
Religion for Machines

Humanism and Dataism as Perspectives on the Design of Machine Learning Systems

Mehmet Aydın Baytaş

Koç University
Istanbul, Turkey
mbaytas@ku.edu.tr

ABSTRACT

In his 2016 book *Homo Deus: A Brief History of Tomorrow*, Yuval Noah Harari introduces “humanism” and “dataism” as “religions” which offer norms and values that can scaffold human responses to technological and cultural developments. This paper introduces these two positions in an attempt to situate the ideas within HCI, for the purpose of uncovering potentials to inform perspectives in human-centered machine learning.

KEYWORDS

Humanism, dataism, data religion, philosophy, religion, values, ethics, values, agency, authority.

INTRODUCTION

In recent years, developments in statistical learning algorithms, computing hardware, and related fields have fueled rapid growth in both research on and deployments of “intelligent” systems. Powered by machine learning (ML), “artificial intelligence” (AI) has already outperformed biological intelligence in many tasks that call for decision-making based on abundant and/or complex (i.e. high-dimensional) data; and shown promise in unexpected ways, e.g. artistic creation [15] and generalizability [23]. Thus, AI is replacing human agency in many domains, as humans relinquish decision-making, in matters both mundane and consequential, to AI. This necessitates that designers and developers of ML-powered systems confront not only technical and practical challenges, but also philosophical

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Emerging Perspectives in Human-Centered Machine Learning, May 4, 2019, Glasgow, Scotland, UK

© 2019 Copyright held by the owner/author(s).

ones involving values and morals. Assuming, deliberately or otherwise, a philosophical stance that scaffolds some ethical, aesthetic, epistemological, or even metaphysical judgments is unavoidable when creating a system meant to replace human agency, even within a limited domain [1, 9, 20].

In the popular 2016 book *Homo Deus: A Brief History of Tomorrow*, Yuval Noah Harari confronts this issue [17]. The author primarily meditates on historical and current trajectories of technological developments (mainly in AI and life sciences), and how they influence the ways in which we reflect and act in the world as individuals and societies. One of the central ideas in the book is *humanism* and *dataism* considered as “religions” – philosophical scaffolds comprising norms and values that inform human responses to the aforementioned developments. I wish to introduce these two positions to inform discussions on perspectives in human-centered machine learning (HCML), as they may lead to implications relevant to various stages of design processes or product lifecycles for ML systems. I draw primarily on Harari’s writings [16–18], while assimilating learnings from other literature as well. Both positions are presented in terms of their premises, and the norms and values they beget, as well as their ideals and shortcomings (or, in Harari’s words, their “dreams and nightmares”).

The fact that Harari introduces these positions as a “religions” deserves some comment. This is due to a rather liberal definition for what a religion entails, which I quote from the author’s earlier book *Sapiens: A Brief History of Mankind* [16]:

Religion can [...] be defined as *a system of human norms and values that is founded on a belief in a superhuman order*. This involves two distinct criteria:

- (1) Religions hold that there is a superhuman order, which is not the product of human whims or agreements. Professional football is not a religion, because despite its many laws, rites and often bizarre rituals, everyone knows that human beings invented football themselves, and FIFA may at any moment enlarge the size of the goal or cancel the offside rule.
- (2) Based on this superhuman order, religion establishes norms and values that it considers binding. Many Westerners today believe in ghosts, fairies and reincarnation, but these beliefs are not a source of moral and behavioural standards. As such, they do not constitute a religion.

Humanism and dataism, as described by Harari, fulfill these criteria. However, other authors have preferred to use the term “philosophy” when referring to the same ideas [10]. Here, I use the terms somewhat interchangeably, based on the sources I refer to.

