On the ethical challenges of innovation in digital health

THOMAS CHRISTIAN BÄCHLE

The evaluation of technological innovation needs to take into account a multitude of different areas, each of which with unique ethical repercussions.

Keywords: autonomy, hybrid responsibility, overdiagnosis, predictive diagnostics, self-governance

INTRODUCTION

The relationship between innovation and ethics is ambiguous. On the one hand, ethical considerations are often said to be an obstacle to social or technological progress, constraining innovative approaches for the sake of being overly cautious. In that vein, codes of conduct, ethical review commissions or public debates on controversial topics such as genome editing might delay the implementation of technological solutions to pressing social and medical problems. On the other hand, while ethical principles reassuringly prevent an anything-goes attitude, they can even come to act as drivers of innovative approaches and even become a decisive factor in the competition for ideas. Of course, neither of these two contradictory viewpoints is valid on its own. Ethical considerations need to go hand in hand with innovation and should not be just an afterthought. It is a truism that any new technology poses a trade-off between benefits and risks. The advent of digital technologies, however, introduces challenges that concern all aspects of life but particularly the sensitive area of health, where a long tradition of ethical principles applies.

These "foundational ethical principles" include non-maleficence (effectively summed up in the famous dictum *primum non nocere*), beneficence (actions that are in a patient's best interest), the respect for a person's autonomy, and justice (in particular, an equal distribution of goods, services or resources); specifically in therapeutic contexts, they also include the principle of fidelity (trust comprising both loyalty and honesty) (Strohm Kitchener & Anderson 2011). When assessing medical and health-related innovations, these principles can be translated into a very general framework of questions that guide ethical concerns:

1. Is the innovative technology harmful? What are potential risks associated with it?
2. What are its potential benefits and is there evidence available to substantiate them?
3. How will this innovation affect the autonomy of affected parties?

DOI: 10.5281/zenodo.3296931

4. How will the innovation affect justice and equity? Will access to the innovative technology be equally distributed? Will it create new inequalities?
5. Which effect will the innovation have on the relationship between patient and physician, therapist or counsellor?

As will be seen in the discussion of the following four issues – the role of data and data analytics, responsibility, (over)diagnosis and autonomy – which are among the most important ethical challenges in the context of eHealth innovations, the answers to these questions are far from obvious.[1]

ANY DATA CAN BECOME "HEALTH-RELATED"

The *universal datafication* and subsequent analysis of large data sets ("big data") of medically relevant information (such as biochemical and genetic information) has been expanded to include physical, social and cultural factors that are creating an unprecedented level of knowledge about each individual. Data repositories maintained by large private companies such as Apple and Alphabet permeate the field of services related to health and well-being. When looking at the issue of data, it is by no means solely the large quantity of data sets that is highly consequential. Digital communications technologies not only allow for representations of all actions, behaviours and (communicative) interactions in minute detail. In addition, these very activities generate additional information that can also be used to accumulate data (Andrejevic 2007). This universal surveillance is based on "extractive processes" that "typically occur in the absence of dialogue or consent, despite the fact that they signal both facts and subjectivities of individual lives (Zuboff 2015, 79). This represents a type of capitalism that introduces "unexpected and often illegible mechanisms of extraction, commodification, and control that effectively exile persons from their own behavior while producing new markets of behavioral prediction and modification" (Zuboff 2015, 75). Based on intricate methods of data analysis used to determine significant patterns – often labelled "artificial intelligence" (AI) or machine learning – this means that any type of data can become *meaningful* in a health-related context, whether it is one's eating, sleeping and drinking habits, personal interests and social ties, or one's appearance or voice. In addition, machine learning analytics can potentially identify previously unknown patterns between variables and create new knowledge about individuals and group affiliations, such as "people who experienced X when they were 10 years old are likely to develop condition Y once they turn 40".

[1] Some of the arguments made in this text were represented in the discussion entitled "Ethics and Innovation" at the "Futures of eHealth" conference which took place on 30 April 2019 in Berlin. Participants were Btihaj Ajana (King's College London), Erwin Böttinger (Digital Health Center, Hasso-Plattner-Institute, Potsdam), Elif Kücüktas (chair), Daniel Strech (Berlin Institute of Health) and Jai Ganesh Udayasankaran (Asia eHealth Information Network).

