. They can help provide the factual basis that corporations, compliance officers, risk officers, and general counsels' offices need to determine whether a certain use of AI can be trusted to comply with their compliance obligations and Digital Ethics.

“AN ESSENTIAL COMPONENT OF TRUST IN A TECHNOLOGY IS TRUST THAT IT SUCCEEDS IN MEETING THE PURPOSE FOR WHICH IT IS INTENDED.”

#### Effectiveness

An essential component of trust in a technology is trust that it succeeds in meeting the purpose for which it is intended. What empirical evidence exists in this regard? For instance, consider privacy, which is such a hot topic these days. AI is increasingly used to identify personal information in vast corporate data repositories, in order to help comply with regimes such as the GDPR and, soon, California's CCPA. If you are a procurement or compliance department, what evidence do you have that the AI system you are about to purchase is actually effective at finding the personal information you are supposed to protect? Saying to a regulator: “I trusted a marketing claim” won't really cut it. Or in HR AI applications: what evidence do you have that the application is effective at avoiding bias?

#### Competence

A second essential component of informed trust in a technological system, especially one that may affect us in profound ways, is confidence in the competence of the operator(s) of the technology. We trust surgeons or pilots with our lives because we know that they have met rigorous accreditation standards before being allowed to step into the operating room or cockpit. No such standards of operator competence currently exist with respect to AI. When it comes to legal and compliance settings, this is not tenable. This area is another topic of focus for our work at the IEEE Global Initiative.



### Accountability


A third essential component of informed trust in a technological system is confidence that it is possible, if the need arises, to apportion responsibility among the human agents engaged, from design to deployment and operation. A model of AI creation and use that cannot hold people accountable will also lack important forms of deterrence against poorly thought-out design, casual adoption, and inappropriate use of AI.

### Transparency

A final key element of informed trust is transparency. Without appropriate transparency, there is no basis for trusting that a given decision or outcome of the system (or its operators) can be explained, replicated, or, if necessary, corrected. I believe that an effective implementation of the transparency principle should ensure that the appropriate information is disclosed to the appropriate stakeholders to meet appropriate information needs.

When it comes to legal and compliance functions in particular, my view is that, if duly operationalized, these four principles allow stakeholders to determine the extent to which they can trust AI to meet certain objectives, or to comply with their institutional ethics.

“A FINAL KEY ELEMENT OF INFORMED TRUST IS TRANSPARENCY. WITHOUT APPROPRIATE TRANSPARENCY, THERE IS NO BASIS FOR TRUSTING THAT A GIVEN DECISION OR OUTCOME OF THE SYSTEM (OR ITS OPERATORS) CAN BE EXPLAINED, REPLICATED, OR, IF NECESSARY, CORRECTED.”



### **Do you expect any regulation regarding ethical use of AI and how do you see that regulation being enforced?**

Like in so many other technological domains, a combination of industry-driven endeavors and regulation will prevail. The balance between these is likely to depend on the societal context. The EU Commission has an AI regulatory agenda, as has the Council of Europe, which has also announced a certification program for AI applications in the law. At the same time, expert industry bodies, such as the IEEE, are developing AI standards. To me, what is essential is that the mechanisms be evidence-based, in particular with respect to the principles we just discussed, absent which trust cannot be achieved.

### **Once we have these standards, how do we make sure that organizations adhere to them? Would there be incentives for organizations to follow ethical practices in AI? If so, what kind of incentives would those be?**

A combination of regulation and market-based incentives will prevail. Consider critical societal functions, such as transportation or medicine: adherence to standards is often imposed by regulations. Regulation will also be needed, in my view, in areas where the imbalance of power between ordinary citizens and corporations is too vast. In the US, for example, when we “click-to-accept” privacy agreements to access an online service, the consent we offer is not the mark of an empowered consumer, just evidence of our loss of agency. But sound standards – what I like to call “the currency of trust” – such as those being developed by IEEE, can accelerate adherence to best practices because the market will naturally gravitate towards products and services that meet trusted standards.

“THERE IS OFTEN NOT A SINGLE RIGHT ANSWER TO COMPLEX ETHICAL QUESTIONS. BUT YOU SHOULD HAVE AN ANSWER THAT YOU CAN STAND BEHIND AND HAVE MECHANISMS TO SHOW THAT YOUR CLAIMS ARE ACTUALLY A TRUE REFLECTION OF YOUR OPERATIONS.”

---

## ACTIONABLE STEPS FOR ETHICAL AI

### **What actionable steps can organizations take today to build and use ethical AI?**

The first step is to define a process. What does it mean to implement Digital Ethics? You need to define what you stand for as an organization – your brand values – and then create a methodology to assess the extent to which your use of AI is currently meeting (or failing to meet) those values. You should also consider the impact of AI on various stakeholders (employees, customers, shareholders, society). From such a gap- and stakeholder-impact analysis, you can assess both where you stand, and define where you want to be. To achieve your objectives, you must develop a methodology that incorporates ethics

as a mechanism for critical thinking and decision making. In doing so, I think it is important to consider what expertise you have, and what expertise you might need to retain, for example, in the discipline of ethics, or in operationalizing principles such as those proposed by IEEE.

There is often not a single right answer to complex ethical questions. But you should have an answer that you can stand behind and have mechanisms to show that your claims are actually a true reflection of your operations. In this respect, the IEEE has set up an Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS), which aims to help companies establish concrete evidence that they meet certain standards of accountability, transparency, and so on, in their use of AI.

## WHERE DOES ACCOUNTABILITY AND RESPONSIBILITY LIE IN AI SOLUTIONS?

---



“In the case of ethics, this is not something where responsibility lies with any particular individual in the company. It is a shared responsibility for all of us.”

.....

– Michael Natusch, Prudential Plc.



“The accountability and responsibility do not lie with the data scientist who develops the algorithm. The business owner in question needs to understand enough about the model to take on that accountability.”

.....

– Paul Cobban, DBS



“An essential component of informed trust in a technological system is confidence that it is possible, if the need arises, to apportion responsibility among the human agents engaged, from design to deployment and operation.”

.....

– Nicolas Economou, H5



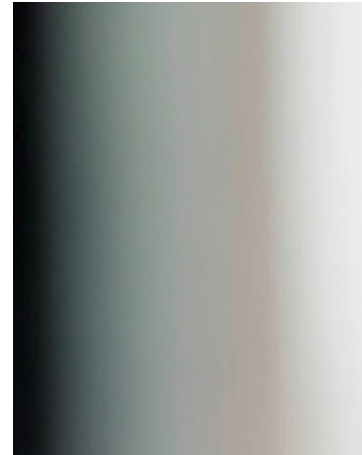
“I think the responsibility must be shared...everyone in an organization has an obligation to respect the privacy of customers or to protect their data.”

.....

– Ryan Budish, Harvard University



**MARCIN DETYNIECKI,**  
Head of Research &  
Development and Group  
chief data scientist



## BUILDING ROBUST AI THROUGH INTERPRETABLE AI IMPLEMENTATION

**M**arcin Detyniecki is head of Research & Development and Group chief data scientist at AXA. With a PhD in Artificial Intelligence from University Pierre et Marie Curie (UPMC) in Paris, he is a professor at the Polish Academy of Science (IBS PAN), and an associate researcher at the Sorbonne University Laboratory of Computer Sciences.

The Capgemini Research Institute spoke with Marcin to understand more about emerging ethical challenges in artificial intelligence and the role of governance frameworks in countering them.



“WE INVEST TIME AND MONEY IN DOING FUNDAMENTAL RESEARCH ON THREE KEY TOPICS: INTERPRETABILITY, FAIRNESS, AND ROBUSTNESS.”

## TOWARDS INTUITIVE, FAIR AND ROBUST AI

### **What are the main issues you are confronting in your work at AXA – in particular around the issue of ethics in AI?**

As chief data scientist of the AXA Group and, above all, head of Research and Development, my role is to produce technical solutions to the challenges facing the insurance industry. The human is very important in the insurance business. If Facebook or Google get a prediction of your appetite for a particular product wrong, it will not change your life. But, in the insurance industry, if you make a wrong prediction, it can have significant repercussions for individuals. Therefore, we invest time and money in doing fundamental research on three key topics: interpretability, fairness, and robustness.

First, interpretability allows you to explain decisions that are made by an algorithm that we called “a black box model,” namely with high accuracy but not explainable. Second, fairness is about mitigating unwanted bias, which may lead to discrimination. Should they come from a non-representative sampling of the population or from an unintentional reconstruction by the AI of protected sensitive attributes, such as religion or race? Finally, robustness is around understanding and fixing the fact that machine learning can be tricked very easily, such as through adversarial attacks, where a minor non-important change of the overall input drastically changes the output of the AI. For instance, changing a few pixels on a “stop sign” image can trick the AI into saying it sees a giraffe.

### **You speak of interpretability of AI, is it the same as transparency?**

We focus our research on interpretability instead of transparency. This is because machine learning tends to produce complex systems. Here, if you bring in transparency, it will enable anybody to see the rules, but you will not necessarily understand anything – especially if you have millions of them. We work on interpretability to make sure that people can understand the impact of decisions made by an AI system. Nevertheless, in general, the widespread use of machine learning and artificial intelligence in our society requires a high level of transparency to ensure that practitioners and users alike are aware of how, when, and why systems behave the way they do.



## Is there a governance mechanism or team that specifically looks at ethical issues in AI?

At AXA, we have an Ethical and Data Privacy Advisory Panel, which addresses the ethical aspect of AI through the lens of data privacy. It is very useful, because more often than not, it is not the AI that has an ethical issue, it is the input (i.e., the data) that poses the challenge. For example, should we use DNA information for pricing or not? Although the answer may seem straight forward it is a bit more complex than expected because, for instance, this information could be used to ensure that a non-curable version of a disease, today excluded of the coverage, would be now covered thanks to that information. These kinds of topics are discussed in this dedicated panel.

Moreover, we initially thought about having a specific code of conduct for AI, but then we decided that it would not be very effective as yet another guide with a generic nature. So, we decided instead to add a specific section on AI to the different internal rules and code of conduct. To drive attention and to cover the eventual transversal gaps, we also created an internal AI charter. We tried to ensure it has proof points and is a living framework that can evolve with time. This charter was an interesting exercise because it is the result of interaction and exchanges on these topics with very different people around the table, who were asked to get aligned on the topic. It has successfully provoked a positive momentum in AXA, which now is even shaping the thinking across industries.

“THE WIDESPREAD USE OF MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE IN OUR SOCIETY REQUIRES A HIGH LEVEL OF TRANSPARENCY TO ENSURE THAT PRACTITIONERS AND USERS ALIKE ARE AWARE OF HOW, WHEN, AND WHY SYSTEMS BEHAVE THE WAY THEY DO.”

## TACKLING ETHICAL CONCERNS IN AI AT AXA

### **Are there any ethical concerns with respect to fairness or interpretability that have been surfacing in your work?**

A first operational case that has brought some attention is considering the creation of very accurate, and thus in some sense, fair insurance product. To achieve this, we could use deep learning, but then the regulator will be not able to audit the way it is done today – since it is not an interpretable algorithm. That’s why it’s very important for us to keep in mind at all stages the ethical aspect and to keep investing in our research activities.

A second operational case is the use of machine learning to detect fraud. The algorithm is trained to flag suspicious people, based on previous examples. The list of suspects is then handed to a human expert who checks for fraudulent activity. The concern was that machine learning provides only a score, for instance eight out of 10 for an individual, but not an explanation. The experts complained that they do not even know what they are supposed to be looking at. This typically hinders adoption. Since this concern was detected, the R&D team has developed tools to provide helpful insights useful for our operators.

## THE ROLE OF REGULATION AND TECHNOLOGY IN BUILDING ETHICAL AI

### **Are there any ethical concerns with respect to fairness or interpretability that have been surfacing in your work?**

A first operational case that has brought some attention is considering the creation of very accurate, and thus in some sense, fair insurance product. To achieve this, we could use deep learning, but then the regulator will be not able to audit the way it is done today – since it is not an interpretable algorithm. That’s why it’s very important for us to keep in mind at all stages the ethical aspect and to keep investing in our research activities.

A second operational case is the use of machine learning to detect fraud. The algorithm is trained to flag suspicious people, based on previous examples. The list of suspects is then handed to a human expert who checks for fraudulent activity. The concern was that machine learning provides only a score, for instance eight out of 10 for an individual, but not an explanation. The experts complained that they do not even know what they are supposed to be looking at. This typically hinders adoption. Since this concern was detected, the R&D team has developed tools to provide helpful insights useful for our operators.

“AT AXA, WE HAVE AN ETHICAL AND DATA PRIVACY ADVISORY PANEL, WHICH ADDRESSES THE ETHICAL ASPECT OF AI THROUGH THE LENS OF DATA PRIVACY.”

“THE BEST WAY TO ENSURE THAT ETHICAL STANDARDS ARE MAINTAINED IS BY ALIGNING THE INTERESTS OF ALL STAKEHOLDERS AROUND THE NOBLE PURPOSE OF WHAT YOU ARE DELIVERING AND THE ASSOCIATED ETHICAL VALUES.”

---

## RECOMMENDATIONS FOR ORGANIZATIONS

### **What three concrete steps would you recommend that organizations take to start embedding ethics into AI systems?**

The first step is to realize that ethical AI is important because it will allow you to drive adoption and develop sustainable technologies that will comply with current and future regulations. Furthermore, being ethical does not necessarily drive up costs. For instance, being bias-free does not mean that you are going to lose money. As a matter of fact, in the insurance case, it will just redistribute the global risk more fairly.

A second step would be to set up a team responsible for implementing ethical AI. This team needs to have a high level of sponsorship because it is an overarching, long-term challenge. Strong sponsorship from senior leaders is important. This team of people must be a multidisciplinary team that can understand the technical issues but also business processes, HR challenges, and compliance.

Lastly, companies need to be patient. The use of AI and its necessary ethical adoption is clearly an opportunity, but improving things and changing processes in society and, in particular, in large companies might cause resistance. The best way to ensure that ethical standards are maintained is by aligning the interests of all stakeholders around the noble purpose of what you are delivering and the associated ethical values. I really think this will happen since long-term sustainability in our complex and ever-changing world implies a necessary alignment between the shareholder, the customers, and also the talent for which you are fighting.



**DANIELA RUS,**  
Director of the  
Computer Science and  
Artificial Intelligence  
Laboratory (CSAIL)

## BUILDING TRUST IN AI-BASED DECISION MAKING BY UNDERSTANDING THE STRENGTHS AND LIMITATIONS OF MACHINES

**D**aniela Rus is the Andrew and Erna Viterbi professor of Electrical Engineering and Computer Science and director of the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT). She serves as the deputy dean of MIT's Schwarzman College of Computing, and as director of the Toyota-CSAIL Joint Research Center. Her research interests are in robotics, mobile computing, and data science. She is known for her work on self-reconfiguring robots, shape-shifting machines that can adapt to different environments by altering their internal geometric structure. She earned her PhD in Computer Science from Cornell University.