HUMANISM

“Theist religions focus on the worship of gods. Humanist religions worship humanity, or more correctly, *Homo sapiens*. Humanism is a belief that *Homo sapiens* has a unique and

sacred nature, which is fundamentally different from the nature of all other animals and of all other phenomena. Humanists believe that the unique nature of *Homo sapiens* is the most important thing in the world, and it determines the meaning of everything that happens in the universe. The supreme good is the good of *Homo sapiens*. The rest of the world and all other beings exist solely for the benefit of this species.” [16]

In previous HCI work, references to humanism have primarily related to a more general turn towards sensitivity to humane values [14, 21] or the intersections of HCI with “the humanities” as disciplines of study [2, 4, 5, 13]. This is, of course, in addition to the popularity of human-centered design [19] within the field, which already builds on humanistic tendencies. More explicitly congruous with Harari’s articulation is Oulasvirta’s (2004) view on humanism, from which he develops a “research strategy” for HCI by philosophically grounding “research goals” and methodology [22]:

The prevailing strategy [in HCI] to find use potentials could be called *technology-driven*. In short, it takes technology as granted and attempts to find some minimum use case that justifies its existence. This can be contrasted by the *humanistic strategy*. Humanism believes in human rationality, creativity, and morality, and recognizes that human values have their source in experience and culture.

Harari further delineates how the fundamental conception of humanism forks into three branches that he terms the “sects” of humanism, based on a disagreement on what exactly “humanity” comprises. The first and prevalent among these is liberal humanism, which holds that “‘humanity’ is a quality of individual humans” and thereby holds up individual freedoms and human rights over collectivist tendencies. The other two are socialist humanism, which subscribes to a collective, rather than individualistic, notion of humanity and promotes equality among humans; and evolutionary humanism, which views humanity as a species that can mutate in aggregate over time into superior (or inferior) forms and seeks to encourage its improvement through promoting superior traits. These three “sects” are closely linked to various political ideologies that have prevailed at different times and places; and conflicts between the sects, as well as certain unforeseen outcomes borne from implementations of these ideologies, have been the source of much tragedy throughout history.

DATAISM

Credited with coining the term *dataism* to refer to the “rising philosophy of the day” is Brooks (2013), who submits the notions “that everything that can be measured should be measured; that data is a transparent and reliable lens that allows us to filter out emotionalism and ideology; that data will help us do remarkable things” as its “cultural assumptions” [10]. He ponders: “In what situations should we rely on intuitive pattern recognition and in which situations should we ignore intuition and follow the data?” Harari’s conception of a “data religion” shares the same premises, but does away with the

open question. At the core of Harari's dataism is the commitment to "follow the data" wherever and whenever possible, and subsequently attain "all the traditional religious prizes – happiness, peace, prosperity, and even eternal life – but here on earth with the help of data-processing technology, rather than after death with the help of supernatural beings" [12]. The abstract of a 2016 talk by the author explains further [12]:

Data Religion believes that the entire universe is a flow of data, that organisms are algorithms, and that humanity's cosmic vocation is to create an all-encompassing data-processing system – and then merge into it. On the practical level Dataists believe that given enough biometric data and enough computing power, you could create an external algorithm that will understand us humans much better than we understand ourselves. Once this happens, authority will shift from humans to algorithms and humanist practices such as democratic elections and free markets will become as obsolete as rain dances and flint knives.

As far as I can tell, dataism has not been used explicitly in HCI as a motivating philosophy. However, it is also my impression that the "data religion" captures many of the covert, latent, unexamined, and implicit motivations that underlie much of the work in AI and ML, as well as in HCI. At the 2018 CHI conference, a similar issue was covered by Browne and Swift through a somewhat satirical performance where an Ouija board-like apparatus was used "for the ritualistic performance of neural network algorithms" in order to establish a "critical distance with which to examine the philosophical and cultural assumptions embedded in our conception of AI" [11].

Scholars who are critical of the increasing prevalence of dataist tendencies and machine agency in our lives often home in on AI's inability to accommodate diversity and reflect on bias as a destructive element. Spiel et al. (2018), for example, note that "the ideal person who uses a fitness tracker is white, able-bodied, mentally stable, already fit and slim" [25]. In my own work using various motion capture and gesture sensing technologies [6–8, 27], I have witnessed firsthand how high-end equipment, commercially successful ML algorithms, impressive interaction designs, and interaction designers' own mental models falter for users with physical or sensory atypicalities.

