Murat Durmus

MINDFUL AI

Reflections on Artificial Intelligence

(THE AI THOUGHT BOOK Reloaded)

MINDFUL AI

REFLECTIONS ON

ARTIFICIAL INTELLIGENCE

(THE AI THOUGHT BOOK Reloaded)

Murat Durmus

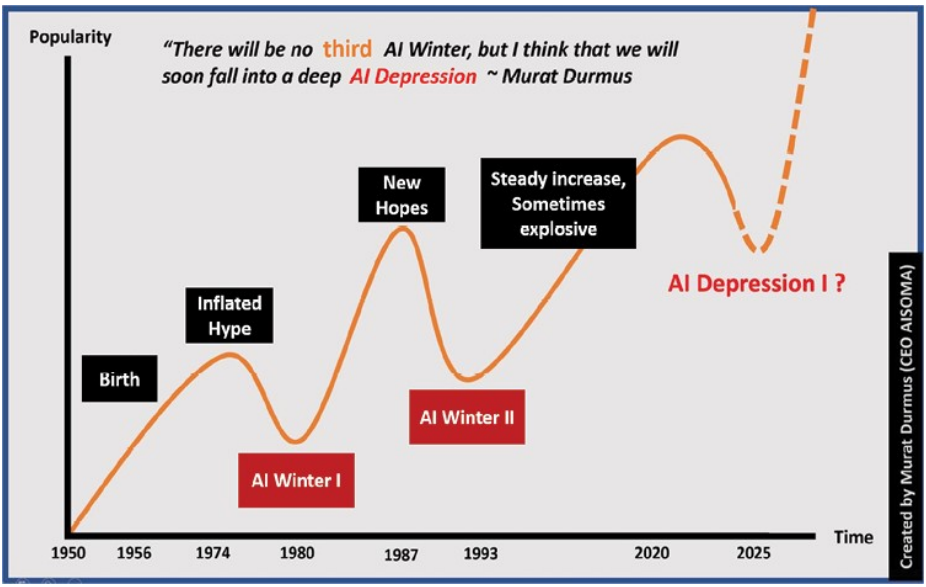
TABLE OF CONTENTS

▶ Many AI systems are still inefficient, need a lot of energy, or must be implemented complicatedly. We need different, more efficient approaches.

▶ Data awareness has not yet developed sufficiently across the world.

▶ Still too little usable (real-time) data is generated. The significant expansion of Edge/IoT/5G devices and infrastructures will continue.

▶ AI is dominated by a few Big-Techs (AI is becoming a power instrument increasingly); This could have adverse effects on AI.



The field of AI is highly ***interdisciplinary*** & ***evolutionary***. The more AI penetrates our life and environment, the more comprehensive the points we have to consider and adapt. Technological developments are far ahead ethical & philosophical interpretations.

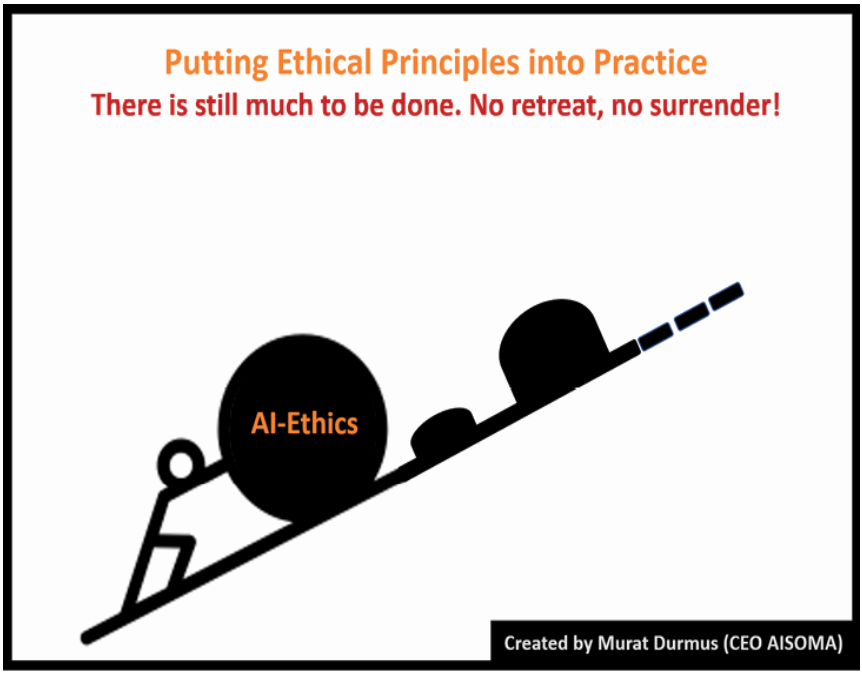
This fact is disturbing,

The AI-Control-Problem

In my opinion, the AI Control Problem is still underestimated and receives too little attention. Only when we have found a satisfactory solution for the control problem should we allow AI to interfere fully in our lives and not before.