The Capgemini Research Institute spoke with Professor Rus to understand the ethical considerations in the design and deployment of AI systems.




## ETHICAL AI CONSIDERATIONS FOR ORGANIZATIONS

### What are the key issues at it relates to ethics in AI?

The ethics problem is broader than just the AI problem. In a system where a machine makes a decision, we want to make sure that the decision of that system is done in a way that ensures people's confidence in that decision. For autonomous decision making, it is important that the machine's decision can be interpreted and explained, so that people get justifications for how the system decided the way it did. So, if someone didn't get a loan, why not? This kind of interpretability is critical. People need to be aware of how these systems work. Additionally, it is critical that the data used to train the system is correct and has checks for biases, because the performance of machine learning and decision systems is only as good as the data used to train them. Altogether, interpretability, explainability, fairness, and data provenance are the critical attributes that ensure trust in the decision of the system.

We can address the ethics problem at multiple levels: technologically, through policy, and with business practices. Technologists, policy makers, and business leaders need to come together to define the problems and chart a path to solutions. As a technologist, I would like to highlight that some of the solutions can be technological solutions: for example, fairness and level of privacy are becoming important metrics for evaluating the performance of algorithms. On the other hand, we also have to be aware that the current solution for machine learning and decision making have errors associated with them. While machines are good at some things, people are good at other things. Removing the person from the decision-making loop entirely will have consequences. My recommendation, therefore, is to have the machines and systems act as a recommender – providing recommendations to decisions and presenting supportive information for those recommendations. But, ultimately, there should be a person making those decisions.



## What is the magnitude of the AI ethics and transparency issue in organizations today?

It is a huge problem. It is not something that the technical community has been thinking about from the very beginning. Computing as a field is very young as compared to other science fields, such as physics. The term artificial intelligence (AI) was coined in 1956 and the artificial intelligence academic discipline started shortly after that. That is barely over 60 years. As compared to other fields of study, the AI field is very young. In the beginning of a new field, people try to lay the foundation of the field to understand the problems that can be addressed, the solutions we can have, and the capabilities that can be introduced as a result of the field of study.

For AI, the focus has been on developing algorithms and systems that enable machines to have human-like characteristics in how they perceive the world, how they move, how they communicate, and how they play games. In recent years, we have started thinking about the societal implications of technology profoundly and seriously. This is a very important issue right now. Our society needs to be positively impacted by what we do.

However, there are cases where organizations and people take a tool that is designed for a certain positive purpose and use it for a negative purpose. How do we address those situations? We can't stop technology from evolving and changing the world, but we need to stop and think about its consequences and come up with policies and provisions that ensure that what gets produced is used for the greater good.

## What are some things that organizations can do today, on a practical level, to work towards having ethical and transparent AI systems?

Organizations should start with understanding the technology. A lot of people use technology without understanding how it works and or how the data impacts the performance of a system. Another action companies can take is to identify their principle for adopting technology – things like fairness, inclusiveness, reliability and safety, transparency, privacy, security, and accountability. Companies should understand what it

“COMPANIES SHOULD UNDERSTAND WHAT IT MEANS TO USE AI FOR GOOD AND INCORPORATE THESE ATTRIBUTES IN THEIR CULTURE AND IN THE EDUCATION OF THE EMPLOYEES.”

---

means to use AI for good and incorporate these attributes in their culture and in the education of the employees. For example, companies can create review panels to make sure that these principles are adopted and, if people have questions about them, they are answered. They can ensure that the latest technological advancements that address the safe use of technology are adopted and incorporated in the operation of the organization.

### **How can we actively prevent biases in AI systems, such as facial and voice recognition, for example?**

The performance of a system is only as good as the data used to train the system. So, if we have a bias in the data, we are going to have bias in the results. There are numerous examples of companies with biased face and voice recognition systems that displayed discriminative behavior. It is important to put in place provisions to make sure that people don't get discriminated against because the data used by the system was biased.

Another type of bias is in over-predicting what is normal and over-emphasizing the expected distribution of

attributes in the data. Therefore, if the data is incomplete, we might not capture critical cases that are very prevalent and that make a difference in how the system operates in the world.

### **CASE IN POINT: SOLVING THE ETHICAL DILEMMAS OF AUTONOMOUS CARS**

#### **Autonomous cars pose a lot of ethical questions. How can the companies involved in designing these cars answer these questions?**

First of all, the companies have to advance the technology. Today there are limitations to what autonomous vehicles can do. For instance, the sensors used by autonomous vehicles are not reliable, they do not work well in rain or in snow, and this causes big limitations for the use cases of self-driving technologies. The companies train and experiment with their car products mostly in Arizona, where it never rains or snows. This is a serious limitation.

**“I WOULD SAY THAT TODAY’S AUTONOMOUS DRIVING SOLUTIONS WORK WELL IN ENVIRONMENTS THAT HAVE LOW COMPLEXITIES, WHERE THINGS DON’T MOVE TOO MUCH, WHERE THERE IS NOT MUCH INTERACTION, AND WHERE THE VEHICLES MOVE AT LOW SPEED.”**



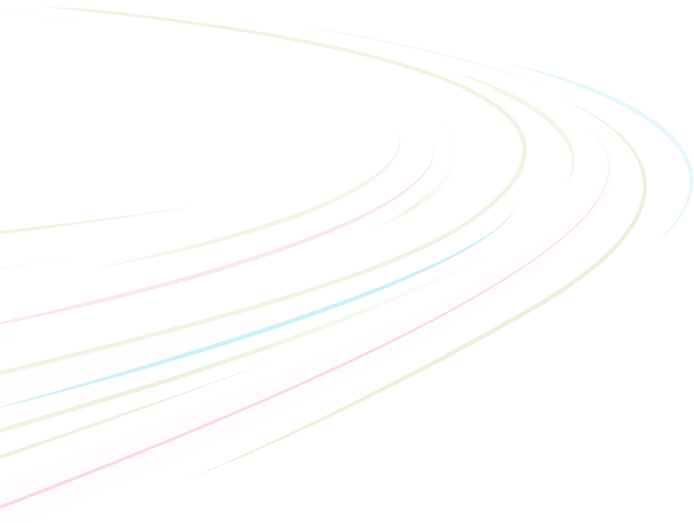
There are other limitations too – the vehicles do not have the ability to respond quickly enough or cope well with high speeds and congestion. They have trouble in human-centered environments because they do not understand human-centric behavior. They have trouble understanding the road context, especially when robot cars are on the same roads as human-driven cars.

These are issues that are being worked on but there are no good solutions yet. I would say that today's autonomous driving solutions work well in environments that have low complexities, where things don't move too much, where there is not much interaction, and where the vehicles move at low speed. So, if you think about three axes – the complexity of the environment, the complexity of the interaction, and the speed of the vehicle – the sweet spot today is around the origin of this system of coordinates. For level-five autonomy, that is providing autonomy anywhere anytime, we need technologies that can address high complexity, high speed, high levels of interaction, in other words we need to push the boundaries along all three axes. While there are many ways of getting there, the most important one is to have reliable technology. Once we have the reliable technology, then we can answer a range of questions. How do we regulate and at what level? If the car has an accident, who is responsible? Is it a manufacturer, is it the programmer, is it the owner? How do we begin to address such issues? These are very important questions.

## ETHICAL AI REGULATION AND CODES OF CONDUCT

### **Do you foresee regulations in the areas of ethics in AI? Is legislation the way ahead or is it counterproductive?**

At MIT, we have one organization that is devoted to studying this question – the Internet Policy Research Initiative (IPRI). IPRI's activities include studying the policy and technology around data use and more broadly the use of algorithms to support decision making. Researchers are looking at what should be regulated and to what extent. There are many deep questions around this issue. Take, for instance, self-driving vehicles. At the moment, we do not have legislation that addresses how to regulate the use of self-driving vehicles. This makes the development of the technology more challenging and slows the rate at which the industry can innovate products around this technology. However, in the case of autonomous vehicles, I believe that regulation is necessary. But, in the US for example, should vehicles be regulated by the federal government or at a state level? How can the policies be coordinated from state to state? These remain open questions, but once we have answers and policies, I think we will see product growth in the space of autonomous vehicles.



“WE MUST LOOK AT TECHNOLOGIES BY INDUSTRY SECTOR AND FIGURE OUT THE POLICIES AND REGULATIONS FOR EACH SECTOR.”

I am giving you a nuanced answer because I don't think there is a single answer to this question. We must look at technologies by industry sector and figure out the policies and regulations for each sector. Altogether, most important is to have the building blocks of trust that can help assure consumers that they have the benefit of innovative products without sacrificing safety, security, fairness, or privacy.

### **How can organizations develop codes of ethics and trust for using machines in decision making?**

There are attributes we should embrace as the basis of a code of ethics and trust. This might include explainability, interpretability, transparency, privacy, or guaranteeing fairness. It might also include descriptions of the data provenance and accountability for the information sources used to build the system. We can think of these

attributes as being generic, cutting across many different industry verticals. Then, each of these attributes would be instantiated to specific questions that are applicable to a vertical. For instance, the attribute that addresses explainability, interpretation, and transparency in the transportation sector might translate to “why did the car crash?” or “was the mistake avoidable in any way?” In the field of finance, this might translate into “why didn't I get the loan?” and in healthcare, “why this diagnosis?” In criminal justice, it might be “is the defendant a flight risk or not?” These are very different questions that address safety, transparency, and explanation for decision making. We can create similar tests for other attributes and verticals. For example, for privacy in finance – we might want to prove that a customer got the best deal without disclosing other consumer data. But how do we show that?

In creating a comprehensive code of ethics, it is important to focus on ensuring consumer confidence in decision making, especially for safety-critical applications. Before a person can drive a car, the person needs to pass a driver's test. Maybe, for AI working on behalf of humans, we need analogs of the driver's test to convince ourselves that the machine operates at a level of trust and robustness with which we are comfortable.

## ROLE OF ACADEMIA

### **What role does academia have to play in ensuring organizations implement ethical practices in AI?**

Academia has a very important role in establishing the foundation and principles, and in highlighting what is important to ponder upon. Academia can also provide support for decision-making processes in various industries. Some of this work falls in the policy space and some of the work falls in the technological space. Machines are better at some things and humans are better at other things. We need to figure out ways of tasking machines and people in ways that makes the most of both worlds so that the collective becomes much more powerful than machines working by themselves or people working by themselves.

“IN CREATING A COMPREHENSIVE CODE OF ETHICS, IT IS IMPORTANT TO FOCUS ON ENSURING CONSUMER CONFIDENCE IN DECISION MAKING, ESPECIALLY FOR SAFETY-CRITICAL APPLICATIONS.”

---

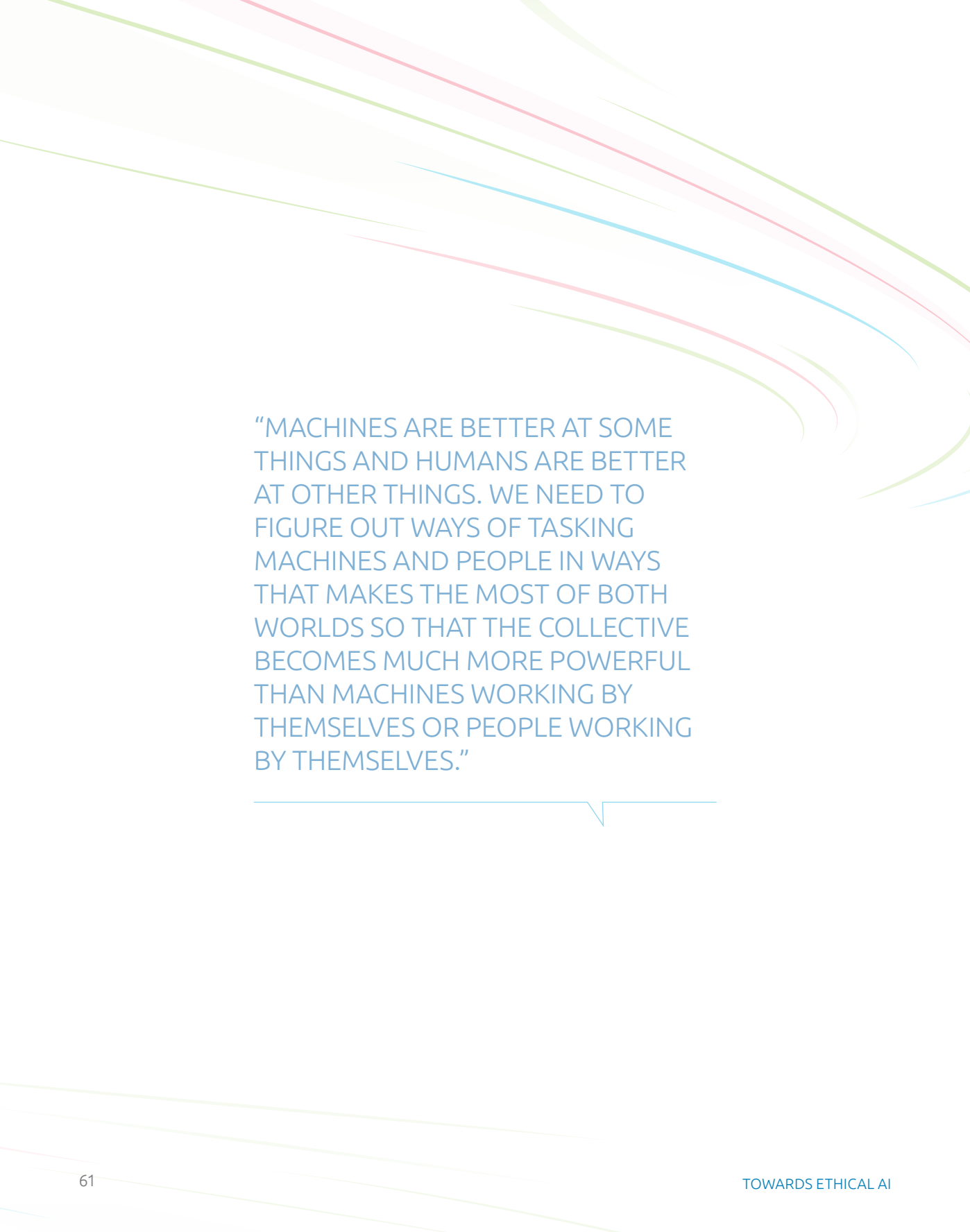
In computer science, the measures defining how well a computer program performs were focused on the time and space required to compute. Now, we are beginning to consider other metrics – for example, what is the fairness of the algorithm? To use metrics such as fairness or privacy, we need to develop mathematical models that allow us to incorporate these properties in the algorithm evaluation. This methodology will result in algorithms that are guaranteed to produce a fair answer, or a system that is guaranteed to preserve privacy. We might even imagine generalizing from fairness metrics to other aspects of human rights.

