二十一世紀的顱相學

Phrenology for the Twenty-First Century

崔利懷

Levi Checketts

摘要

儘管 Lyreskog 等人堅持認為人工智能腦機介面 (ABTs) 的使用需要新的“概念和框架”，但他們所依據的人的定義和人類繁榮模式似乎本身就值得商榷。雖然自主權或隱私權的問題很重要，但這些問題不應該從屬於這些技術本身所蘊含的價值觀以及技術在發展過程中所宣導的意識形態。ABTs 技術超出嚴格的治療用途之上的應用得到了超人類主義者的支持，他們所認可的人類學偏向於神經典型主義，傾向於用工具理性的方法來進行道德

崔利懷，香港浸會大學宗教及哲學系助理教授；應用倫理學研究中心研究員，中國香港。
Levi Checketts, Research Fellow, Centre for Applied Ethics; Assistant Professor, Department of Religion and Philosophy, Hong Kong Baptist University, Hong Kong, China.

《中外醫學哲學》XXI:2 (2023 年) ：頁 35-41。
International Journal of Chinese & Comparative Philosophy of Medicine 21:2 (2023), pp. 35-41.

評價。因此，我認為，ABTs 的新“概念和框架”需要與批判性的聲音進行對話，以免 ABTs 成為本世紀版本的顱相學。

At the conclusion of their essay, Lyreskog et al. suggest that bioethicists should anticipate that “the concepts and frameworks of [their] discipline, which [they] have built and utilised over the last 60-odd years, are likely to be up for refurbishment” due to the advances of artificially intelligent brain control interfaces (ABTs) (28). I agree with this assessment, though I anticipate that I entirely disagree with how they would “refurbish” the field of bioethics. They note well that the development of ABTs demand moral reflection given their potential and their danger. And while they denote many (but certainly not all) dangers ABTs carry with them, they do little to contend what actual good these technologies portend. The dangers of inserting commercially-developed, black-box machine-learning enhanced computer devices into people’s brains *demands* strong justification in line with the principle of double effect (or is this one of the concepts to be refurbished?). They articulate, in my count, exactly one case of a positive therapeutic application of brain control interfaces in the well-known case of drug-resistant epilepsy treatment (19). They do give some suggestions about *potential* avenues where ABTs *may* have notable benefits (14), but these are not given any sort of due moral reflection to contextualize whether and to what degree they will comply with existing moral principles, such as nonmaleficence. Rather, they are just suggested as hopeful “benefits” of ABTs.

Ultimately, the arguments *in favor* of ABTs fall on whether or not one accepts the moral vision, that is, the philosophical anthropological assumptions, of Lyreskog et al. Consider their discussion of autonomy: the authors deliberately avoid defining autonomy (16), but it is clear their understanding entails a sense of individual responsibility and freedom from external pressure. On the other hand, Anita Ho (2023) recently shows how much of the “medical-AI”-autonomy discourse suffers from a narrow anthropological view of humans-as-consumers. The promise that AI will enhance autonomy of patients is entirely contingent upon one accepting autonomy as a liberal individualistic notion of non-interference, not a question of making decisions in light of one’s greater social and relational context. Because human beings, even moral philosophers at University of Oxford, do not emerge whole cloth from the ether but rather from social and cultural contexts, autonomy must be understood as part of a broad image of the human person in relation to their setting. This version of autonomy is outside of the individualistic framework Lyreskog et al. adhere to in their moral discussion.

On the other hand, AI is built without any assumption of the broader human world. Hubert Dreyfus (1994) initially argued over fifty years ago that the problem with making "intelligent" machines was that computers lack a world. The broader set of assumptions that contextualize our entire understanding of ourselves in relation to others constitutes a totality which AI inherently lacks. Rather, at present AIs run on limited *models* of the world to perform a limited set of tasks. What this means for ABTs is significant. While Lyreskog et al. do well to point out how an interfering device trained to accomplish one task according to an algorithmic model independent of broader human contexts can "lead to feeling alienated or estranged from one's mental content" (17), they fail to ask the bigger question demanded of these technologies: which models are the ones being prescribed for human mind in ABTs, and are these models really something we should *want*?

