Guaranteeing Agency in the Age of Neural Prosthetics

Executive Summary

Sector: Healthcare & Neuro-Technology **Proposed Safeguard:** The "Neuro-Escrow Standard"

In the next decade, Artificial Intelligence will dissolve the functional barrier between "abled" and "disabled." In the future, by leveraging "*Generative Motor Policies"* — transformer-based models that would translate noisy neural activity into fluid robotic movement — we may restore physical agency to the more than 1.6 million Americans with limb loss [1] and tens of millions more worldwide living with paralysis. This is not merely a medical treatment, but an economic and civic shift that would allow individuals with quadriplegia, ALS, and combat-related amputations to re-enter the workforce and navigate the world with autonomy previously thought impossible.

However, the current trajectory of proprietary biology poses a catastrophic risk: *obsolescence of the self*. As demonstrated by the March 2020 wind-down of Second Sight Medical Products [2], which left hundreds of blind patients with unsupported and failing retinal implants, tethering the human body to cloud-based, proprietary servers creates a single point of failure. If a manufacturer fails or changes its business model, users may find their own limbs locked and their autonomy converted into a subscription they can no longer afford.

To mitigate these risks, we must decouple the function of the device from the fate of the company. This essay proposes the Neuro-Escrow Standard, a regulatory framework that requires the source code and model weights for any essential neural prosthetic to be held in a public trust. Should the manufacturer cease support, this code automatically enters the public domain. This grants the user the permanent right to repair their own body.

Guaranteeing Agency in the Age of Neural Prosthetics

The year is 2029. Julian, a 24-year-old software engineer, sits in a coffee shop in Cambridge drinking his coffee and coding. Four years ago, after a severe car accident, a spinal cord injury left him paralyzed from the neck down. Today, he is typing code at roughly 90 words per minute, a plausible extrapolation from speech and handwriting Brain-Computer Interface (BCI) benchmarks already achieved [3,4], and holding a hot espresso. His movements look indistinguishable from anyone else's. Julian is not biologically healed; he is "bridged."

A high-bandwidth BCI implanted in his motor cortex streams raw neural spikes to an AI model running on his exoskeleton. The AI doesn't just listen, it anticipates, using a Generative Motor Policy to smooth his noisy intent into fluid, graceful motion. For Julian, the AI is part of him. It is his hands, his legs, his freedom, and his hope.

Then, as he gets ready to stand up, the notification appears on his retinal display.

"SERVER UNREACHABLE. LICENSE VALIDATION FAILED. MOTOR FUNCTIONS SUSPENDED."

The miracle vanishes. Julian's arms lock in place. The cup falls from his hand and shatters on the floor. He is paralyzed again. Not because of biology, but because the startup that owns his subscription just entered administration.

This scenario is the inevitable conclusion of our current trajectory in neuro-technology. As we witness the rapid strides occurring in the field of BCI, the science of establishing direct communication pathways between the brain and external devices, the stakes are becoming tangible. We see this most clearly with Neuralink, the neurotechnology company pioneering high-bandwidth implantable chips that allow users to control digital devices through neural activity alone [5]. As these proprietary AI models evolve to become the essential "operating system" for the human body, we face a profound ethical divergence. We are building a future where the disabled can walk, run, and work again, but only as tenants in their own bodies. To realize the true "net positive" of AI in healthcare, we must answer a question that has never existed in human history: When the software crashes, who owns the limb?