I can't say that we have a clear solution, but this is why the topic remains an area of research. How technologists, policy makers, and company leaders come together and incorporate their different objectives into something that encourages innovation for the greater good and enforces positive and constructive application of technology requires a level of understanding of policy, technology, and business. Co-training in technology, policy, and regulation law should be part of our future processes. Academia has an important role to play here.

### **What is the current focus of research in this field? How likely is it that the research will help solve issues on ethics and transparency?**

I don't think we have a silver bullet right now. This space remains a very important and exciting area of research. I believe a step forward is to identify the right attributes to be checked when involving machines in decision making. New approaches to trustworthy and robust machine learning engines will lend transparency and performance guarantees to the systems.

Advancing fields such as homomorphic encryption and understanding how to deal with bias in data is also very important. We have advanced technology to the point where we produce quintillion bytes of data every day, but in a world with so much data, everyone can learn everything about you. So how can we maintain privacy? Well, the field is working to develop technologies, such as differential privacy and homomorphic inscriptions, which will enable computation on encrypted data. When machines will be able to perform computations without decrypting data, we will have the benefits of the data-driven computation without revealing what is in each of those individual records. This is an example of a technological solution that could have a profound impact on the use of data in the future. Other solutions will necessarily have to be at the intersection of policy and business and technology.



“MACHINES ARE BETTER AT SOME THINGS AND HUMANS ARE BETTER AT OTHER THINGS. WE NEED TO FIGURE OUT WAYS OF TASKING MACHINES AND PEOPLE IN WAYS THAT MAKES THE MOST OF BOTH WORLDS SO THAT THE COLLECTIVE BECOMES MUCH MORE POWERFUL THAN MACHINES WORKING BY THEMSELVES OR PEOPLE WORKING BY THEMSELVES.”

---

## HOW CAN REGULATIONS BE FRAMED FOR ETHICAL AI?

---



“The most important thing is to have building blocks of trust that can help assure consumers that they have the benefit of innovative products without sacrificing safety, security, fairness, or privacy.”

– Daniela Rus, MIT CSAIL



“It’s about finding the right balance: It wouldn’t make sense to have a regulation in place that would make it impossible to develop AI solutions here in Europe.”

– Saskia Steinacker, Bayer



“It involves discussing the issues across the industry, putting in place some guidelines initially, and getting feedback to see how that operates before we cement in any regulation.”

.....

– Paul Cobban, DBS



“More conversations about the best legislation should start now ... we need both self-regulation and legislation as they are complementary tools.”

.....

– Luciano Floridi, University of Oxford



“It is extremely important that we adopt an approach where industry and institutions are being asked for their views.”

.....

– Cecilia Bonefeld-Dahl, DigitalEurope

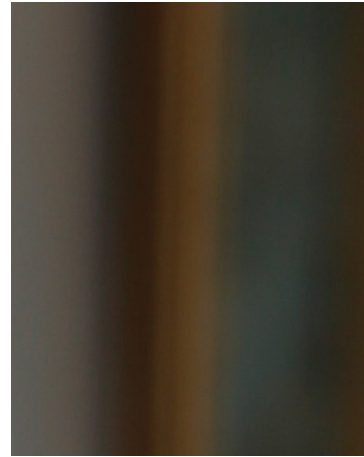


**LUCIANO FLORIDI,**  
Professor of  
Philosophy and  
Ethics of Information

## THE VIRTUOUS CIRCLE OF TRUSTED AI: TURNING ETHICAL AND TRANSPARENT AI INTO A COMPETITIVE ADVANTAGE

**L**uciano Floridi is the professor of Philosophy and Ethics of Information at the University of Oxford, and the director of the Digital Ethics Lab of the Oxford Internet Institute. Outside of Oxford, he is faculty fellow of the Alan Turing Institute (the national institute for data science) and chair of its Data Ethics Group.

The Capgemini Research Institute spoke with Professor Floridi to understand more about the philosophy underpinning ethical and transparent AI.







“THE VALUE OF VISIBILITY IS OFTEN UNDERESTIMATED. IT’S A STEP TOWARDS ACCOUNTABILITY.”

## THE KEY ISSUES FACING ORGANIZATIONS IN ETHICAL AND TRANSPARENT AI

### **What is the magnitude of the challenge when it comes to AI and ethics in large businesses?**

My whole career has been spent saying this is big. It’s big because we are finally seeing the maturity of this significant information transformation. With the invention of the alphabet we could record information, and with the invention of printing, we could not only record but disseminate that information. Today, with computers, we can automate the recording and dissemination of information.

We will be feeling the effects of what we are doing now for centuries to come, in the same way we are still feeling the effects of the Gutenberg revolution. I am not sure that organizations fully realize yet the enormity of this challenge.

### **Some companies are setting up ethics boards. Is this one way in which organizations can tackle this challenge?**

It’s one of many ways in which the situation can be improved. Companies need to understand the problem and then design policies to deal with what is happening. For example, the external advisory board that Google set up to monitor for unethical AI use was a good step in the right direction. Of course, it is not the only step that needs to be taken; we need to make sure all possible efforts are explored to find the right approach. If the top 500 companies in the world were to create an ethics advisory council tomorrow, I would be happy. This would bring more awareness, more engagement, and more visibility to the issue. The value of visibility is often underestimated. It’s a step towards accountability.

One major risk is having companies become tired or skeptical of any approach to technological development, especially in AI. They start retreating behind the wall of pure legal compliance. That is the future I do not want to see.

### **How do you build greater awareness of the need for ethical and transparent AI among companies of all sizes?**

I think there are two critical strategies. First, leading by example is crucial. Smaller companies or companies less engaged need to see large companies taking responsibility for ethical AI. These smaller companies will want to be on the right side of the divide.

Second, clarifying that “good business means good society and good society means good business” is so important. A company needs to understand that doing the right thing is a win-win situation. It’s good for business and it’s good for society. If you look at the ecosystem within which a large company is operating, in the long run, the healthier


that ecosystem, the better the company will perform. That ecosystem requires financial and social investment. This approach needs a long-term vision that is over and above the quarterly return. A company must ask itself, do I want to be here for the next decade? For the next century?

## TRUST AND COMPETITIVE ADVANTAGE

### **As organizations implement AI systems, how do you think they can gain the trust of consumers and their employees?**

I think trust is something that is very difficult to gain and very easy to lose. One classic way of gaining trust is through transparency, accountability, and empowerment. Transparency so that people can see what you are doing; accountability because you take responsibility for what you are doing; and empowerment because you put people in charge. You say to them, “you have the power to punish me, or you have the power to tell me that something was not right.”

“A COMPANY NEEDS TO UNDERSTAND THAT DOING THE RIGHT THING IS A WIN-WIN SITUATION. IT’S GOOD FOR BUSINESS AND IT’S GOOD FOR SOCIETY.”



Transparency is perfectly reasonable, achievable, and should be expected. Until you understand what exactly is going on in your system, you must go back to the drawing board. If an engineer were to say that they couldn't do something in the early stage of development, that product likely should not be released. Imagine if a nuclear physicist creates a nuclear system and they are not quite sure how it is going to behave, but still puts it on the market. This would be insane.

### **Can an organization that focuses on being ethical in their AI systems gain competitive advantage in the long run?**

Absolutely! There is intangible value in brand reputation, credibility, and trustworthiness that will drive this advantage. Competition is necessary to this scenario. If there is no competition, there is less accountability and less need to be transparent.

### **How can academia play a role in ensuring organizations implement ethical AI?**

Academia can add value and help a lot if engaged properly. To my mind that means allowing academia to conduct independent not-for-profit research for the advancement of our understanding and knowledge. A focus on scholarly and/or scientific understanding is part of the solution. We need this ingredient in the whole strategy.

I like the idea that around the same table you have experts from academia, experts from research and development in the industrial world, policymakers, experts from NGOs, and representatives from startups and civil society. Academia has a duty to provide advice and insight to support technological and business development that improves society.

“TRANSPARENCY IS PERFECTLY REASONABLE, ACHIEVABLE, AND SHOULD BE EXPECTED. UNTIL YOU UNDERSTAND WHAT EXACTLY IS GOING ON IN YOUR SYSTEM, YOU MUST GO BACK TO THE DRAWING BOARD.”

“MOST LARGE ORGANIZATIONS TODAY ACROSS THE UNITED STATES AND EUROPE ARE TALKING ABOUT ‘DUTY OF CARE’ AND AI (I.E., THE DUTY TO TAKE CARE TO REFRAIN FROM CAUSING ANOTHER PERSON INJURY OR LOSS).”

## ETHICAL AI REGULATION AND STANDARDS

### **Are organizations prepared for eventual regulation in ethical AI?**

I think organizations are preparing and expecting it. Most large organizations today across the United States and Europe are talking about “duty of care” and AI (i.e. the duty to take care to refrain from causing another person injury or loss). We also hear a lot about the need for clear normative frameworks in areas such as driverless cars, drones, facial recognition, and algorithmic decision-making guidelines in public-facing services such as banking or retail. I shall be surprised if we will have this conversation again in two years’ time and legislation hasn’t already been seriously discussed or put in place.

More conversations about what the best legislation is should start now. I am happy with the first step we’ve taken at the European Union level, with a high-level expert group on the ethics of AI (disclosure: I am a member). I think this will help the development of not only the technology, but also normative rules and legal frameworks.

### **Do you think organizations can self-regulate or is legislation necessary?**

Putting this question as an either/or is common, but I reject that. We need both self-regulation and legislation as they are two complimentary tools. To win a tennis game, you need to play according to the rules, this is the law, but you also need to develop your skills through discipline and training, and have a winning strategy, and that is ethics and self-regulation. For example, there is no legislation today that forces a company to publish open source software for AI solutions, for example. While I think this would be a good idea, it would need to be done carefully, because it could also be misused. I like the idea that a company sees publishing and making their own software available as a matter of default as opposed to say, “the law doesn’t require it, therefore we’re not going to do it.”

## HARVARD UNIVERSITY

**RYAN BUDISH,**  
Assistant Director  
for Research



## MAKING AI ACCOUNTABLE: CREATING OPENNESS AND TRANSPARENCY IN AI

**R**yan Budish is an assistant director for Research at the Berkman Klein Center for Internet and Society at Harvard University. This research center's mission is to explore and understand cyberspace, with Ryan's main focus being policy and legal analysis.

The Capgemini Research Institute spoke with Ryan to understand more about accountability in AI, ethical challenges, and the risks of these advanced technologies.



## ACCOUNTABILITY AND TRANSPARENCY IN AI

### What are the important themes that underpin ethics in AI?

“Ethics” as a term has a very specific meaning and body of scholarship behind it. But broadly speaking, organizations should be concerned about issues such as fairness, accountability, and transparency in AI. I think there is also a growing recognition of the importance of human rights as organizations deploy and use AI.

We need to ensure that AI is acting in such a way that we can hold it accountable and also respond if we determine it is acting in a way that we don't believe to be consistent with our values and/or laws. There are multiple complementary approaches to doing this. For example, one can take a top-down, system-wide approach defining ex ante the standards by which we want to hold these systems accountable, which could be ethical, normative, or political standards. One can also take a bottom-up, generative approach looking at individual instances of technologies or applications and asking whether that

specific AI system is operating in the way that we want it to. These approaches work together. On the micro-level you're ensuring that a particular system is operating in the way that the designers intended, without unintended, harmful effects. And at the macro-level ensuring that the system is operating in accordance with broad, system-wide standards that policymakers, ethicists, or society as a whole has put in place.

We see these various approaches play out on issues such as the use of lethal autonomous weapon systems. At the societal level there is vibrant debate about whether such systems should be banned outright as outside of the bounds of what a society accept or tolerate. Simultaneously, at the organization level several large AI companies have established their own set of principles, guidelines, or standards limiting the kinds of uses for which they'll sell their technology.



### **If something goes wrong, such as a consumer or media backlash, who can be held accountable?**

I don't think AI is a special case. AI technology is not being deployed into a vacuum, but rather it's being deployed into areas that already have quite a bit of laws and regulations.

For example, imagine if an autonomous vehicle doesn't perform as it should and there is an accident where someone is hurt or dies. It's not the case that, because AI was involved, no one knows what to do. In fact, there are legal liability regimes that already exist. There are regulations about consumer product safety and vehicle safety, and if any of those were violated, there is potential liability and recourse against the auto manufacturer and/or their suppliers. There are lots of tools that are already available, in most cases, the main issue is how those tools can be leveraged to properly ensure accountability.

### **Technology companies have been hit with AI and ethics questions recently, but what other types of organizations will be affected?**

Outside of the companies that are leading AI development, there are two categories of organizations that are facing similar, but not identical challenges: public sector and governments on the one hand, and non-AI or even non-technology companies on the other. Among these two groups, there is a lot of enthusiasm for trying to use AI technologies, but also a growing recognition that there is a lot of potential risk that comes with it. For example, the potential for AI to behave in a discriminatory way if biased data is fed into the machine learning system.

I think there's a general understanding of this challenge but not enough knowledge of what to do about it. There are a lot of high-level principles promoting things like "AI should respect human rights" or "AI should not be discriminatory," but there's real sense that they don't necessarily know how to bridge the gap between these high-level principles and what's happening on the ground.

“THERE WILL BE SOME INSTANCES WHERE HAVING EXPLAINABLE AI WILL BE SO IMPORTANT THAT WE WILL BE WILLING TO ACCEPT ANY COMPROMISE WITH HOW ACCURATE THE SYSTEM MIGHT BE. THERE WILL BE OTHER LOWER-RISK CIRCUMSTANCES WHERE MAYBE AN EXPLANATION IS NOT AS IMPORTANT, AND SO WE CAN TRY TO HAVE GREATER ACCURACY.”

## KEY CHALLENGES AND RISKS IN AI

### What do you see as the greatest challenges and risk in AI and ethics?

I think the biggest challenge right now is the information asymmetry that exists between the people who are creating these AI technologies and the people who need to decide whether to use them, and how to use them.

Procurement professionals – who have long decided what kind of computers or software to buy – are now being asked to determine what kind of AI systems to purchase. Or they are being asked to make decisions about what type of data to give to third parties to create AI systems. In many cases, these people are not necessarily well prepared to assess the risk and opportunities of a particular AI system.

### What kind of risks do traditional organizations face?

The risk is entirely dependent on where the AI system is being used. Some applications will have a very low risk, but there are others that will have huge substantial risk. For example, in autonomous vehicles, there is potential for risk to public safety. In the criminal justice system, there is a risk of unfairly incarcerating people for longer than they should be or letting potentially dangerous criminals on to the streets when they should be in jail. In healthcare, there could be a risk of improper diagnoses.

## DELIVERING ETHICAL AND TRANSPARENT AI

### Do you think ethical and transparent AI is realistic?

I think organizations have a lot of incentive to pay attention to it. Given that there is a growing understanding of the potential risks that AI systems can present, I think there is a desire to try to deploy these systems in a way that respects human rights, that preserves human dignity, and that is fair, accountable, and transparent.

