Opinion

AI & Neuroscience

Brain-to-AI Adaptive Feedback Systems The Next Frontier of Human–Machine Symbiosis

Rachel Eglinton *
University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
*: All correspondence should be sent to: Dr. Rachel Eglinton
Author's Contact: Dr. Rachel Eglinton, PhD, E-mail: rachel_eglinton1@gmail.com
DOI: https://doi.org/10.15354/si.25.op349
Funding: No funding source declared.
COI: The author declares no competing interest.
AI Declaration: The author affirms that artificial intelligence did not contribute to the process of preparing the work.

Brain-to-AI adaptive feedback systems refer to a class of technologies in which neurophysiological signals from human brains are used in real time to adapt the behavior of artificial intelligence systems, creating closed‐loop feedback that can adjust according to the mental, emotional, or cognitive state of the user. These systems sit at the intersection of brain-computer interfaces (BCIs), neurofeedback, affective computing, adaptive learning, and AI, and promise to transform domains ranging from education and rehabilitation to human–machine collaboration and mental health. But with great promise come profound technical, ethical, and societal challenges: issues of signal fidelity and latency; interpretability and trust; individual variability; data privacy and autonomy; potential for bias and misuse. In this opinion piece I explore the potential benefits of brain-to-AI adaptive feedback systems, the key obstacles they face, and the governance, design, and value judgments that must guide their development if they are to enhance human well-being rather than undermine it.

Keywords: Brain–Computer Interface; Adaptive Artificial Intelligence; Neurofeedback; Human–AI Interaction; Cognitive Augmentation

Science Insights, October 30, 2025; Vol. 47, No. 4, pp.1987-1990.

THE IDEA that machines should adjust their behavior to human mental states is not new, but what is new is how far our technological capacity has advanced. Modern brain-to-AI adaptive feedback systems can monitor neural or neurophysiological activity—via EEG, fNIRS, or related sensors, sometimes combined with peripheral biometrics—and feed that into AI algorithms that adapt in real time: changing interface difficulty, pacing, modality, prompts, stimulus, or even form of feedback (Koelewijn et al., 2021; Valeriani et al., 2022). A learner whose engagement drops might be given easier tasks; someone stressed might be redirected with breathing cues; someone suffering from attention deficit might get visual or auditory modifications to keep them on track. The feedback loop thus becomes continuous and dynamic, not just reactive. Early

empirical results, for example in neuroadaptive tutoring systems, show increased subjective engagement and better emotional regulation (Angulo et al., 2024; Baradari et al., 2025).

The potential applications are broad. In education, such systems may allow adaptive tutoring that is responsive to fatigue, distraction, emotional overload; in therapy or rehabilitation, they can tailor interventions to enable recovery or compensate for impairment with greater sensitivity; in mental health they may help with anxiety, depression, attention disorders; in human-computer interaction they may make devices less frustrating, more accessible. For people with severe motor or communication disabilities, brain-AI systems already enable control of prosthetics, cursors, or robotic limbs via decoding of neural signals (Belwafi & Ghaffari, 2024).

Yet despite promise, these systems confront serious challenges. The technical ones are formidable: real-time neuro-signal acquisition is noisy, subject to artifacts, and often requires trade-offs in comfort, invasiveness, and cost. Machine learning models that use these signals must generalize across individual differences in brain anatomy, physiology, and cognitive style; they must do so with minimal latency so that feedback is meaningful and not disruptive (Kostas & Rudzicz, 2020). There is also the "ground truth" problem—how do we reliably know what mental state we are measuring, whether attention, engagement, or emotional valence, and how these translate into appropriate adaptive actions? Misclassifications or mis‐inferences risk causing frustration, dissuading users, or worse, reinforcing negative states.

Beyond technical issues, there are ethical, psychological, and social dimensions. One major risk is loss of autonomy. If an AI system begins to anticipate and steer one's cognitive states—"nudging" intentionally or not—users may come to rely on it, possibly de-emphasizing their own self-regulation (Laitinen & Sahlgren, 2021). Who decides what the "optimal" mental state is? What if what is optimal for one purpose (e.g. productivity) conflicts with what is optimal for another (well-being, creativity, rest)?

Data privacy and security loom large. Brain data are among the most sensitive possible: unique, hard to anonymize, possibly revealing more about traits and disorders than the user intends. Systems that store or transmit such data must guard against misuse (commercial, political, discriminatory) (Jwa & Poldrack, 2022). There is also the risk of bias—if training data for neuroadaptive systems is skewed demographically, cognitively, culturally, then the adaptive responses may systematically favor some users over others, exacerbating existing inequalities.

Another issue involves trust, transparency, interpretability. Users must have some sense of how the system works—why it adapts the way it does. If AI behavior seems arbitrary, or pairings of brain state → action are mysterious, trust will suffer. Worse, users may mistake AI suggestions for truths about themselves, internalizing wrong assumptions (Weld & Bansal, 2019). Studies suggest that outcome feedback (showing users outcomes of system decisions) can increase trust more than mere interpretability (Afroogh et al., 2024).