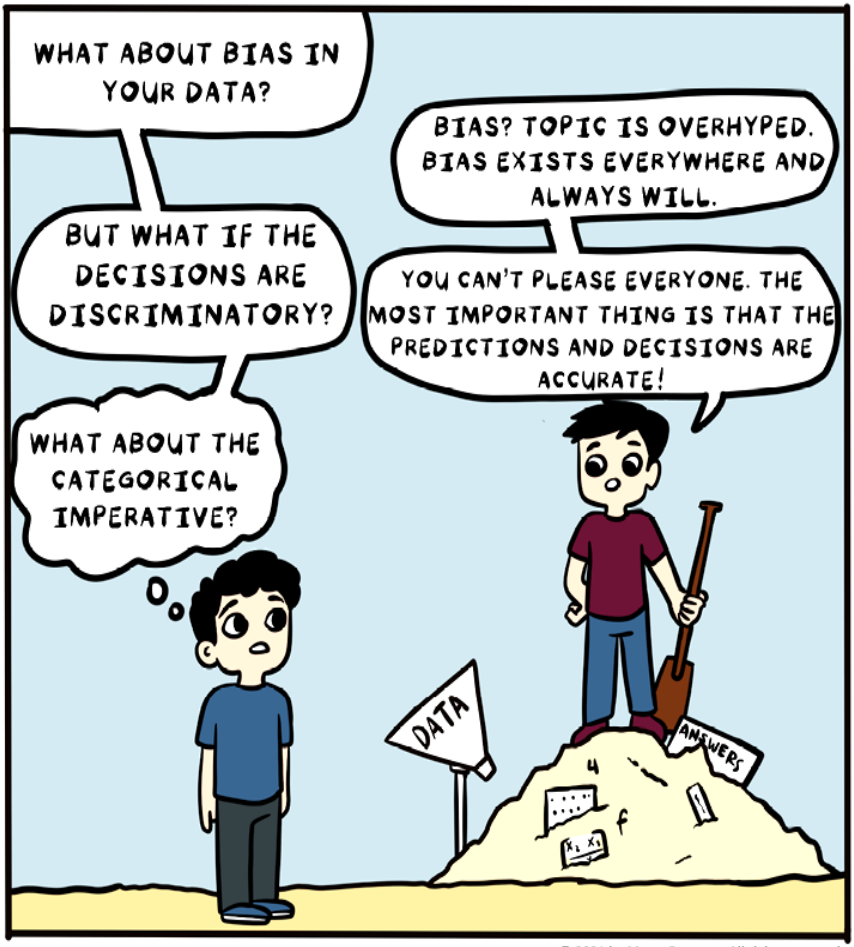
There is still a long way to go.

AI ETHICS



Discussions about AI Ethics are still primarily conducted in academic circles. But one can already see that many companies are earnestly dealing with it. One thing that seems clear to me:

EXPLAINABLE AI (XAI)

