UNBOXING THE BLACK BOX

EPISTEMIC AND ETHICAL CONSIDERATIONS OF KNOWLEDGE PRODUCTION BY MACHINE LEARNING ALGORITHMS IN HEALTHCARE

MARTIJN LOGTENBERG

paper submitted in partial fulfillment
of the requirements for the course of
knowledge production in health and medicine
as part of the program of health humanities
at the school of humanities and digital sciences
of tilburg university

STUDENT NUMBER

2067986

INSTRUCTOR

prof. dr. Jenny Slatman

LOCATION

Tilburg University
Department of Culture Studies
School of Humanities and Digital Sciences
Tilburg, The Netherlands

DATE

October 22, 2022

NUMBER OF WORDS (INCLUDING FOOTNOTES AND CAPTIONS)

4,011

NUMBER OF WORDS (EXCLUDING FOOTNOTES AND CAPTIONS)

3,494

UNBOXING THE BLACK BOX

EPISTEMIC AND ETHICAL CONSIDERATIONS OF KNOWLEDGE PRODUCTION BY MACHINE LEARNING ALGORITHMS IN HEALTHCARE

MARTIJN LOGTENBERG

1 INTRODUCTION

Machine learning (ML) studies[1], outperforming clinicians, are increasingly prevalent. Such approaches frame the aim of providing better healthcare as a function of ever-increasing diagnostic accuracy, predictive power or treatment efficacy. Through the availability of *big data* from the real world, ML approaches become more viable, complicating our understanding within the traditional Evidence-Based Medicine (EBM) paradigm.
On the other hand, the ML approach does raise questions regarding privacy, practicalities (e.g incorporating algorithms into clinical practice)[2], epistemic responsibility and accountability. Although ML is often claimed to improve objectivity, this is not uncontested. The black-box nature of many ML algorithms raises questions regarding the societal necessity of our *trust in numbers*, as well as what counts as *knowledge*.

In this paper, I elaborate on the changing dynamics of knowledge production that accompany the rise of the ML approach.

RQ *How does the rise of the ML approach change dynamics on how we ought to produce knowledge in healthcare and medicine?*

Accordingly, this involves an elaboration on three concepts regarding knowledge production, which are covered respectively.

SQ1 *How does the rise of the ML approach complicate the traditional EBM hierarchy?*

[1] The use of the term ML might slightly deviate from the traditional broad meaning. I consider most conventional practices, in both data mining and model building, to provide a fair overview of what ML considerations are. My arguments, consequently, may not focus on the ML approach as a whole.

[2] Even though privacy and practical concerns currently are among the largest issues with ML deployment, I ignore this to provide space to think about the philosophical concepts relating to this approach. Therefore, I assume the ML solutions discussed are to be implemented with *sufficient safeguards and social infrastructure*.

SQ2 *How does the rise of the ML approach relate to our interpretation of trust in numbers?*

SQ3 *How does the rise of the ML approach relate to the concept of understanding?*

Within each of the three SQs, I explain the discussed concept and elaborate on the relationship and issues with the ML approach.

2 THE ML APPROACH AND EBM HIERARCHY

EBM has been the conventional way of *knowing* for decades. It prioritises objective and standard procedures, where RCTs and meta-analyses are on top of creating this knowledge hierarchy (Solomon, 2016). This hierarchy is visualised in Figure 1.

