Superintelligence, Superintimate

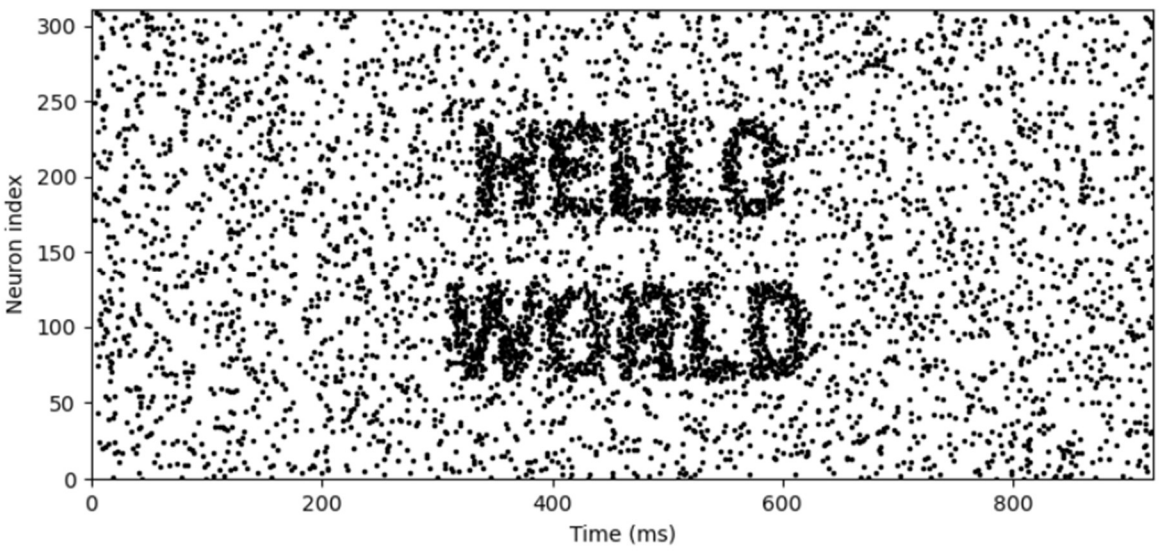

Figure 1. When neurons talk, AI listens. Each dot in this raster plot represents the ‘spiking’ of a single neuron; only in their ensemble activity does meaning emerge. Generated in the Brian2 neural network simulator.

Executive Summary

Within the next decade, one of the most transformative uses of artificial intelligence will be in neural interfacing: systems that read from and write to the human brain via brain computer interfaces (BCIs). I focus on a specific clinical application: AI-driven neural decoding and closed-loop neuromodulation that restores communication to people who cannot speak, grants environmental control to people who cannot move, and dynamically tunes neuromodulation for those whose brains oscillate between pathological states. The focus is on clinically supervised systems—implanted or medically monitored devices used in neurology, neurosurgery, psychiatry, and rehabilitation—not consumer gadgets or speculative enhancement.

The benefits are empirically grounded. For people with locked-in syndrome or advanced motor neuron disease, AI-mediated BCIs have moved the frontier of communication from letter-by-letter spelling to conversational text and natural-sounding speech[1–3]. For movement disorders, adaptive deep brain stimulation already uses neural biomarkers and algorithmic control to stabilize symptoms and reduce side effects in ways that static stimulation cannot[4]. For neuroscience, AI-based decoders and encoders connect distributed brain activity to the representational spaces used in modern language and vision models, offering new tools for understanding how the brain computes[5].

The risks are equally novel. Intrusive sensing and stimulation threaten mental privacy, cognitive liberty, and psychological continuity. Ienca and Andorno argue that neurotechnology exposes gaps in existing human-rights frameworks and motivate a family of “neurorights” to mental privacy, mental integrity, cognitive liberty and psychological continuity[6]. Ligthart and colleagues trace how such ideas are beginning to filter into international law and national constitutions[7]. AI compounds these concerns by enabling subtle inferences about internal states and by inserting opaque decision-making into stimulation and control. There are also familiar clinical risks: surgical harm, device reliability, model drift, and mis-decoding intentions with serious consequences.

My central thesis is that there is a strong moral and practical case for applying AI to neural interfacing—provided we treat these systems as high-stakes infrastructure governed by neurorights, strong data protections, and user-centric design. For people who are cognitively intact but unable to speak or move, the baseline is not a benign status quo but a condition many describe as a form of living imprisonment. In that context, declining to use AI-enabled BCIs when they can plausibly restore communication is not ethically neutral.

