

Ethical Challenges of Artificial Intelligence

By Kenneth Wacks March 1, 2021



Introduction

I have been attending the Consumer Electronics Show (CES) online¹ where artificial intelligence (AI) is becoming a selling feature for consumer products. AI is being portrayed as endowing devices with lots of smarts that provide the user with a personalized experience. In this article I introduce AI and probe the recent concerns about ethical issues in developing AI algorithms.

What is AI?

In the 1970s while studying engineering and computer science at MIT, I attended lectures by Professor Marvin Minsky, a pre-eminent AI researcher. He was predicting that AI would enable a computer to perform as well in math as a college student. Is this the goal of AI?

I was amazed to learn that predictions about the prowess of AI began in the 1950s. According to Wikipedia, in the early 1950s there were various names for the field of “thinking machines”: cybernetics, automata theory, and complex information processing. In 1955 the Rockefeller



Foundation funded a summer seminar at Dartmouth University for about 10 participants to discuss “thinking machines.” The organizer, Prof. John McCarthy of Dartmouth (and later Stanford University), together with Prof. Minsky, Prof. Claude Shannon of MIT and Bell Labs (a pioneer in information theory), and Nathaniel Rochester of IBM, proposed funding for research into a new discipline they called “artificial intelligence.”

In 1955, Prof. McCarthy explained AI in his grant request:

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. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The proposal also discussed computers, natural language processing, neural networks, the theory of computation, abstraction, and creativity. The current definition of AI from the Encyclopedia Britannica [<https://www.britannica.com/technology/artificial-intelligence>] is:

Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience. Since the development of the digital computer in the 1940s, it has been demonstrated that computers can be programmed to carry out very complex tasks—as, for example, discovering proofs for mathematical theorems or playing chess—with great proficiency. Still, despite continuing advances in computer processing speed and memory capacity, there are as yet no programs that can match human flexibility over wider domains or in tasks requiring much everyday knowledge. On the other hand, some programs have attained the performance levels of human experts and professionals in performing certain specific tasks, so that artificial intelligence in this limited sense is found in applications as diverse as medical diagnosis, computer search engines, and voice or handwriting recognition.

AI and international standards

The International Electrotechnical Commission (IEC) is composed of 89 countries that have been developing electrical standards to promote world trade since 1906. Now that AI is starting to be incorporated into controller software, the IEC formed a special committee called a Systems Evaluation Group (SEG) to consider how AI will impact the development of standards.



IEC SEG 10, *Ethics in Autonomous and Artificial Intelligence Applications*, was established in March 2019 to develop guidelines for IEC committees on ethical aspects related to autonomous and/or AI applications. The following is an excerpt from an IEC report on the first SEG 10 meeting:

Ethics and societal concerns, such as trustworthiness, privacy, security and algorithm bias, are hot topics, as the world embraces AI in many aspects of daily life. Important questions need to be answered to ensure AI technologies deployed in homes, hospitals, schools, universities, workplaces, factories and other public spaces are safe and secure, and that decisions made by autonomous and AI systems are fair and beneficial for all.

What are the AI ethical issues?

In 2020 IEC SEG 10 conducted a survey among more than 200 IEC technical committees about potential ethical issues in the development of standards that incorporate elements of AI. I chair the ISO/IEC² committee whose scope is to develop “Standards for home and building electronic systems in residential and commercial environments to support interworking devices (IoT-related) and applications such as energy management, environmental control, lighting, and security.”

The survey of IEC SEG 10 introduced the AI challenge and related questions:

The rapid development of Autonomous and Artificial intelligence techniques and their applications (AAA for short) brings new ethical considerations. IEC SEG 10 is undertaking this survey amongst the relevant stakeholders of AAA, aiming at summarizing AAA ethical requirements in different scenarios.