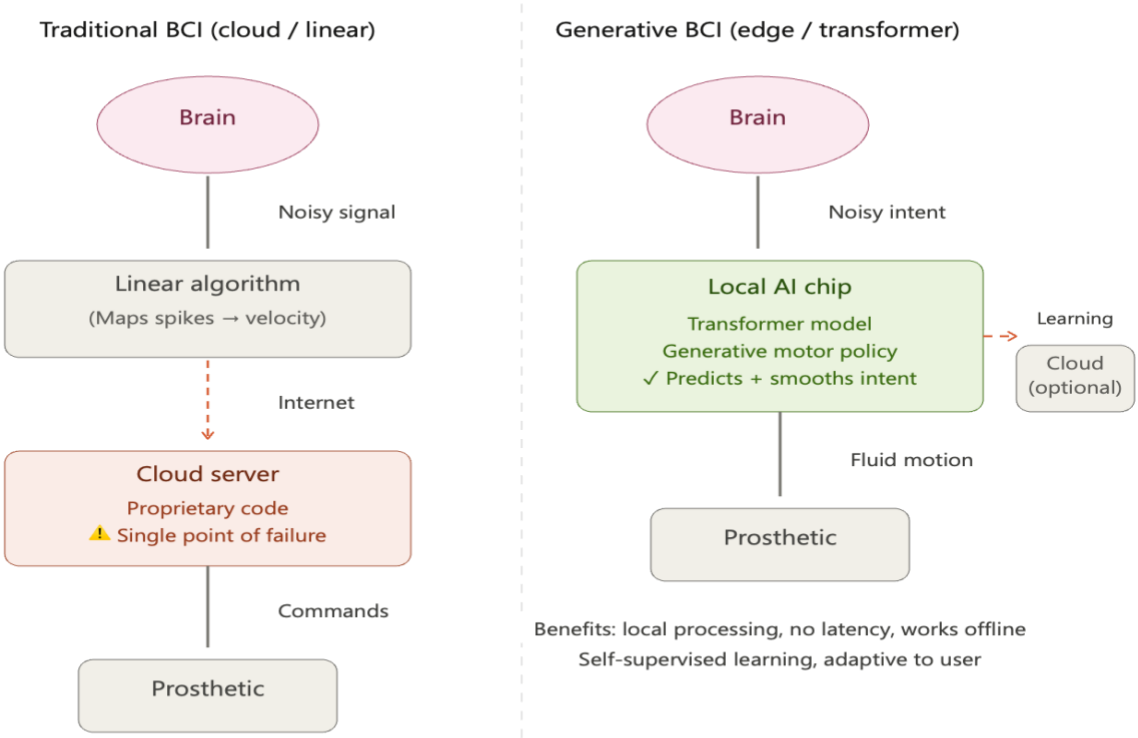
The healthcare sector stands to gain the most profound "net positive" impact from AI because it is the only domain where technology does not merely optimize systems but fundamentally redefines the limits of human biology. AI offers tangible restoration to those who have suffered catastrophic accidents, turning permanent disability into temporary injury. For decades, the field was stagnant. Prosthetics were "unintelligent mechanics" or passive devices that acted as dead weight. To operate a robotic arm, a user had to engage in myoelectric switching, manually toggling muscles in their chest or back to trigger simple open/close movements. This process was cognitively

exhausting, non-intuitive, and frustratingly slow, leading to a device abandonment rate of roughly 26% for body-powered and 23% for myoelectric upper-limb prostheses [6], with rejection remaining stubbornly high in contemporary cohorts [7]. AI eliminates this cognitive friction by replacing manual toggling with seamless, thought-driven intent.

Generative Motor Policies represent the specific architectural shift, illustrated in Figure 1, that would solve this cognitive friction. This technology would represent the convergence of robotics and the Foundation Model architecture, large models pre-trained on broad data and adaptable to many downstream tasks [8], that powers Large Language Models (LLMs). Building on breakthrough research in Vision-Language-Action (VLA) models, which co-fine-tune web-scale vision-language models with robot demonstrations so that actions are emitted as text tokens [9], and emerging World Models, such future systems would treat physical movement as a language. Just as an LLM predicts the next word in a sentence based on context, a Generative Motor Policy would simulate the next millisecond of muscle contraction based on intent. By training on large-scale "cross-embodiment" data — training data aggregated across different robot bodies so a single policy transfers between morphologies [10] — the model learns the fundamental "grammar" of the human body. Recent surveys of foundation-model-driven robotics document this paradigm shift across perception, planning, control, and human-robot interaction [11]. We are at the inflection point for this technology today because of the Transformer architecture [12]. Unlike previous neural networks (RNNs), which struggled with long sequences, Transformers can attend to the "physics of the cup," the "geometry of the hand," and the "noisy neural signal" simultaneously. When Julian thinks "grab," the model doesn't just amplify the signal; it would simulate the optimal trajectory to fulfill that intent, filling in the gaps to create fluid action from noisy thought. It would effectively act as a semantic translator, converting the vague intent of the mind into the precise syntax of the machine.

