

Nano-totalitarianism: Brain-computer interfaces, fusion databases, and the weaponisation of consciousness

John Kendall Hawkins¹

1 Independent scholar, San Diego, California, USA. Correspondence to: jkhawkins@sandiego.edu

Abstract

Debate about interrogational torture still returns to the ticking-bomb scenario, despite the weight of evidence that it neither works nor withstands moral scrutiny. The post-9/11 record, from Abu Ghraib to the CIA black sites, already marked a decisive shift toward psychological forms of coercion. The technologies now emerging make a different kind of coercion possible. This paper argues that defence-funded brain-computer interfaces, of which Synchron's Stentrode is the clearest current example, combined with large language models and fusion databases, could support what I call *nano-totalitarianism*: the convergence of pervasive personal data, neurodata, adaptive AI, and technologically mediated coercion in ways that could penetrate or continuously condition the inner sphere of the person. Operationalised, this is a self-sustaining regime in which a detainee is held in continuous psychological distress through fabricated scenarios, interference with memory, and synthetic social contact. Such a system would turn neuroplasticity into an instrument, laying down durable neural pathways of fear, while automated monitoring assesses compliance in real time. Its lineage runs from MKULTRA's psychochemical experiments to the extraction logic of surveillance capitalism, and its endpoint is a panopticon inside the skull. The essay proceeds anticipatorily, distinguishing throughout between demonstrated current capability, plausible convergence of existing components, and speculative future possibility.

Keywords: neuro-totalitarianism, surveillance, ChatGPT, BCI, fusion databases

Introduction

After September 11, 2001, a distinct form of torture became common: punitive torture, intended to inflict severe emotional damage while leaving the body unmarked. Popular culture followed, with films and television normalising and at times celebrating the practice (Delehanty & Kearns, 2019). The controversy over *Zero Dark Thirty* (2012), arriving alongside the disclosures from Abu Ghraib and Guantánamo, opened a public argument about whether such methods are lawful or effective, and exposed how thoroughly entertainment, propaganda, and policy had run together.

The twenty-first century has produced a sharper turn. Neurotechnology and large-scale data systems now make it possible to monitor, manage, and potentially torture a person at

the level of thought. Through programmes such as DARPA's Next-Generation Nonsurgical Neurotechnology, and through commercial developers including Neuralink and Synchron, military and intelligence agencies are funding interfaces that read from and, at the far end, write to neural activity. Paired with large language models, these devices might let a paralysed patient speak or browse by intention alone. The same pairing could open an era of interrogation in which psychological coercion is continuous, adaptive, and difficult to observe from outside.

Fusion databases are the third component. These are consolidated repositories drawing on financial records, communications, biometrics, social media, and increasingly neural data. Edward Snowden, formerly an NSA and CIA contractor, has

argued that such systems produce a permanent record for every individual, enabling lifelong profiling and targeting (Snowden, 2019). Joined to a neural interface and a language model, that record becomes the raw material for psychological operations built against one specific person. The mind becomes a cell with no door.

A note on register. Throughout what follows, I distinguish three levels of claim: demonstrated current capability, plausible convergence of existing components, and speculative future possibility. The essay is anticipatory in character, and it draws on the governance-oriented framework developed by Istace and Costas Trascasas (2024) for reasoning about neurotechnologies whose full coercive potential has not yet been demonstrated but whose components are already in place. Distinguishing the three registers is what allows anticipatory analysis to proceed without collapsing into either technological credulity or dystopian projection.

This paper examines the evolution from historical torture to the nascent menace of neurotechnological and data-driven psychological torment. It looks at the technical aspects, the moral and legal issues, and the role of fusion databases in creating a new kind of “nano-totalitarianism” that threatens the fundamental essence of human freedom and dignity.

Torture, power, and the history of control

From physical pain to mental warfare

The myth of exile, and the Tree of Knowledge at its centre, sits near the foundation of Western culture. It is a story about forbidden knowledge and the authority that forbids it. What Genesis prohibits is not an act but an acquisition, and the prohibition establishes a relation between divine power and human curiosity that Hegel would later render as dialectic (Hegel, 1807/1977) and Foucault as biopower (Foucault, 1976/1978).

Torture has always concerned the control of information: forbidden, subversive, or heretical. Eden anticipates the interrogation room, where punishment is applied to extract a truth or to break a refusal. Oppenheimer’s invocation of the Bhagavad Gita at the first detonation, “Now I am become Death, the destroyer of worlds,” names the same vertigo in a secular register. The physics that produced medical isotopes also incinerated two cities. A brain-computer interface that restores speech to a paralysed patient can also take speech from a detained one. The symmetry is not hypothetical. Surveillance technologies sold as protective infrastructure have repeatedly been turned to coercion, from telegraph interception to metadata collection to the export of spyware platforms.

