[The original, German version of this article appeared on January 16, 2022, as part of the AI column “Aus dem Maschinenraum der KI” in the economics section of the German Sunday newspaper Welt am Sonntag, p. 26.]

[Translated with www.DeepL.com/Translator (free version) - AI technology made in Europe, see https://en.wikipedia.org/wiki/DeepL_Translator, and subsequently polished and modified by the author.]

Messages from the AI engine room

An Oath for European values in AI

Artificial intelligence (AI) has a social responsibility and should be reminded of it, even if it is through a symbolic oath.

When a baby babbles "Ma-Ma" or "Pa-Pa" for the first time, it is a great moment for the parents, even though still a lot has to happen before this babbling turn into meaningful sentences. A complex learning process begins, first instinctively and later supported by school and loving care. This is different for apes. Although they learn highly complex skills and even communicate, they do not have a complex language.

Language is probably the most important "invention" of human mankind. More than any other feature, language makes us human. It enables our social coexistence, gives rise to culture and science, and provides us with a construct for thinking and dreaming. The ability to precisely teach machines our language using AI methods therefore opens almost unlimited possibilities.

AI systems have reached this point in 2021. They have started to write and speak well enough to be used commercially on a larger scale beyond keyword recognition. A true revolution in AI, and thus a future revolution in research in general, business, and everyday life.

Often, however, what reads and sounds reasonable turns out to be gibberish upon closer inspection. Nevertheless, many "AI parents" are excited. And rightly so! Three or four years ago, no one believed that AI systems would be able to babble in perfect grammar, logical word problems or even formulating some jokes; with texts that are "plausible" at a surprisingly high level.

However, a lot still needs to happen. For example, Language AI lacks a normative compass. For example, they parrot social prejudices and stereotypes. That they do so should not surprise us. They reflect our collective writings and statements, and with them our norms, values, views, and narratives that we have uploaded to the Internet and used to train the Language Ais in first place.

Fortunately, the first technical approaches to making AI systems more ethically sound, socially aware, and culturally inclusive are already emerging. But it turns out that morality and ethics are at least as “knotty for machine as it is for humans” as New York Times’ Cade Metz argued. This must be an incentive for researchers and developers to meet the social responsibility that AI has.

As early as 1939, in his play "Life of Galileo," Bertold Brecht wished that natural scientists would develop something like the Hippocratic Oath of physicians. Unfortunately, Brecht

could not yet include AI, which was only founded later in 1956. But the necessity of such an oath has been urged again and again since then. Here is our shortened proposal, adapted from the oath for natural scientists due to Werner A.P. Luck who was a German chemist and co-founder of the German Chapter of the Society for Social Responsibility in Science:

„I will try to the best of my knowledge and ability to use my AI skills for the benefit of all mankind, all living beings, and the environment. I will strive to never harm or injure people unnecessarily, nor will I aid and abet such harm or injustice. Even under threat, I will not use my AI knowledge to violate human rights and civil liberties. I will share my AI knowledge and pass on findings without deceiving or concealing for self-interest. I will show respect and gratitude to colleagues and students, and take care of my own health, welfare, and abilities to perform AI development at the highest level. I solemnly pledge this, freely and upon my honor.“

The basic prerequisite for any AI deployment is integrity and technical excellence. Only through tireless work and intensive study can we create transparent and robust AI systems that best meet the current state of the art and where no carelessness or lack of competence compromises the outcome or even knowingly violates our values.

Because one thing is clear, AI will radically transform our society in the coming years and, with its complex capabilities, brings novel ethical challenges for researchers, engineers, entrepreneurs, and users who are shaping this future. Our universities, research institutes and the trend-setting innovative companies need an environment that allows these challenges to be addressed. Such an oath would be an important step toward this responsibility.

Kristian Kersting is Professor of AI and Machine Learning at TU Darmstadt, co-director of the Hessian Center for AI (hessian.ai), and winner of the "German AI Award 2019". His AI column "Aus dem Maschinenraum der KI" appears regularly in the German Sunday newspaper Welt am Sonntag. **Jonas Andrulis** the winner of the "German AI Award 2021", founder and CEO of Aleph Alpha, the European AI research company that, among other things, technologically develops and operates the new generation of large-scale language AI systems.