To understand the context of the questionnaire, let us first clarify our perspective about the subject – What does “ethics” or “ethical” mean? And what aspects and nuances make any AI Algorithm/methodology trustworthy and what are the contours of ethics in an AI enabled/powered product, system and/or solution to qualify it as trustworthy. In this questionnaire we start off with setting the context and enumerating the prerequisite that a trustworthy AI-methodology should be (a) lawful, (2) ethical, and (3) robust.

Requirements to obtain trustworthy AI include the following:

- It should have human agency and oversight;
- It should be technically robust and safe;
- There must be respect for privacy and quality and integrity of data;
- It should be transparent;
- It should be non-discriminatory and fair;
- It should strive for societal and environmental wellbeing;
- It must be accountable.



Our position on AI ethics

My colleagues and I on the ISO/IEC home and building standards committee³ responded to the IEC SEG 10 survey by submitting the following letter to offer our position on AI ethics. I appreciate the thoughtful collaboration of Dr. Linda M. Zeger, CEO of Auroral LLC, and Dr. Timothy Schoechele, CEO of Smart Home Labs.

Dear IEC SEG 10:

Thank you for the important questions you raise in your survey of ethics in autonomous and artificial intelligence applications (AAA). We would like to address ethical issues that pertain to the Home Electronic System (HES). HES is the title of the 50+ standards we have developed in JTC 1/SC 25 for the fields of home and building systems, also known as “smart homes” and “building automation systems.” We also offer some general comments and issues to consider related to risk, as well as the transparency requirement listed for trustworthy AI.

Risk

The risks described here could be mitigated with innovative methods for data selection and cleansing, as well as for algorithm development, monitoring, and evaluation. Here are financial requirements and risks for smart homes with ethical ramifications. These have potential applicability to analogous types of risks in other use cases such as autonomous driving and manufacturing.

1. **A data risk** is that the data obtained and used in AI algorithms may be of poor quality and/or quantity and that the cleansing done on the data may be insufficient to improve the resulting inaccuracies and statistical biases in the outcome of AI algorithms. Potential resulting harmful effects include those listed below in Risk 3.
2. **A data, financial, and environmental risk** arises from the resources required for AI processing. The volume of data collected, stored, curated, transported, and used in an AI algorithm, particularly if that AI algorithm is computationally intensive, may require a large quantity of digital resources for AI processing. This processing incurs OPEX (operating expense) financial costs, harmful emissions from the energy required by information and communications technology (ICT) equipment to process these data, as well as human resources needed. (This risk may be more likely in the aggregate from a large number of smart homes, as well as from manufacturing, autonomous vehicles, etc. than from a single smart home.)
3. **A technological risk** is that the AI systems perform in a manner not intended by the designers, either due to unrecognized inaccuracies in AI algorithm inferences and/or sub-optimally designed or improperly used AI systems, or due to



unrecognized inaccuracies in AI algorithm inferences and/or sub-optimally designed or improperly used AI systems, or due to cybersecurity attacks from malicious actors gaining access to and controlling the system in an adversarial manner. The harmful effects that may result include:

- a) In addition to compromising privacy, cybersecurity attacks could threaten the safety, health, finances, etc. of people by controlling the AI system so it intentionally makes poor or incorrect decisions resulting in harmful actions.
- b) In energy-management applications a key goal of our standards is to reduce greenhouse gases and other harmful pollutants from smart houses. However, an increase in pollutants and financial costs may instead result from unintended or deliberate inaccuracies introduced into AI algorithm inferences and actions.
- c) In energy-management applications there may be a potential risk that AAA could affect electrical safety and/or grid stability.

The risk of unintended consequences is encompassed in Risk 3 (above). The history of technology and innovation demonstrates that inventors rarely, if ever, comprehend what they are actually inventing or how it will be applied in society. Creating autonomous and/or intelligent systems or devices is particularly dangerous because the systems may escape the control of the creator, and of society.⁴ An example of this risk of unintended consequences is how search engine technology by Google reshaped the business model of the IT industry around collection of data and surveillance.⁵ Technologies are always ultimately socially constructed—from the telephone to the bicycle. We are forced to ask, why do we want to create “autonomous” or “intelligent” systems and who are they intended to benefit? Are they being created to benefit venture capitalists, shareholders, etc.—or do they bring real value, and to whom? Much of what is proposed as AAA may accrue into the former category.