And, in general, I think it's realistic. But of course, there are compromises between explainability and how accurate a system might be. There will be some instances where having explainable AI will be so important that we will be willing to accept any compromise with how accurate the system might be. There will be other lower-risk circumstances where maybe an explanation is not as important, and so we can try to have greater accuracy.

### What role can team diversity play in removing bias and discrimination in AI systems?

Improving diversity is incredibly important. In my opinion, one of the things that must happen is that the people who are being impacted by AI technologies must play a bigger role in helping to develop and govern those technologies. AI technologies cannot just be developed in a few places in the world and exported everywhere else. We need greater diversity in terms of the people who are developing the technologies. We need more diverse datasets to go into developing those technologies. We need greater understanding of the implications of these technologies for both good and bad across the public sector all around the world, so that information asymmetry is not an obstacle to good policymaking. I think that there is no one place that diversity and inclusion must be improved, but rather it must be addressed throughout the landscape.

“A GOOD FIRST STEP IS LOOKING AT THAT LANDSCAPE OF PRINCIPLES AND EMERGING STANDARDS AS A WAY TO BEGIN TO UNDERSTAND AND THINK CRITICALLY ABOUT BOTH THE POTENTIAL RISKS AND BENEFITS OF AI SYSTEMS.”

---

## RECOMMENDATIONS FOR ORGANIZATIONS

### **What concrete steps do organizations need to take to build and use ethical and transparent AI?**

There is a lot that organizations can do. There are high-level principles that exist and emerging standards they can adopt. A good first step is looking at that landscape of principles and emerging standards as a way to begin to understand and think critically about both the potential risks and benefits of AI systems.

The second step is to understand what gaps exist in their ability to address those risks and opportunities. For example, organizations can examine their resources and talent. Do organizations have data scientists on staff? If they do not have data scientists in their organization, how can they partner with local universities to develop a pipeline of data scientists? Are there people who can help them audit their datasets and, if not, where can they find those people? Are there people within the organization who understand how AI systems work? If not, can they partner with computer scientists and computer engineers at local universities?

### **Where do you think ethical AI accountability and responsibility should lie within private, non-technology organizations?**

I don't think there is one place. I think the responsibility must be shared, similar to the approach that organizations have taken for issues like human rights and privacy. Everyone in an organization has an obligation to respect the privacy of customers or to protect their data. Certainly, organizations have created positions like chief privacy officer to help ensure that the right policies and systems are in place. But the responsibility itself lies with everyone. The same goes for human rights. No one gets a free pass for violating human rights from the lowest person in the company all the way up to the senior executives and the board. The question of behaving ethically is similar in that I don't think the responsibility lies with any one position.

## HOW CAN BIAS IN AI BE TACKLED?

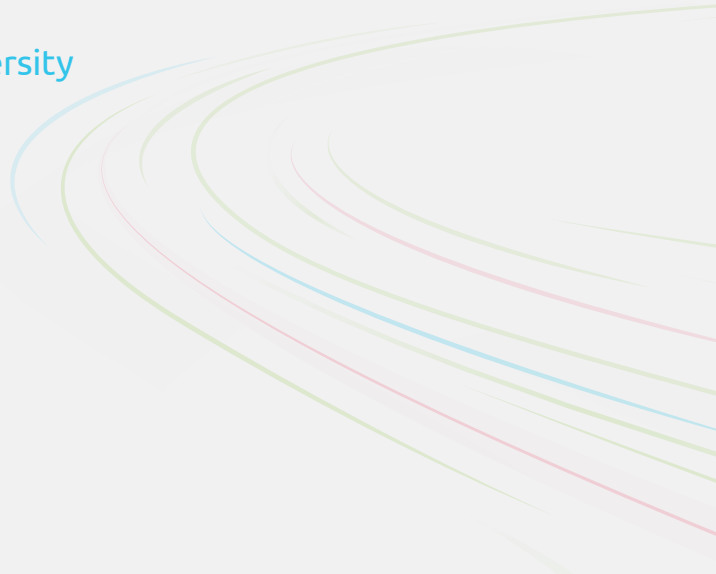
---



“The people who are being impacted by AI technologies must play a bigger role in helping to develop and govern those technologies.”

.....

– Ryan Budish, Harvard University





“It is important to put in place provisions to make sure that people don’t get discriminated against because the data used by the system was biased.”  
.....

– Daniela Rus, MIT CSAIL



“I think diversity in every day interactions is extremely important for an AI team, because you are not going to ask yourself questions that somebody from a different background would come up with.”  
.....

– Michael Natusch, Prudential Plc.

## DIGITALEUROPE

**CECILIA BONEFELD-DAHL,**  
Director General

## UNLOCKING THE TRUE POTENTIAL OF AI THROUGH ITS ETHICAL IMPLEMENTATION

**C**ecilia Bonefeld-Dahl is director general of DIGITALEUROPE, the digital technology industry association that represents over 35,000 digital companies in Europe. She is a member of the European Commission's High-Level Expert Group on Artificial Intelligence, a board member of the European Commission's Digital Skills and Jobs Coalition, and a board member of the European Parliament-led European Internet Forum.

The Capgemini Research Institute spoke with Cecilia to understand more about the state of ethical and transparent AI in Europe.





“ ... IT IS GOOD TO HAVE REGULATION AND GUIDELINES, BUT THEY NEED TO ALIGN WITH INDUSTRY AND COMPANIES.”

## DIGITALEUROPE AND ETHICAL AI

### Can you tell us about DigitalEurope and its mandate in ethics and AI?

DIGITALEUROPE is the biggest association of tech in the world. We represent 36,000 tech companies in Europe, and we have 40 associations around the European territory. We also have a chamber of big global companies, such as SAP, Siemens, Bosch, Schneider, Microsoft, and Google. So, you can call it a collaboration partner, where we work with them to shape regulation on tech in Europe.

## REGULATION AND GUIDELINES ON ETHICAL AI

### Could you tell a bit more about the guidelines published by the EU's high-level expert group and what is currently happening in that space?

When the GDPR was launched, it had been discussed in the political environment for about seven or eight years. But, once it finally came into force, many companies, especially the SMEs, were not ready. The change in legislation slowed down European industry to a high degree. Learning from that, we realized that it is good to have regulation and guidelines, but they need to align with industry and companies. After developing the guidelines, we have launched a pilot where they are being tested by different companies, public institutions, and representatives from civil society. We are now collecting feedback on how we should implement these guidelines, and this runs until the end of 2019.



In parallel, we also have a series of sector workshops, where we look at different areas – such as the public sector, manufacturing, and health. We basically look at the high-level expert group’s guidelines and recommendations, and how we can implement ethical and trustworthy AI. We are listening closely and taking it on the ground to test the right way to work with it. If we don’t do it this way, we might just slow down innovation by dropping things on people that might not fit into the way they work. So, it is extremely important that we adopt an approach where industry and institutions are being asked for their views. It gives them an opportunity of working with something like trustworthy AI in a way that it is coherent with the real world.

### **Are there any milestone dates that the expert group has put for moving toward regulation?**

It is not only about regulation but also looking at whether regulation is necessary or not. It is also about keeping an open mind and looking at existing regulations. The overall goal is not just about implementing trustworthy AI, but also about boosting its uptake and getting a competitive start on how to do this in Europe. So, the next big step is to have all the feedback, understand how we can work with trustworthy AI, and – if changes are needed - how we can handle those changes.

“THE OVERALL GOAL IS NOT JUST ABOUT IMPLEMENTING TRUSTWORTHY AI, BUT ALSO ABOUT BOOSTING ITS UPTAKE AND GETTING A COMPETITIVE START ON HOW TO DO THIS IN EUROPE.”

---

### **How do you think the approaches to regulation or guidelines on trustworthy AI will differ for Europe compared to the US and China?**

It is basically about creating an approach to AI where we are all sure that its application is for the benefit of people, companies, and society as a whole. And I think in most cases, it is. For example, we can do amazing stuff with artificial intelligence in health, preventive medicine, and predicting life-threatening diseases before they break out. We need to make sure that the development is pushed in a direction where it is for good for the society and the company. So, the whole idea is to create a feeling of safety and trust around AI and its benefits.

### **What is your position with respect to regulations in this space?**

I want to be sure that the companies and institutions sit down and look at their environment and the current laws and see if there are any missing links. Let us do a thorough exercise and ask people, and if we find something that is missing, let us add it in. If we find something that we need to interpret – or give clear guidance on how to apply the rules – let us do it. But let's not do this simply for the sake of it. So, I am not against regulation, I just want things to be done right.

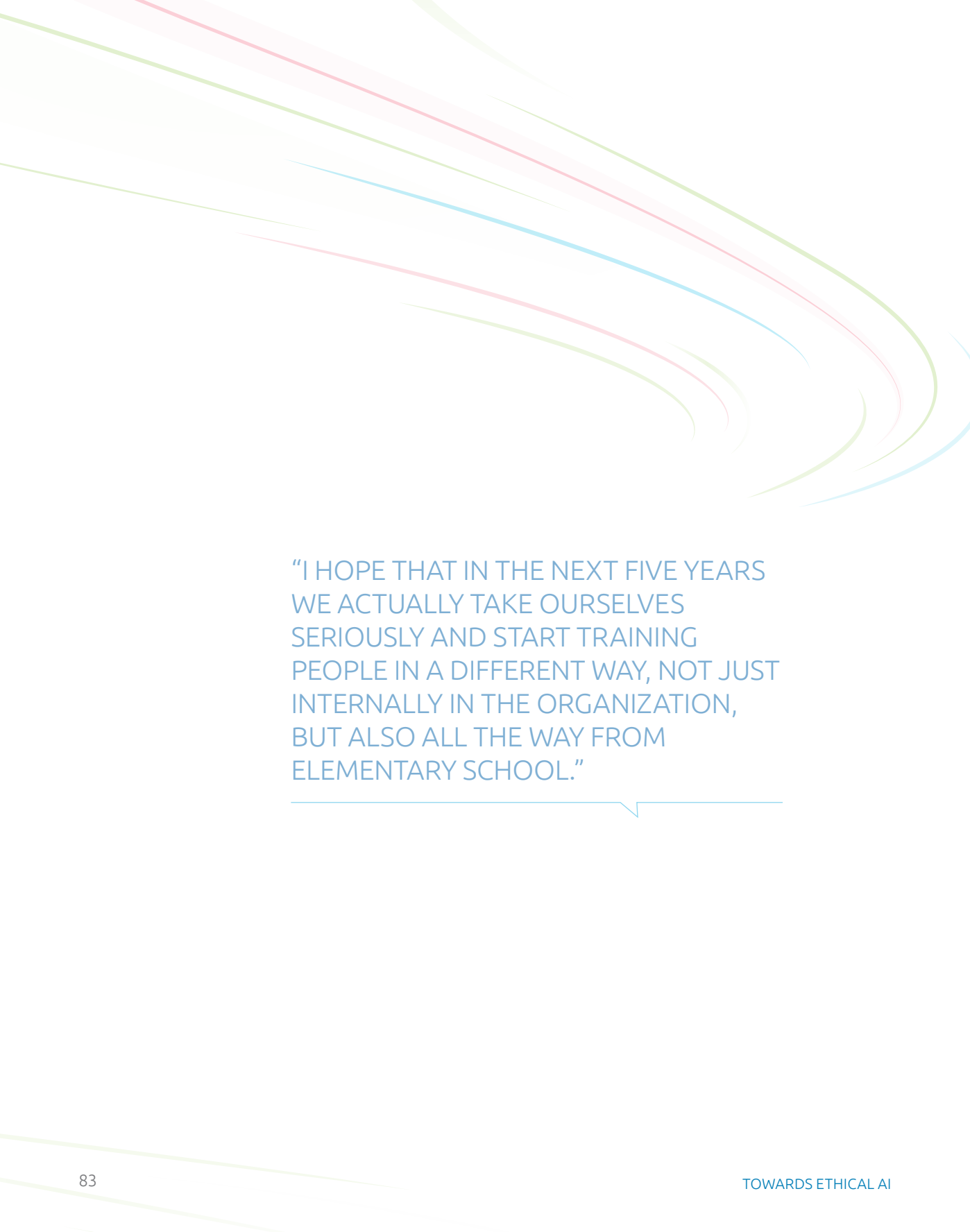
### **Do you think self-regulation in the form of organizations crafting their own code of conduct or policies around ethics in AI could be useful?**

My first response is that a tool in itself is not bad, it depends on how the tool is used. So, self-certification is something that has been done for cybersecurity. And, giving people responsibility for their own actions is a very European thing, which seems to work very well and can be really powerful. But we need to be sure that we do not just talk about tools without knowing how to apply them. It will take at least around a year before we know exactly what the results are. So, I would say "give it time."

### **Team diversity is one way to tackle bias in AI. What approaches can organizations adopt to have more diversity in their AI teams?**

First of all, it is about creating an interest. The Commission has put a lot of money in the DigitalEurope program toward education of AI and cyber specialists. And, we also have more projects with people who retrain secretaries and PAs into cyber and AI specialists.

It is also about teaching people right from elementary school. I think we have failed in discussing education and technology. We just started talking about it 10 years ago. I hope that in the next five years we actually take ourselves seriously and start training people in a different way, not just internally in the organization, but also all the way from elementary school.



“I HOPE THAT IN THE NEXT FIVE YEARS WE ACTUALLY TAKE OURSELVES SERIOUSLY AND START TRAINING PEOPLE IN A DIFFERENT WAY, NOT JUST INTERNALLY IN THE ORGANIZATION, BUT ALSO ALL THE WAY FROM ELEMENTARY SCHOOL.”

---

## WHAT IS THE FUTURE OF ETHICAL AI?

---



“In creating a comprehensive code of ethics, it is important to focus on ensuring consumer confidence in decision making, especially for safety-critical applications.”

.....  
– Daniela Rus, MIT CSAIL



“The first step is to realize that ethical AI is important because it will allow you to drive adoption and develop sustainable technologies that will comply with current and future regulations.”

.....

– Marcin Detyniecki, AXA



“You need to define what you stand for as an organization and then create a methodology to assess the extent to which your use of AI is currently meeting those values.”

.....