To understand why this shift is imminent, we must distinguish between decoding the signal and inferring intent. Traditional BCIs rely on linear algorithms that map neural spikes directly to velocity. Figure 1 is a schematic illustration; today's clinical BCIs typically run on local desktop PCs, but proposed commercial deployments increasingly assume cloud connectivity. These systems are fragile; they require daily recalibration and struggle with the shifting nature of the brain. The next generation of AI introduces Self-Supervised Learning, in which models extract supervision from the data itself [13]. Much like a Large Language Model predicts the next word in a sentence, a future class of transformer-based models would predict the next movement in a sequence. This shift from command-following to intent-prediction is what makes the technology viable for widespread deployment within the decade.

Figure 1: Traditional vs generative BCI architecture



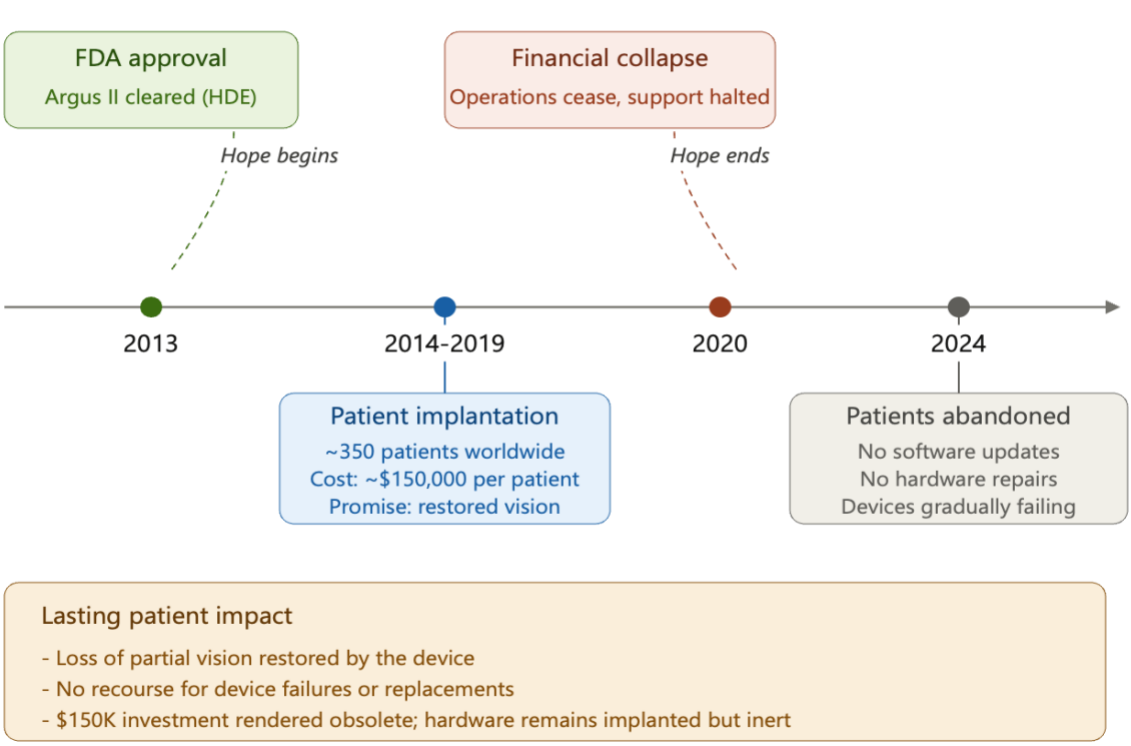
This technology's impact extends beyond spinal cord injuries to reach the approximately 1.6 million Americans living with limb loss, a figure projected to double by 2050 [1], including thousands of veterans from recent conflicts [14]. Traditional prosthetics are passive mechanics that essentially act as fancy springs. They cannot decide to run, jump, or stabilize on uneven terrain. AI changes this dynamic. A veteran with a smart amputation doesn't just get a plastic foot; they get a robotic limb that predicts terrain. Using the same transformer architecture described above, the leg can identify stairs or gravel via sensors and adjust its torque in milliseconds. This results in a return to active duty, competitive sports, or simply the ability to run and take care of their own child.