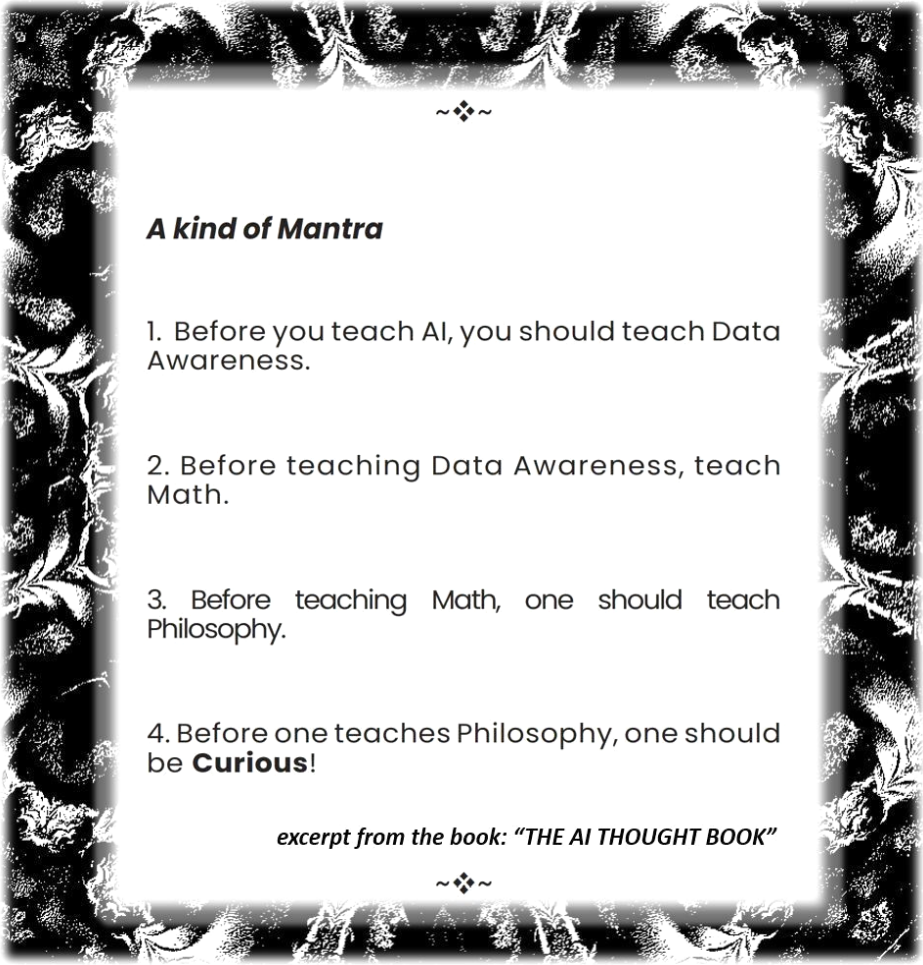


Demonstrating explanations only for the correct method or class is misleading and insufficient. Furthermore, this approach can create false confidence in the explanation method and the black box. This situation can occur when saliency maps (In computer vision, a saliency map is an image that shows each pixel's unique quality) are the explanations because they tend to highlight edges, thus providing similar answers for each class. These explanations could be identical even if the model is always wrong.

Algorithmic Transparency

Algorithmic transparency may help mitigate ethical issues such as fairness or accountability, but it also creates ethically essential risks. Too much openness in the wrong context can destroy the positive development of AI-enabled processes. Everyone should know that the idea of fully transparent algorithms should be weighed carefully. We still have significant challenges ahead, as we have to find a balance between security and transparency

A kind of Mantra



A BRIEF HISTORY OF ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is a growing discipline of sixty years that encompasses a range of sciences, theories, and techniques (including mathematical logic, statistics, probabilities, computational neurobiology, computer science, and philosophy) that aim to mimic the cognitive abilities of humans. Its developments are closely related to those in computer science. They have resulted in computers being able to perform increasingly complex tasks that previously could only be assigned to humans.

However, this automation is still a long way from human intelligence in the strict sense, which has criticized the term among some experts. The final stage of their research (a “strong” AI, i.e., the ability to contextualize very different specialized problems completely autonomously) is not comparable to current achievements (“weak” or “moderate” AI, extremely efficient in its training domain). “Strong” AI, which so far exists only in science fiction, would require advances in basic research (not just performance improvements) to be able to model the world as a whole.

Since 2010, however, the discipline has experienced a resurgence, mainly due to significant improvements in computer processing power and access to vast amounts of data.

Promises, renewed and sometimes fantasized concerns complicate an objective understanding of the phenomenon. Nevertheless, brief historical recollections can help situate the discipline and inform current debates.

1940-1960: The birth of AI

The period between 1940 and 1960 was strongly marked by the combination of technological developments (whose accelerator was World War II) and the desire to understand how to bring together machines and organic beings' workings. For Norbert Wiener, a cybernetics pioneer, the goal was to unite mathematical theory, electronics, and automation as "a whole theory of control and communication, both in animals and machines." Shortly before, the first mathematical and computer model of the biological neuron (formal neuron) had already been developed in 1943 by Warren McCulloch and Walter Pitts.