– Nicolas Economou, H5

## CAPGEMINI PERSPECTIVE

---

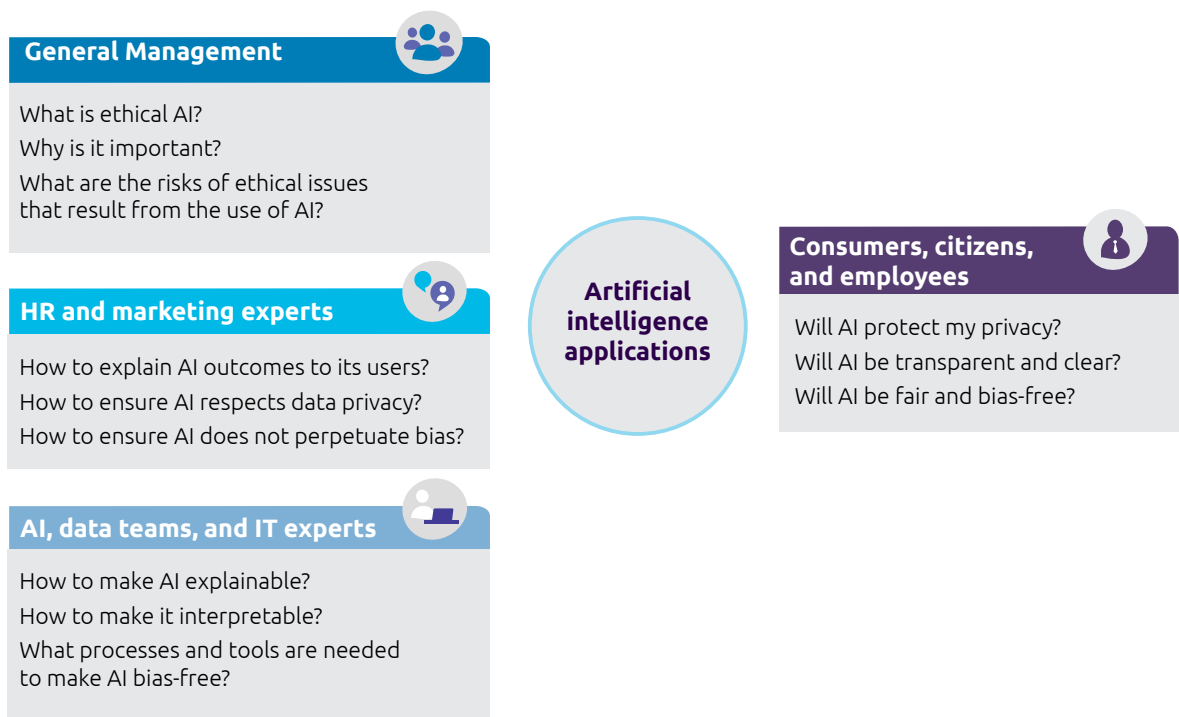
An abstract graphic featuring a blue paper airplane with a white circular cutout, flying towards the right. The background is a dark blue gradient with several curved, glowing lines in shades of blue, green, and red, suggesting motion and technology. The overall aesthetic is modern and futuristic.

# WHY ADDRESSING **ETHICAL** QUESTIONS IN **AI** WILL BENEFIT ORGANIZATIONS

## INTRODUCTION

Artificial intelligence offers a huge opportunity for businesses and the economy, but significant questions are being raised about the ethical issues surrounding this technology. To examine these questions more closely, we have undertaken this research to understand the current relationship between AI applications and their users. As Figure 1 shows, we examined both sides of the debate – from a business and end-user perspective.

**Figure 1.** Scope of our research on ethics in AI



Source: Capgemini Research Institute analysis.

We surveyed over 1,500 industry professionals from 500 organizations; over 4,400 consumers; and conducted in-depth interviews with over 20 industry executives, academics, and startup entrepreneurs (see the research methodology at the end of the paper for more details).

Through this paper, we want to demonstrate:

1. Why it is important to pursue ethics in AI from a business perspective. Our analysis shows that individuals (both consumers and employees) trust organizations that they perceive to be using AI ethically and are willing to advocate for them.
2. Why most organizations have encountered ethical issues in AI over the last two to three years.
3. How organizations can start to address ethics in AI more proactively. We therefore identified clear and actionable first steps while designing, developing, and using AI applications.



## WHAT IS AI?

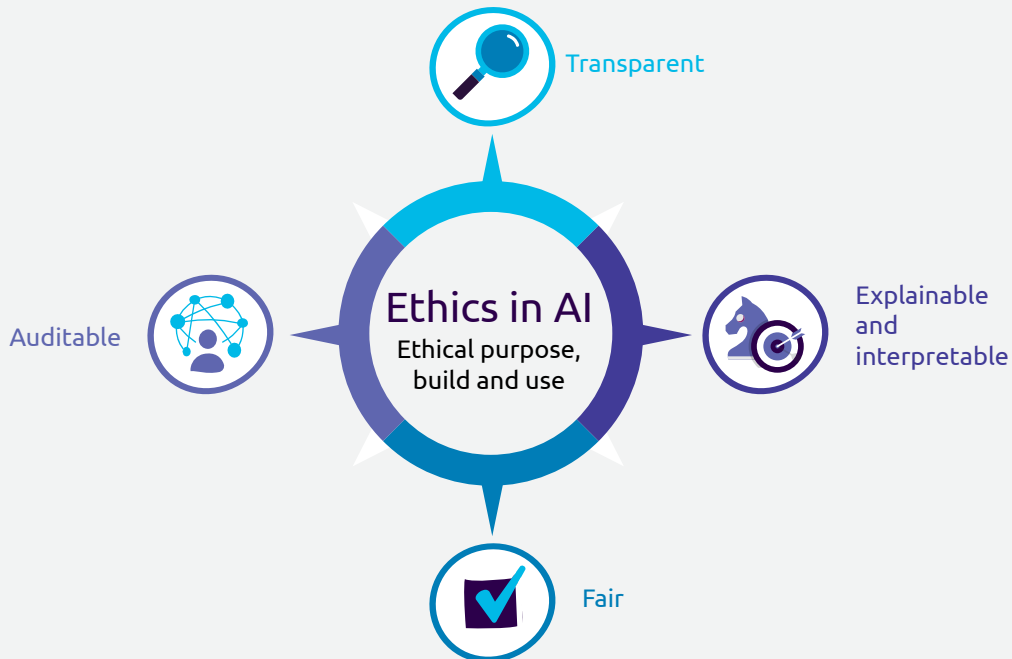
Artificial intelligence (AI) is a collective term for the capabilities shown by learning systems that are perceived by humans as representing intelligence. Today, typical AI capabilities include speech, image and video recognition, autonomous objects, natural language processing, conversational agents, prescriptive modeling, augmented creativity, smart automation, advanced simulation, as well as complex analytics and predictions.

## WHAT DO WE MEAN BY ETHICS IN AI?

According to the European Commission, the ethics of AI is a sub-field of applied ethics and technology that focuses on the ethical issues raised by the design, development, implementation, and use of AI.<sup>1</sup>

Key components of ethical AI include:

- Being ethical in its purpose, design, development, and use
- Transparent AI: AI where it is clear, consistent, and understandable in how it works
- Explainable AI: AI where you can explain how it works in language people can understand
- Interpretable AI: AI where people can see how its results can vary with changing inputs
- Fair AI: AI that eliminates or reduces the impact of bias against certain users
- Auditable AI: AI that can be audited, allowing third-parties to assess data inputs and provide assurance that the outputs can be trusted.



Source: Capgemini Research Institute.

1. European Commission High-Level Expert Group on Artificial Intelligence, "Ethics Guidelines for Trustworthy AI," April 2019.

## GETTING ETHICS IN AI RIGHT WILL BENEFIT ORGANIZATIONS

*“Trust is something very difficult to gain and very easy to lose. But a classic way of gaining trust, with AI interactions in particular, can be summarized in three words: transparency, accountability, and empowerment. That means transparency so that people can see what you are doing; accountability because you take responsibility for what you are doing; and empowerment because you put people in charge to tell you if something you did was not right or not good.”*

— Luciano Floridi, professor of Philosophy and Ethics of Information and director of Digital Ethics Lab, Oxford Internet Institute, University of Oxford<sup>2</sup>

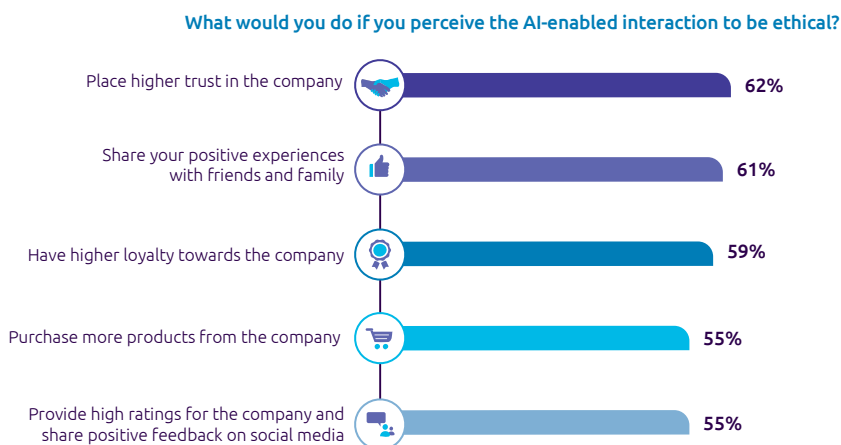
Our previous research on AI’s role in the customer experience established that nearly three-quarters of consumers say that they are aware of having interactions enabled by AI.<sup>3</sup> They see great benefits in these interactions – greater control, 24/7 availability, and convenience. With our current research, we see that organizations can build on these benefits if consumers perceive AI interactions to be ethical.

## ETHICAL AI INTERACTIONS DRIVE CONSUMER TRUST AND SATISFACTION

Ethical AI interactions earn consumer trust and build satisfaction. Three in five consumers who perceive their AI interactions to be ethical<sup>4</sup> place higher trust in the company, spread positive word of mouth, and are more loyal (see Figure 2). A positive perception can also have a tangible impact on the top line as well. Over half of the consumers we surveyed said that they would purchase more from a company whose AI interactions are deemed ethical.

The Net Promoter Score (NPS®) provides a measure of the positive impact of ethics in AI. Our research found that organizations that are seen as using AI ethically have a 44-point NPS® advantage over those seen as not.

**Figure 2.** Consumers become advocates when they perceive AI interactions as ethical



Source: Capgemini Research Institute, Ethics in AI consumer survey, N = 4,447 consumers.

- All quotes in this paper that are not attributed to a public source come from interviews directly conducted by the Institute.
- Capgemini Research Institute, “The Secret to Winning Customers’ Hearts with Artificial Intelligence: Add Human Intelligence,” July 2018.
- We explained to the individuals we surveyed what ethical AI interactions mean (for example, ethical in purpose, transparent, explainable, fair, etc.) and gave them examples of use of AI that may result in ethical issues to gauge their perception.

62% OF CONSUMERS WILL PLACE HIGHER TRUST IN THE COMPANY IF THEY PERCEIVE AI-ENABLED INTERACTIONS AS ETHICAL

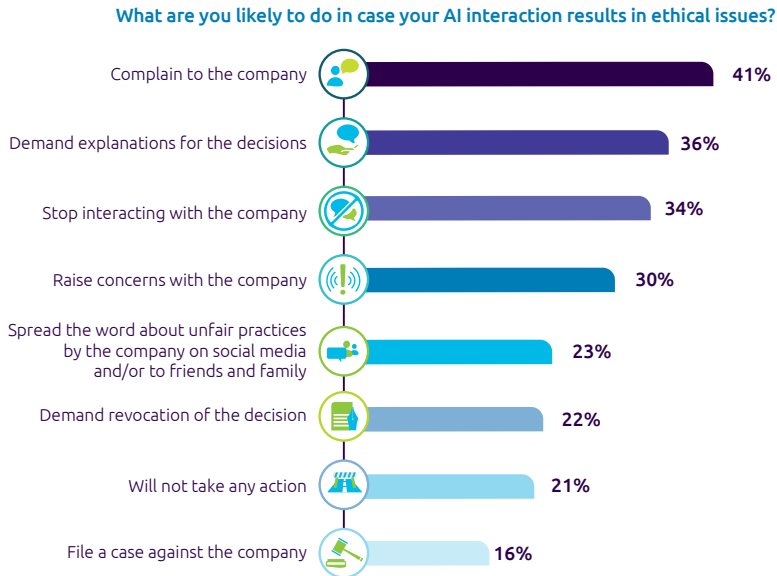
### AI INTERACTIONS THAT ARE PERCEIVED AS UNETHICAL CAN HARM BUSINESS AND DAMAGE BRAND REPUTATION

While there is an ethical responsibility on teams to put in place the checks and balances for fairness and transparency, there is also a significant business case for it. *“If a data science team is working on a machine learning project that will be affecting humans, I think that they have both ethical and commercial responsibility to do basic*

*disparate impact analysis,”* says Patrick Hall, senior director of Product, H2O.ai, an open source machine learning and artificial intelligence platform. *“Even if you personally don’t care about fairness – and you think it’s some kind of far-out liberal cause – you’ll do financial and reputational harm to your employer if your model is discriminatory.”*

If consumers decide to take action when their AI interaction results in ethical issues, (see insert in executive summary for examples of ethical issues in AI) as Figure 3 shows, nearly two in five consumers would complain to the company and demand explanations, a third of them (34%) can even stop interacting with the company – potentially causing loss of business and negative word of mouth.

**Figure 3.** AI interactions resulting in ethical issues can backfire on organizations



Source: Capgemini Research Institute, Ethics in AI consumer survey, N = 4,447 consumers.

# MOST ORGANIZATIONS HAVE ENCOUNTERED ETHICAL ISSUES IN AI OVER THE LAST 2–3 YEARS

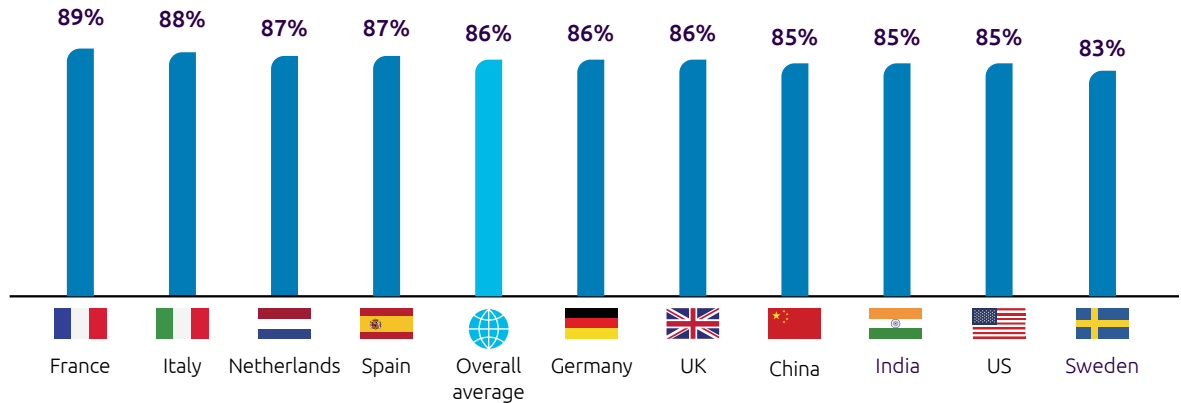
EXECUTIVES IN NINE OUT OF TEN ORGANIZATIONS BELIEVE THAT ETHICAL ISSUES HAVE RESULTED FROM THE USE OF AI SYSTEMS OVER THE LAST 2–3 YEARS

Our research shows that executives have witnessed at least one instance of the use of AI systems where ethical issues have resulted and close to half of consumers say they have experienced the impact of an ethical issue:

- 86% of executives say they are aware of instances where AI has resulted in ethical issues
- 47% of consumers say they have experienced the impact of an ethical issue
- 77% of executives are uncertain about the ethics and the transparency of their AI systems.