The "net positive" impact is not just moral but fiscal. Currently, the lifetime cost of care for a young person sustaining high tetraplegia (C1–C4) at age 25 exceeds $5.8 million in 2022 dollars [15], predominantly driven by the need for round-the-clock physical assistance. By restoring motor function, AI prosthetics have the potential to invert this economic dynamic. We are not just

reducing the cost of care; we are unlocking the productivity of a workforce that has been systematically sidelined. In an aging society facing labor shortages, the ability to convert care dependency into economic contribution represents a massive, quantifiable social dividend. Furthermore, the psychological benefit of restoring physical capability is directly linked to the restoration of identity.

However, every coin has two sides, and the same is true of this technology. While the capabilities of this technology are promising, we must recognize the danger in treating these devices like consumer electronics. We accept that phones have a shelf life; when an old phone stops getting updates, we buy a new one, but you cannot swap out a brain implant like a SIM card. We have already seen the preview of this dystopia. In 2020, Second Sight Medical Products, a pioneer in bionic eyes, suffered a financial collapse (Figure 2). The company stopped developing the technology. Hundreds of patients with Argus II retinal implants were left stranded [2]. Their bionic eyes still worked physically, but the software ecosystem vanished. When parts broke, there were no repairs. When the software glitched, there was no patch. They were left with dark, useless metal in their skulls.

Figure 2: The Second Sight collapse - a case study in device abandonment



Scale this up to 2030, and the risks compound. We can imagine tens of thousands of people relying on proprietary AI models to walk. We face the risk of ransomware, where a hacker locks the cloud

server and demands payment to unlock 50,000 pairs of legs. We face the risk of subscription traps, where a company introduces surge pricing for mobility. Most likely, we face the risk of bankruptcy. If a startup fails and the servers go dark, the user is blocked.

A common critique of open standards in medical devices is security. Critics ask: If the code for walking is public, can't a malicious actor hack a user's legs? This relies on the fallacy of "Security through Obscurity," the mistaken belief that hiding code makes it safe. History proves otherwise. In 2017, the FDA issued a voluntary firmware recall covering approximately 465,000 Abbott (formerly St. Jude Medical) implantable pacemakers after researchers demonstrated authentication-bypass and RF-wake vulnerabilities [16,17]. Because the code was closed, the vulnerabilities remained hidden in firmware for years. The obscurity protected the vulnerability, not the patients.

By contrast, an open-standard architecture benefits from the "immune system" of global verification. We see this in the OpenAPS (Open Artificial Pancreas System) movement, where a community of patient-hackers built an open-source automated insulin delivery system. In the 2022 CREATE randomized controlled trial published in *The New England Journal of Medicine*, the OpenAPS algorithm was proven safe and effective [18]. Open-source code allows for transparency, so bugs can be identified and patched by a global network of engineers faster than any single corporate team could manage. In the context of neural prosthetics, transparency is not a risk factor. It is the only way to ensure that the code running our bodies is robust enough to survive the real world.

We cannot rely on the benevolence of tech companies to secure human agency. We must implement it at the policy layer. To ensure the net impact of neural prosthetics is positive, we must adopt a regulatory framework that treats the code running a human body differently from the code running a thermostat. I propose the implementation of the Neuro-Escrow Standard (NES), a two-pillar policy safeguard designed to prevent the obsolescence of the self.

The first pillar states that we need to have right-to-repair trigger in place. Currently, medical device code is protected as a trade secret. The NES would mandate that any company seeking FDA approval for a Class III neural interface must deposit their source code, model weights, and code-signing keys into a federal public trust escrow. This escrow remains sealed as long as the company supports the device. However, a "Trigger Event," such as bankruptcy, discontinuation of support, or server shutdown, automatically unlocks the escrow. The code enters the public domain immediately. This guarantees that a community of open-source developers can patch, update, and maintain the software, ensuring that no patient is ever orphaned by a balance sheet.

