HHAI2022: Augmenting Human Intellect
S. Schlobach et al. (Eds.)

doi:10.3233/FAIA220204

Open, Multiple, Adjunct. Decision Support at the Time of Relational AI

Federico CABITZA [a,1] and Chiara NATALI [b]
[a] *University of Milano-Bicocca, IRCCS Istituto Ortopedico Galeazzi*
[b] *University of Milan, University Vita-Salute San Raffaele*

Abstract. In this paper, we consider some key characteristics that AI should exhibit to enable hybrid agencies that include subject-matter experts and their AI-enabled decision aids. We will hint at the design requirements of guaranteeing that AI tools are: *open*, *multiple*, *continuous*, *cautious*, *vague*, *analogical* and, most importantly, *adjunct* with respect to decision-making practices. We will argue that especially *adjunction* is an important condition to design for. *Adjunction* entails the design and evaluation of *human-AI interaction protocols* aimed at improving AI usability, while also guaranteeing user satisfaction and human and social sustainability. It does so by boosting people's cognitive motivation for interacting analytically with the outputs, reducing overreliance on AI and improving performance.

Keywords. Relational Artificial Intelligence, Decision support, Machine Learning, Interaction protocols, Usability

Nearly 25 years ago, Giorgio De Michelis [1] wrote a little book in Italian, titled "Aperto, molteplice, continuo: gli artefatti alla fine del Novecento" (Open, multiple, continuous: artifacts at the end of the Twentieth century), where he adopted a phenomenological stance in regard to the design of artifacts, in particular digital artifacts, and their use. Today, we believe that these aesthetic categories should be taken again in consideration in regard to the design of Human-AI Interaction models. We will make a point that, in order to both exhibit artificial intelligence (i.e. autonomy in producing effective behaviors in front of partly unexpected situations) and promote augmented intelligence (in decision makers facing the very same unexpected situations), ML-based decision support systems must be: open, multiple, continuous, cautious, vague, analogical and adjunct.

An **open** system is configured as an open loop, capable of updating its reference data and, consequently, its correlative models, so as to cope with ever-changing environments and mitigate the risk of errors due to concept drift. [2]

Multiple systems provide users with several complementary indications or even possibly identical and diverging pieces of advice by different competing models, instead of single pieces of advice and clear-cut categories.

A **continuous** decision support system allows for the exploration of the causal factors, possible explanations and effects on their output, deriving from a full range of small (counterfactual) differences in the digital representation of those instances and cases.

A **cautious** system expresses a judgment only when its confidence is sufficiently high, or above a threshold that depends on task criticality, the risk of failure, or users' expertise and preferences, abstaining in all other cases. [3].

[1] Corresponding Author: Federico Cabitza; E-mail: federico.cabitza@unimib.it

A **vague** system, as in the case of multiplicity, does not limit itself in providing one best option, but rather promotes reflection in expert users by proposing multiple pertinent classes for the case at hand, guaranteeing high confidence in that the list or interval of values given contains the right answer, like in *conformal prediction* settings.
Analogical systems try to foster analogical thinking in experts by presenting to them the most (or the least) similar cases to the case at hand, according to their correlative models and some similarity metric [4], and by inviting the users to reflect on what answer such similarity (or dissimilarity) could suggest [5].
Reflective systems, instead, promote reflection by matching their advice with questions that challenge users about their own confidence, and promote counter-factual reasoning or the pursuing of alternative options.
Finally, a theory of **adjunction** invites to focus on the process-oriented and relational aspects of the joint action of humans and machines working together. This entails the evaluation of *human-plus-machine* systems as a whole, recognizing both the cooperative nature of decision-making [6] and the distributed nature of cognition. In adjunction, human-AI interaction protocols are conceived to purposefully move the AI support to the background or to a role of "second opinion" giver [7] after that an official (and registered) decision has been already made by single human decision makers or by small teams of decision makers, who take ultimate responsibility for their decisions. [8]