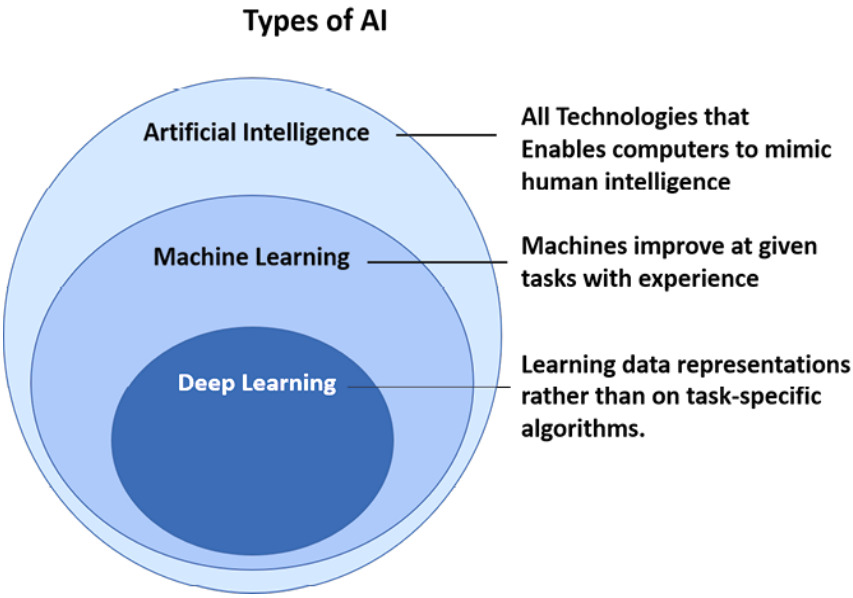
THREE ESSAYS FOR THE FUNDAMENTAL UNDERSTANDING OF AI

Artificial Intelligence: An attempt to define and distinguish

The development of Artificial Intelligence can be seen as the latest wave of automation since industrialization. While in the late 19th and early 20th centuries, the focus of automation was mainly on the substitution of physical human work by machines, artificial intelligence is the attempt to recreate human-like structures of perception and decision-making (to enable machines to perform specific (cognitive) tasks as well as, or even better than, a human being). A clear definition of the term artificial intelligence does not exist until today. AI can be defined as follows:

"The designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment."

(Russell and Norvig 1995)



A distinction is made between a strong and weak AI. Weak artificial intelligence (AI) aims to solve concrete, clearly defined application problems; This is done based on mathematical methods (algorithms) especially developed and optimized for the individual requirement. Therefore, weak AI is designed to support people in a specific activity.

These rule-based systems are primarily designed to perform clearly defined tasks without understanding problem-solving. This form of AI is already used in many areas, such as character and image recognition, individual

5 Variations of Artificial Intelligence

According to an unofficial consensus, the birth of artificial intelligence as an independent research project can be dated to the summer of 1956, when John McCarthy at Dartmouth College, where he belonged to the Mathematical Department, was able to persuade the Rockefeller Foundation to finance an investigation" The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it". In addition to McCarthy (who was a professor at Stanford University until 2000 and is responsible for the coining of the term "artificial intelligence"), several other participants took part in the historical workshop at Dartmouth: Marvin Minsky (former professor at Stanford University), Claude Shannon (inventor of information theory); Herbert Simon (Nobel Prize winner in economics); Arthur Samuel (developer of the first chess computer program at world champion level); furthermore half a dozen experts from science and industry, who dreamed that it might be possible to produce a machine

for coping with human tasks, which, according to the previous opinion, require intelligence.

The Manifesto of Dartmouth (written at the dawn of the AI age) is irritating and blurred. It is unclear whether the conference participants believed that machines would think or behave as if they could imagine one day. Both possible interpretations allow the word "simulate." Written and oral reports on the meeting support both positions. Some participants were concerned with studies of networks of artificial neurons, which, they hoped, could, in some sense, recreate the biological neurons of the brain. While others were more interested in producing programs that should behave intelligently, regardless of whether the principles underlying the plans bear any resemblance to the functioning of the human brain. This gap between the paradigms

Thinking = the way the brain does it,

&

Thinking = the results that the brain produces.