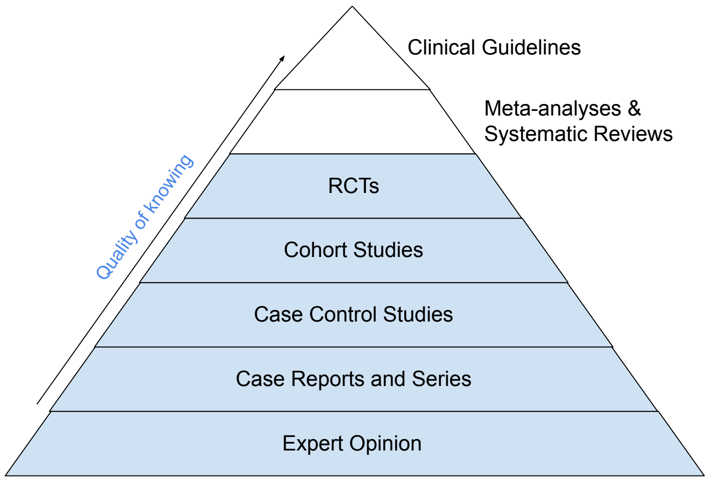

Figure 1: Visualisation of the EBM hierarchy

It is good to note here, that EBM and ML approaches are culturally more similar than one might expect. ML approaches have often been deployed to improve the diagnostic accuracy of the clinician (Grote & Berens, 2020). This subordinates the clinicians' expertise, which is similar to what is pursued through the EBM hierarchy. The purpose remains the standardisation of medical decisions and, in this sense, the most significant difference may be found between the practical deployment of the two approaches (i.e. RCTs for medicine efficacy and ML for diagnostic prediction or accuracy), rather than discarding the aim of EBM completely.

Even though this may have some truth to it, still this means the mechanism to create this knowledge is fundamentally different between the two approaches. Conventional ML approaches tend to focus on large cohort, secondary and real-world data (often called *big data*). This conflicts with the EBM understanding of reliable information, which prioritises RCTs on *controlled* and *randomised* groups[3], and subordinates the big data cohorts as lower in the hierarchy. This paradigm difference can be explained as (1) a critical perspective on EBM and (2) a means toward personalised medicine.

2.1 *ML approach as an answer to RCT critics*

Stegenga (2018) names the two main arguments for RCTs. Firstly, it is claimed RCTs are less susceptible to overestimation because of randomisation, although critics claim this is not practically true (cf. Benson & Hartz, 2000; Stegenga, 2018). Secondly, it is claimed RCTs are less susceptible to confounding factors, although critics claim such confounders can still be present in RCTs (Stegenga, 2018).
Critiques on RCTs are brought by Sanson-Fisher et al. (2007). Firstly, generalisability in RCTs may be limited through studying in the ideal circumstances, small sample sizes or selection of participants in specific contexts. Secondly, the dominance of RCTs discourages alternative study designs. This second point becomes a problem because RCTs often have short-term scopes and conventionally ask less innovative research questions (Sanson-Fisher et al., 2007). Therefore, I argue that the *dominance* of RCTs in the EBM structure is something undesirable we ought to prevent by including multivariate study designs. Additionally, we can see instances where RCTs are impractical, impossible or unethical. However, even if an RCT is deemed ethical, they still withhold treatment from the control group.

The ML approach claims several advantages and, thereby, provides some answers to critics of RCTs. Before going into those differences, I perceive the implementation of randomisation in the ML approach as important enough to elaborate on it and visualise it in Figure 2.

[3] I am aware that EBM and RCTs are not interchangeable. However, the dominance of RCTs in meta-analyses, as well as, the weight given to *randomisation* and *control* concepts within RCTs, compose the most fundamental differences between the ML approach and EBM.

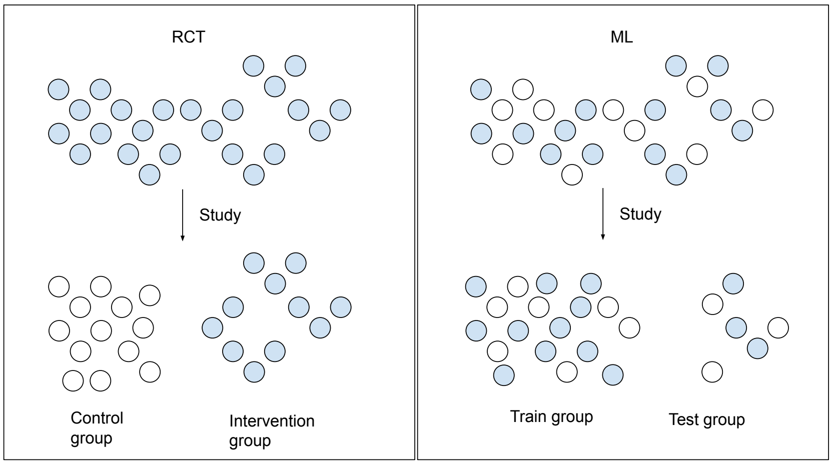

Figure 2: Difference in the implementation of randomisation between RCTs and the ML approach

Where RCTs randomly assign persons to treatment or control groups, the ML approach implements randomisation differently. In the ML approach, two real-world groups are defined randomly, one for building a model and one for validation. Both groups ought to be similar, and thus do not differ in received treatment. The fundamental assumption is: If the model truly predicts the factor Y (e.g. whether a disease progresses), based on factors X (e.g. individual characteristics and received treatment) then the model should also be able to predict this for data it has not seen yet.
This is fundamentally different on two points: (1) ML restrains from actively assigning treatment and is, hence, more appropriate when this freedom is not given to the researchers and (2) ML does not measure the efficacy of treatment directly, but predicts, for example, the progress of the disease[4]. Based on variation between individual characteristics and which treatment was used, the model learns which combinations are effective in treating disease. Note that this is a different paradigm which does not lend itself to direct comparisons and focuses on the individual rather than the treatment.