The former question goes ultimately unanswered in the text, but not without some implicit articulations. In reading the article for the first time, I found myself pondering frequently whether the authors assume that mind and brain are ultimately the same, and, if so, how they assume mind works as a biological process. The question is only briefly raised on page 11, and in this case, it is tied only to the question of identity. While identity is itself a pressing concern, it is not the entire extent of the question of mind. But more concerning is the authors' blithe dismissal of the concerns raised by this perennially vexing philosophical question. Citing positively the work of David Chalmers and Andy Clark, the authors propose an "increasingly popular take on [the threat of ABTs to patients' identity] relies on not limiting the boundaries of the self and identity to one's immediate physical and/or psychological continuity, but allows the extension of oneself – of one's mind, to be precise – to external objects" (21). Ignoring briefly the fact that the authors never clearly define what mind is, so the "precision" of their clarification is merely an obfuscation, the bigger worry is that they dismiss the very real concern about how untested invasive automated technologies may cause irrevocable harm to people's minds and sense of self.

It is at this point that underlying philosophical assumptions of at least some of the authors become apparent. One of the authors, Savulescu, is a major proponent of human enhancement, especially what he calls "moral enhancement" (Savulescu and Persson 2012). In this regard, Savulescu adheres to the transhumanist philosophy, a small range of philosophical positions which all tie to the proposition that human beings should use science and technology to direct our own evolution. While some transhumanists, such as James Hughes or Steve Fuller, have strong altruistic and pro-social views of what

transhumanism should look like, its most pronounced spokespersons, such as Max More, Ray Kurzweil or even Nick Bostrom, favors an individualistic liberal approach (or even libertarian, as is the case for Extropianism founder More).

There is not space in this short response to tease out the various philosophical arguments or varieties of transhumanism. At the risk, then, of overgeneralizing, I note that transhumanists *tend* to think that modern industrial technologies are generally good and should be pursued. A specific philosophy embraced by many transhumanists and AI proponents called patternism further asserts that the brain is merely a pattern-finding machine, entirely like a computer, with the mind merely the "software" running on it. In other words, the way a digital computer function is entirely, without exaggeration, the way our minds work, with only a challenge in figuring out how the different "hardware" (or "wetware" for the human brain) components can interface.

The assumption of Lyreskog et al. that ABTs will actually interface with our minds, and that they may "enhance" us (23–24) suggests a very typical transhumanist orientation to the opening question, i.e., their philosophical anthropology. The question of enhancement is predicated upon a series of assumptions not always made explicit. The first is that augmenting certain human capacities would be both desirable and an actual improvement for human beings' overall wellness. The second is that a computer interface could somehow enhance human capacities. This is itself predicated upon a third assumption, namely that the structure of both the physical brain and the more ephemeral mind can fully interface with a computer. To some degree, of course, we see this legitimated, as in the case of epilepsy patients experiencing an "enhancement" of their life experience. However, this is a strictly limited application; the device used to control seizures is not intended to accelerate brain pathways or expand memory or some other science-fiction fantasy. Whether "enhancement" is intended to mean some sort of aid to the internal structure of the brain (such as regulating neurotransmitters) or is meant to directly interface with our "thoughts" is unclear in the article itself, but clearly a patternist philosophy will find this congenial to its own position. The addition of artificial intelligence to BCIs suggests that the artificially intelligent is somehow meant to connect with the naturally intelligent and not merely serve as a feedback regulator

As I have argued elsewhere (2024), the underlying worldview that informs much of AI research, and especially its status as a philosopher's stone for *every* problem we face, reduces to instrumental rationality. Algorithms are mathematical models, meaning that they have to reduce all content to numerical data. From a mathematical

standpoint, the "best" option amounts to the most efficient. Thus, AIs are designed to find the most quantitatively efficient (i.e. mathematically justifiable) solution. This is not inherently wrong, but it is a limited way to approach the world. All phenomena which we might consider non-reductive *have* to be reduced to numerical data for algorithmic processing. Therefore beauty, wonder, transcendence, bliss, happiness, and so on, must be translated to pure numbers. Thus also, we must evaluate things not based on qualitative differences but rather on quantitative differences, which is to say there will be only numerically better and worse solutions. Difference cannot be appreciated as itself an important but intractable reality; it must be obliterated because in the context of instrumental rationality, different is indicated numerically as deviation, which is to say deficiency.