Socially, there is the question of what effects widespread use will have on human cognition. If attentional lapses are immediately corrected by AI, do people lose capacity to endure boredom or distraction, skills essential for many tasks? If emotional regulation is offloaded, do we risk atrophy of internal coping mechanisms? And more broadly: who benefits from these systems? If engineered primarily for commercial or institutional ends (more efficient workers, better test takers), might they reinforce forms of control rather than liberation?

Given these tension points, what design principles and governance frameworks should guide the responsible development of brain-to-AI adaptive feedback systems? First, it is essential to respect human dignity and agency: systems should be opt-in, allow user override, provide users with meaningful control over how, when, and to what extent adaptation occurs. Second, transparency: users should know what signals are being collected, what inferences are being drawn, and how adaptation works. Third, privacy and security: brain signals should be treated as highly sensitive data, with high standards of encryption, minimal retention, and robust consent regimes. Fourth, fairness: ensure representative training / calibration datasets; build systems that adapt to many brain types and cognitive styles; monitor for bias continuously. Fifth, usability and inclusivity: adaptivity should reduce (not increase) user frustration; systems should be accessible in cost, hardware, comfort; ensure adaptation does not mean increasing burden of calibration or maintenance. Sixth, empirical evaluation not just for engagement or self‐report, but for long-term cognitive, emotional, social outcomes.

Regulatory and policy frameworks also deserve attention. It is unlikely that market forces alone will ensure the right trade-offs. Regulators may need to treat brain data under similar or stronger protections as medical or biometric data. Standards bodies may need to define norms for accuracy, reliability, safety. Ethical review should accompany deployments in sensitive areas (mental health, education) as in clinical interventions. Public discourse should involve stakeholders: end users, neuroscientists, ethicists, technologists, policymakers.

In many ways brain-to-AI adaptive feedback systems represent a microcosm of deeper questions about how humans want to relate with intelligent machines: do we want assistive partners? Controllers? Guardians of our inner life? Without reflection, there is risk that the "assistive partner"—if designed poorly—becomes a subtle controller. But if designed thoughtfully, such systems could help people extend their capacities: learn more deeply, recover more fully, live more resiliently in an increasingly complex world.

The future is neither a predetermined utopia nor an inevitable dystopia. As brain-to-AI adaptive feedback systems mature, our collective choices about values, priorities, and governance will matter enormously. We must insist that these systems serve human flourishing in all its richness, not just efficiency, optimization, or profit. ■

Received: May 14, 2025 | Revised: July 26 2025 | Accepted: September 04, 2025

References

Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. Humanities and Social Sciences Communications, 11(1), 4044. DOI: https://doi.org/10.1057/s41599-024-04044-8

Angulo, J. A., Dapena, E., & Aguilar, J. (2024). Emotions as implicit feedback for adapting difficulty in tutoring systems based on reinforcement learning. Education and Information Technologies, 29(16), 21015. DOI: https://doi.org/10.1007/s10639-024-12699-8

Baradari, D., Kosmyna, N., Petrov, O. V., Kaplun, R., & Maes, P. (2025). NeuroChat: A neuroadaptive AI chatbot for customizing learning experiences. ACM Transactions on Interactive Intelligent Systems, 1, 1–29. DOI: https://doi.org/10.1145/3719160.3736623

Belwafi, K., & Ghaffari, F. (2024). Thought-controlled computer applications: A brain–computer interface system for severe disability support. Sensors, 24(20), 6759. DOI: https://doi.org/10.3390/s24206759

Jwa, A. S., & Poldrack, R. A. (2022). Addressing privacy risk in neuroscience data: From data protection to harm prevention. Journal of Law and the Biosciences, 9(2), lsac025. DOI: https://doi.org/10.1093/jlb/lsac025

Koelewijn, A. D., Audu, M. L., del-Ama, A. J., Colucci, A., Font-Llagunes, J. M., Gogeascoechea, A., Hnat, S. K., Makowski, N., Moreno, J. C., Nandor, M. J., Quinn, R. D., Reichenbach, M., Reyes, R.-D., Sartori, M., Soekadar, S. R., Triolo, R. J., Vermehren, M., Wenger, C., Yavuz, U. Ş., … Beckerle, P. (2021). Adaptation strategies for personalized gait neuroprosthetics. Frontiers in Neurorobotics, 15, 750519. DOI: https://doi.org/10.3389/fnbot.2021.750519

Kostas, D., & Rudzicz, F. (2020). Thinker invariance: Enabling deep neural networks for BCI across more people. Journal of Neural Engineering, 17(5), 056008. DOI: https://doi.org/10.1088/1741-2552/abb7a7

Laitinen, A., & Sahlgren, O. (2021). AI systems and respect for human autonomy. Frontiers in Artificial Intelligence, 4, 705164. DOI: https://doi.org/10.3389/frai.2021.705164

Valeriani, D., Santoro, F., & Ienca, M. (2022). The present and future of neural interfaces. Frontiers in Neurorobotics, 16, 953968. DOI: https://doi.org/10.3389/fnbot.2022.953968

Weld, D. S., & Bansal, G. (2019). The challenge of crafting intelligible intelligence. Communications of the ACM, 62(6), 70–79. DOI: https://doi.org/10.1145/3282486