The universal collection of personal data raises obvious issues such as data security and confidentiality but also entails much less visible consequences, such as a slow change in our understanding of what it means to be healthy and the general tendency to overdiagnose or overtreat. The collection of data is sometimes intentional (with self-tracking devices and lifestyle apps) but often happens inadvertently, with a smartphone being used as a universal tracking device. This raises the question of control of the potentially health-related data that is being collected, stored and analysed by large private companies. Although it is customers who buy and own a device and its materiality, the software, the application and most importantly the data thus generated is not held by them (Fairfield 2017).[2] This is particularly noteworthy, since the data is produced by each *individual's* actions and corporeal physicality but is monetised by third parties (Ajana 2017). "Paying off" people for the data they generate only offers a purely economic solution to an economic problem but will likely further deepen social divides, since it leads to an even stronger commercialisation of the data that may overlook social, legal or ethical perspectives (Prainsack 2017). This is also the case when the more affluent individuals, groups or populations are able to pay for any service they want while their data is kept private and secure. Creating a business model that is both viable and "ethical" and that factors in issues such as self-determination, responsibility and social power structures is still an unresolved challenge.

The health-related data that is generated via commercial self-tracking devices must be differentiated from the data that is collected, stored and processed in biomedical research contexts. From an ethical perspective, however, similar issues are concerned, namely informed consent, privacy (including anonymisation and data protection), control of personal data and the questionable objectivity of "big data" analyses (Mittelstadt/Floridi 2016). Ensuring informed consent in biomedical research promises transparency and even citizen empowerment, with people taking control of their own data. In reality, however, people often do not understand what they are consenting to. In the case of data and machine learning tools, this concerns diagnostics systems that – given their self-learning and automatic nature – have a future scope of knowledge (diagnostic, research or otherwise) that cannot be fully understood at present. Will and should the informed consent include these unknowns? How can this complexity be made transparent to patients and physicians in order to ensure informed decisions?

Regarding the issue of anonymisation of data, data security and complete anonymisation cannot be guaranteed even in non-profit research contexts. In practice and principle, data can be re-identified or stolen from the most secure clinical data warehouses. However, the cry for a total anonymisation of data in research settings might arguably proceed from a false premise. When data analysis turns out to have a positive impact on diagnostics and

[2] Technically, from a legal perspective, personal data is not subject to property rights and hence cannot be "owned".

therapies, the benefits outweigh the risks. Since other risks are accepted in society, the intense debates on data safety risks might even seem disproportionate. When there is clear evidence for a direct personal benefit, the normative stance on total anonymisation (or its modern core value of privacy) as a principle could even lose its validity and persuasiveness. A context-specific "micro ethics" might lend ethical legitimacy to health-related surveillance when evaluating its necessity, purpose and intent (Sewell & Barker 2007).

HYBRID AGENCY LEADS TO HYBRID RESPONSIBILITY

Hopes are high for the implementation of so-called AI in diagnostics or medical decision-making: it is likely that well-trained algorithms will one day outperform a human physician in applying the latest, most relevant medical research, in providing the most accurate diagnoses or in making the best decisions when choosing the most beneficial therapy. They might be deployed in order to reduce the number of errors made by humans in healthcare. Even though algorithms bear an inherent risk of making incorrect determinations, in many other areas and applications they make fewer mistakes than humans. The self-learning nature that is characteristic of machine learning systems, which are an authority in their own right in decision-making processes, effectively turns them into "autonomous moral agents" (Wallach & Allen 2009). This gives rise to an ethical question that is being fiercely debated at present: when the use of algorithm-based decision-making systems leads to incorrect conclusions that have detrimental consequences for the health of the individual patient or groups, who should be held responsible?

The most straightforward answer to this is probably that the company providing the algorithm should be held responsible for inaccurate results, ineffective therapies and the dire consequences they entail. Also, the certification body – in other words those who test and evaluate innovations – should share the responsibility for faulty technologies.

Yet the responsibility question becomes more complicated when the treating physician double-checks the results or decisions provided by the algorithmic system, which should always be the case. How much of the responsibility should she or he bear? Medical associations demand that the final decision must always lie with the doctor. The potential that is currently attributed to automated decision-making systems in the medical context leads to an unprecedented scenario: what if algorithms can prove that a physician's bad decision had harmful consequences for a patient?

So should the responsibility ultimately lie with the individual patient, who can decide whether to trust the doctor or the technology? If the beneficence of machine learning based decision-making and diagnostics can be statistically proven, what ethical principle should apply if patients choose to renounce these possibilities? Should doctors abandon diagnostic tools when a patient wishes them to do so?

A continuity argument might emphasise that the question of responsibility is not unprecedented, since previous innovations in the field of diagnostics or treatment (e.g. MRI, CT) also gave rise to concerns about unforeseeable risks and long-term side effects. Yet, with autonomous moral agents such as algorithmic decision-making systems, there is reason to believe that the innovative technology in question is categorically different from previous ones. This primarily concerns their considerably heightened degree of agency.