Human-centered AI, when aimed at smoothing out every instance of friction from our course of action, harbors the risk of engendering a gradual yet unavoidable deskilling and degradation of the human attributes we value most in decision making: autonomy, intuition, and accountability [9,10]. If AI does have a detrimental influence on the attitude and learning processes of users, changing our minds, as users, is simpler than changing the AI itself. Raising awareness of the risks of automation is more straightforward than creating an ever more explicable, ethical or responsible AI, whatever this might mean. While efficient AI aims to accelerate workflows and reduce relational friction, Adjunct AI can be given an opposing duty: slowing decision-makers down, making task fulfillment difficult or cumbersome, or even hindering people from performing a certain action [10]. The main goal behind these *programmed inefficiencies* [11] is fostering constructive distrust [12] by arousing critical thinking, shattering the false impression of objectivity provided by algorithms, seeding questions about the outcome, nudging the user to look for more conclusive proof and fostering a sense of personal responsibility. Such cognitive-forcing functions would boost people's cognitive motivation for interacting analytically with the outputs, reducing overreliance on AI and improving performance. [13]

In this contribution we presented some essential design-oriented concepts, and argued about their deeper significance for the design of effective, satisfactory and sustainable human-AI interaction. Instead of evaluating technology in isolation, we should consider the interaction protocol as a whole, assessing the entire socio-technical system that adopts and deploys the AI, in terms of efficiency, efficacy, the satisfaction of both users and those affected, human sustainability and cost-effectiveness. The attributes *open*, *multiple*, *continuous*, *cautious*, *vague*, *analogical*, *reflective* and *adjunct* provide us a sufficiently narrow and practical list of system capabilities in order to evaluate, design and even legally define human - AI interaction protocols through which a *humachine system* can exhibit some form of hybrid intelligence that is functional to some aim and sustainable in the long run.

References

[1] De Michelis G. Aperto, molteplice, continuo: gli artefatti alla fine del Novecento. Zanichelli, Milano; 1998.
[2] Zenisek J, Holzinger F, Affenzeller M. Machine learning based concept drift detection for predictive maintenance. Computers & Industrial Engineering. 2019;137:106031.
[3] Campagner A, Cabitza F, Ciucci D. Three–way classification: Ambiguity and abstention in machine learning. In: International Joint Conference on Rough Sets. Springer; 2019. p. 280-94.
[4] Keane M. Analogical mechanisms. Artificial Intelligence Review. 1988;2(4):229-51.
[5] Baselli G, Codari M, Sardanelli F. Opening the black box of machine learning in radiology: can the proximity of annotated cases be a way? European Radiology Experimental. 2020;4(1):1-7.
[6] Sloman S, Fernbach P. The Knowledge Illusion: Why We Never Think Alone. Penguin; 2017.
[7] Cabitza F. Biases affecting human decision making in AI-supported second opinion settings. In: International Conference on Modeling Decisions for Artificial Intelligence. Springer; 2019. p. 283-94.
[8] Skitka LJ, Mosier K, Burdick MD. Accountability and automation bias. International Journal of Human-Computer Studies. 2000;52(4):701-17.
[9] Frischmann B, Selinger E. Re-engineering humanity. Cambridge University Press; 2018.
[10] Farindon P. Cabitza, Federico. In: Pelillo M, Scantamburlo T, editors. Cobra AI: Exploring Some Unintended Consequences. MIT Press; 2021. p. 87-104.
[11] Cabitza F, Campagner A, Ciucci D, Seveso A. Programmed inefficiencies in DSS-supported human decision making. In: International Conference on Modeling Decisions for Artificial Intelligence. Springer; 2019. p. 201-12.
[12] Hildebrandt M. Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. Theoretical Inquiries in Law. 2019;20(1):83-121.
[13] Buçinca Z, Malaya MB, Gajos KZ. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction. 2021;5(CSCW1):1-21.