The ML approach does possess characteristics that solve issues brought up by RCT critics. The major disadvantage of RCTs is the challenge of generalisability, which is partly tackled by the ML approach which does not use ideal circumstances but real-world data and is often able to include more subjects for less cost since most data is readily available (Lee & Yoon, 2017). Although selection bias is still a real threat to the ML approach, it is mitigated when data is drawn from multiple sources, which is easier

[4] Since it is the most intuitive case, I stick to research done to test *safe treatment* to treat *disease*. I am aware this both simplifies and limits the concepts discussed.

to do with secondary data. Additionally, this can be extended to cases where no sampling is needed because the sample is the whole population (*N=All*). However, this is still a controversial topic in academic literature (cf. Harford, 2014).
On top of this, since the data is cheaper and data sources are more variable, this may incentivise researchers to ask different and more innovative questions. Furthermore, data sources used for ML studies are often collected for other use (e.g. institutions providing care), this provides observation advantages similar to those of observational cohort studies (e.g. observation length).

There is a debate about whether RCTs or the ML approach are more susceptible to confounders. On the one hand, RCTs have smaller sample sizes and *the law of large numbers* gives the ML approach an advantage. On the other hand, an RCT takes place in a controlled environment where lots of factors are kept stable (cf. Lee & Yoon, 2017; Nguyen et al., 2017). This remains impossible for the ML approach.

2.2 *ML approach as the mean toward truly personalised medicine*

Solomon (2016) provides us with a charming description of personalised medicine: "The ethos of personalised medicine is that it rejects the supposedly one-size-fits-all or *cookbook* therapeutics of evidence-based medicine." However, in practice, EBM strategies are still utilised in personalised medicine to study subpopulations. I argue that a *subpopulation* is not the same as *personalised treatment*.[5]

The ML approach offers an alternative paradigm to use each individual's combination of factors responsible for the variation of treatment to move towards truly personalised medicine. There are two reasons why this is possible with the ML approach: (1) big data and (2) structure to learn complex relationships, mostly through the rise of deep learning. Since big data is not exclusive to the ML approach, I focus on the interaction of big data and the model interactions, rather than both arguments separately.
Note, however, that there is one fundamental difference between tradi-

[5] One definition of *personalised medicine* assumes that genetic information carries all variational information and the deterministic nature of the individual (Savard, 2013). This poses both moral questions about individual responsibility, but may also be too restrictive. Most common diseases are not caused by only genetic factors but are caused by a complex interplay of genetic, lifestyle and environmental factors (Feiler et al., 2017). I do not think it is useful to get into that debate, but I use a perspective including other information as well. This does not impact the case made, as it is applicable to the *only genetic information* definition as well.

tional personalised medicine (TPM) and the ML variant. In TPM, we seek causal relationships between the individual's characteristics and the disease variation. The ML approach is not fit for obtaining causal relationships (Wilkinson et al., 2020).

Personalised medicine has struggled to find success in a number of areas. This is due to the complexity of factors relating to disease progression, risk and treatment efficacy, thereby, making it ineffective to study single factors (e.g. cancer; Cheng & Zhan, 2017). The nature of the ML approach, on the other hand, provides us with a mechanism to research a multitude of factors simultaneously, including their relationship (Zhang et al., 2018). On the most fundamental level, we can think of TPM as being burdened by an *understandability requirement*. This however is built on the assumption that one cannot study personalised medicine without understanding the outcome. ML models do not provide any understanding to the researcher but may be able to learn an understanding themselves (this is a point I will return to). Since we can test for the extent a model has learned the complexity, this may be a plausible approach to catch the underlying complexity. In this sense, we can think of the ML approach as folding the complexity in a model. This difference is visualised in Figure 3.