**Figure 4.** Nearly nine in ten organizations across countries have encountered ethical issues resulting from the use of AI

In the last 2–3 years, have the below issues resulting from the use and implementation of AI systems, been brought to your attention? (percentage of executives, by country)



We presented over 40 cases where ethical issues could arise from the use of AI, to executives across sectors. We asked them whether they encountered these issues in the last 2–3 years.

Source: Capgemini Research Institute, Ethics in AI executive survey, N = 1,580 executives, 510 organizations.

Many of these issues are not intentional but are the result of not having the right checks and balances during the development and deployment of these systems. According to the chief digital officer of a large European consumer products firm we interviewed, organizations currently lack the processes to check for unintended impact of using AI: *“More often than not, AI bias does not come from the people who program the algorithm – they have honest intentions. I think that the bias often comes from the data you feed into*

*the system. This is because the data is basically historical data and historical data is not devoid of bias – it just shows you what consumers did in the past. Therefore, it is best to not fully depend on a historical view of data, but also factor in the socio-economic context.”*

As Table 1 shows, the most common issue from executives’ perspective is in the banking sector, where they think that banks use machine-led decisions without disclosure.

**Table 1.** Top ten ethical issues resulting from use of AI: by executive awareness

Top ten issues across all sectors (in decreasing order of share of executives who have encountered these issues)	
1	Over reliance on machine-led decisions without disclosure in the banking sector
2	Collecting and processing patients’ personal data in AI algorithms without consent
3	Biased/unclear recommendations from an AI-based system for diagnosis/care/treatment
4	Over reliance on machine-led decisions without disclosure in the insurance sector
5	Citizens objecting to use of facial recognition technology by police force for mass surveillance
6	Limiting access/discriminatory pricing of services/products due to consumers’ race/gender, etc. (originally part of the company’s target group)
7	Discriminatory pricing of insurance policies due to consumers’ demographic profiles
8	Processing patients’ personal data in AI algorithms for purposes other than for which it was collected
9	Customers demanding reasoning/clarity behind a decision taken by an AI algorithm to deny credit
10	Citizens objecting to the collection and use of their personal data such as biometrics by an AI system

Source: Capgemini Research Institute, Ethics in AI executive and consumer survey, N = 1,580 executives, 510 organizations; and 4,447 consumers; sample size for each sector is different – banks: 124, healthcare: 54, insurance: 127, public sector: 74.

## THE PRESSURE TO IMPLEMENT AI IS FUELING ETHICAL ISSUES

When we asked executives why ethical issues resulting from AI are an increasing problem, the top-ranked

reason was the pressure to implement AI (see Figure 5). This pressure could stem from the urgency to gain a first-mover advantage, acquiring an edge over competitors in an innovative application of AI, or simply to harness benefits that AI has to offer.

**Figure 5:** Top reasons behind ethical issues in AI






Source: Capgemini Research Institute, Ethics in AI executive survey, N = 1,580 executives, 510 organizations.

**34%** OF EXECUTIVES IDENTIFIED PRESSURE TO URGENTLY IMPLEMENT AI WITHOUT ADEQUATELY ADDRESSING ETHICAL ISSUES AS ONE OF THE TOP ORGANIZATIONAL REASONS FOR BIAS, ETHICAL CONCERNS, OR LACK OF TRANSPARENCY IN AI SYSTEMS

We probed the second-ranked reason, that ethical issues were not considered when building AI, in a separate question. The finding is consistent – about one in three organizations (37%) report to focus significant attention

on ethical issues when implementing AI systems, and only about four in ten organizations (44%) are prepared to mitigate ethics issues in AI.

**Table 2:** Top organizational reasons identified for bias, ethical concerns, or lack of transparency in AI systems – by function (in decreasing order of importance)

	 <b>General management and ethics professionals</b>	 <b>HR and marketing professionals</b>	 <b>AI, data, and IT professionals</b>
1	Pressure to urgently implement AI without adequately addressing ethical issues	Pressure to urgently implement AI without adequately addressing ethical issues	Lack of ethical AI code of conduct or ability to assess deviation from it
2	Ethical issues were not considered while constructing AI systems	Lack of resources (funds, people, technology) dedicated to ethical AI systems	Lack of relevant training for developers building AI systems
3	Lack of a diverse team with respect to race, gender, etc. developing AI systems	Lack of relevant training to developers building AI systems	Ethical issues were not considered when constructing AI systems
4	Lack of ethical AI code of conduct or a deviation from it	Ethical issues were never considered when constructing AI systems	Pressure to urgently implement AI without adequately addressing ethical issues
5	Lack of relevant training for developers building AI systems	Lack of ethical AI code of conduct or to assess a deviation from it	Lack of resources (funds, people, technology) dedicated to ethical AI systems

Source: Capgemini Research Institute, Ethics in AI executive survey, N = 1,580 executives, 510 organizations.

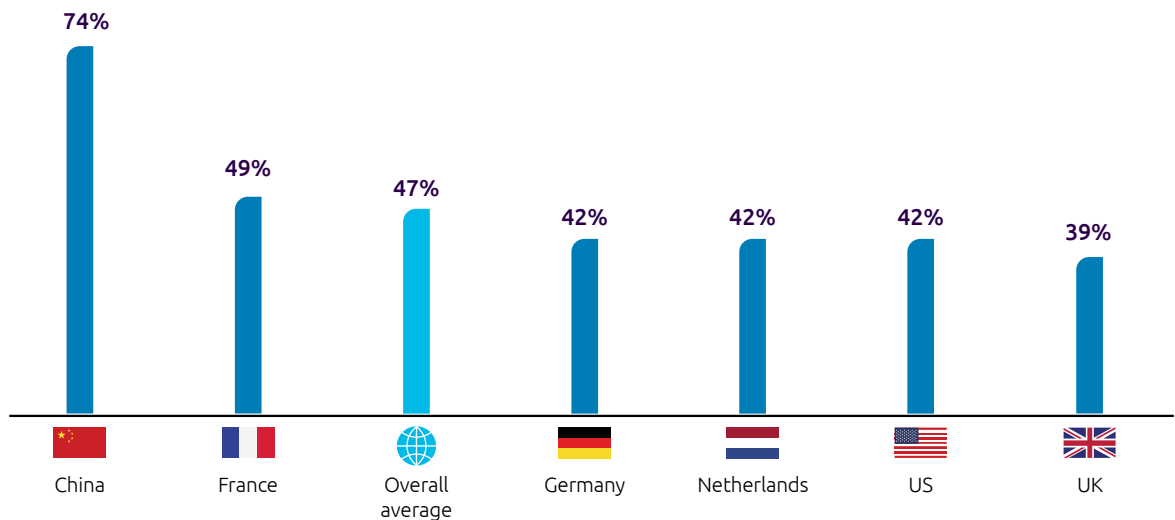
**47%** OF CONSUMERS BELIEVE THEY HAVE EXPERIENCED AT LEAST TWO TYPES OF USE OF AI THAT RESULTED IN ETHICAL ISSUES IN PAST 2–3 YEARS

## CLOSE TO HALF OF CONSUMERS FEEL THEY HAVE BEEN EXPOSED TO ETHICAL ISSUES RELATED TO AI

As we have seen, close to half of consumers (47%) feel they have been exposed to more than two instances of ethical concerns resulting from the use of AI in the last two to three years (see Figure 6). To probe this matter, we gave survey respondents specific instances of unethical practices, e.g. reliance on machine-led decisions without disclosure (see appendix at the end of the report for more details).

**Figure 6.** Close to half of the consumers believe to have been exposed to some use of AI that resulted in ethical issues

### Share of consumers who believe they have experienced at least two types of use of AI that resulted in ethical issues in past 2–3 years – by country



Source: Capgemini Research Institute, Ethics in AI consumer survey, N = 4,447 consumers.

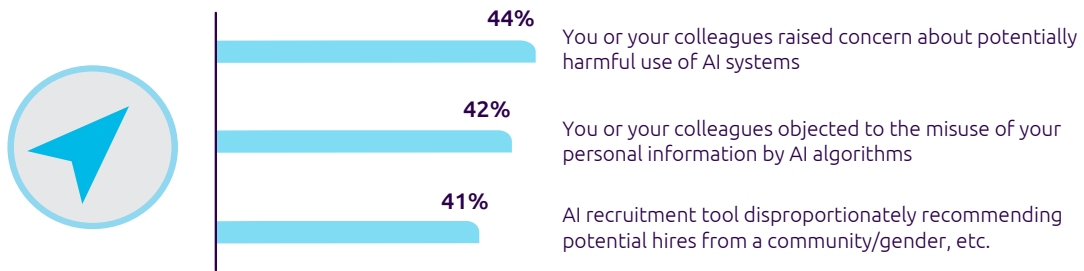


## TWO IN FIVE EMPLOYEES HAVE EXPERIENCED ETHICAL ISSUES THEMSELVES OR SEEN IT WITH THE GENERAL PUBLIC

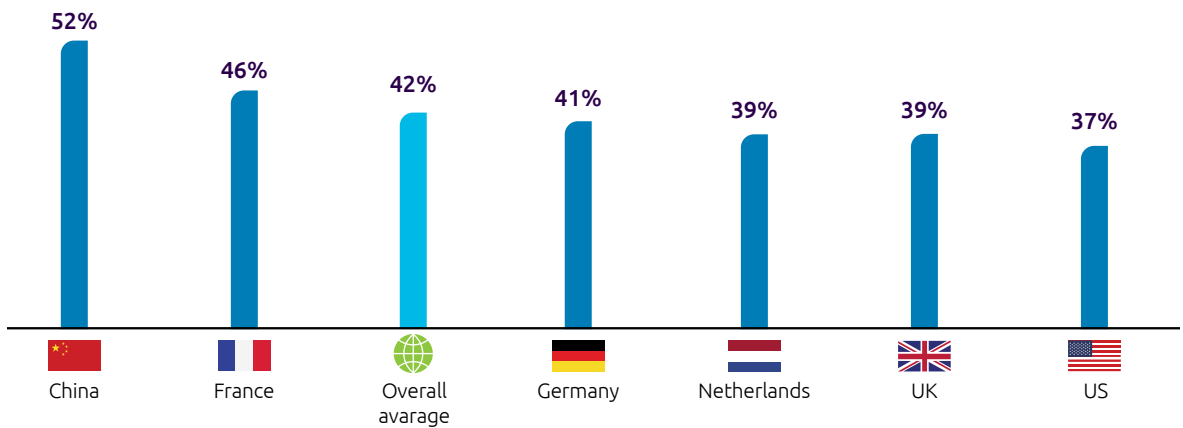
At least 40% of employees have come across some form of AI use that resulted in ethical issues (see Figure 7). Not only are employees aware about ethical issues in AI, they are also raising concerns about the use of such systems. As we found in our research, 44% of employees have raised concerns about the potentially harmful use of AI systems and 42% of employees have objected to the misuse of personal information by the AI systems and 41% of employees have objected to the misuse of personal information by the AI systems.

**Figure 7.** Employees are raising concerns about potential ethical issues resulting from the use of AI systems

**As an employee, have you experienced the following issues in your interactions with organizations?**



**Average share of employees who believe to have experienced use of AI by their organization that resulted in ethical issues – by country**



Source: Capgemini Research Institute, Ethics in AI consumer survey, N = 3,288 employees.

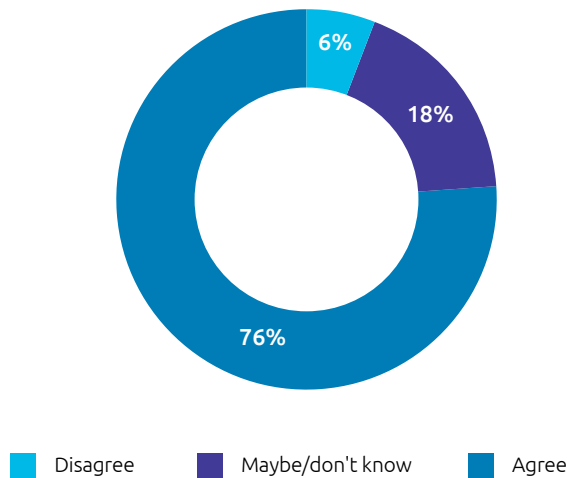
## CONSUMERS WANT REGULATION ON THE USE OF AI

Given that many consumers feel they have experienced ethical issues, it is not surprising that over three-quarters (76%) expect new regulations on the use of AI (see Figure 8). This is fueled in part by rising awareness of ethical issues in AI, as well as the positive perception of recent data privacy regulations, such as GDPR. *"Today, we don't*

*really have a way of evaluating the ethical impact of an AI product or service,"* says Marija Slavkovic, associate professor at University of Bregent. *"But it doesn't mean that regulation or law will not catch up with bad actors eventually. Organizations or individuals may not be inherently evil, but we as a society need to develop a way in which we can systematically evaluate the ethical impact of AI to account for negligence."*

**Figure 8.** Consumers want regulations on the use of AI

**Do you think there should be a new law or regulation to regulate the use of AI by organizations?**  
(percentage of consumers)



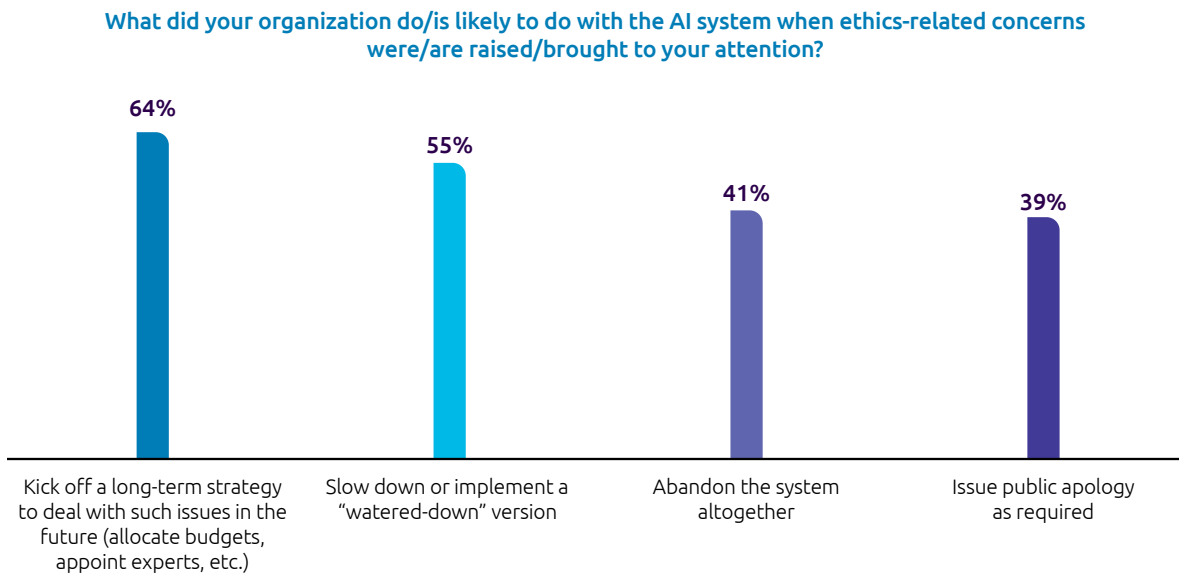
Source: Capgemini Research Institute, Ethics in AI consumer survey, N = 4,447 consumers.

