Our Code of Ethics for AI

Capgemini

Our Code of Ethics for AI

AI is a general-purpose technology that can affect entire economies, and which is spreading very visibly beyond the business area to areas of daily life. A challenge facing both business and society today is how to optimize the opportunities offered by AI technology, whilst addressing the risks and fears that AI may generate.

Since its foundation, Capgemini places ethics at the center of its activity. As a leader in digital transformation, we are committed to the adoption of AI in a way that delivers clear benefits from AI technologies within a trusted framework, by building a Code of Ethics for AI.

Our ethical culture drives our vision of AI, guided notably by 5 of our core Values: Honesty, Trust, Boldness, Freedom, and Modesty. These Values work together to inform our approach. Boldness drives us to act as entrepreneurs, identifying and pursuing the opportunities presented by innovation in this field. We aspire to increase Freedom by empowering, complementing and augmenting human cognitive, social and cultural skills, giving people more say over how they live their lives. Modesty keeps us mindful of the need to mitigate risks, building solutions that are robust, safe, and human-centric. Honesty underpins our commitment to transparency, and to creating solutions that are accountable and controllable. We consider Trust to be an essential basis for long-standing interdependent relationships with clients, users, and all members of our ecosystem; the value we place on Trust drives our efforts to create AI that protects privacy and ensures equal access rights and fair treatment.

Our Code of Ethics for AI guides our organization on how to embed ethical thinking in our business. It is illustrated by concrete examples from projects or solutions that we deliver. Reference to its principles stimulates ethical reasoning and is intended to launch an open-ended process of discussion within the company, with our clients, and with all stakeholders.

Our Code of Ethics for AI concerns both the intended purpose of the AI solution, and the way we embed ethical principles in the design and delivery of AI solutions and services to our clients.

It should be read in combination with applicable legislations with which Capgemini is, of course, committed to comply.

AI definition

Artificial Intelligence (AI) is a collective term for the capabilities shown by learning systems that are perceived by humans as representing intelligence.

These intelligent capabilities typically can be categorized into Machine Vision & Sensing, Natural language processing, Predicting & Decision-making, and Acting & Automating.

Various applications of AI include speech, image, audio and video recognition, autonomous vehicles, natural language understanding and generation, conversational agents, prescriptive modelling, augmented creativity, smart automation, advanced simulation, as well as complex analytics and predictions.

Technologies that enable these applications include automation, big data systems, deep learning, reinforcement learning and AI acceleration hardware.

At Capgemini, we believe that human ethical values should never be undermined by the uses made of AI by business. We want AI solutions to be **human-centric**, which we define as follows:

1. **AI with carefully delimited impact** – designed for human benefit, with a clearly defined purpose setting out what the solution will deliver, to whom.
2. **Sustainable AI** – developed mindful of each stakeholder, to benefit the environment and all present and future members of our ecosystem, human and non-human alike, and to address pressing challenges such as climate change, CO_2 reduction, health improvement, and sustainable food production.
3. **Fair AI** – produced by diverse teams using sound data for unbiased outcomes and the inclusion of all individuals and population groups.
4. **Transparent and explainable AI** – with outcomes that can be understood, traced and audited, as appropriate.
5. **Controllable AI with clear accountability** – enabling humans to make more informed choices and keep the last say.
6. **Robust and safe AI** – including fallback plans where needed.
7. **AI respectful of privacy and data protection** – considering data privacy and security from the design phase, for data usage that is secure, and legally compliant with privacy regulations.