DISCUSSION

While the presentation (both Harari's and mine) may suggest that humanism and dataism are incompatible, contrasting belief systems, my opinion is that this is not so. They share significant common ground (as many religions do), in that both are ultimately meant to serve human needs and desires; even though the two philosophies (and their various flavors) may disagree in terms of which particular needs and desires are to be prioritized, as well as on what scale.

I find these ideas of interest to the HCML community because, in my view, efforts to advance HCML, including those that suppose humanist intentions, often in effect serve the dataist agenda. Based on the definitions above, democratized, unbiased, explainable, intelligible, user-friendly ML is not a humanist goal, but rather, a dataist one, in that it ultimately seeks to increase the agency and authority of machines over our lives. Effectively, HCML and related efforts may increase the *trust* we have in the decisions made by AI on our behalf, while in turn giving AI the means to explain itself, i.e. convince us, when we disagree with it.

It seems to me that the tendency to assign more and more agency and authority to machines is also reflected in the approaches and language we have recently been adopting; as we progress towards a paradigm where we would like to assume the role of “teachers” who nurture intelligence, rather than programmers or engineers who build to machines [24, 26].

I recognize the possibility that the above may be interpreted as an adversarial critique of ML, but I would like to make it clear that it is quite the opposite. My intention has been to provoke honest philosophical reflection, which I hope will promote clarity and precision in thought and language, thereby serving to advance the field. In future work I would like to explore how design, development and evaluation of ML artifacts may be informed in practice by such philosophical considerations—e.g. through critical design [3], or case studies applying humanism and dataism as lenses for critiquing particular ML artifacts—and explore other philosophies as alternatives to the two introduced here.

REFERENCES

- [1] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. The moral machine experiment. *Nature* 563, 7729 (2018), 59.
- [2] Amid Ayobi, Paul Marshall, and Anna L. Cox. 2016. Reflections on 5 Years of Personal Informatics: Rising Concerns and Emerging Directions. In *CHI EA '16*. ACM, New York, NY, USA, 2774–2781. <https://doi.org/10.1145/2851581.2892406>
- [3] Jeffrey Bardzell and Shaowen Bardzell. 2013. What is “Critical” About Critical Design?. In *Proc. CHI '13*. ACM, New York, NY, USA, 3297–3306. <https://doi.org/10.1145/2470654.2466451>
- [4] Jeffrey Bardzell and Shaowen Bardzell. 2016. Humanistic HCI. *interactions* 23, 2 (Feb. 2016), 20–29. <https://doi.org/10.1145/2888576>
- [5] Jeffrey Bardzell, Shaowen Bardzell, Carl DiSalvo, William Gaver, and Phoebe Sengers. 2012. The Humanities and/in HCI. In *CHI EA '12*. ACM, New York, NY, USA, 1135–1138. <https://doi.org/10.1145/2212776.2212405>
- [6] Mehmet Aydın Baytaş, Asım Evren Yantaç, and Morten Fjeld. 2017. LabDesignAR: Configuring Multi-camera Motion Capture Systems in Augmented Reality. In *Proc. VRST '17*. ACM, New York, NY, USA, Article 86, 86:1–86:2 pages. <https://doi.org/10.1145/3139131.3141778>
- [7] Mehmet Aydın Baytaş, Yücel Yemez, and Oğuzhan Özcan. 2014. Hotspotizer: End-user Authoring of Mid-air Gestural Interactions. In *Proc. NordiCHI '14*. ACM, New York, NY, USA, 677–686. <https://doi.org/10.1145/2639189.2639255>
- [8] Mehmet Aydın Baytaş, Damla Çay, Asım Evren Yantaç, and Morten Fjeld. 2017. Motion Capture in Gesture and Sign Language Research. Poster presented at the DComm conference Language as a Form of Action.
- [9] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. 2016. The social dilemma of autonomous vehicles. *Science* 352, 6293 (2016), 1573–1576. <https://doi.org/10.1126/science.aaf2654>

