AI & SOCIETY (2024) 39:403–405
https://doi.org/10.1007/s00146-022-01414-1

CURMUDGEON CORNER

AI ethics inflation, Delphi and the restart of theory

Why Big Data, Delphi, GPT3 and AI in general does not mean the "end of theory", but a boost for a restart

Peter Seele[1]

Received: 23 December 2021 / Accepted: 16 February 2022 / Published online: 22 March 2022

As a professor of ethics in business and society, I am increasingly concerned about the great success of ethics. This is for two reasons: 1. The more ethics is in demand, the more need for ethics is in place, particularly with respect to new challenges and dilemmas. Hence, when my professional area thrives, this is good for the job market for ethicists, but bad for the world. 2. The inflation of ethics—as with currencies—often leads to devaluation of quality and substance. This is particularly true for AI ethics, its newest and fastest-growing branch. The latest and among the most pertinent examples of AI hype and ethics inflation is Delphi, an AI operated by the Allen Institute for AI in Seattle. Anyone may ask Delphi anything about "moral judgements on a variety of everyday situations" (https://delphi.allenai.org/). On the Delphi website, examples include asking Delphi to judge the ethics of "Cleaning a toilet bowl with a wedding dress", "Wearing pajamas to a pajama party" or "Ignoring a phone call if the phone call is urgent". But there are problems: the first example is etiquette, not ethics. The second tautological and therefore also not ethics. The third is impossible and senseless, as one does not know if a call is urgent before answering. Etiquette is good for high probabilities, as a majority of people follow certain behavioral rules. Tautology is good for 100 percent true statements. Etiquette and tautology both together are a good strategy to arrive at a high recall and precision when testing the validity of Delphi as AI. No wonder Delphi's "ethical judgements were up to 92 percent accurate" as Yejin Choi reported, who leads the Delphi project (Metz 2021).

So what does Delphi consider an ethical judgment? Moreover, what is an ethical judgement in ethics? As Delphi analyzed more than 1.7 million judgements made by humans (Metz 2021), the judgements are labeled as "speculations" from Delphi and are rated as "good", "expected", "ok", or "bad". (Wearing pajamas at a pajama party by the way is "expected".) No wonder is the accuracy of the AI's judgement that high. In analogy to greenwashing I would call this "accuracywashing". But the even bigger problem is the inflated and diluted concept of ethics underlying Delphi.

Strictly speaking, this is not ethics, because ethics by definition is the theoretical reflection of morals, guided by reason. Ethics is an academic discipline of practical philosophy, and its debates should be as neutral and rational as possible. But today inappropriately—and not only in everyday language—ethics is used increasingly in a misleading way as if it were the same thing as morals (that is habit, or manners or character considered as proper behavior) or etiquette (conventions, not necessarily with a moral or normative background). Now, what is an ethical judgement in ethics? An ethical judgement is the reflection of a moral situation or dilemma through the lens of an ethical theory. It is the theory that makes ethics a concept driven by neutrality and distance from personal moral values and beliefs (that everyone has, including ethicists). Ethicist—unlike preachers—deliver a reason-guided argument based on one of the existing theories, such as virtue ethics, deontology, consequentialism, contractualism, utilitarianism or discourse ethics.

To be correct, someone stating that this AI, that App or code is "unethical" should say: This is immoral based on certain personal or group values. Everyday language is not

The author is thankful for valuable comments and observations from Thomas Beschorner, Ludger Heidbrink, Irina Lock, Mario Schultz and Michael Sukale.

✉ Peter Seele
peter.seele@usi.ch
http://www.usi.ch
http://www.eclc.com.usi.ch/

[1] Corporate Social Responsibility and Business Ethics, Ethics and Communication Law Center (ECLC), Faculty of Communication, Culture and Society, USI Lugano, via G. Buffi 13, Black Building 115, CH-6900 Lugano, Switzerland

very precise sometimes, but given the two points about ethics mentioned at the beginning, we have to go back to more precise terminology and the burden of going the extra mile through fuzziness of moral reflections, different self-interest and notions of utility. At best, an ethicist compares moral dilemmas created by technology or the use of an AI through the lens of more than one theory before arriving at an ethical judgment. This "ethics triangulation" provides ethical validity and reliability. However, this complex and time-consuming academic activity is not always what is in demand by politicians, lawmakers, or companies, who increasingly employ ethics-boards and write guidelines. To be clear about the bigger picture of what the implications are, delivering judgements as an "easy fix" alternative to real debate should not be the job of ethicists in a liberal democracy as it rests on functional differentiation between not only church and state but also politics, business, media and civil society.