Building on our core Values, we believe that the design and delivery of AI solutions should be guided by these seven principles, aligned with the "Ethics Guidelines for Trustworthy AI" issued in 2019 by the independent High-Level Expert Group on AI set up by the European Commission.
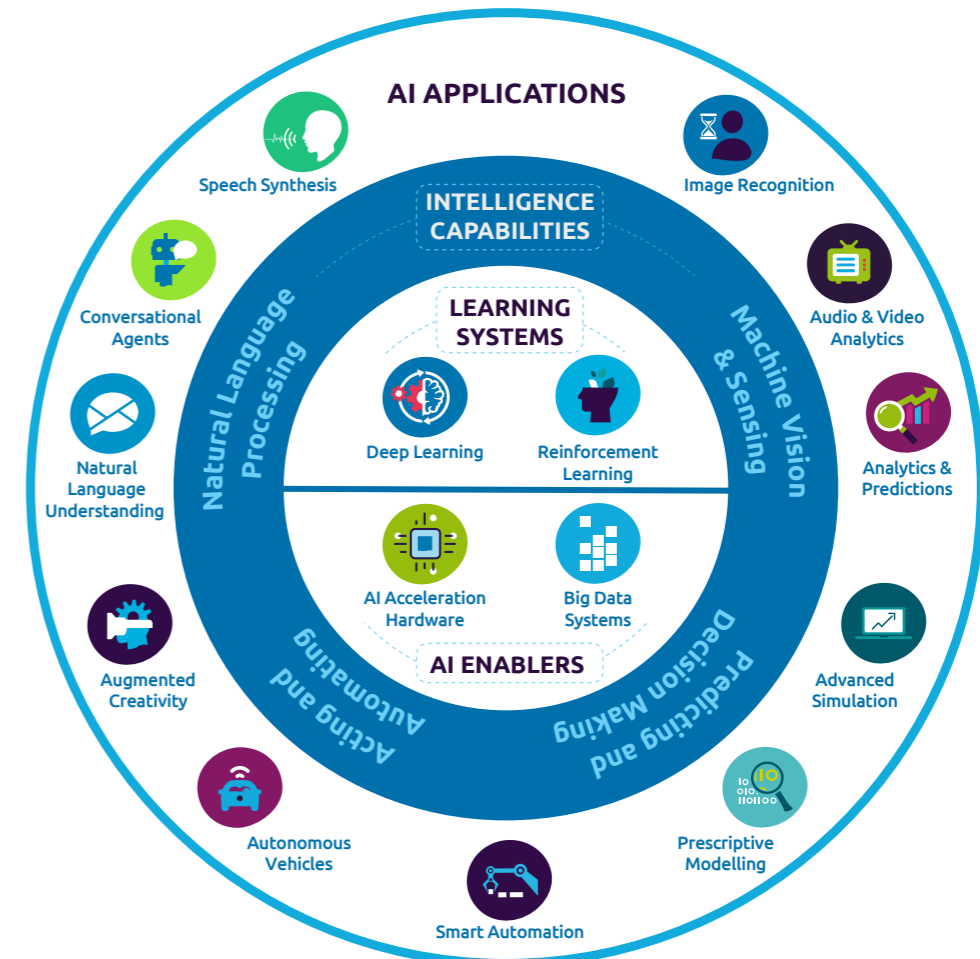
1. AI with carefully delimited impact

● ETHICAL CHALLENGE

The very first and fundamental ethical question to be considered is the intended purpose of the AI solution and its impact on humans. Like any general-purpose technology, AI solutions can equally enable and negatively affect human fundamental rights.

● CAPGEMINI'S RESPONSE

Capgemini cares about the intended purpose of AI solutions; our solutions must be mindful of the impact on humans and respect universal fundamental rights, principles and values, in particular the Universal Declaration of Human Rights and the UN Global Compact. AI must focus on improving life for humans and should neither exacerbate existing harm nor create new harm for individuals.

The intended purpose of an AI application – what the AI solution will deliver, for whom, and to whom – must be clearly defined, and AI should be used according to its intended purpose. To this end, we are transparent about the intended purpose with our various stakeholders, notably the end users, and include appropriate provisions in our agreements, clearly describing the use for which the technology is intended.

As AI is a highly evolutive technology, we believe that assessing the impact of AI solutions, notably on individuals, is important to help identify the overall impact, i.e. the likely benefits against the foreseeable risks, such as social impact or potential risk deriving from inadequate or inappropriate use. Assessing the potential impact that a new technology can have before adopting it helps identify undesired side-effects and consequent ethical risks and helps mitigate them.