ARTICLES

Why we need to Regulate the use of AI Technologies

The potential benefits of AI for our societies are manifold, from improved medical care and knowledge discovery to better and more efficient education. However, given the rapid technological development of AI, we should be especially mindful of this technology because significant risks are not far from where great opportunities open up. While most AI systems pose little to no risk, specific AI systems create risks that must be addressed to avoid undesirable effects on people and society. For example, the opaque nature of many algorithms can generate uncertainty and bias and hinder the effective enforcement of existing security and fundamental rights legislation.

Note:

Although existing legislation provides some protection, it is not sufficient to address the unique challenges that AI systems may pose.

The proposed regulations will:

- address the risks explicitly posed by AI applications;
- submit a list of high-risk applications;
- establish precise requirements for AI systems for high-risk applications;
- set specific obligations for AI users and providers of high-risk applications;
- submit a conformity assessment before the AI system is placed in service or on the market;
- propose enforcement after such an AI system is placed on the market;
- Propose a governance structure at the European and national levels.

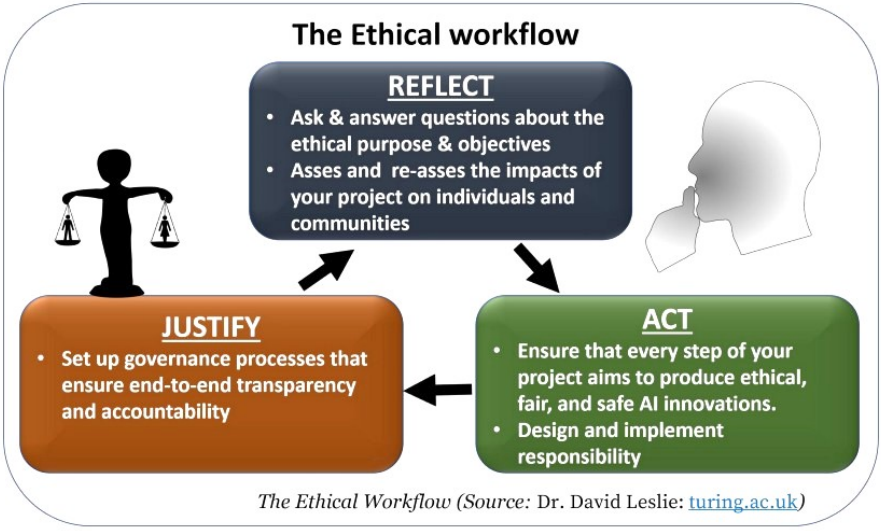
(Based on the recommendations of the European Commision)

Regulatory and legislative measures are needed to address these challenges. This is the only way to ensure a well-functioning market for AI systems where the benefits and risks are adequately addressed. This includes applications such as biometric identification systems, Deep Fakes, or AI

Do Companies need a Chief AI-Ethics Officer?

The world we live in is becoming more and more data-driven; This is causing companies to use more and more AI techniques such as machine learning and deep learning. It seems to be the only “efficient” way to get control over the data and generate value for the company relatively quickly. But, of course, future competitiveness also plays a significant role.

The task of the Chief AI Ethics Officer (CAIEO) should not be primarily technical. Instead, it should sensitize data scientists, machine learning engineers, and developers to ethical issues. The whole process of sensitization should be part of every data-driven project. I mean that the ethical workflow should be firmly integrated into the respective process models and phases.



Source: "Understanding artificial intelligence ethics and safety" by the Alan Turing Institute

In the long term, AI may lead to 'breakthroughs' in numerous fields. From basic and applied science to medicine and advanced systems. However, as well as great promise, increasingly capable intelligent systems create significant ethical challenges. The issues discussed deal with impacts on: human society; human psychology; the financial system; the legal system; the environment and the planet; and impacts on trust.

The Criminal Potential of Artificial Intelligence

I omitted this article to denigrate AI or to stir up fears (I believe AI will bring more benefits to humanity than any other technology to date) but to point out the dangers AI can pose and how it can be abused.

AI can be implicated in crime in several ways. Most obviously, AI could be used as a tool for crime, using its capabilities to facilitate actions against real-world targets: predicting the behavior of people or institutions to discover and exploit vulnerabilities; generating fake content for extortion or to damage reputations; performing acts that human perpetrators cannot or will not perform themselves for reasons of danger, physical size, speed of response, etc. Although the methods are new, the crimes themselves may be traditional in nature - theft, extortion, intimidation, terror.

Alternatively, AI systems themselves may be the target of criminal activity: Circumventing protective systems that stand in the way of a crime; evading detection or prosecution of crimes already committed; causing trusted

or critical systems to fail or misbehave cause harm or undermine public trust.