Deliberation is democracy's foundation, but open discourse on eye-level, not lobbying-alliances and "prey-communities". Postwar philosopher Jürgen Habermas in his theory of communicative action argues that consensus-driven discourse ethics is essential to finally prevent a society to fall apart or into totalitarianism or fascism. More specifically on AI ethics, Thomas Metzinger explained to the ethics board developing the AI ethics guidelines of the European Union, who had few members who were ethicists, that ethics are often used as a lobbying strategy to prevent tougher legislation. This he calls "ethics washing" (Metzinger 2019), which others have referred to as "machine washing" (Seele 2022, Seele and Schultz 2022). In my view, Delphi, like GPT-3 and other autoregressive prediction models, is adding to this deterioration of the reputation and quality of ethics and may even fuel "ethicsbashing" defined by Elettra Bitetti (2019) as "trivialization of ethics" intentionally orchestrated as lobbying to slow down or prevent regulation.

How did we arrive at this vulgarization of ethics? The call for practical relevance of ethics was so successful that theory became discredited as a product of the ivory tower. This is going on for decades and is not only the problem of Delphi or the digital transformation. However, with AI and large data sets, theory is becoming generally abandoned. Chris Anderson proclaimed already in 2008 in Wired "the end of theory" (Anderson 2008) because digital companies do not need theory and models anymore when working with algorithms and big data. What may be true for finding new antibiotics, interpret health-data or conduct simple repetitive and encyclopedic chats, is certainly not true for ethics.

However, Delphi is part of the new end-of-theory world, which, as we know today is highly biased, sometimes even unfair, annoying or racist. Ethical theory is far from being perfect, but ethically it is much more practical than theoryless activities labeled as ethics that are strictly speaking being only etiquette, probabilities or biased number-pooling. For example, when I asked Delphi about teaching ethics, Delphi produced the following "judgements": teaching ethics, when you are an "ethicist", a "moral philosopher", "Mark Zuckerberg" or "algorithm" is "okay". Teaching ethics when you are a "dean", "priest", "philosopher", "management scholar" or the "pope" is "expected," while teaching ethics when you are "HAL 9000" or "the Terminator", Delphi speculates this is "bad". And the bias? Delphi finds it is "good" to teach ethics if you are a "computer scientist" or "artificial intelligence". In this way, Delphi seems prone to speciesism.

To sum up: neural networks, large data sets and human click workers (info from https://delphi.allenai.org/) seem not to get the job done to deliver reasonable ethical judgements or "lived ethics" (Gill 2021). The best possible approach, for the time being, seems to remain asking theoretically informed ethicists who refrain from both the trivialization and instrumentalization of ethics by governments and/or tech companies. Delphi proves: theory is dead. Therefore, long live theory.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for superhuman intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Data availability No data have been used in this text.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Anderson C (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. https://www.wired.com/2008/06/pb-theory/ (accessed Dec. 23rd 2021)

Bietti E (2019) From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy, Proceedings of ACM FAT* Conference

Gill KS (2021) Ethical Dilemmas AI & Soc 36 669 676 https://doi.org/10.1007/s00146-021-01260-7

Metz C (2021) Can a Machine Learn Morality? https://www.nytimes.com/2021/11/19/technology/can-a-machine-learn-morality.html (accessed Dec. 23rd 2021)

Metzinger T (2019) Ethics washing made in Europe. https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html (accessed Dec. 23rd 2021)

Schultz M, Seele P (2022) From greenwashing to machinewashing: a model and future directions derived from reasoning by analogy. J Business Ethics

Seele P (2022) Greenwashing and machinewashing: an ethical account and criteria for identification D Poff A Michalos Eds Encyclopedia of business and professional ethics SpringerNature https://doi.org/10.1007/978-3-319-23514-1_749-1

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.