Technology – in both an instrumental and conceptual sense – is often used to delegate responsibility from human agents. However, social consequences are never caused by human or technological agents alone but emerge in their interaction. Agency, therefore, is hybrid, distributed between human and non-human agents.[3] The ethical relevance of this argument translates hybrid agency into the notion of hybrid (or shared) responsibility, since accountability, responsibility or morality can no longer be attributed to one defined entity (such as an "individual", a "subject" or "a machine"), but is distributed within dynamic networks of agency (Mittelstadt et al. 2016). The challenge of assigning agential responsibility in this network of "autonomous" technologies, technology producers and designers, and technology users becomes ever greater. Assuming a co-evolution of humans and technology in effect means that ethics and innovation co-evolve as well.

THE DILEMMA OF OVERDIAGNOSIS – IS IGNORANCE BLISS (AND WHO DECIDES THAT?)

The increasing use of machine learning tools in diagnostics, such as visual analytics in radiology, is very likely to produce more accurate results. At first glance, this seems to represent an overall improvement: the more accurate and the earlier a disease is diagnosed, the more effective the treatment and the more positive the patient outcomes. Also, these tools can easily be equally distributed, providing access to high-quality diagnostics with no expensive human expert necessary, even on a global scale. However, in applying these tools, a somewhat paradoxical line of conflict emerges between innovation and ethical principles.

Let us assume that a disease can be diagnosed with 100% accuracy and at a very early stage. With no symptoms being present, a hitherto perfectly healthy individual becomes – as a result of the diagnostic procedure (Mol 2002) – a patient, an individual who has no symptoms yet. Since a condition has been "discovered", however, there is an imperative to treat it. Depending on how invasive the treatment is, there is a greater

[3] Theoretical paradigms such as science and technology studies (e.g. Latour 1993) take this as an instrumental factor in the construction of social reality.

risk of it decreasing the patient's quality of life.[4] This problem may be aggravated by the use of so-called *predictive diagnostic* tools that make probability-based predictions about potential future medical conditions. Just knowing that they are likely to become ill has an impact on individuals' well-being. Should treatment therefore be started based on risk assumptions alone?

Hence, another responsibility concerns the principle of the long-term beneficence of AI tools rather than their accuracy and effectiveness. In other words, assessing their diagnostic validity is not enough, since it is crucial to address questions of clinical utility: do patients live longer, do they experience a higher quality of life? Otherwise, there is the risk of overdiagnosis, leading to overtreatment, with the sensitivity and accuracy of AI diagnostics causing more harm than good.

The responsibility for this decision cannot be left in the hands of either the individual physician and patient or with the provider of a diagnostics systems. These innovations pose the much more general question on what our collectively defined social and cultural limits for their use should be. In practice, it could be argued that the responsibility lies with those who test the long-term consequences of the diagnostic tools, who should conduct "fair" tests of whether an individual can expect to live a better life with or without the AI diagnostic tool.

The problem is to define the criteria that are put in place – what does "more good than harm" mean in practice? – and to determine who decides on them. This could even lead to a scenario in which diagnostic tests would have to be made deliberately less accurate or unavailable in situations where the quality of life would decrease for the majority of patients. What if one specific patient was part of a minority for whom quality of life increases? Who should decide on whether a test can be conducted or its results be discarded? For which groups would the test be defined as too precise? Who would determine when the variable "quality of life" attained a sufficient value to justify making the diagnostic reality known to physicians or patients? Should it be physicians who weigh the benefits against the diagnostic reality? Should it be patients who are given the opportunity to make these decisions for themselves?

Once these both highly ambiguous and consequential decisions are made, it might even be possible to automate and delegate them to the technology, for example, in the form of a test that only works when certain social and cultural variables are factored in and when an overall positive outcome is expected.

[4] The example of the well-known PSA test, which is used to identify a prostate-specific antigen (PSA) with the objective of detecting undiagnosed prostate cancer, is a case in point here. Despite the fact that it provides more accurate diagnoses, it is unclear whether using the PSA test reduces the mortality rate. This uncertainty is due to the side effects of surgery.

In the end, the ethical questions concern human agency and the criteria for socially, historically and culturally specific constructs such as "benefit", "harm" or "quality of life", which might differ greatly regarding the range of affected cultures, nations or social groups. Who has the authority to determine the social norms that are inscribed into the technology and the best practices for their deployment?

AUTONOMOUS PATIENTS, RESPONSIBLE INDIVIDUALS

The issue of knowledge is closely linked to individual autonomy, since it is always interwoven with power structures. Algorithmic agency – especially the agency emerging from self-learning systems – is not independent of social or cultural norms but is rather reflective of them in the results or decisions they produce. Biases are instilled into them, which includes the definitions of what is considered to be "healthy". Health is not just the absence of disease or the product of the ability to prevent or cure diseases. It is a contextual, value-laden, socially and culturally embedded construct.