**76%** OF CONSUMERS EXPECT NEW REGULATIONS ON THE USE OF AI

## EXECUTIVES ARE STARTING TO REALIZE THE IMPORTANCE OF ETHICAL AI AND ARE TAKING ACTION WHEN ETHICAL ISSUES ARE RAISED

Our research shows that 51% of executives consider that it is important to ensure that AI systems are ethical and transparent. Organizations are also taking concrete actions when ethical issues are raised. As Figure 9 shows, more than two in five executives report to have abandoned an AI system altogether when an ethical issue had been raised.

**Figure 9.** When ethical issues are raised, organizations are taking action to address the concerns



Source: Capgemini Research Institute, Ethics in AI executive survey, N = 1,580 executives, 510 organizations.

**51%** OF EXECUTIVES CONSIDER THAT IT IS IMPORTANT TO ENSURE THAT AI SYSTEMS ARE ETHICAL AND TRANSPARENT

## FIRST STEPS TO PROACTIVELY ADDRESSING ETHICS IN AI

Ethics is a challenging and complex topic that does not stand still, and requires that organizations constantly update their approach. As a start point, organizations need to build a long-term strategy that involves all departments focused on AI planning, development, and deployment, with the goal of building ethics-by-design principles into their AI systems. This requires:

- Understanding the ethical vision of the company and trade-offs against other aspects of company strategy.
- Putting in place policy and governance to give these goals weight within the organization.
- Operationalizing the strategy across the systems, people and processes involved in developing and using AI systems.

Scotiabank, for example, has set a vision for its interactive AI systems – they need to improve outcomes for customers, society, and the bank. The bank also monitors systems for unacceptable outcomes and to ensure there is accountability for any mistakes, misuse, or unfair results.<sup>5</sup>

On the basis of our extensive primary research and our conversations with industry experts, startup executives and leading academics in this field, we suggest a three-pronged approach to building a strategy for ethics in AI that embraces all key stakeholders:

- 1. General management:** CXOs, business leaders, and those with trust and ethics in their remit who will be responsible for laying the foundational practices and processes for ethical AI, including defining an ethical purpose for using AI
- 2. The customer and employee-facing teams such as HR, marketing, communications, and customer service** – who are responsible for designing the finality and intent of the use of AI in their processes and tasks, and are accountable of deploying AI ethically for users
- 3. AI, data, and IT** leaders and their teams, who will be responsible for the ethical technology design, development, deployment, and monitoring of AI systems (see Figure 10).

5. The Globe and Mail, "The AI revolution needs a rulebook. Here's a beginning," December 2018.

**Figure 10.** Formulating an ethical AI strategy

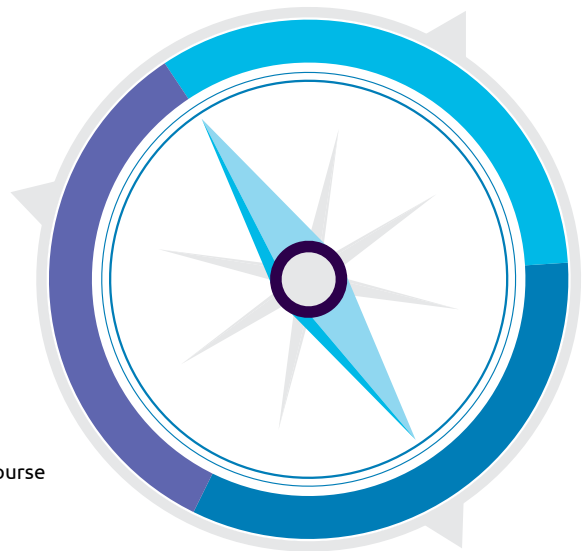
### For CXOs, ethics, and business leaders

- Lay down a strong foundation with a strategy and code of conduct for ethical AI
- Develop policies that define acceptable practices for the workforce
- Build awareness of ethical issues across the organization
- Create ethics governance structures and ensure accountability for AI systems
- Build diverse teams to inculcate sensitivity to ethical issues.



### For customer- and employee-facing teams such as HR and Marketing

- Ensure ethical usage of AI systems
- Educate and inform users to build trust in AI systems
- Empower users with more control and ability to seek recourse
- Proactively communicate on AI issues internally and externally to build trust.



### For AI, data, and IT teams

- Make AI systems transparent and understandable to gain users' trust
- Practice good data management and mitigate potential biases in data
- Use technology tools to build ethics in AI



Source: Capgemini Research Institute analysis

## 1. FOR CXOs, BUSINESS LEADERS, AND THOSE WITH A REMIT FOR TRUST AND ETHICS

**Establish a strong foundation with a strategy and code of conduct for ethical AI:** Leadership teams need to start by developing a long-term strategy and code of conduct for ethical AI that gives the rest of the organization a roadmap to follow and boundaries to respect. This requires financial investment in building ethical AI – for example, through research and development, hiring external experts, investing in external events and committees and technological investments. But it also means giving departments sufficient time so that data and customer-facing teams do not feel rushed to deploy AI systems without checking the ethical implications of data, data models, algorithms, and the AI systems as a whole.

The code of conduct is a statement on AI ethics that recognizes the trade-offs involved and guides the organization on how to prioritize ethical questions against other business objectives.

*“The first step is to define a process,”* says Nicolas Economou, co-chair, Law Committee of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, and CEO of H5, a Silicon Valley-based consultancy and technology firm.

*“What does it mean to implement digital and AI ethics? Beyond what is legal or not – and therefore what you must comply with – you need to determine what you stand for as an organization – what are your brand values? These values should exist whether you use AI or not. You can then define your ethical AI code on the basis of what you stand for; what that implies for how you think of the impact of your decisions on your company, employees, customers, and society at large; and, as a result, determine what kind of AI practices you can deem conformant to your ethics.”*

- Organizations can develop a code by drawing on widely acknowledged frameworks, such as the “Ethics Guidelines for Trustworthy AI,” from the European Commission’s High-Level Expert Group (HLEG) on Artificial Intelligence. “The EU HLEG has taken a first step towards the development of good normative rules around ethical AI that organizations can adapt.” says Luciano Floridi, professor of Philosophy

- and Ethics of Information and director of Digital Ethics Lab, Oxford Internet Institute, University of Oxford. In our research, only 16% of organizations were “highly influenced” by external benchmarks when designing or implementing ethical and transparent AI systems.
- To build a code of conduct, organizations also need to collaborate with others, including academic institutions and government and regulatory bodies. However, they also need to involve diverse stakeholder groups, encompassing employees, customers and the wider society within which they operate.
- Some organizations have already developed AI principles and ethical guidelines or are in the process of doing so. These organizations are also taking active steps to ensure that the code of conduct is translated into ground-level action. Telefónica, for example, which published its ethical guidelines for AI applications last year, says that it will assess all projects that include AI in accordance with its guidelines.<sup>6</sup> It will apply these principles as rules in all markets in which it operates, throughout its value chain, and across partners and providers. In addition, Standard Chartered Bank is developing a framework to ensure fairness, ethics, accountability and transparency (FEAT) in the Group’s use of AI.<sup>7</sup>

**Develop policies that define acceptable practices for the workforce and users of AI applications:** An organization’s ethical code of conduct must be translated into practice through policies that combine the goals of the ethics statement along with applicable regulations and industry best practice. These policies must define boundaries for the workforce, giving them a framework in which to operate and ensuring they know what is acceptable practice. For instance, a large, US-based healthcare organization that we spoke to has designed strict policies to restrict access to customer data, which is classified based on its sensitivity. As part of their approach, it has a rigorous approval process for when personally identifiable and protected customer health information can be accessed for use in AI applications. Those requesting access must have a strong justification for why they need to use a data set.

6. Telefónica press release, “Telefónica is committed to using Artificial Intelligence with integrity and transparency,” October 2018.

7. Standard Chartered, “Annual Report 2018,” March 2019.

**Build awareness of ethical questions across the organization:** Organizations must build awareness across functions and organizational layers – of ethics, transparency, explainability, interpretability, and bias in AI systems. In particular, teams building and deploying AI systems need to be fully aware of these issues if they are to mitigate any ethical risks or weaknesses.

Danya Glabau, faculty member at the Brooklyn Institute for Social Research, points out that organizations will need to build their employees' skills and understanding in what will be a new field for many. *"Thinking about employee education, organizations need to rethink what skills, knowledge, and experiences they expect employees to bring in,"* she says. *"These skills may not follow the typical engineering or executive pathway. Organizations need to think how they can build the teaching of these skills into their employee learning programs, so that there are resources available for employees who are thinking about ethics in AI and other such issues."*

**Create ethics governance structures and ensure accountability for AI systems:** Leadership teams also need to create clear roles and structures, assign ethical AI accountability to key people and teams and empower them. Key steps can include:

- Adapting existing governance structures to build accountability within certain teams. For example, the existing ethics lead (e.g., the Chief Ethics Officer) in the organization could be entrusted with the responsibility of also looking into ethical issues in AI
- Creating new roles, such as AI ethicists – potentially with a background in business ethics, compliance, and also with an understanding of how that applies to AI – and AI leads who can be held accountable for good AI practices
- Assigning senior leaders who would be held accountable for ethical questions in AI
- Building internal/external committees responsible for deploying AI ethically, which are independent and therefore under no pressure to rush to AI deployment. *"Organizations need to ask the question as to how will they ensure that AI is accountable,"* says Ryan Budish, assistant director of Research at Berkman Klein Center for Internet & Society at Harvard University. *"One way to do it is through a top-down, system-wide approach where the organization thinks about the sort of standards needed to hold these systems accountable. These could be ethical standards or normative standards or political standards. There could be any number of perspectives that shape the standards used to hold these systems accountable. Such standards can also emerge bottom-up, in a more organic, iterative fashion."*

"WHAT DOES IT MEAN TO IMPLEMENT DIGITAL AND AI ETHICS? BEYOND WHAT IS LEGAL OR NOT – AND THEREFORE WHAT YOU MUST COMPLY WITH – YOU NEED TO DETERMINE WHAT YOU STAND FOR AS AN ORGANIZATION – WHAT ARE YOUR BRAND VALUES? THESE VALUES SHOULD EXIST WHETHER YOU USE AI OR NOT. YOU CAN THEN DEFINE YOUR ETHICAL AI CODE ON THE BASIS OF WHAT YOU STAND FOR; WHAT THAT IMPLIES FOR HOW YOU THINK OF THE IMPACT OF YOUR DECISIONS ON YOUR COMPANY, EMPLOYEES, CUSTOMERS, AND SOCIETY AT LARGE; AND, AS A RESULT, DETERMINE WHAT KIND OF AI PRACTICES YOU CAN DEEM CONFORMANT TO YOUR ETHICS."

**— NICOLAS ECONOMOU,**

CO-CHAIR, LAW COMMITTEE OF THE IEEE GLOBAL INITIATIVE ON ETHICS OF AUTONOMOUS AND INTELLIGENT SYSTEMS, AND CEO OF H5

**Build diverse teams to ensure sensitivity towards the full spectrum of ethical issues:** To ensure that algorithms are bias-free, and AI systems are as ethical as possible, it is important to involve diverse teams. For example, organizations not only need to build more diverse data teams (in terms of gender or ethnicity), but also actively create inter-disciplinary teams of sociologists, behavioral scientists and UI/UX designers who can provide additional perspectives during AI design. *“I think this is very important because many of these systems will be implemented in different areas for people with different backgrounds,”* says Christoph Luetge, director of the TUM Institute for Ethics in Artificial Intelligence at Technical University of Munich.

*“You cannot just assume that you are dealing with some specific group of people only. And, as we know from many famous examples, there might be a problem where systems, for example, assume that people only have a certain skin type. It’s very important to involve people with many different backgrounds right from the start.”* GE Healthcare, for instance, has committed to employing a diverse workforce in teams working on AI. Eighty percent of their data science team have a minority background, 44% who sit outside the US, and 26% are women.<sup>8</sup> They also have

“YOU CANNOT JUST ASSUME THAT YOU ARE DEALING WITH SOME SPECIFIC GROUP OF PEOPLE ONLY. AND, AS WE KNOW FROM MANY FAMOUS EXAMPLES, THERE MIGHT BE A PROBLEM WHERE SYSTEMS, FOR EXAMPLE, ASSUME THAT PEOPLE ONLY HAVE A CERTAIN SKIN TYPE. IT’S VERY IMPORTANT TO INVOLVE PEOPLE WITH MANY DIFFERENT BACKGROUNDS RIGHT FROM THE START.”

### — CHRISTOPH LUETGE,

DIRECTOR OF THE TUM INSTITUTE FOR ETHICS IN ARTIFICIAL INTELLIGENCE AT TECHNICAL UNIVERSITY OF MUNICH

mature data practices – all data is clearly sourced and how it can be used is clearly specified.

## 2. FOR CUSTOMER- AND EMPLOYEE-FACING TEAMS SUCH AS HR AND MARKETING

**Ensure ethical usage of AI systems:** Customer- and employee-facing teams, such as marketing, communications and HR, must ensure AI systems are transparent, explainable, and free of bias for end users.

Working in collaboration with the AI, data and IT teams, they must be empowered, from the first design of the AI application, to define finality and intent of an AI application very clearly, and the corresponding transparency towards end users. The finality and intent as defined by business users would serve as the cornerstone for the proper design, development, and testing phase of the AI application, including any possible impact on users.