In this essay, I situate AI-enabled neural interfaces within healthcare, review the state of the art in decoding and neuromodulation, analyze benefits and beneficiaries, examine risks, and make the affirmative case for cautiously ambitious deployment under robust safeguards.

1. Introduction

Deep within folds of grey and white tissue lies the totality of our subjective reality – and, for many patients, the last intact refuge of the self. That is why disorders of the brain can be so intimately devastating, and why technologies that touch neural circuits carry unusual moral stakes: they can restore agency where the body has become an unreliable vehicle for the mind. This promise became tangible in 2016 at EPFL, when a paralyzed monkey took its first steps: its leg reanimated by a stimulating probe that translated cortical activity into commands for its own nerves. The breakthrough wasn't enabled by a better electrode array, but by the machine learning systems that decoded the brain's activity in real time. In the decade since, a growing share of advances in neurotechnology can be credited, in part, to artificial intelligence. We now stand at a critical juncture: if we proceed carefully, AI could restore autonomy to those suffering from conditions once deemed intractable; if we don't, we could forfeit sovereignty over the mind. In this essay, I argue that neural interfaces carry both the highest stakes and the greatest near-term impact among AI applications in the coming decade.

2. Sector: Brain Computer Interfaces

Since early work linking simple electroencephalography (EEG) rhythms to cursor movements, brain computer interfaces (BCIs) have evolved into complex systems that translate central nervous system signals into commands for external devices or neuromodulation patterns[8]. In their modern instantiation, BCIs have accrued substantial support and optimism from researchers and clinicians alike: they represent a major advance in neurology and rehabilitation for patients with otherwise intractable communication and motor deficits[9]. Importantly, the systems of interest here are clinically supervised, often surgically implanted, and targeted at serious neurological and psychiatric conditions: they belong alongside pacemakers and cochlear implants, not consumer wearables.

BCIs may be divided into those which read-out and those which write-in (with an emerging class targeting bidirectional modulation). Both capabilities already exist in partial form: the NEJM neuroprosthesis and related systems published in *Nature* show that spiking or high-gamma activity can be translated into rapid, conversational communication[2,3,10], while adaptive deep brain stimulation platforms demonstrate that neural biomarkers can guide real-time adjustments to stimulation in Parkinson's disease and other movement disorders[4]. The near-term question is therefore not whether AI will become a mainstay of BCIs, but how deliberately we will shape its role and under what ethical, legal, and social terms.

3. Benefits

3.1 Restoring communicative agency

For individuals with locked-in syndrome, advanced amyotrophic lateral sclerosis, or severe brainstem stroke, communication is often limited to a few words per minute via eye-tracking, letter boards, or residual facial movement. AI-driven BCIs that decode attempted or imagined speech can provide major improvements in throughput while allowing users to form sentences in their own idiolect and style. Willett et al. show that intracortical recordings paired with deep learning can approach conversational text communication in attempted speech decoding[2]. Metzger and colleagues complement this approach with high-density electrocorticography (ECoG), showing real-time decoding that supports text, synthetic audio approximating a patient's pre-injury voice, and a photorealistic avatar producing corresponding facial expressions[3].

Non-invasive systems remain lower bandwidth, but Tang et al. demonstrate continuous semantic decoding from fMRI by combining a language-model prior with an encoding model that predicts BOLD responses[11]. Decoding is strongly individual-specific, and participants can intentionally disrupt it by thinking different thoughts – meaning controllability and mental privacy are engineering constraints, not merely downstream ethics.

As decoding becomes faster and more expressive, the benefits extend beyond message rate: restoring prosody, facial expressivity, and conversational turn-taking moves communication toward genuine social interaction. This affects psychological well-being, family relationships, and participation in work or education. Case reports emphasize the emotional significance of hearing a synthetic voice approximating a patient's pre-injury timbre and seeing "their" facial expressions again[3].

3.2 Precision neuromodulation for neurological and psychiatric disease

Movement disorders such as Parkinson's disease, dystonia, and essential tremor impose substantial disability and economic cost. Continuous DBS provides symptom relief but can generate dyskinesias, mood changes, and cognitive effects when stimulation is not well tuned. Adaptive DBS aims to deliver precise stimulation when pathological oscillations emerge, reducing side effects while maintaining symptom control[4]. AI can enhance this by learning person-specific mappings from neural biomarkers and behavioral telemetry (for example, wearables capturing tremor) to stimulation parameters, enabling more subtle and individualized control policies.