In situations where there is any doubt about a potential risk of affecting fundamental rights, a fundamental-rights impact assessment will be undertaken to ensure that such a risk is eliminated.

2. Sustainable AI

● ETHICAL CHALLENGE

Beyond the direct impact on humans and human society, other beings and the environment can be impacted by AI solutions. The challenge goes beyond guiding "human-friendly AI", to ensuring "Earth-friendly AI". As the scale and urgency of the economic and health impacts from our deteriorating natural environment grows, we have an opportunity to look at how AI can help transform traditional sectors and systems to address climate change, deliver food and water security, build sustainable cities, and protect biodiversity and human wellbeing. Furthermore, AI cannot support a sustainable future if it is not itself sustainable by design.

● CAPGEMINI'S RESPONSE

AI systems should benefit all human beings. This means that their design and development should take into careful consideration the social and societal impacts. Design and development must also be mindful of future generations, the environment, and all beings – human and nonhuman alike – that make up our ecosystem. They must be considered as stakeholders throughout the AI solution's life cycle, so that AI solutions are sustainable and environmentally friendly.

We support AI to address challenges in societal areas as diverse as climate change and CO_2 reduction, digital literacy and inclusion, environmental protection, health improvement, and sustainable food production.

CASE

Global demand for food is anticipated to increase by 60% by 2050. Today, much of the world's population is fed by small-scale farmers, primarily in developing countries, using rudimentary farming practices. This agricultural inefficiency is exacerbated by a complex value chain and a lack of resources and connectivity, so there is a strong need for a wider package of yield-optimizing and risk-decreasing services for these small-scale farmers. Project FARM, created at Capgemini's Applied Innovation Exchange (AIE) Collaboration Zone (CoZone) in the Netherlands, aims to address these issues.

The Project FARM platform uses AI to determine farming patterns through big data, generating insights from the data to make recommendations. It uses machine learning to make the platform applicable at scale, by connecting it with cell phones. This solution has been built in collaboration with Agrics, a social enterprise operating in East Africa, which provides local farmers with agricultural products and services on credit.

3. Fair AI

● ETHICAL CHALLENGE

In order to be effective, AI needs to learn from historical data. The more data, the more accurate an AI system will be in terms of categorizing, predicting, prescribing, and overall decisioning. However, training data for machines, notably statistics, may reflect an organizational or individual perspective on a given subject matter, or a historical picture of reality. This perspective may be biased or incorrect, as data can include various forms of bias, resulting in extrapolations that can conflict with or undermine current trends and desired evolutions, gradually building up over time. This can result in discrimination against certain population groups based on gender, ethnicity, or similar social factors.

Likewise, unfair biases and discrimination can be built in the algorithms themselves, by design and development teams lacking appropriate diversity.

● CAPGEMINI'S RESPONSE

We embed diversity and inclusion principles throughout the entire AI system's life cycle:

- AI design and development teams must be built as diverse teams, with diversity in terms such as gender and ethnicity, but also discipline, for multiple perspectives during AI design, and sensitivity to the fullest spectrum of ethical issues.

- We seek to identify any unfair bias likely to lead to discrimination and inappropriate results in the context of decision making, and present possible correction scenarios to remove them.

- We will advise clients to put in place an oversight process to analyze and address the system's purpose, constraints, requirements, and decisions in a clear and transparent manner.

- AI systems must entail and ensure equal access rights and treatment by people (regardless of ethnicity, disability, age, religious belief, sexual orientation, or other personal characteristics).

- As an alternative to – potentially biased – historical training data, generated (i.e. synthetic) data or off-the-shelf industry data should be considered.

CASE

Capgemini Invent has developed SAIA – Sustainable Artificial Intelligence Assistance – a demonstrator designed to show our approach to prevent discrimination and make AI decisions transparent throughout the AI life cycle. It identifies potential biases and analyzes bias behavior. It also provides recommendations on ways to correct algorithm biases and simulates the impact of these corrections.