The stakes of not having this mechanism are already measurable. When Medtronic could not patch its MiniMed 508 and Paradigm insulin pumps in 2019, the FDA stated plainly that the

manufacturer was “unable to adequately update” the devices, forcing roughly 4,000 U.S. patients to transfer to alternative pumps [19]. At least Medtronic still existed and could substitute hardware. When Second Sight wound down in March 2020, that door closed. Roughly 350 Argus II recipients worldwide now carry a permanently frozen version of a sophisticated implant: in 2013 Barbara Campbell’s device went dark on a New York subway platform mid-commute and never returned; her four years of partially restored vision vanished between two trains. The source code was never released.

There is also a clear historical parallel outside medicine that proves the escrow model works. When Microsoft killed its PlaysForSure DRM authentication servers, customers lost access to music they had legally purchased, the files would not play once the servers went dark. The lesson, learned at the cost of consumer trust, was that any device dependent on a remote license server is rented, not owned. Enterprise software learned this lesson earlier: today, virtually every enterprise SaaS contract over a certain dollar value includes a source-code escrow clause precisely so the customer retains operational continuity if the vendor fails. NES proposes that we extend to human bodies a protection we already extend to corporate procurement.

Open-source medical communities have proven capable of stepping in once code becomes available. OpenAPS users had working closed-loop insulin delivery years before any commercial system reached market, and the algorithm was later validated in a peer-reviewed 2022 NEJM randomized trial [18]; Tidepool Loop, another patient-built open-source system, secured FDA 510(k) clearance in 2023. A code-escrow trigger does not guarantee continued service. It guarantees the possibility of continued service, which is precisely what is currently foreclosed.

The second pillar would focus on local-first sovereignty, which entails that we must reject the cloud-dependent architecture for vital biological functions. Edge computing is the paradigm in which computation is performed on or adjacent to the end device rather than in centralized datacenters, to satisfy latency, bandwidth, privacy, or availability constraints that cloud architectures cannot meet [20,21]. The NES would therefore require that the Minimum Viable Function, the subset of operations necessary for basic embodied agency — specifically the ability to walk, grasp, and speak — must run entirely on-device via Edge Computing. Cloud connectivity can be used for optimization or learning new skills, but it cannot be a dependency for operation. If the internet cuts out or the server fails, the legs must still walk. This Local-First architecture ensures that the user’s agency is physically located on their body, not rented from a data center. When a device shifts from a cloud-tethered state to an MVF architecture, only some of the following “enhancement services” would be suspended to protect the user’s primary mobility.

These suspended enhancements may include the ability to pull real-time firmware updates derived from aggregate user data, where a user might miss a new model version that improves balance on

icy surfaces but retains their current "local" best-fit model, as well as the ability to preload architectural layouts of specific buildings to prime gait. Other suspended features include the capacity to retrain models as neural signals shift over months and real-time health monitoring by manufacturers. Ultimately, this local-first architecture ensures that the user's agency is physically located on their body, not rented from a data center, ensuring that a minimum viable function can never be compromised by a company's balance sheet or a server failure. This technical decoupling is more than a design preference; it is the necessary bridge to a future where cybernetic restoration is as reliable as biological health.

The transition from biological bodies to cybernetic ones is inevitable. The technology is already here. The policy is what falls behind. If we do nothing, we risk creating a future where the disabled are liberated from their physical limitations only to be shackled by legal and economic ones. We risk a world where a declined subscription payment causes a literal loss of footing. But if we implement safeguards like the Neuro-Escrow Standard, we can mitigate this risk. We can build a future where neural prosthetics are not corporate appliances that we use, but open instruments that we control. By guaranteeing that the code governing our bodies is as durable and accessible as the bodies themselves, we ensure that AI does not simply manage disability but truly erases it. We unlock a future where Julian does not drop his cup, because his hands, finally and permanently, belong to him.

References

[1] Ziegler-Graham, K., MacKenzie, E. J., Ephraim, P. L., Travison, T. G., & Brookmeyer, R. (2008). Estimating the prevalence of limb loss in the United States: 2005 to 2050. *Archives of Physical Medicine and Rehabilitation, 89*(3), 422–429.