Take, for instance, the authors' proposal to use ABTs to "correct" attention-deficit hyperactive disorder (ADHD) in children (25–28). The assumption underlying this proposal is that ADHD is a deficiency, one that must be corrected in children. This reflects the ableism transhumanists have sometimes been accused of. To assert that neurotypicalism is the ideal outcome, that children need to have ADHD, a *different* way of experiencing cognition within the world, "corrected" to a neurotypical standard, is to enforce an ideology that quantifies cognitive function against a presumed baseline of "normal" functioning and derivation outside as deficiencies. The ideology assumed here is that the neurotypicalism of bourgeois intellectuals, the same ideology underlying much of the educational system in Western countries generally where ADHD is most perceived as a "problem," is the base line for health. And here the long-standing question of enhancement versus therapy stands out in full relief: an ADHD child does not experience natural pain or suffering, or natural deficiency by virtue of their neurodivergence—they merely experience a different way of encountering the world. Just as previous decades have seen medical activism from women and people of color to challenge the hegemonic health assumptions enshrined in a profession long-dominated by white men following white male models of health, neurodivergent activism should make us reluctant to listen to anyone who proposes to use dangerous experimental technologies to "cure" children who have no health deficiencies.

Ultimately, of course, it may be in the interest of a child with ADHD, or at least their parents trying to raise them in a world that prizes neurotypicalism above neurodivergence, to use ABTs to give their children an "advantage" in society, just as it would be in the interest of any parent who can to genetically engineer their children to excel at whatever traits late capitalist society deems valuable. But we ethicists *ought* to be quick to point out how bleakly dystopian this

model is and how dangerously close this treads to outright eugenics. We should not forget that last century's "race science" was a philosophical distortion of the advances of modern science that framed scientific descriptions of genetic differences as moral evaluations and enshrined the accepted neurotypical measure of IQ as somehow a moral evaluation instead of a poor measurement of excelling at Western bourgeois epistemological standards. At its peak, this morally repugnant abuse of science focused on the human brain through phrenology, arguing that some people were less human due to brain shape and size. And lest we think this nightmare is ancient history, we ought to note that genetic testing has led to a near eradication of all people with Down Syndrome in Iceland as neurotypicalism has led to selective abortions and a devaluation of the mentally handicapped seen as less than human.

By way of conclusion, then, the question of what new "concepts and frameworks" bioethicists ought to use as AI prognosticators try to force their unimaginative vision into our bodies should be one informed *more* by critical voices than those peddling dangerous invasive devices. Medical practitioners and bioethicists need to become *more* critical than the general public about the fantastic promises made by computer scientists who insist that we should all want to uncritically stick computer hardware in our brains. We must ask about the motivations and interests of people like Elon Musk, hawking his as-yet disastrous Neuralink as an augmentation for the future. We must pay attention to the voices of neurodivergent activists and patients' rights groups. We must follow the way monied interests dominate these conversations and shape public opinion. We must emphasize that while our understanding of the brain has developed tremendously in recent decades, there is still too much that we do not know, especially about the interplay of mind and brain. Undoubtedly, we *will* find important medical uses for ABTs, perhaps in repairing degenerative or damaged tissue. But in this, we must emphasize the dangers that invasive implants pose for tissue as vital as the brain and the need for caution, humility and openness for technologists and biomedical researchers in deploying therapeutic ABTs.

參考文獻 References

萊瑞斯科、佐赫尼、辛格、薩烏萊斯庫：〈與機器一起思考：腦機介面技術〉，《中外醫學哲學》，2023 年，第 XXI 卷，第 2 期，頁 11–34。David M. Lyreskog, Hazem Zohny, Ilina Singh, Julian Savulescu. “The Ethics of Thinking with Machines: Brain-Computer Interfaces in the Era of Artificial Intelligence,” *International Journal of Chinese & Comparative Philosophy of Medicine* 21, no. 2 (2023): 11–34.

Checketts, L. 2024. *Poor Machines: Artificial Intelligence and the Experience of Poverty*. Minneapolis: Fortress Press.

Dreyfus, H. 1992. *What Computers Still Can’t Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.

Ho, A. 2023. *Live Like Nobody Is Watching: Relational Autonomy in the Age of Artificial Intelligence Health Monitoring*. Oxford: Oxford University Press.

Savulescu, J., and Persson, I. 2012. “Moral Enhancement, Freedom, and the Good Machine.” *The Monist* 95, no. 3: 299–421.