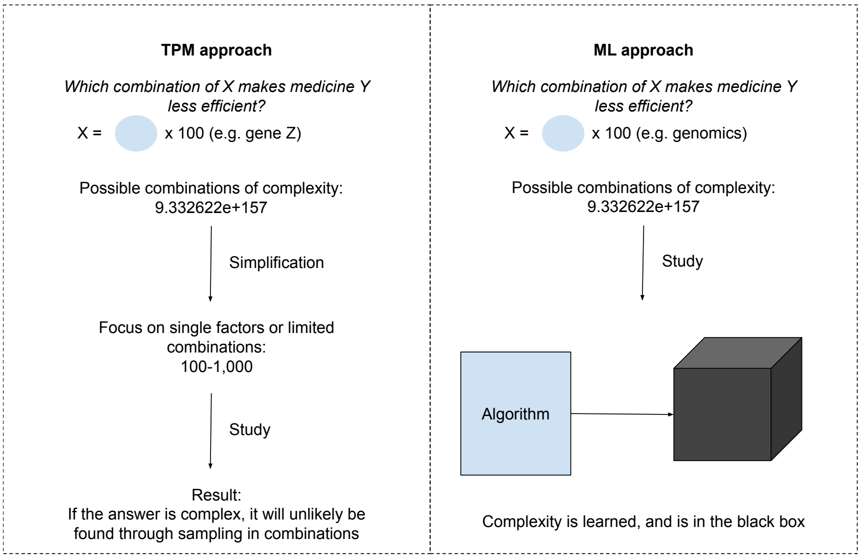

Figure 3: TPM vs ML approach related to combinations of factors to study

The ML approach is, hence, more effective in learning complexity than human researchers. This increases the window of opportunity in which researchers may find factors associated with treatment variation, thereby providing more opportunities to tailor treatment to each specific individual.

The drawback is that the answer is now no longer *out there*, but folded in the black box. This raises two questions: (1) 'Do we need to know what is in the black box?' and (2) 'If so, are we able to access and understand the black box?'. One might argue that as long as we know the model understands the mechanisms, it is enough. However, this raises some epistemic and ethical concerns which I will return to, but at least requires *trust in models*, which I believe to be a move beyond our *trust in numbers*.

3 THE ML APPROACH AND OUR TRUST IN NUMBERS

Porter (1995) shows that, historically, the statistical society pursued the provision of numbers and allow them to speak for themselves. This trust in numbers was important to make facts free of opinions. Therefore, the public would receive thoroughly and self-explanatory facts as the most important source of knowledge.

Therewith, the main characteristics of numbers are that they provide knowledge, authority and, hence, have the capacity to be seen as objective, unbiased and fair.

Firstly, the need for self-explanatory facts is one that is grounded in the core values of democracy. Secondly, making decisions on numbers, to some extent, shifts the accountability of a decision towards those numbers (Porter, 1995). Interpretability, and the associated trust, are therefore necessary to give science a function within society.

The ML approach has a clear disadvantage here. Today the most promising architecture in the ML approach is the Artificial Neural Network, a model that lacks interpretability[6]. An example is visualised in Figure 4.

[6] I am aware that *interpretability* is an ambiguous concept. For this paper, I do not talk about decomposability, simulatability or other related concepts. I mostly refer to *explainability*, which I define as the provision of human-readable justification over each individual model prediction.

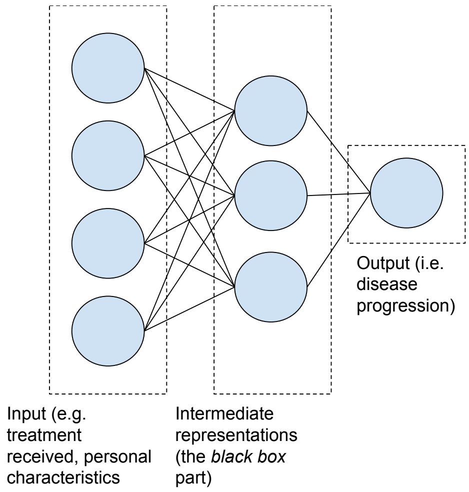

Figure 4: Example of an Artificial Neural Network (feedforward architecture). *Note:* In reality, intermediate representations are significantly larger than three nodes.

Artificial Neural Networks generate intermediate representations within the model through mathematical transformations. These transformations are (1) composing a complex structure, with no decomposable meaning, and (2) learned by the model itself, and thus not interpretable.

Han and Liu (2022) classify the need for interpretability as a *key challenge* within clinical sciences since the impacts on medical decision-making tend to be larger. Therefore, an upcoming branch within the ML approach is explainable Artificial Intelligence (XAI), with the ethos of explaining *why a model works*, instead of only quantifying *how good it works*.
XAI techniques are mostly grounded in proxies outside the black box model (i.e. the agnostic approach). Therefore, explanations are generated through, for example, having a simpler model explain the black box model outcome or changing factors and see how the outcome of a model changes accordingly. Although these strategies do provide an explanation, they do not provide the true reasoning behind the black box model (Lipton, 2018). Given that it is not possible to provide the real reasoning behind a black box model, one may question whether trust can only be earned through interpretability. A prevalently made case is that models in medical practice often have huge impacts on human lives and that, therefore, we cannot accept solely relying on models when they are unable to incorporate a clinician's expertise or communicate in another way (Kim, 2015, pp. 17-18). However, although it sounds intuitive that one could only trust something that is able to explain itself, this may not be true. I will show how explainable models can still be unfair. Therewith, I prove that interpretability

ought not to be *necessary* for trust in models.

Society has seen many examples of how unfairness can be built into models, which led to faulty decision-making. In some cases, models have been disfavouring specific groups, by discriminating on factors that are not desirable (e.g. ethnicity or gender). The reason this happens is that models are built upon real-world data, and when this data implies human bias, the model will copy the human bias in its own decision-making.
Such biased data is not always explicit. As an example, data can be biased through discriminatory societal norms, such as in the case of predictive policing models. These models deploy more police officers in areas where previous crime has been high. However, the crime rate itself is also a proxy for the amount of deployment in an area since once one deploys more officers in an area, one will find more crime *even if the crime rate is lower than in other neighbourhoods* (Richardson et al., 2019). If such confounding factors are not taken into account by the model, they would not show up in its explanations. Therefore, interpretability does not mean full understanding or fairness, and therefore may not deserve our trust.

This problem becomes most evident when we think of a model's *accountability capacity*, specifically in the context of healthcare where decision-making has serious impacts on individuals. If a model outperforms a clinician, less harm is done to fewer people. This might make it a utilitarian net positive outcome. Moreover, clinicians may also lack interpretability, specifically for patients with low health literacy.
On the flip side, the problem is two-fold, namely (1) both the clinician and the creator are not able to fully account for a model's output and (2) ML models will make some mistakes, leaving a patient harmed. Therefore, so far one would find few advocates for shifting accountability (Smith, 2021). The *epistemic responsibility* remains with the clinician.

A more sceptical view can be taken to explain trust in models. Accordingly, the amount of trust is not earned by fairness but by low *perceived* risk, experiential performance and interpretability (Hengstler et al., 2016). Note how this would mean that interpretability is still necessary for our sociological trust, even if it does not always increase its fairness. Therefore, trust in models is not something rational but is grounded in a sense of objectivity as perceived by the clinician and the patient. Along that line, the main characteristics of models are that they have authority and hence are *seen* as objective, unbiased and fair. Consequently, this move beyond trust in numbers plays with the theory of how trust is earned, and if people are able to trust *what a model says*. This is a sociological question, rather than a philosophical one, but Elish and Boyd (2018) show how ML limitations are often neglected and society shapes an alternative truth on which their trust is built.

4 THE ML APPROACH AND THE CONCEPT OF UNDERSTANDING

I leave the sociological question aside and assume the theory of perceived truth is necessary for our trust processes. In the preceding section, I have shown an example of how the ML approach suffers from confounders in the same way as the statistical approach. This illustrates that the ML approach does not solve a potential lack of truth in the statistical approach. Furthermore, the ML approach changes the fundamental[7] scientific concept of understanding.

Where statistics provide society with *open* knowledge through interpretable results, this cannot be said from the ML approach. As posed earlier, the question then is: 'Do we need to understand the black box?', for which I already raised ethical concerns, but one can pose an epistemic case as well. As many philosophers have mechanised, meaning is grounded in symbols in the real world. This is intuitively illustrated by the 'Chinese Room' argument (CRA). In this thought experiment, Searle receives a set of Chinese characters and translates them into English characters, with the help of a Chinese cataloguing system (Searle, 1980). As argued by Searle, this system nor himself *understands* Chinese. This also extends to clinical practice.

> "In fact, computers don't *know* or do anything besides producing the product of their program designers' instructions. A program's *recommendation* for medical therapy is simply a set of symbols representing the prior thoughts of its programmer. In itself, the running program understands nothing about medicine, health, illness, or death. It does not even know that it is a program!" (Luger, 2021, pp. 116)

Grounding is, hereby, necessary for understanding. This is something that can currently only be offered by humans, specifically clinicians who have sufficient knowledge of clinical practice to ground knowledge. Thereby,

[7] Another concept changed by the ML approach would be replicability. However, as I believe this is more of a misunderstanding of these concepts, I do not cover it in this paper. This misunderstanding is grounded in the idea that *data sharing* is fundamental for replication (Kitzes et al., 2018, pp. 3-7; Tenopir et al., 2011). I believe this creates *replication without difference*, which misses the ethos of replication altogether (Schmidt, 2009). Replicating the experiment with the same data is not replication, it is just *reproducing* the exact same experiment.

we can say ML models do not understand anything. Implementing black box models hence creates a situation in which neither the model nor the clinician understands anything. We may wonder if this is desirable; if understanding is necessary for knowledge or if knowledge can be something that is framed into *what a model tells us*. I believe the latter would be hard to argue.

Searle's CRA is primarily criticised on two points. Firstly, for misunderstanding the concept of a system, where supposedly Searle and the cataloguing system both from the system and therefore understand Chinese together. Note, that this requires us to include the clinician in the system as well. This would be true if the clinician can then actually read the black box.
Secondly, for misunderstanding the concept of mental states and brains (Harnad, 2001). A computer's mental state, here supposedly provides us with the same level of understanding through computational transformations as a human brain would. This refers to the *mind-body connection*. One way to refute this is by pointing at examples where the mind and the body collaborate in unconscious cognition (e.g. one does not remember the numbers of a phone number, but is able to dial the number because their fingers recognise the pattern). Moreover, cognition can be thought of as knowing what an object is through experience and generalising, instead of calculation (Mitchell, 2019). I believe it, therefore, remains impossible for the ML approach to break the *barrier of meaning*.

5 CONCLUSION

In this paper, I showed how the ML approach has changed the dynamics regarding knowledge production in healthcare on three points.
Firstly, I have framed the ML approach as both an answer to RCT criticism as well as a move towards personalised medicine. By doing this, I have aimed to sketch the window of opportunity the ML approach offers us to provide more variability in our research designs, increase generalisability and enable effective implementation of true personalised medicine. By those means, the ML approach offers knowledge for us to utilise and implement in practice, increasing the quality of research and healthcare. This requires a reconsideration of the traditional EBM hierarchy, providing more recognition for big data.
Secondly, I have framed the ML approach as a move beyond trust in numbers towards trust in models. Literature often introduces interpretability as a key concept here, but I argued why XAI methodology does not provide true interpretability. The debate on the need for interpretability gives us

three answers: it is needed for democracy, fairness and accountability. I am aware I did not provide any answer to this debate by stating that current technology is not able to take on epistemic responsibility and providing a sociological answer instead on how trust can be formed through alternative truths. I believe this is closer to practice, although I do not believe this is a sufficient answer to the debate.
Thirdly, I have shown how the ML approach changes concepts regarding truth on the fundamental axis of understanding. This paper argues that implementing the ML approach brings us into a situation where neither the clinician nor the model understands anything.

Thereby, I have both made a case as to why the ML approach should be implemented to increase the window of opportunity while breaking down that same case myself in the two latter SQs. That may leave the reader unsatisfied, but the truth is that I am unsatisfied too. Since many believe we are at the start of a breakthrough, it is worrying that there are no answers to the ethical and epistemic questions this paper leaves.
Future literature should therefore focus on to what extent we need understanding and accountability in our *ways of knowing*. This may provide a satisfying answer to the underlying question 'Is the rise of the ML approach desirable in healthcare?'.

REFERENCES

Benson, K., & Hartz, A. J. (2000). A comparison of observational studies and randomized controlled trials. *The New England Journal of Medicine*, *342*, 1878-1886. doi: 10.1056/NEJM200006223422506

Cheng, T., & Zhan, X. (2017, 3). Pattern recognition for predictive, preventive, and personalized medicine in cancer. *EPMA Journal*, *8*, 51-60. doi: 10.1007/s13167-017-0083-9

Elish, M. C., & Boyd, D. (2018, 1). Situating methods in the magic of big data and ai. *Communication Monographs*, *85*, 57-80. doi: 10.1080/03637751.2017.1375130

Feiler, T., Gaitskell, K., Maughan, T., & Hordern, J. (2017, 1). Personalised medicine: The promise, the hype and the pitfalls. *The New Bioethics*, *23*, 1-12. doi: 10.1080/20502877.2017.1314895

Grote, T., & Berens, P. (2020, 3). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, *46*, 205-211. doi: 10.1136/medethics-2019-105586

Han, H., & Liu, X. (2022, 1). The challenges of explainable ai in biomedical data science. *BMC Bioinformatics*, *22*, 443. doi: 10.1186/s12859-021-04368-1

Harford, T. (2014). Big data: A big mistake? *Significance*, *11*, 14-19. doi: 10.1111/j.1740-9713.2014.00778.x

Harnad, S. (2001). Minds, machines and searle 2: What's right and wrong about the chinese room argument. In M. Bishop & J. Preston (Eds.), *Essays on searle's chinese room argument.* Oxford University Press.

Hengstler, M., Enkel, E., & Duelli, S. (2016, 4). Applied artificial intelligence and trust-the case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, *105*, 105-120. doi: 10.1016/j.techfore.2015.12.014

Kim, B. (2015). *Interactive and interpretable machine learning models for human machine collaboration* .

Kitzes, J., Turek, D., & Deniz, F. (2018). *The practice of reproducible research: case studies and lessons from the data-intensive sciences*. Univ of California Press.

Lee, C. H., & Yoon, H. J. (2017, 3). Medical big data: Promise and challenges. *Kidney Research and Clinical Practice*, *36*, 3-11. doi: 10.23876/j.krcp.2017.36.1.3

Lipton, Z. C. (2018, 6). The mythos of model interpretability. *Queue*, *16*, 31-57. doi: 10.1145/3236386.3241340

Luger, G. F. (2021). *Knowing our world: An artificial intelligence perspective*. Springer. doi: 10.1007/978-3-030-71873-2

Mitchell, M. (2019, 2). Artificial intelligence hits the barrier of meaning.

Information (Switzerland), *10*, 51-53. doi: 10.3390/info10020051

Nguyen, T. L., Collins, G. S., Lamy, A., Devereaux, P. J., Daurès, J. P., Landais, P., & Manach, Y. L. (2017, 4). Simple randomization did not protect against bias in smaller trials. *Journal of Clinical Epidemiology*, *84*, 105-113. doi: 10.1016/j.jclinepi.2017.02.010

Porter, T. (1995). The political philosophy of quantification. In *Trust in numbers: The pursuit of objectivity in science and public life* (p. 73-84). Princeton University Press.

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYUL Rev. Online*, *94*, 192-233.

Sanson-Fisher, R. W., Bonevski, B., Green, L. W., & D'Este, C. (2007, 8). Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventive Medicine*, *33*, 155-161. doi: 10.1016/j.amepre.2007.04.007

Savard, J. (2013, 6). Personalised medicine: A critique on the future of health care. *Journal of Bioethical Inquiry*, *10*, 197-203. doi: 10.1007/s11673-013-9429-8

Schmidt, S. (2009). Shall we really do it again? the powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*, 90-100. doi: 10.1037/a0015108

Searle, J. (1980). The chinese room. *Minds, Brains, and Programs*.

Smith, H. (2021, 6). Clinical ai: opacity, accountability, responsibility and liability. *AI and Society*, *36*, 535-545. doi: 10.1007/s00146-020-01019-6

Solomon, M. (2016, 3). Medicine and society: On ways of knowing in medicine. *CMAJ*, *188*, 289-290. doi: 10.1503/cmaj.150673

Stegenga, J. (2018). Evidence in medicine. In *Care and cure: An introduction to the philosophy of medicine* (p. 103-124). The University of Chicago Press.

Wilkinson, J., Arnold, K. F., Murray, E. J., van Smeden, M., Carr, K., Sippy, R., . . . Tennant, P. W. (2020, 12). Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, *2*, e677-e680. doi: 10.1016/S2589-7500(20)30200-4

Zhang, S., Bamakan, S. M. H., Qu, Q., & Li, S. (2018, 8). Learning for personalized medicine: A comprehensive review from a deep learning perspective. *IEEE Reviews in Biomedical Engineering*, *12*, 194-208. doi: 10.1109/RBME.2018.2864254