End-user testing, drawing on a small set of pilot users, can help weed out adverse effects before the system goes into operation. Up-front action can ensure that AI systems do not result in major ethical problems later. For example:

1. Human-resources professionals must put in place measures for AI systems used in the following areas:
  - Recruitment: that they are fair and non-discriminatory in selecting potential hires from all communities, gender, age, or race/ethnicity
  - Performance reviews: that systems show no bias towards any particular community, gender, and race/ethnicity
  - Workplace surveillance: that systems have the full consent of employees and their purpose is made clear to employees before deployment
  - Collecting and processing employee data: that systems operate with the consent of employees

8. GE Healthcare, “Ethics in healthcare aren’t new, but their application has never been more important,” October 2018.



“ORGANIZATIONS DEPLOYING AI SYSTEMS ARE ACCOUNTABLE FOR NOT ABUSING THE INFORMATION THAT THEY HAVE OR THE TRUST THAT THEY HAVE BEEN GIVEN. IF THAT HAPPENS, USERS SHOULD HAVE CHANNELS TO RAISE ALARM AND COMMUNICATE THAT THEY ARE NOT HAPPY. FOR EXAMPLE, CONSIDER A USER WHO STRUGGLES WITH INFERTILITY THAT KEEPS GETTING ADVERTISEMENTS FOR DIAPERS. THIS COULD BE A PROBLEM FOR THE USER, AND SHE SHOULD HAVE INSTRUMENTS TO OBJECT TO BEING IN THIS TARGET GROUP. USERS HAVE THE RIGHT TO DEMAND THIS AND ORGANIZATIONS HAVE A RESPONSIBILITY TO PROVIDE SUCH CHANNELS.”

— MARIJA SLAVKOVIK,  
ASSOCIATE PROFESSOR AT UNIVERSITY OF BERGEN.

2. Marketing and communication professionals must ensure that AI applications:
  - Explain their workings and any outcomes when end-users request that explanation
  - Inform end-users that they are interacting with human-like chatbots and not humans
  - Are not limiting access to, or not pricing services/products differently, because of the demographics of customers
  - Do not use biased or sexist words or phrases (e.g., by chatbots/voice bots)
  - Make legally compliant use of personally identifiable data, such as medical records and biometrics, and ensuring that people understand why the data is being collected and how it will be used.

#### **Educate and inform users to build trust in AI systems:**

The customer- and employee-facing teams have the task of building user trust. Organizations can build trust with consumers by communicating certain principles: the use of AI for good, valuing human autonomy, and respecting the end consumer’s rights. Organizations should also seek to inform users every time they are interacting with AI systems. In other words, the system should not pose as a human when interacting with users. They should also educate users about potential misuse and the impact and risks of using AI systems in an environment where ethical questions are not addressed proactively.

#### **Empower users with more control and the ability to seek recourse:**

Organizations should empower individuals with the ability to access, share, and seek clarity on the use of their data. This means building policies and processes where users can ask for explanations of AI-based decisions. Consumers want more transparency to when a service is powered by AI (75% in our survey) and know if an AI is treating them fairly (73%).

Effective feedback channels for users will also help them to reach out to organizations for:

- sharing feedback or grievances
- reviewing AI decisions
- seeking explanations.

All of these will help in building greater consumer trust. *“Organizations deploying AI systems are accountable for not abusing the information that they have or the trust that they have been given,”* says Marija Slavkovik, associate professor at University of Bergen. *“If that happens, users should have channels to raise alarm and communicate that they are not happy. For example, consider a user who struggles with infertility that keeps getting advertisements for diapers. This could be a problem for the user, and she should have instruments to object to being in this target group. Users have the right to demand this and organizations have a responsibility to provide such channels.”*

### **Proactively communicate on AI issues internally and externally to build trust:**

Communication is a bedrock of transparency and trust and can be a useful vehicle to build trust among all stakeholders, especially consumers, employees, and citizens. The communications and marketing team must adopt a well defined strategy to drive internal and external communications to convey the concrete steps the organization is taking to build ethics and transparency into AI applications. Currently organizations adopting AI across their business operations consider it a competitive advantage to be “early movers.” To maintain this competitive edge, organizations often do not necessarily communicate openly on their use of AI and its impact on end users or employees. We consider that “smart movers” could turn being transparent in their use of AI into a competitive advantage by thinking “people-first,” and getting end users on their side by being as transparent as possible about their use of AI.

## 3. FOR AI, DATA, AND IT TEAMS

### **Make AI systems transparent and understandable to gain users' trust:**

Systems need to be transparent and intuitive for users and business teams. The teams developing the systems should provide the documentation and information to explain, in simple terms, how certain AI-based decisions are reached and how they affect an individual. These teams also need to document processes for data sets as well as the decision-making systems. When we asked consumers what long-term actions would convince them that companies are using AI ethically, close to eight out of ten opted for “Providing explanations for AI decisions in case I request it.” Close to an equal number opted for “Informing me about the ways in which AI decisions might affect me.”

### **Practice good data management and mitigate potential biases in data:**

While general management will be responsible for setting good data management practices, it falls on the data engineering and data science and AI teams to ensure those practices are followed through. These teams should incorporate “privacy-by-design” principles in the design and build phase and ensure robustness, repeatability, and auditability of the entire data cycle (raw data, training data, test data, etc.). The AI practitioners need to:

- ensure that data is sourced ethically and in line with what regulation permits
- check for accuracy, quality, robustness, and potential biases, including detection of under-represented minorities or events/patterns

- build adequate data labelling practices and review periodically
- store responsibly, so that it is made available for audits and repeatability assessments
- constantly monitor results produced by models as well as precision and accuracy, and test for biases or accuracy degradation.

Good data management must also involve creating checks and balances to mitigate AI bias. Teams need to particularly focus on ensuring that existing datasets do not create or reinforce existing biases. For example:

- Identifying existing biases in the dataset through use of existing AI tools or through specific checks in statistical patterns of datasets
- Being mindful of not creating a selection bias on the data when developing algorithms
- Exploring and deploying systems to check for and correct existing biases in the data set before developing algorithms
- Conducting sufficient pre-release trials and post-release monitoring to identify, regulate, and mitigate any existing biases.

### **Use technology tools to build ethics in AI:**

One of the problems faced by those implementing AI is the black-box nature of deep learning and neural networks. This makes it difficult to build transparency and check for biases. Increasingly, some companies are deploying tech and building platforms which help tackle this. IBM's AI OpenScale, for instance, gives explanations on how AI models make decisions, and detects and mitigates against bias in the datasets.<sup>9</sup> There are many other open source tools that use AI to detect existing biases in algorithms and check the decisions and recommendations that AI systems provide. These AI tools mean companies can check their data sets and algorithms and make corrections as necessary. For example, ZestFinance which helps lenders use machine learning to deploy transparent credit risk models, developed its “ZAML Fair” tool to help reduce the disparity that affects minority applicants for credit.<sup>10</sup> Some startups are also building AI-based tools that are able to look into AI systems to make them more explainable and interpretable.<sup>11</sup> Organizations can use these tools to check their AI practices, mitigate biases, and build transparency into their AI systems.

9. Forbes, “IBM Wants To Make Artificial Intelligence Fair And Transparent With AI OpenScale,” October 2018.

10. Forbes, “ZestFinance Using AI To Bring Fairness To Mortgage Lending,” March 2019.

11. H2O.ai website.

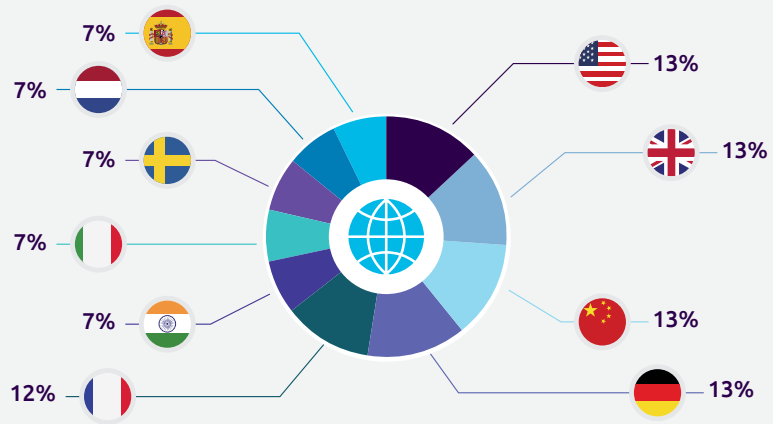
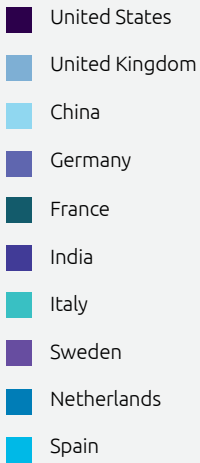
## CONCLUSION

AI offers significant benefits for organizations with the right vision, planning, and approach to implementation. Our research shows that proactively addressing ethical questions in AI from the start is now a critical step to ensuring the adoption of artificial intelligence at scale in organizations. Organizations adopting an “ethics-by-design” approach for AI will earn people’s trust and loyalty and greater market share compared to their peers. At the same time, they will stand to gain by preemptively averting significant risks from a compliance, privacy, security, and reputational perspective. Today, however, ethics in AI often does not get the attention it deserves even though some organizations are starting to take action. It is critical for organizations to establish strong foundations for why AI applications are built and how they are used. Organizations need to build accountability – for all teams involved – to infuse ethics into AI applications, by design, from day one. “Getting it right” with ethics in AI is still to be fully defined, and will evolve with the progressive adoption of AI across businesses and organizations, as well as with technology innovation. The first steps in this report will help organizations kick-start this journey towards ethical AI systems and practices.

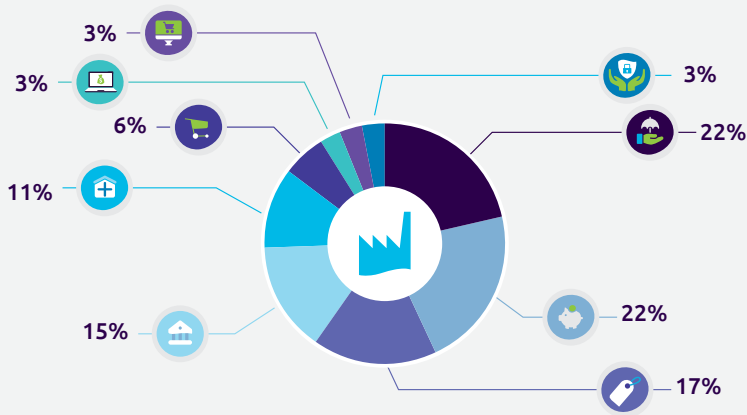
# RESEARCH METHODOLOGY

**Executive Survey:** We surveyed over 1,580 executives from large organizations (with more than \$1 billion in annual revenues each) in ten countries between April 2018 and June 2019. The executives were drawn from three broad groups:

1. General management/strategy/corporate
2. AI, data, and IT
3. HR/Marketing, with one executive from each group per organization.

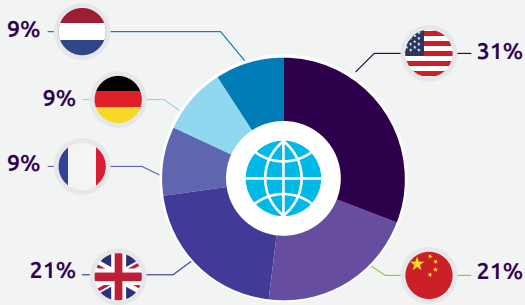


Country distribution of executives

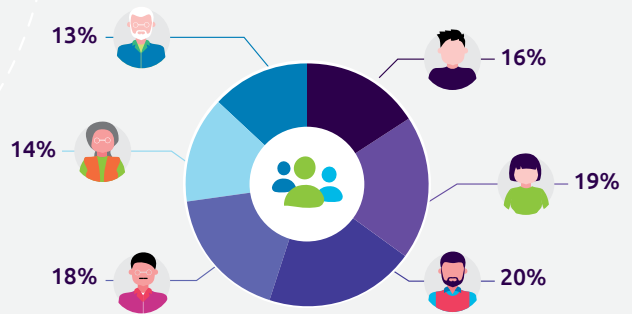


Industry distribution of executives

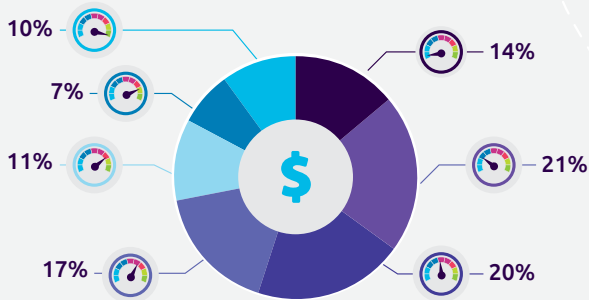
**Consumer Survey:** We surveyed 4,400 consumers from six countries. All of these consumers had had some form of AI interaction in the past so that they could relate to ethics-, trust-, and transparency- related issues in AI.



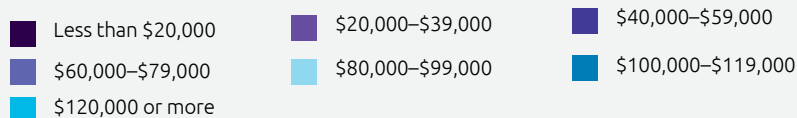
**Country distribution of consumers**



**Age distribution of consumers**



**Annual income of**



We also conducted in-depth interviews with more than 20 academics, industry experts, and entrepreneurs.

# Appendix

Top three AI use cases that caused ethical concerns among consumers – by country

Country	AI use case
China	Healthcare – Processing personal data in AI algorithms for purposes other than for which it was collected
	Public sector – Collection and use of personal data (such as biometrics) by an AI system without consent
	Insurance – Reliance on machine-led decisions without disclosure
France	Healthcare – Collecting and processing personal data in AI algorithms without consent
	Healthcare – Processing personal data in AI algorithms for purposes other than for which it was collected
	Insurance – Reliance on machine-led decisions without disclosure
Germany	Public sector – Collection and use of personal data (such as biometrics) by an AI system, without consent
	Healthcare – Biased/unclear recommendations from an AI-based system for diagnosis/care/treatment
	Insurance – Reliance on machine-led decisions without disclosure
Netherlands	Public sector – Denied, without any explanation, aid/public benefits based on an AI algorithm's decision
	Insurance – Premium was set by an AI system based on race/ethnicity/income without any explanation
	Healthcare – Collecting and processing personal data in AI algorithms without consent
UK	Healthcare – Collecting and processing personal data in AI algorithms without consent
	Healthcare – Processing personal data in AI algorithms for purposes other than for which it was collected
	Insurance – Reliance on machine-led decisions without disclosure
US	Insurance – Reliance on machine-led decisions without disclosure
	Public sector – Collection and use of personal data (such as biometrics) by an AI system without consent
	Healthcare – Processing personal data in AI algorithms for purposes other than for which it was collected



# CONVERSATIONS

---

## GUEST CONTRIBUTORS



SASKIA STEINACKER



MICHAEL NATUSCH



PAUL COBBAN



NICOLAS ECONOMOU



MARCIN DETYNIECKI



DANIELA RUS



LUCIANO FLORIDI



RYAN BUDISH  
HARVARD UNIVERSITY



CECILIA BONEFELD-DAHL