Beyond movement disorders, there is growing interest in closed-loop neuromodulation for depression, obsessive–compulsive disorder, chronic pain, and epilepsy. AI models that characterize patient-specific neural states could support "precision psychiatry," where stimulation is targeted not merely to anatomical loci but to distributed network states associated with symptom

clusters[8,12]. AI-enabled neural interfaces also offer new paradigms for rehabilitation. BCIs can drive exoskeletons or virtual environments in ways that promote neuroplasticity after stroke or spinal cord injury. When augmented with adaptive decoders, they can adjust task difficulty and feedback based on estimates of engagement, fatigue, or frustration derived from neural and behavioral signals, potentially improving adherence and outcomes[8]. Proof-of-concept work has even mapped dynamic fMRI into rough reconstructions of imagined scenes (e.g., “Cinematic Mindscapes”), hinting – speculatively – at future cognitive prostheses for memory encoding and retrieval, attention stabilization, or language production in conditions such as traumatic brain injury or aphasia [11,13,14].

3.3 Beneficiaries

The immediate beneficiaries of AI-based BCIs are patients with severe motor, speech, or movement disorders, for whom current assistive options remain low-bandwidth and fragile. The neurosurgeons, neurologists, psychiatrists, and rehabilitation clinicians who devote their lives to the betterment of these patients also gain more precise tools. Importantly, if governed correctly, the downstream gains extend to the health system: therapies that restore function or stabilize symptoms can reduce long-term disability, intensive caregiving needs, and institutionalization. Conversely, if governance is weak, the main beneficiaries could be private firms monetizing neural data and devices, with patients bearing risk. The distribution of benefits therefore depends critically on policy choices, not only technical ones.

From a scientific perspective, AI-driven BCIs act as high-dimensional probes of neural computation. Decoders that map neural activity to linguistic or visual representations effectively serve as “model organisms” for theories of representation and transformation in cortex. Foundation models trained on language or images provide candidate representational spaces; fitting them to neural data tests whether those spaces align with biological coding[5]. Similarly, closed-loop neuromodulation experiments that perturb specific network states and observe behavioral and experiential changes can constrain causal models of brain function. In this sense, AI’s application to neural interfaces is not only therapeutic, but an engine for cognitive neuroscience, potentially accelerating discoveries of future clinical interventions.

4. Risks

4.1 Mental privacy and cognitive liberty

AI-enhanced neural interfaces sharpen longstanding philosophical questions about the opacity of the mind. Ienca and Andorno argue that brain data should be considered an especially sensitive category because it can reveal not only health status but also preferences, intentions, and potentially aspects of personality[6]. Ligthart et al. survey the emerging "neurorights" discourse and show that mental privacy, mental integrity, and cognitive liberty are increasingly invoked by international organizations and national legislatures as distinct values that may warrant special legal protection[7]. AI decoders heighten these concerns because they are probabilistic and can infer latent states from patterns not easily inspected by humans. Non-invasive semantic decoders, for example, could theoretically reconstruct rough content of inner speech or imagined narratives. Tang et al. explicitly test this and show that successful decoding requires both extensive participant-specific training and cooperative engagement; without cooperation, the model fails, and cross-subject generalization is poor[11]. This is reassuring, but future improvements in modeling and sensing could erode some of these practical barriers. The underlying ethical concern is not omnipotent "mind-reading" but incremental encroachment: employers, insurers, or states might incentivize or coerce adoption of neural monitoring ostensibly for safety or productivity, gradually normalizing access to mental states that were previously private.

4.2 Autonomy, agency, and dependence

Closed-loop BCIs blur the boundaries of agency. If a system continuously decodes affective state and adjusts stimulation or environmental feedback, who is "in control" of the resulting behavior? Some ethicists worry about "techno-paternalism," where systems nudge users toward designer-selected goals under the guise of therapy or optimization. Others note risks of over-reliance, where users defer to algorithmic outputs in ways that diminish their own self-trust. Inner-speech decoding BCIs illustrate both sides. A recent Stanford study shows that inner monologue can be decoded with reasonable accuracy from implanted electrodes, but the system is designed to activate only when the participant silently thinks a specific passphrase ("chitty chitty bang bang"), precisely to prevent inadvertent decoding[15]. This kind of explicit, user-controlled gating respects autonomy, but similar systems without such gates could be misused.