4. Transparent and explainable AI

● ETHICAL CHALLENGE

The complexity of AI may amplify the "black box" concern. A "black box" is a device, system, or program that allows input and output to be seen but gives no view of the processes or workings between the two. For example, in tools using artificial neural networks, hidden layers of nodes process the input and pass their output to subsequent layers of nodes, while in deep learning, an artificial neural network "learns" autonomously by pattern recognition. As with a human brain, one cannot see the output between layers, how data has been analyzed, or what has been "learnt" – one sees only the conclusion. Where a conclusion needs to be checked and justified, because it is unexpected, incorrect, or problematic, it can therefore be highly challenging to understand. This is of greater concern where AI functionality plays a role in high-stakes decision-making areas, such as banking, justice or health, where the potential impact of decisions is more serious.

● CAPGEMINI'S RESPONSE

AI must be transparent: its capabilities and purpose should be openly communicated. Decisions based on AI solutions should be explainable, with the degree of explicability dependent on the context and severity of the consequences if the output is erroneous. The data sets and processes used for the AI solution should also be documented to allow for traceability and, if required, for auditability.

When interacting with an AI interface, individuals should be aware that they are communicating with a machine, and should not be misled into thinking otherwise, while being informed of the AI capabilities and limits. Individuals interacting with AI should be made clearly aware about the purposes of the AI system, how it works, with whom the information may be shared, the impact of the AI solution and any potential impact on their rights, if any, in relation to the AI system at stake.

We will advise our clients to ensure with us that systems developed are explainable, especially regarding data selection and treatment, notably weightings. We will endeavor to indicate the limits that can exist in the understanding of their functioning.

Working with technologies that we understand and control, we will provide documentation and training to users to explain the logic behind the functioning of the AI and to indicate the limits of understanding and testing scenarios, in a manner adapted to the different stakeholders potentially concerned.

CASE

A world-leading bank for which AI is expected to become a significant part of operations needed to understand and compare the most popular AI explainability methods. Capgemini designed a 4-step approach, comparing the explainability models on several axes (quality of results, smoothness, source dataset impact, and consistency between the results of each method). By putting explainability at the heart of the project, the client will ensure that innovation through AI is properly understood and actionable for its teams..

5. Controllable AI with clear accountability

● ETHICAL CHALLENGE

While the control responsibilities in any IT system depend on organizing accountability, several aspects complexify this with regard to AI. The production environment itself often involves many discrete contributors, including highly specialized third parties, rendering in-built controllability and oversight more difficult. Moreover, in a legal environment largely based on the assumption of human agents, AI systems depend to a greater or lesser degree on AI-driven intelligence, autonomous agents, and autonomous decision-making. Determining responsibility for AI outcomes is further complexified by techniques such as deep learning, which can make systems hard to control and outputs difficult to explain. While individuals need to know who will be answerable in case of malfunction, or should a system have unintended consequences or cause harm, tying an AI's actions or decisions to a human, or group of humans, can therefore present a considerable challenge.

● CAPGEMINI'S RESPONSE

Humans should keep the last say on AI, and we should design AI solutions in such a way that AI cannot learn to circumvent the controls and voluntary interruptions by humans.

The design of AI solutions should protect the human's autonomy and decision making. As such, AI solutions should help humans make more informed choices.

To ensure such respect, humans should be part of the AI governance mechanism in such a way that they always keep control over AI. Appropriate measures should be implemented from the AI solution's design phase, with the appropriate level of human oversight depending on the AI solution application area and potential risks.

AI systems cannot be the subject "per se" of legal responsibility for their own functioning, so the AI system design should embed accountability rules (to identify who is responsible for what) and trackability principles, allowing the AI-based decision-making process to be explained and audited, thus helping to identify and prevent future mistakes or bias.

To achieve this objective, we will advise our clients that it implies to define and clearly identify together roles and responsibilities amongst the different actors involved in the design, manufacturing, integration, deployment and operation chain, including the designer of the AI solution, the data provider, and the company that adopts the AI solution or the final user. This would enable appropriate allocation of liability and effective recourse when needed.

CASE

A major energy supplier wanted to improve the management of inbound requests from customers to enhance its service quality. Capgemini developed and integrated semantic analysis for emails into the information system, allowing to automatize dispatches to the right team. Accountability was secured by a reference book setting out the business rules and AI processing guidelines. It will be regularly updated to allow for clear accountability over time.

6. Robust and safe AI

● ETHICAL CHALLENGE

Like any tools or systems, those utilizing AI must be fit for their intended purposes, and resilient and secure from a technical perspective. As AI uptake increases, so does the scope for potential impact, and the need to also consider the broader social and environmental context in which AI-based tools and systems operate. From this arises the challenge to foresee measures to safeguard against any risks, such as unlikely mishaps or malevolent intent, that might prevent the AI from delivering the desired benefits.

● CAPGEMINI'S RESPONSE

Robustness should be embedded throughout the life cycle of the AI systems, from the design and development to the deployment and use over their lifetime. AI systems should include, when achievable, fallback plans in case of failure of the AI system itself (e.g. allowing to adjust rule-based logic or even switch to human control, to avoid any wrong output), as well as being accurate, reliable and having reproducible results, to the extent allowed by applicable laws.

CASE

Sogeti, part of Capgemini, was requested to help a public social insurance agency to better leverage the sensitive data at its disposal in order to provide a secure, efficient service to the agency's clients. A 6-step proof of concept was implemented to design, train, and control machine-learning algorithms. This ensured the robustness and safety of the model, by allowing original data to be preserved and data set performances to be evaluated, for referential integrity.

7. AI respectful of privacy and data protection

● ETHICAL CHALLENGE

With AI taking off, the need for data is greater than ever, much of it driven by consumers. The opportunity for greater freedom presented by easily accessible data brings with it a related risk to data protection and informational privacy. Lengthy user agreements can tempt consumers to click "accept" without checking what rights they are giving away, while companies can be tempted to feed consumer and vendor data into advanced, AI-fueled algorithms, without the awareness and approval of the affected consumers and employees. Facial recognition, voice identification systems and smart home-appliances collect data about when we come and go; while many such functions provide a helpful service, the potential risks they carry are not trivial: Seemingly anonymized data can be de-anonymized by AI. Data collected can also enable tracking, monitoring, people profiling, and behavior prediction. By raising analysis of personal information to new levels of power and speed, AI magnifies our ability to use – and misuse – personal information, presenting a challenge for privacy and data protection.

● CAPGEMINI'S RESPONSE

In agreement with the client, we must ensure, that we will put in place all the necessary means for our current perimeter of responsibility, to contribute to the Clients' global AI objectives in terms of compliance with privacy regulations, data protection and proper data governance.

AI and data protection are compatible as long as data protection and cybersecurity are taken into account from the design phase of any AI project. Besides ensuring full respect for privacy and data protection laws and regulations, adequate data governance mechanisms should also be put in place. In practice, this means that any AI project would need to ensure that only data that are strictly necessary are collected and processed. Indeed, the data collected and used shall be proportionate, accurate, and processed in a secure manner.

Individuals will be provided with the relevant level of information on how their data is processed and they should be provided with appropriate means to exercise their rights as may be required by law.

The above ethical principles on AI aim to create an ethical culture for AI and represent a commitment to trustworthiness and ethical quality in services relating to AI.

AI solutions developed by Capgemini are based on ecosystems that may be composed of several third parties. Capgemini is committed to selecting and working with third-party providers that commit to comply with ethical principles.

Our Code of Ethics for AI defines areas for attention of ethical importance. These are being completed operationally by the development of tools integrated into service offerings and methodologies that comply with the ethical principles set out in the present code. Our professionals (such as AI solution architects and project managers) are trained to fully apply the code "by design" in all their AI-related engagements.

Values & Ethics

Capgemini
Place de l'Étoile – 11, rue de Tilsitt
75 017 PARIS – France

Values & Ethics | Talent Capgemini

Learn more about us at

www.capgemini.com