AI could also provide context for a crime. The fraudulent activity could depend on the victim believing that a certain AI functionality is possible when it is not - or that it is possible but not used for the fraud.

Of course, these categories are not mutually exclusive. As in the adage about catching a thief, an AI system attack may itself require an AI system to be carried out. The fraudulent simulation of nonexistent AI capabilities could be executed using other AI methods that exist.

Crimes vary enormously. They may be directed against individuals or institutions, businesses or customers, property, government, the social fabric, or public discourse. They may be motivated by financial gain, acquisition of power, or change in status relative to others. They may enhance or damage reputations or relationships, change policy, or sow discord; such effects may be an end in themselves or a stepping stone to a broader goal. They may be committed to mitigate or avoid punishment for other crimes. They may be driven by a desire for revenge or sexual

SOME SIGNIFICANT ACHIEVEMENTS IN THE FIELD OF AI SINCE 2010

2010

DeepMind Technologies is founded

A British AI company acquired by Google in 2014 and is part of Alphabet Inc. DeepMind Technologies' most amazing products are the Neural Turing Machine, AlphaFold, Wavenet and WaveRNN, and AlphaGO. In 2014, DeepMind received the "Company of the Year" award from Cambridge Computer Laboratory.

IBM's Watson computer beats human champions on game show Jeopardy

Watson is an interrogative computer system capable of answering questions posed in natural language. It was developed as part of IBM's DeepQA project by a research team led by study director David Ferrucci. Watson was

named after IBM's founder and first CEO, industrialist Thomas J. Watson. The computer system was originally developed to answer questions on the quiz show Jeopardy! In 2011, the Watson computer system competed against champions Brad Rutter and Ken Jennings on Jeardy! and won the first prize of $1 million.

2011

The Google Brain Project

The project was first launched in 2011 as a part-time research project by Google employees Jeff Dean, Greg Corrado, and Stanford University professor Andrew Ng. The project first received significant attention in June 2012, when a computer cluster of 16 thousand computers designed to replicate the human brain early recognized a cat based on YouTube images.

Apple introduced SIRI on the iPhone 4s

Apple launched SIRI as the first speech assistance program with the iPhone 4S. Speech Interpretation and Recognition

APPENDIX: GLOSSARY

Artificial intelligence

Artificial intelligence (AI) is intelligence exhibited by machines instead of the natural intelligence of humans and animals, which includes consciousness and emotionality. The distinction between the first and second categories is often made clear by the choice of an acronym. 'Strong' AI is usually referred to as AGI (Artificial General Intelligence), while attempts to emulate 'natural' intelligence are referred to as ABI (Artificial Biological Intelligence). Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and performs actions that maximize its chance of successfully achieving its goals. Colloquially, the term "artificial intelligence" is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem-solving."

Artificial General Intelligence (Strong AI)

Artificial General Intelligence (AGI) is an intelligent agent's hypothetical ability to understand or learn any intellectual task that a human can. It is a primary goal of some artificial intelligence research and a common theme in science fiction and futurology. AGI may also be referred to as strong AI, full AI, or general intelligent action. Some academic sources reserve the term "strong AI" for computer programs that can experience sentience, self-awareness, and consciousness. It is speculated that today's AI is still decades away from AGI.

Weak artificial intelligence (weak AI)

Weak artificial intelligence (weak AI) is an artificial intelligence that implements a limited part of the mind, or as narrow AI, is focused on a narrow task. In the words of John Searle, it would be "useful for testing hypotheses about the mind, but it would not really be mind." Contrast this with strong AI, which is defined as a machine capable of applying intelligence to any problem, rather than just a specific issue, which is sometimes considered a prerequisite for consciousness, sentience, and mind.

The ethics of artificial intelligence

The ethics of artificial intelligence is the branch of technology ethics that deals specifically with artificially intelligent systems. It is sometimes divided into a concern with humans' moral behavior as they design, make, use, and treat artificially intelligent systems, and a problem with machines' behavior, machine ethics. It also includes the question of a possible singularity due to superintelligent AI.

The European ethics guidelines for trustworthy AI[5]

On 8 April 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence. This followed the publication of the guidelines' first draft in December 2018 on which more than 500 comments were received through an open consultation.

According to the Guidelines, trustworthy AI should be:

(1) lawful - respecting all applicable laws and regulations

[5] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai