Machine Learning and the End of Theory: Reflections on a Data-Driven Conception of Health

Guersenzvaig, Ariel

Erstveröffentlichung / Primary Publication
Konferenzbeitrag / conference paper

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

MACHINE LEARNING AND THE END OF THEORY

REFLECTIONS ON A DATA-DRIVEN CONCEPTION OF HEALTH

Guersenzvaig, Ariel
Elisava Barcelona School of Design and Engineering, UVic-UCC
Barcelona, Spain
aguersenzvaig@elisava.net

KEYWORDS

machine learning, health, theory, normative concepts

DOI: 10.34669/wi.cp/4.5

ABSTRACT

Taking the notion of health as a *leitmotif*, this paper discusses some conceptual boundaries for using machine learning—a data-driven, statistical, and computational technique in the field of artificial intelligence—for epistemic purposes and for generating knowledge about the world based solely on the statistical correlations found in data (i.e., the "End of Theory" view). The thrust of the argument is that prior theoretical conceptions, subjectivity, and values would—because of their normative power— inevitably blight any effort at knowledge-making that seeks to be *exclusively* driven by data and nothing else. The conclusion suggests that machine learning will neither resolve nor mitigate the serious internal contradictions found in the "biostatistical theory" of health—the most well-discussed data-driven theory of health. The definition of notions such as these is an ongoing and fraught societal dialogue where the discussion is not only about what *is*, but also about what *should be*. This dialogical engagement is a question of ethics and politics and not one of mathematics.

1 INTRODUCTION

An influential argument in favor of using artificial intelligence (AI) for epistemic purposes can be found in Chris Anderson's essay "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" (Anderson, 2008), where he argues that big data and the AI tools used to process them offer a new way of understanding the world based on the statistical correlations between data. Correlations make causal explanations —i.e., human-made (conceptual) causal models and theories—unnecessary for scientific progress. Anderson does not just propose the use of AI to computationally support scientific discovery and theory generation; he wants AI to take the lead because "science can advance even without coherent models." Data-driven discovery is also defended by the astrophysicist Kevin Schawinski: "Let's erase everything we know about astrophysics. To what degree could we rediscover that knowledge, just using the data itself?" (Cited in Falk, 2019). His "generative" approach represents a much weaker—yet more plausible—version of Anderson's argument. Schawinski (et al., 2018) concedes that human insight is still required for high-level interpretation, which enables an expert to make sense of the discoveries. For some, an instantiation of this perspective can be found in the case of *AlphaFold*, an AI system that has been able to accurately predict the 3D structure of a protein, thus solving one of the great contemporary challenges of biology (Heaven, 2020).

For space reasons, I will not explore the view in detail nor the various epistemological questions that emerge from it (for a detailed treatment, see, e.g., Casacuberta and Vallverdú, 2014). Rather, I shall engage with the thrust of the argument—very succinctly laid out in the previous paragraph—indirectly and try to scrutinize whether data makes theories and previous conceptions truly redundant. To structure the discussion, I will examine whether I can use machine learning (ML)—possibly AI's most popular technique nowadays—to resolve or at least mitigate some of the internal contradictions found in a well-discussed data-driven theory of health that seeks to define what health is—the "biostatistical theory" of health—which I will first briefly introduced below.

2 CONCEPTS OF HEALTH AND TELLING WHO'S HEALTHY

Before jumping to machine learning, I will summarily consider some relevant aspects around the notion of "health," which will be the *leitmotif* here. First and foremost: there is no consensus on what health is. In the Western literature on health, we find, on the one hand, "naturalist" theories, whereby health is a value-free notion that is determined by empirical facts. On the other, we find "normativist" theories, whereby health is essentially value laden. I will briefly outline the naturalist view, which pursues a descriptive goal like Anderson's: to derive knowledge from statistical data.

Possibly the most vigorously debated naturalist perspective on health is the "biostatistical theory" proposed by Christopher Boorse (1977; 2014). This theory rests on a nonnormative understanding of biological function and a statistical notion of the concept of "normality." For Boorse, health and disease are nothing more than biological states. In this sense, to say that an organism is healthy is to describe a natural fact and not to make an assessment of it in terms of good or bad, desirable or undesirable, and so on. Boorse states "if diseases are deviations from the species biological design, their recognition is a matter of natural science, not evaluative decision" (1977, p. 543).

As will be made clear shortly, Boorse's biostatistical theory fits nicely with the end of theory proposed by Anderson. Boorse (1977, p. 542) maintained that health is the "statistical normality of function" and that "the normal is the natural" (1977, p. 554). Diseases are "internal states that depress a functional ability below species-typical level" (1977, p. 542; 2014, p. 684). An organism is thus healthy when its functioning conforms to its natural design and function. Boorse's theory is much richer than I can cover here, alas, yet the upshot is that health is the fitness of an organism to perform its normal functions with statistically normal efficiency under typical conditions.

Typical levels for a species are those close to the statistical mean (Boorse, 1977, pp. 558–559). Although "normal" levels could be determined statistically for the whole species, from a clinical perspective, it would be impossible to conduct a comparison at a species level. Hence, a smaller reference class is needed. Since species design seems to be contingent on sex, age, and race, the statistical abstractions should be made from reference classes smaller than species (Boorse, 1977, p. 558). To assess the normality of a biological state for a subgroup within a species, Boorse needs some sort of benchmark of normality. To determine whether a particular organism is healthy in relation to the species-typical level Boorse introduces the notion of a "reference class."

A reference class is "a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species" (1977, p. 555). Examples of reference classes would be "a 35-year-old white woman" or "a neonate of Aymara ancestry." In short, according to Boorse, if we want to establish the health of a neonate's heart, we should compare it to the hearts of other neonates, factoring in sex and race, and not to an average adult human heart, as an adult with the constant heart rate of a neonate would be considered diseased, and vice versa.

However reasonable and clinically necessary reference classes may be, Boorse undermines himself methodologically by introducing them—and rather evidently so. This is an objection noted by Elselijn Kingma (2010): It is not clear *why* it would be appropriate from a naturalist, nonnormative perspective to factor sex, age, and race in when calculating normality and not other criteria. There are no empirical facts that determine that "neonates" represent an appropriate reference class, but "people with beards" or "children with dental cavities" are not. Indeed, both beards and caries are statistically frequent. What's more, even allowing for sex to be partially constituted by some empirical indicators, such as testosterone levels, its status as a full-blown natural category has been a hotly debated issue since the 1990s (see Butler, 1990).

Kingma convincingly shows that Boorse cannot justify his choice of appropriate reference classes without involving value judgements and prior theories and conceptions of health. If Boorse's theory seeks to stand independently of normative knowledge, it should be able to offer a value-free explanation of which criteria constitute an appropriate reference class. It is not enough to assert that "sex," race," and "age" are (the) appropriate reference classes. In other words, for the biostatistical theory to be truly naturalistic, the required reference classes must be determined and justified neutrally and empirically objectively without underlying value judgements. And for his critics, this is what Boorse's biostatistical theory fails to achieve.

Would it be possible to use ML to "end the theory" that is inherent in the above-mentioned reference classes? Could ML release the biostatistical theory from the insidious values, conceptual models, and theories that sabotage its quest for nonnormativity? If we succeed in this task, Anderson's views on the end of theory would become more compelling and Boorse's work would be free from its internal methodological contradictions. Above all, a truly naturalistic theory of health would be closer to hand.

3 MACHINE LEARNING TO THE RESCUE?

Why "sex," "age," and "race" instead of other criteria? Kingma asked. Fortunately, given the possibilities of machine learning, we could virtually limitlessly extend the range of reference classes beyond these three. Certainly, an ML system could use any attribute of the human body that can be incorporated into database tables: from eye color to bone density to hair thickness to lung capacity to weight. For instance, an ML system could then be trained with anthropometric data: skull shape and volumetric measurements, abdominal circumference, limb alignment, eye color, and so on. To train the system, we would first need to label the input data so that the system could develop a model from it and assign an output label for a new value, i.e., a result in terms of "healthy" or "diseased."

Alas, this would not satisfactorily address Kingma's objection regarding prior normativity and subjectivity, which would still be detected in the defining training variables. "Why are *these* signs used and not others?," we might ask. Why skull shape or eye color? The only difference is that instead of having three criteria without atheoretical justification, we would possibly have many more.

Perhaps there might be a way out of this tangle. Given the sheer number of reference classes that could be defined, it would be conceivable to make health assessments by considering randomly chosen classes. If we were to work with a reference class based on *multiple* data sources, this could perhaps bring us closer to assessing an individual's health status in a non-normative way.

This seems technically viable thanks to “random sampling.” The system could be trained with anthropometric data, medical history data, and clinical signs data from people labelled as healthy or diseased. In this case, though, we would only use a fraction of the available data, randomly selecting which database tables are considered or left out in the construction of the model. So, the system could include data about the swelling (or lack thereof) of the lymph nodes but leave out data about blood pressure or head circumference.

However, this does not remove the normative influence. Indeed, while the choice of classes to be used as a benchmark would be random, the pool of reference classes the system could choose from would be normatively and theoretically justified. The very choice to use anthropometric and clinical database tables is itself a decision that is based on the value judgements and prior theories that underpin the judgement about what database tables to include. And this does not bring us any closer to having to accept that prior theory has become unnecessary thanks to machine learning.

Still, a defender of the end-of-theory view might retort: If the problem is in the choosing, what if the system could be trained with available data of *any* kind? Perhaps nonmedical data, which is prima facie neutral about health (e.g., high school grades, social network activity, data from tax returns, parking violations records, etc.) could be used. All that would be necessary is to train the system with a dataset containing the nonmedical records of healthy and diseased groups or individuals. After a while, the system would detect salient features in the data and identify connections between the medical and the nonmedical data.

However, another evident problem emerges, that of circularity. In supervised machine learning, the putatively neutral data needs to be connected to health (or to a proxy thereof) to be able to generate a result. In the same way that a bird-identification app needs to be taught (through labels) what different types of birds look like to be able to assign a label to an image of a new bird, the health system would still need to find patterns in the parking records or in the high-school grades belonging to healthy or diseased individuals or groups. Yet to train the system in this manner, we would necessarily require a prior conception of health and disease to label the data. It is precisely this that enables individuals or groups to be classified as either healthy or diseased! And this manifestly violates the very theory-free approach we are trying to achieve.

Undoubtedly, supervised machine learning can be used to make valuable assessments of health based on large volumes of data once the appropriate reference classes have been defined and the data has been labelled, but it is far from making prior theory redundant. On the contrary, it highlights how data is intertwined with theory.

4 THEORY PRECEDES DATA

A defender of the end-of-theory position might claim that unsupervised machine learning is the way to go as, it would not be tainted by circularity. Indeed, unsupervised ML does not need labelling up front, and the reference classes could emerge as clusters from the data alone thanks to correlations. This is a point also made by Anderson (2008), for whom big data "allow us to say: 'Correlation is enough.' We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot."

It is certainly possible to cluster data without labels, for instance, by using profiling to detect patterns or structures present in the data that have not been previously hypothesized. Through techniques such as profiling, classifications (i.e., clusters) can be made without the need for causal models or other theoretical explanations. What's more, the system could deal with vast purely numerical vectors, whereby the original attributes—i.e., the column labels—in the database would not even need to be explicit at all. Only the numbers would be required.

We could generate reference classes thusly. Yet, some questions remain. Would these clusters truly *precede* theory and models? Would the classes be naturalist in the strict sense? To answer these questions, the discussion needs to turn to the nature of data itself. Since the issue is vast, I will be content to outline reasonable doubts about the possibility of unsupervised ML being able to generate naturalist, *atheoretical* reference classes.

The first aspect to bear in mind is that data are not directly and neutrally incorporated into systems as if they were a mirror of empirical reality. Data need to be collected and processed in order to be computationally readable. This first step already implies a reduction of the complexity of the world to a few database fields. This reduction, contrary to the intentions of naturalists, is marked by values such as efficiency, effectiveness, cost-effectiveness, budget limitations, and so on. No researcher simply "throws numbers" into a computer system.

Bowker and Star (2000) famously showed how classification systems shape and are themselves shaped by perspectives on the world and by social interactions. Data are not a Platonic entity. Data are a construct that is *made appropriate* to the systems and classification schemes in which they are incorporated according to some goals or purposes. Categories and attributes make some aspects visible while making others invisible. They are never a mere naturalist reflection of reality.

There is a second problem: misrepresentation in data selection ("sampling bias"), a common problem in datasets. An example is found in artificial intelligence systems that aim to assist dermatologists in the detection of skin cancer. These systems exhibit great potential by achieving levels of prediction comparable or superior to that of dermatologists (Esteva et al., 2017; Fink et al., 2020). However, one grave problem from the perspective of justice is that they are much more accurate with light skin than with dark skin, which likely has to do with the datasets employed to train the systems to recognize potential moles (Adamson and Smith, 2018).

So, we see that approaches that start from data might yield reference classes, but there is no theory-free way of assessing the validity of these classes to determine that these are adequate from the statistical, clinical, and justice perspectives. We necessarily need auxiliary theories to align the three perspectives, starting from the very beginning at the data collection stage.

Relatedly, there is a third obstacle in the road to naturalist, theory-free reference classes: Artificial intelligence systems are characteristically affected by structural biases that go beyond sampling bias. Consider the gender bias that not only plagues medical data but medicine itself. Its history shows a *structural* lack of interest in women’s health. Let’s review a few examples. Eight of the 10 prescription drugs that were withdrawn from the US market in the period 1997–2001 posed greater health risks to women than to men (USGAO, 2001). Diseases are ignored when they do not affect men, as in the case of endometriosis (Huntington and Gilmour, 2005). Procedures and therapies might have distinct effects on men and women, yet this can go unnoticed for many years until women are included in controlled trials (Ridker et al., 2005). As happened with COVID-19 vaccines, the effects of medical interventions on menstruation seem to be an afterthought. Indeed, changes to period patterns and vaginal bleeding are not included among the common side effects of COVID-19 vaccination listed by the UK’s regulatory agency MHRA, yet these events are reported to be frequent shortly after vaccination (Male, 2021).

In short, bias in machine learning is first and foremost a matter of justice and structural inequalities; it is not only a technical issue of statistical representativeness. There is a vast literature related to how race, gender, age, educational level, cognitive abilities, and many other vectors of unfairness (e.g., Benjamin, 2019; Eubanks, 2018) interact with datasets and algorithms. It would be irresponsible to accept and trust reference classes generated by a ML system *as is* without further assessment. This assessment necessarily requires auxiliary theories, for example of justice.

Lastly, and fourth, models in the social sciences can change the basic coordinates they describe (Blakeley, 2020). An example is the way “the economy” is measured with *prima facie* neutral indicators such as gross domestic product, the unemployment rate, or the Dow Jones Index, while other indicators—such as the humanity of labor, the impact of economic activities on the environment, or extreme inequalities—are not considered. Taking these as relevant indicators is a choice motivated by political and moral views.

To exemplify this, consider the body mass index (BMI), which is a measure of body fat based on height and weight that categorizes a person along a continuum from underweight to obese. The higher the BMI, the stronger the risk of suffering from heart failure (Khan, 2018). Arguably, the very existence of this model reflects scientific, cultural, political, social, aesthetic, and even religious values and perspectives present in society. Even when the data are not statistically biased *per se*, it's important to be aware that neither the index nor the data are atheoretical in the strict sense, as they determine what counts as an important or promising indicator for detecting risk. To illustrate the difference, consider an example related to the issue of cardiovascular risk. The "social determinants of health" perspective—unlike the BMI—pays primary attention to systemic and structural parameters, such as access to good transportation, education, and housing, which can also be positively or negatively linked to heart disease and stroke (see e.g., WHO, 2010).

5 CONCLUSION

I have explored a series of problems with the "end of theory" view. Prior theory, subjectivity, and values blight the naturalistic effort. The necessary labelling required for training data in supervised ML systems introduces an element of circularity that is unacceptable from a naturalistic point of view. At the same time, assessing the appropriateness of a reference class determined by unsupervised machine learning and profiling techniques requires prior theoretical conceptions of health. ML systems are prone to suffering from sampling and structural biases. These biases are often the result of prior theories and values, which are expressed in the data itself. Previous theories and values are also necessary to recognize and mitigate these problems.

What's more, we do not simply expect an ML system to generate reference classes, which is a computationally trivial task of finding correlations between different variables. Rather, what we expect are *clinically relevant* correlations that enable the system to generate *adequate* reference classes that are also fair and equitable. If we wish to accept a reference class as adequate, we should deem it insufficient to just establish a positive correlation between two or more variables *per se*. We should require explanatory justifications in terms of how and why the system defined a particular reference class (Casacuberta et al., 2022). Yet, unsupervised machine learning systems are said to operate as a "black box" (Holm, 2019), which makes it difficult to comprehend how and based on what reasons the algorithm generates an output.[7]

[7] The link between concepts, explainability, and explanatory justifications merits a richer discussion, but alas, due to space limitations I cannot discuss this matter in further detail. I refer the interested reader to Casacuberta et al., 2022, where my associates and I engage with these themes in depth.

The thrust of medical deliberations is about when to attribute a particular evaluative concept (i.e., healthy or diseased) to a biological state. Take the case of osteoporosis. While its diagnosis largely depends on a quantitative assessment of bone mineral density, the clinical significance of osteoporosis lies in the fractures that arise. The causes of these fractures are multifactorial. To assess the risk of fracture there is a myriad of methods, with different input variables and models that generate different risk estimations (Kanis et al., 2017).

Different conceptions of health enable individuals (medical professionals, patients, citizens in general, and so on) and collectives (such as governments, international and local organizations, patients associations, and so on) to offer reasons in favor or against calling a state or condition "healthy" or "diseased." These critical deliberations have profound implications. The most obvious one is their influence on the contents of classificatory standards, such as the International Statistical Classification of Diseases (ICD). As an illustration, consider the fact that, from the second millennium BC onwards, hysteria was considered a diagnosable physical disease affecting women especially. In 1980, hysterical neurosis was deleted from the DSM, the standard classification of mental disorders (Tasca et al., 2012). No medical professional uses the term "hysterical" anymore. Hysteria is more a reflection of Victorian gender dynamics and oppressive attitudes toward women than anything else. Yet the effects of hysteria once having been an *official* female disease linger on and are suffered by women all over the world on a daily basis.

It is because of this that these critical engagements also fuel the emancipatory collective struggles that seek to remove diagnoses from the official classificatory manuals like ICD and DSM. Besides hysteria, another example concerns the diagnoses that once defined widely prevalent aspects of human sexuality, such as homosexuality, as a mental disorder (Drescher, 2010). Both the classification as a disease as well as the resistance against it reflect scientific perspectives and changing societal views. There is no end of theory.

Defining fraught and value-laden notions such as health is an ongoing project where the discussion is not only about what *is* but also about what *should be*. This dialogical engagement is a question of ethics and politics, not one of finding positive correlations between data; it is not a question of mathematics.

That machine learning won't save naturalism about health from its internal conflicts does not mean that all normatively engaged conceptions of health are equally coherent or comprehensive. Nor does it entail that the search for objectivity must be abandoned—this claim naturally deserves further elaboration, but alas, I lack the space to do so. Suffice it to say that objectivity can still be obtained by evaluating which of any number of "competing theories is more fruitful, better at resolving certain dilemmas, or more able to subject its rival to an effective immanent critique" (Blakeley, 2019).

To end, I wish to present some broader reflections and a call to action. While a fully blown injunction against the use of ML systems for epistemic purposes seems unwarranted, we ought to avoid giving these systems the final word in determining value-laden notions such as health, privacy, gender, trustworthiness, criminality, education, and so on. This is not least because this task requires genuine *judgment*—understood as "deliberative thought, ethical commitment and responsible action"—something no current AI system is capable of (Cantwell Smith, 2019: XV, p. 82). Neither should we—for the sake of neutrality, science or efficiency—abdicate the competence to determine these meanings to the creators and deployers of AI systems rather than society at large. To do so would be to deny the public the possibility of participation, yet public reasoning and discussion are the key to the digital future. Safeguarding the agentic ability to interpret and evaluate the world is a way of retaining fundamental epistemic agency. In other words, we must preserve the public's power to make judgements about what these fraught, normative notions mean. But there's more: If epistemic agency is to be safeguarded, what must also be preserved is the human capacity to discuss what "normality" looks like, that is, discussing what inherently contestable and time-bound reference classes should be the basis for making evaluations related to those normative notions.

6 ACKNOWLEDGEMENTS

The author wishes to thank Sara Pedraz and David Casacuberta, who provided comments on a draft version of this paper. The presentation of this work at the Weizenbaum Conference has been partially funded by the BBVA Foundation through the research project for SARS-CoV-2 and COVID-19 Research in Humanities (Detección y eliminación de sesgos en algoritmos de triaje y localización para la COVID-19).

7 REFERENCES

1. Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. *JAMA Dermatology*, *154*(11), 1247-1248.
2. Anderson, C. (2008). *The end of theory: The data deluge makes the scientific method obsolete*. Wired Magazine. Retrieved 15/01/22 from https://www.wired.com/2008/06/pb-theory/
3. Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim Code. Cambridge: Polity.
4. Blakeley, J. (2020). We built reality: How social science infiltrated culture, politics, and power. Oxford: Oxford University Press.
5. Blakeley, J. (2016). Alasdair MacIntyre, Charles Taylor, and the Demise of Naturalism. Notre Dame: University of Notre Dame Press.
6. Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science, 44*(4), 542-573.
7. Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy, 39*, 683–724.
8. Bowker, G. C., and Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge: MIT Press.
9. Butler, J. (1990). *Gender Trouble*. London: Routledge.
10. Casacuberta, D., & Vallverdú, J. (2014). E-science and the data deluge. *Philosophical Psychology, 27*(1), 126-140.
11. Casacuberta, D., Guersenzvaig, A., & Moyano-Fernández, C. (2022). Justificatory explanations in machine learning: for increased transparency through documenting how key concepts drive and underpin design and engineering decisions. *AI & Society*.
12. Cantwell Smith, B. (2019). *The Promise of Artificial Intelligence*. Cambridge: MIT Press.
13. Drescher, J. (2010). Queer diagnoses: parallels and contrasts in the history of homosexuality, gender variance, and the diagnostic and statistical manual. *Arch Sex Behav.* 39(2):427–60.
14. Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118.
15. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. New York: St. Martin's Press.
16. Falk, D. (2019). How Artificial Intelligence Is Changing Science. Quanta Magazine. Retrieved 15/01/22 from https://www.quantamagazine.org/how-artificial-intelligence-is-changing-science-20190311/
17. Fink, C., Blum, A., Buhl, T., et al. (2020). Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas. *Journal of the European Academy of Dermatology and Venereology, 34*(6), 1355-1361.
18. Heaven, W. D. (2020). DeepMind's protein-folding AI has solved a 50-year-old grand challenge of biology. *MIT Technology Review*. Retrieved 10/01/2022 from https://www.technologyreview.com/2020/11/30/1012712/deepmind-protein-folding-ai-solved-biology-science-drugs-disease/
19. Holm, E.A. (2019). In defense of the black box. *Science, 364*(6435), 26-27.

20. Huntington, A., & Gilmour, J. A. (2005). A life shaped by pain: Women and endometriosis. *Journal of Clinical Nursing*, *14*(9), 1124-1132.

21. Kanis, J. A., Harvey, N. C., Johansson, H., et al. (2017). Overview of Fracture Prediction Tools. *Journal of clinical densitometry: the official journal of the International Society for Clinical Densitometry*, *20*(3), 444–450.

22. Khan, S. S., Ning, H., Wilkins, J. T., et al. (2018). Association of Body Mass Index With Lifetime Risk of Cardiovascular Disease and Compression of Morbidity. *JAMA cardiology*, *3*(4), 280–287.

23. Kingma, E. (2014). Naturalism about health and disease: Adding nuance for progress. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, *39*(6), 590-608.

24. Male, V. (2021). Menstrual changes after COVID-19 vaccination *BMJ*, 374:n2211.

25. Ridker, P.M., Cook, N.R., Lee, et al. (2005). A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *New England Journal of Medicine*. *352*(13):1293–1304.

26. Schawinski, K., Turp, D.M., & Zhang, C. (2018). Exploring galaxy evolution with generative models. *Astronomy & Astrophysics*, 616, L4.

27. Tasca, C., Rapetti, M., Carta, M. G., & Fadda, B. (2012). Women and hysteria in the history of mental health. *Clinical practice and epidemiology in mental health*, 8, 110–119.

28. USGAO – United States General Accounting Office. (2001). *Drug Safety: Most Drugs withdrawn in Recent Years had Greater Health Risks for Women*. Washington, DC: US Government Publishing Office.

29. WHO - World Health Organization. (2010). *A conceptual framework for action on the social determinants of health.* World Health Organization. Retrieved 15/01/22 from https://apps.who.int/iris/handle/10665/44489