- [10] David Brooks. 2013. The Philosophy of Data. *The New York Times* (2013). <https://nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>
- [11] Kieran Browne and Ben Swift. 2018. The Other Side: Algorithm As Ritual in Artificial Intelligence. In *CHI EA '18*. ACM, New York, NY, USA, Article alt11, alt11:1–alt11:9 pages. <https://doi.org/10.1145/3170427.3188404>
- [12] Centre for Research in the Arts, Social Sciences and Humanities (CRASSH) at the University of Cambridge. 2016. The Rise of Data Religion. <http://www.crassh.cam.ac.uk/events/26948> Accessed: 2018-06-18.
- [13] Wen Chen, David J. Crandall, and Norman Makoto Su. 2017. Understanding the Aesthetic Evolution of Websites: Towards a Notion of Design Periods. In *Proc. CHI '17*. ACM, New York, NY, USA, 5976–5987. <https://doi.org/10.1145/3025453.3025607>
- [14] Batya Friedman and David Hendry. 2012. The Envisioning Cards: A Toolkit for Catalyzing Humanistic and Technical Imaginations. In *Proc. CHI '12*. ACM, New York, NY, USA, 1145–1148. <https://doi.org/10.1145/2207676.2208562>
- [15] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [16] Yuval Noah Harari. 2014. *Sapiens: A Brief History of Humankind*. Harvill Secker.
- [17] Yuval Noah Harari. 2016. *Homo Deus: A Brief History of Tomorrow*. Harvill Secker.
- [18] Yuval Noah Harari. 2018. *21 Lessons for the 21st Century*. Jonathan Cape.
- [19] ISO 9241-210:2010 2010. *Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. Standard. International Organization for Standardization, Geneva, CH.
- [20] Zachary C. Lipton. 2016. The Mythos of Model Interpretability. In *Proc. 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*.
- [21] Michael Massimi and Andrea Charise. 2009. Dying, Death, and Mortality: Towards Thanatosensitivity in HCI. In *CHI EA '09*. ACM, New York, NY, USA, 2459–2468. <https://doi.org/10.1145/1520340.1520349>
- [22] Antti Oulasvirta. 2004. Finding Meaningful Uses for Context-aware Technologies: The Humanistic Research Strategy. In *Proc. CHI '04*. ACM, New York, NY, USA, 247–254. <https://doi.org/10.1145/985692.985724>
- [23] S. J. Pan and Q. Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (Oct 2010), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- [24] Patrice Y. Simard, Saleema Amershi, David Maxwell Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang, and John Wernsing. 2017. Machine Teaching: A New Paradigm for Building Machine Learning Systems. *CoRR* abs/1707.06742 (2017). arXiv:1707.06742 <http://arxiv.org/abs/1707.06742>
- [25] Katta Spiel, Fares Kayali, Louise Horvath, Michael Penkler, Sabine Harrer, Miguel Sicart, and Jessica Hammer. 2018. Fitter, Happier, More Productive?: The Normative Ontology of Fitness Trackers. In *CHI EA '18*. ACM, New York, NY, USA, Article alt08, 10 pages. <https://doi.org/10.1145/3170427.3188401>
- [26] Xiaojin Zhu. 2015. Machine Teaching: An Inverse Problem to Machine Learning and an Approach Toward Optimal Education. In *AAAI*. 4083–4087.
- [27] Advıye Ayça Ünlüer, Mehmet Aydın Baytaş, Oğuz Turan Buruk, Zeynep Cemalcılar, Yücel Yemez, and Oğuzhan Özcan. 2018. The Effectiveness of Mime-Based Creative Drama Education for Exploring Gesture-Based User Interfaces. *International Journal of Art & Design Education* 37, 3 (2018), 353–366. <https://doi.org/10.1111/jade.12136>