Digital Society (2023) 2:32
https://doi.org/10.1007/s44206-023-00061-3

EDITORIAL

What Can AI Learn from Medicine?

Federica Russo[1]

Published online: 2 August 2023

Since OpenAI released ChatGPT on 30 November 2022, there has been a burgeoning discussion about—and use of—the tool. Or rather, the *tools* since ChatGPT is fast growing into newer and better-performing versions. Lots of ink has already been spilt discussing its potential, inherent methodology, costs (both in terms of energy needed to run the algorithm and of the cheap human labour to train it), and pitfalls. Likewise, in the realm of policy and governance, the decision of the Italian 'Garante' for Privacy has added another layer of complexity to the conversation, with their reaction temporarily stopping the use of ChatGPT in Italy, pending OpenAI addressing specific issues about privacy. Here, I am only scratching the surface of a debate that quickly becomes very technical and specialised about ethico-legal aspects of digital technologies.

No doubt, the release, dissemination, and use of ChatGPT are of great interest to *Digital Society*; in co-piloting its immense generative capacities, we find ourselves profoundly affecting our infosphere in real-time. The changes that new digital technologies, such as ChatGPT, are pervasive and not entirely foreseeable. Likewise, efforts to regulate, predict, or control these effects are regularly met with a sceptical rejoinder insisting that such technologies, and by extension, the changes they bring about, are inevitable (often from those standing to gain the most from a *lack* of regulation). But I submit that this is an opportunity for AI to learn from medical methodology and drug regulation. Here is why.

Research and interventions in medicine and pharmacology are highly regulated fields. At the level of methodology, both fields are heavily regulated with strict protocols about the set-up and development of randomized controlled trials (RCTs). Likewise, the marketing of drugs (including drug retrieval) and administering drugs or other medical interventions for individual therapy are highly regulated. All this regulation is clearly not foolproof. However, it is undeniable that medicine and pharmacology have come a long way since the scandals and mistakes of the twentieth century, which motivated such regulatory developments – I will recall only the first trial on streptomycin and the thalidomide scandal as exemplars of how we had to learn the hard way about how to set up a methodologically sound RCT, to anticipate

✉ Federica Russo
f.russo@uu.nl

[1] Freudenthal Institute, Utrecht University, Utrecht, Netherlands

hazards, and to quickly retract a product once harmful effects are reported. What could AI learn from medicine, then?

To begin with, there is an important methodological lesson that comes not only from medicine but from non-AI-based sciences in general, one that has been powerfully encapsulated in the maxim: 'garbage in – garbage out'. Long before we called science 'data science' and the (mistaken) announcement that Big Data would make 'theory' obsolete, philosophers and methodologists of science have discussed the conditions under which we can trust the results of scientific models (especially quantitative models); that is how we can infer the correct results from the analysed *data*. In any scientific field, a crucial part of scientific training is about data: what data we need for what purpose; whether we can generate, collect, and measure data accurately; whether our methodological approaches are well suited to the data being analysed; whether running a different method on the same data set will produce different results; etc.

Data, too, is a crucial aspect of ChatGPT which, as well known by now, is based on a class of models developed within machine learning called 'large language models' (LLM). I will not rehearse here explanations of how they work, other than that they are *not* based on anything that requires attention to the inputted data. The point of an LLM is that the *larger* the data set, the better the model (or so we are told). But the larger the input, the smaller our control over the quality of data. In short, if we follow the maxim of philosophers of science, we are in a 'garbage in – garbage out' situation. This is why, to simplify things a bit, it has been reported that ChatGPT sometimes provides misleading and even false information, 'invents' academic references or biographical notes, or reproduces bias in its outputs. My point is not that we should abandon large language models altogether, but that we must reflect on what exactly should go (or not) in this 'large' basket. In other words, we miss a question about *purpose*.

Purpose is important not just to ensure the quality of output based on the quality of input, but also to foresee domains of application *and* prospective users. AI and LLM developers may want to learn from medicine again. Here 'medication guides', which require drug developers, regulators, and clinicians to pause and ask who should (not) be taking a drug, under what conditions, and how the drug should be administered, etc., are a good role model. In the context of AI, these questions become: How should we train ChatGPT if it is to be used in an educational context? Or in a media context? What level of expertise would be required for a user to use ChatGPT? Is there a minimum age to competently and wisely use ChatGPT? Clearly, there are many more questions about purpose and use. The stakes, I think, are very high because by allowing unrestricted use of a technology, such as ChatGPT, without any anticipation of its users, we run the risk of exposing users to unknown impacts on both individual and social levels. To be sure, there is nothing new under the Sun. The unreflective release and use of social media such as Facebook seemed unproblematic at that time, while now we worry about its effect on democratic processes, manipulation, the attention span of users, etc.

The problem is that, at this point in time, nothing obliges OpenAI, or any other big tech company, to significantly revise their modelling approach or specify its purpose and intended users. Admittedly, following the model of medicine and drug

development would mean considerably slowing down the process and increasing the costs with no guaranteed financial benefit. If you think that regulating ChatGPT (or the like) is science fiction, a similar suggestion has been expressed by the editorial board of the Financial Times.[1] But, if we regulated the development and marketing of drugs (a lucrative enterprise comparable to developing digital technologies), we can certainly regulate AI too. It is a *choice*, just as it was to introduce Traffic Laws, driving licenses and minimum age to drive, seatbelt obligations, and the list may go on. To put it in the jargon of philosophers of technology, we are emphatically *not* passive victims of technological determinism; instead, we continuously make choices, and *non*-regulation counts as a choice too.

Here is why I submit that this discussion *should* happen within *Digital Society*. Digital society is a choice, and so is any step we take in developing and releasing new (digital) technologies. Some choices are within the remit of big tech companies, whilst others are the preserve of individual developers. Some choices are in the hands of institutions touching upon legal and governance aspects and are political in character, (the EU is taking a stance with the development of the AI Act and the GDPR before that). Other choices, instead, are in our hands, qua *users* of (digital) technologies. Whilst we may perceive these choices as being less significant than an EU directive, their political character is not to be underestimated.

[1] https://www.ft.com/content/7ba3e97b-d930-4f96-8365-f840eaabf523, accessed 18 July 2023.