4.3 Safety, robustness, and clinical risk

Invasive neural interfaces carry surgical risks: infection, hemorrhage, hardware failure. BCIs that require chronic implants therefore need clear evidence of benefit and robust hardware design. AI adds another layer of risk: model drift, adversarial examples, unexpected failure modes under distribution shift, and the opacity of deep networks. A mis-decoded command in a communication interface may be embarrassing; a mis-inferred state in a stimulation system could worsen mood

or motor symptoms. Regulatory guidance for software as a medical device and for adaptive algorithms is still evolving. Without requirements for continual post-market surveillance, interpretability, and fail-safe modes (for example, fallback to conventional DBS parameters), there is a risk that systems will be deployed faster than they can be properly monitored.

4.4 Equity and global justice

Neural interfaces are likely to be expensive, at least initially. There is a risk of creating a stratified system in which wealthy patients in high-income countries receive AI-augmented BCIs, while others are left with minimal care. Data governance also raises equity concerns: large, high-quality neural datasets collected from early adopters could become valuable proprietary assets, reinforcing economic concentration. The patient population who most benefit from this technology are necessarily disadvantaged in broader society, and may lack a political voice: it is therefore extremely important that these patients are not merely research subjects but co-designers and beneficiaries. Achieving this requires deliberate policy, including public funding, equitable trial recruitment, and pricing models that reflect long-term societal benefits.

4.5 Dual use and militarization

Finally, AI-enhanced BCIs have clear dual-use potential. Military and intelligence contexts may be interested in systems that enhance vigilance, stress resilience, or coordination, as well as in tools for interrogation or behavioral prediction. Reports of national strategies for BCIs that explicitly include enhancement and military applications illustrate that some states already view neural interfaces as strategic technologies[16]. Because many implants depend on wireless telemetry and software updates, they also create a cyber-physical attack surface: compromised devices could exfiltrate sensitive neural data or alter stimulation parameters. This “brainjacking” risk has been discussed explicitly for DBS, and regulators increasingly treat cybersecurity as a core safety requirement for networked medical devices[17]. These applications lie outside healthcare but could influence public trust and normative expectations. If BCIs are widely associated with surveillance or coercion, public resistance could slow or block even clearly beneficial medical uses.

5. The affirmative case: why AI-enabled neural interfacing will improve healthcare

The case for deploying AI in neural interfacing rests on three pillars: unmet clinical need, technical tractability of safeguards, and the availability of emerging normative frameworks.

First, the unmet clinical need is enormous. Neurological and psychiatric disorders are major contributors to global disability-adjusted life years, and many of the most devastating manifestations – locked-in syndrome, advanced ALS, severe Parkinson's, treatment-resistant depression – are poorly addressed by current therapies. High-bandwidth BCIs and precision neuromodulation directly target this gap, and evidence from speech BCIs and adaptive DBS suggests large benefits in patients who have exhausted standard options[2–4].

Second, many of the most concerning risks are amenable to concrete technical and legal mitigations. Decoders often require person-specific training and cooperative engagement[11], and system designers can amplify this "privacy by default" structure by requiring explicit activation signals, local on-device decoding, and cryptographic protections that ensure raw neural data never leave the device unencrypted. As a regulatory requirement, any software update to decoding or stimulation policies should be logged, independently auditable, and rollback-capable, so clinicians and users can detect drift and revert to a validated state after adverse events. The use of mental passphrases in inner-speech BCIs is a good example of how design choices can institutionalize respect for cognitive liberty[15].

Third, the normative and legal groundwork for neurorights is advancing rapidly. Conceptual foundations for rights to mental privacy, cognitive liberty, and mental integrity are increasingly discussed by international bodies and some national legislatures[6,7]. If codified alongside medical device regulation, they can constrain high-risk uses (for example, compulsory neural monitoring) while permitting restorative clinical applications under strict consent and governance.

The key is to treat AI-enabled neural interfaces as part of a broader system that includes technical architecture, clinical practice, law, and public deliberation. A plausible governance architecture would include: legally recognized neurorights; mandatory user-controlled activation and deactivation mechanisms; strict limits on reuse of neural data outside the clinical context; public or non-profit stewardship of large neural datasets; independent ethics and safety boards for high-risk trials; and international norms that proscribe coercive or military uses. None of these are trivial, but all are compatible with existing regulatory tools and international human-rights regimes.