[2] Strickland, E., & Harris, M. (2022, February 15). Their bionic eyes are now obsolete and unsupported. *IEEE Spectrum*.

[3] Willett, F. R., Kunz, E. M., Fan, C., Avansino, D. T., Wilson, G. H., Choi, E. Y., et al. (2023). A high-performance speech neuroprosthesis. *Nature, 620*(7976), 1031–1036.

[4] Card, N. S., Wairagkar, M., Iacobacci, C., Hou, X., Singer-Clark, T., Willett, F. R., et al. (2024). An accurate and rapidly calibrating speech neuroprosthesis. *New England Journal of Medicine, 391*(7), 609–618.

[5] Musk, E., & Neuralink. (2019). An integrated brain-machine interface platform with thousands of channels. *Journal of Medical Internet Research, 21*(10), e16194.

[6] Biddiss, E. A., & Chau, T. T. (2007). Upper limb prosthesis use and abandonment: A survey of the last 25 years. *Prosthetics and Orthotics International, 31*(3), 236–257.

[7] Salminger, S., Stino, H., Pichler, L. H., Gstoettner, C., Sturma, A., Mayer, J. A., Szivak, M., & Aszmann, O. C. (2022). Current rates of prosthetic usage in upper-limb amputees — have innovations had an impact on device acceptance? *Disability and Rehabilitation, 44*(14), 3708–3713.

[8] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

[9] Zitkovich, B., Yu, T., Xu, S., Xu, P., Driess, D., Ichter, B., et al. (2023). RT-2: Vision-language-action models transfer web knowledge to robotic control. *Proceedings of the 7th Conference on Robot Learning* (PMLR Vol. 229, pp. 2165–2183).

[10] Open X-Embodiment Collaboration, O'Neill, A., Rehman, A., Gupta, A., et al. (2024). Open X-Embodiment: Robotic learning datasets and RT-X models. *IEEE International Conference on Robotics and Automation (ICRA)*.

[11] Khan, M. T., & Waheed, A. (2025). Foundation Model Driven Robotics: A Comprehensive Review. *arXiv preprint arXiv:2507.10087*.

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.

[13] Ye, J., & Pandarinath, C. (2021). Representation learning for neural population activity with Neural Data Transformers. *Neurons, Behavior, Data Analysis, and Theory, 5*(3), 1–18.

[14] Reiber, G. E., McFarland, L. V., Hubbard, S., Maynard, C., Blough, D. K., Gambel, J. M., & Smith, D. G. (2010). Servicemembers and veterans with major traumatic limb loss from Vietnam war and OIF/OEF conflicts. *Journal of Rehabilitation Research & Development, 47*(4), 275–297.

[15] National Spinal Cord Injury Statistical Center. (2023). *Traumatic spinal cord injury facts and figures at a glance*. University of Alabama at Birmingham.

[16] U.S. Food and Drug Administration. (2017). *Firmware update to address cybersecurity vulnerabilities identified in Abbott's (formerly St. Jude Medical's) implantable cardiac pacemakers: FDA safety communication*.

[17] Baranchuk, A., Refaat, M. M., Patton, K. K., Chung, M. K., Krishnan, K., Kutyifa, V., Upadhyay, G., Fisher, J. D., & Lakkireddy, D. R. (2018). Cybersecurity for cardiac implantable electronic devices: What should you know? *Journal of the American College of Cardiology, 71*(11), 1284–1288.

[18] Burnside, M. J., Lewis, D. M., Crocket, H. R., Meier, R. A., Williman, J. A., Sanders, O. J., et al. (2022). Open-source automated insulin delivery in type 1 diabetes. *New England Journal of Medicine, 387*(10), 869–881.

[19] U.S. Food and Drug Administration. (2019). *Certain Medtronic MiniMed insulin pumps have potential cybersecurity risks: FDA safety communication*.

[20] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal, 3*(5), 637–646.

[21] Satyanarayanan, M. (2017). The emergence of edge computing. *Computer, 50*(1), 30–39.