Transparency, Social Equity, and Bias

Transparency in terms of acceptable applications, performance ability, uncertainty, potential biases, environmental costs, risks, and limitations of AI algorithms and automation is essential. It is important that robust clearly-defined metrics be designed for and tailored to each use case and presented with the accuracies and limitations of these metrics, so users and others can understand the corresponding benefits, drawbacks, uncertainties, appropriate and inappropriate uses, and limitations of AAA systems.

AI systems, and particularly neural networks, are notoriously inscrutable in their internal structure. Transparency, in terms of sharing the models underlying the AI, is clearly important for the many algorithms that make financial, medical, legal, and other decisions affecting individual people. However, AI may also be applied in other non-personal areas, such as in predicting energy supply and demand or weather forecasts, on which companies may base their business models for creating or using proprietary



algorithms. In addition, non-transparent AI algorithms may perform better in some cases than transparent explainable algorithms, so a question arises: when is it acceptable for AI algorithms to be non-transparent or proprietary?

An intellectual property question is: who is considered the owner of an original work (such as art, writing, software or medicine, for example) that is created by or with the help of an AI algorithm? A related ethics question to consider for cases in which reverse engineering of an AAA algorithm is technically feasible: when, and to what extent, is it acceptable, to reverse-engineer a proprietary AI algorithm to obtain the exact model or an approximation of it? There are cases in which such reverse engineering has been performed to expose societal biases in AI algorithms, but it could also be maliciously employed to steal proprietary algorithms that were used for non-personal purposes. Any appropriate use of proprietary AI algorithms inherently necessitates revelation of some, usually small, amount of information about the algorithm. How much information about the underlying proprietary models is acceptable for end users or others to obtain and should there be restrictions on how they use it?

Another consideration for the ethics of AAA for HES is the requirement that the HES energy conservation standards achieve societal and environmental well-being in a non-discriminatory and fair manner. Technical solutions for these AAA systems should be sought that are affordable and accessible so that people everywhere have access to them. A question is: why do we need AAA in Home Systems, and whose interest are they to serve?

In applying AI algorithms, it is important to consider that such systems have been shown to reflect the biases of their designers, creators, and/or trainers. Current examples in the news are 1) racial bias in facial recognition systems and their use by police and surveillance; 2) bias in AI systems by courts or institutions for sentencing of offenders; and 3) medical diagnosis and recommendation systems for maternity practices.

Please let us know if you have any questions, or would like to discuss any of these points.

Sincerely,

Linda M. Zeger, SC 25/WG 1 expert
Kenneth Wacks, SC 25/WG 1
convenor
Timothy Schoechle, SC 25/WG 1
secretary



© Copyright 2021 Kenneth P. Wacks

Dr. Kenneth Wacks has been a pioneer in establishing the home systems industry. He delivers clear and practical advice to manufacturers and utilities worldwide on business opportunities, network alternatives, and product developments in IoT and AI for home and building systems. The United States Department of Energy appointed him to the GridWise® Architecture Council to guide the electric industry toward smart grids. For further information, please contact Ken at +1 781 662-6211; kenn@alum.mit.edu; www.kenwacks.com

Notes

[1] I will report on the virtual CES in a future *CABA Journal* article.

[2] ISO is the International Organization for Standardization. ISO and IEC collaborate on standards related to information technology (IT). Standards for home and building systems are being developed by a group of experts from Africa, Asia, Australia, Europe, and North America in the committee that I chair.

[3] The official designation of the standards committee I chair is ISO/IEC JTC 1/SC 25/WG 1, “Home Electronic System.”

[4] This was a key theme of the classic novel by Mary Shelly, *Frankenstein; or, the Modern Prometheus*, 1818, which initiated the literary genre we call Science Fiction.

[5] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future in the new Frontier of Power*, 2019.