Cultures of self-tracking and self-optimisation, together with discourses of risk management and mitigation in health and biomedical research, are to be understood as part of a larger neo-liberal market logic (Rose 2007). The universal datafication of our social lives makes it easy to implement more effective systems of patient surveillance and control. With the rising importance of data, patients are slowly transforming into "data subjects" who are kept under permanent surveillance (Lyon 2007); this creates new social and cultural categories, such as the various labels of being healthy, being "at risk" or being "not yet ill". This type of data categorisation always has social consequences as it is based on sorting and excluding certain groups or people. Even if you as an individual are not directly affected, your personal data might be part of a sorting mechanism or decision-making process.

Medical (self-)surveillance and tracking reinforce the idea that it is each individual's responsibility to stay healthy by improving his or her own lifestyle, nutrition and general fitness. Against the background of collective responsibility and solidarity, behaviour that is deemed irresponsible ceases to be an option (Rosengarten 2005), as can be observed in campaigns of obligatory vaccinations, health screenings or the debates on the donations of data or organs. Being a "responsible and healthy subject" may be in conflict with respecting an individual's desire and freedom to not know, perhaps even culminating in a form of resistance to the expectation to be productive, reproductive and healthy for the common good.

This type of self-governance can be a burden, promoting a machinistic image of the self, and it might profoundly transform our understanding of the concepts of health and illness altogether. When we are incessantly at risk of becoming ill, the actual state of well-being becomes secondary.

CONCLUSION

Evaluating technological innovation in terms of their risks and benefits in the field of digital health is not an easy task, since we need to take into account a multitude of different areas, each of which has unique ethical repercussions. Practices of data collection, analysis and surveillance come with both positive and negative implications for individuals and social groups; hybrid forms of agency and distributed responsibility make it increasingly difficult to ascribe accountability to individual stakeholders or the technology; knowledge can be empowering but also reflects and reinforces existing and often oppressive power structures, culminating in the requirements of self-surveillance and self-governance in seemingly liberal societies.

None of these ethical challenges can be solved by focusing on technologies alone, as these will undoubtedly evolve in accordance with existing social structures in the present system, which is dominated by a free-market logic constituted by stakeholders such as insurers, pharmaceutical and technological companies. Key to addressing the ethical issues is uncovering the underlying social problems that are merely reflected in – but never solved by –innovative technologies.

BIBLIOGRAPHY

Andrejevic, M. (2007). *iSpy: Surveillance and Power in the Interactive Era.* Lawrence: University Press of Kansas.

Ajana, B. (2017). Digital Health and the Biopolitics of the Quantified Self. *Digital Health, 3*(1), 1. doi:10.1177/2055207616689509

Fairfield, J. A. (2017). *Owned. Property, privacy, and the new digital serfdom.* Cambridge: Cambridge University Press.

Latour, B. (1993). *We have never been modern.* Harvard University Press.

Lyon, D. (2007). Everyday surveillance: personal data and social classifications. In S. Hier, P. Sean, & J. Greenberg (Eds.), *The Surveillance Studies Reader* (pp. 136-146). Maidenhead: Open University Press.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21.

Mittelstadt, Brent Daniel/Floridi, Luciano (2016): The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics, 22*(2), 303–341. doi:10.1007/s11948-015-9652-2.

Mol, A. (2002). *The Body Multiple. Ontology in Medical Practice.* Durham/London: Duke University Press.

Prainsack, B. (2017). *Personalized Medicine. Empowered Patients in the 21st Century?*. New York: NYU Press.

Rose, N. (2007). *The Politics of Life Itself. Biomedicine, Power and Subjectivity in the Twenty-First Century.* Princeton/Oxford: Princeton University Press.

Rosengarten, M. (2005). The Matter of HIV as a Matter of Bioethics. In M. Shildrick, & R. Mykitiuk (Eds.), *Ethics of the Body. Postconventional Challenges* (pp. 71-90). Cambridge, Mass./London: MIT Press.

Sewell, G., & Barker, J. R. (2007). Neither good, nor bad, but dangerous: Surveillance as an ethical paradox. In S. P. Hier & J. Greenberg (eds.), The surveillance studies reader (pp. 354–367). Maidenhead: Open University Press.

Strohm Kitchener, K., & Anderson, S. K. (2011). *Foundations of ethical practice, research, and teaching in psychology and counselling.* New York, NY: Routledge/Taylor & Francis Group.

Wallach, W., & Allen, C. (2009). *Moral machines. Teaching robots right from wrong.* Oxford: Oxford University Press.

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology 30*, 75–89. doi:10.1057/jit.2015.5.