Under such conditions, the outcome we may expect from deploying AI-driven neural interfaces in healthcare is predominantly therapeutic. The alternative – foregoing or slowing these developments – would leave millions of patients without access to lifechanging treatments, and would unfortunately cede the field to actors less constrained by ethical scrutiny.

6. Conclusion

AI applied to neural interfacing is not a marginal application. It is already enabling new forms of communication and neuromodulation, and the stakes for autonomy, dignity, and human flourishing are unusually high. Within a decade, integrated AI-enabled neural interfaces could plausibly become routine in specialized clinical centers.

These technologies both respect and threaten the mind’s special status. They can restore communicative agency and self-expression to people whose bodies can no longer serve as reliable vehicles for their minds, yet they can also expose and shape mental states in ways that undermine privacy and autonomy. The response should be neither rejection nor uncritical embrace, but deliberate design and governance: decoders and control systems that structurally encode consent, neurorights embedded in law, and a priority on restorative therapies over enhancement. The choice is not whether AI will be applied to neural interfacing, but how, and under whose terms. Superintelligence matters most when it meets the superintimate – where the boundary of the self is at stake.

References

1. Card, N. S. *et al.* An Accurate and Rapidly Calibrating Speech Neuroprosthesis. *N. Engl. J. Med.* **391**, 609–618 (2024).

2. Willett, F. R. *et al.* A high-performance speech neuroprosthesis. *Nature* **620**, 1031–1036 (2023).

3. Metzger, S. L. *et al.* A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* **620**, 1037–1046 (2023).

4. Bronte-Stewart, H. M. *et al.* Long-Term Personalized Adaptive Deep Brain Stimulation in Parkinson Disease: A Nonrandomized Clinical Trial. *JAMA Neurol.* **82**, 1171 (2025).

5. Wang, R. & Chen, Z. S. Large-scale foundation models and generative AI for BigData neuroscience. *Neurosci. Res.* **215**, 3–14 (2025).

6. Ienca, M. & Andorno, R. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sci. Soc. Policy* **13**, 5 (2017).

7. Ligthart, S. *et al.* Minding Rights: Mapping Ethical and Legal Foundations of 'Neurorights'. *Camb. Q. Healthc. Ethics* **32**, 461–481 (2023).

8. Awuah, W. A. *et al.* Bridging Minds and Machines: The Recent Advances of Brain-Computer Interfaces in Neurological and Neurosurgical Applications. *World Neurosurg.* **189**, 138–153 (2024).

9. Deng, Q., Fu, Z., Ma, N. & Wang, B. Application and future directions of brain-computer interfaces in neurological disorders: Technological advances, clinical practices, and challenges. *Brain Hemorrhages* **6**, 306–314 (2025).

10. Moses, D. A. *et al.* Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. *N. Engl. J. Med.* **385**, 217–227 (2021).

11. Tang, J., LeBel, A., Jain, S. & Huth, A. G. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat. Neurosci.* **26**, 858–866 (2023).

12. Andelman-Gur, M. M. & Fried, I. Consciousness: a neurosurgical perspective. *Acta Neurochir. (Wien)* **165**, 2729–2735 (2023).

13. Chen, Z., Qing, J. & Zhou, J. H. Cinematic Mindscapes: High-quality Video Reconstruction from Brain Activity. Preprint at https://doi.org/10.48550/arXiv.2305.11675 (2023).

14. Takagi, Y. & Nishimoto, S. High-resolution image reconstruction with latent diffusion models from human brain activity. Preprint at https://doi.org/10.1101/2022.11.18.517004 (2022).

15. Kunz, E. M. *et al.* Inner speech in motor cortex and implications for speech neuroprostheses. *Cell* **188**, 4658-4673.e17 (2025).

16. Kosal, M. & Putney, J. Neurotechnology and international security: Predicting commercial and military adoption of brain-computer interfaces (BCIs) in the United States and China. *Polit. Life Sci.* **42**, 81–103 (2023).

17. Pycroft, L., Boccard, S. G., Fitzgerald, J. J., Green, A. L. & Aziz, T. Z. Brainjacking: cyber-security risk in deep brain stimulation. *Brain Stimul. Basic Transl. Clin. Res. Neuromodulation* **10**, 359 (2017).