MKUltra was the Cold War’s version of the same transgression: a covert, quasi-divine attempt to take possession of the human mind. Its methods included involuntary LSD administration, electroshock, sensory deprivation, and psychological torture (Kinzer, 2019). Its personnel came in part from Nazi and Japanese war criminals brought to the United States under Operation Paperclip. Stephen Kinzer documents the CIA’s recruitment of Kurt Blome, formerly director of Nazi biological warfare research, who had “developed aerosol delivery systems for nerve gas, to be tested on inmates at Auschwitz; bred infected mosquitoes and lice for experiments at Dachau and Buchenwald; and produced gas to kill thirty-five thousand tuberculosis patients in camps across Poland” (Kinzer, 2019, p. 17). The recruitment of General Shiro Ishii, the army surgeon who directed Unit 731, went further still:

Ishii’s victims were slowly roasted by electricity, hung upside down, locked into high-pressure chambers until their eyes popped out, spun in centrifuges, infected with anthrax, syphilis, plague, and cholera, forcibly impregnated to provide infants for vivisection, bound to stakes to be incinerated by flamethrowers, and slowly frozen to observe hypothermia’s progress (Kinzer, 2019, p. 30).

American opinion would later register horror at waterboarding. What Ishii had done, and what the United States acquired from him in exchange for his immunity, did not enter that accounting.

The programme’s LSD experiments, conducted on subjects including Allen Ginsberg, Ted Kaczynski, and James “Whitney” Bulger, proceeded under the cover of scientific research. MKUltra never produced the Manchurian candidate it was chasing, but it established the state’s standing interest in the manipulation of mind as the most efficient route to information. The institutional lineage matters for what follows. MKUltra was a CIA programme; the current work is DARPA’s, which is to say military. That shift belongs to an era of cyberwarfare, in which knowing the adversary has become a continuous mission requirement rather than an episodic one.

Warnings about the security state’s expanding authority were issued early and largely disregarded. Eisenhower’s farewell address of 1961 urged the nation to guard against “the acquisition of unwarranted influence, whether sought or unsought, by the military-industrial complex” (Eisenhower, 1961). The Church Committee demonstrated in 1975 that the NSA’s capabilities could be turned on the American population at large. Speaking on national television that August, Church warned that the technology could “enable it to impose total tyranny,

and there would be no way to fight back,” and that such agencies had to be held under law and proper supervision “so that we never cross over that abyss. That’s the abyss from which there is no return” (Church, 1975). Both anticipated what Snowden documented four decades later (Snowden, 2019).

The ticking bomb fallacy in action

After September 11, the ticking-bomb scenario supplied the justification. Extraordinary rendition, the network of black sites, and Guantánamo Bay became the institutional form of an argument that information must be obtained at any cost. Zero Dark Thirty dramatised the premise and implied that it worked. The Senate Select Committee on Intelligence concluded otherwise in 2014, finding that enhanced interrogation had produced little intelligence of value (SSCI, 2014).

Abu Ghraib marked a shift from physical coercion to psychological display. Photographs recorded sexual humiliation, forced nudity, and the deliberate desecration of cultural and religious identity. The techniques echoed those the CIA had sanctioned, but the camera introduced something new: cruelty performed for a record (Hersh, 2004; Human Rights Watch, 2005).

Hersh’s reporting established institutional involvement, showing that military intelligence had not merely permitted these acts but encouraged them (Hersh, 2004). The psychological damage was the object rather than the byproduct. What Abu Ghraib demonstrated about ordinary people acting under institutional protection restated, in an actual prison, what the Stanford Prison Experiment had found in a simulated one (Haney et al., 1973).

Israel’s Lavender system in Gaza shows what fusion databases do at scale. According to investigative reporting by +972 Magazine and Local Call corroborated by named military sources, telecommunications metadata, facial recognition, and movement patterns were combined to generate target lists automatically, with reported error rates of ten to fifteen per cent (Abraham, 2024; Hawkins, 2025). The associated “Where’s Daddy?” protocol timed strikes for the moment targets returned to their family homes at night. The database was not an aid to the weapon. It was the weapon. The evidentiary status of these accounts is that of investigative journalism, not of official finding, and that qualification should be kept in view.

Torture recorded, distributed, and consumed marks the arrival of a condition in which perception itself is the contested ground. What began as forbidden fruit has become a secular preoccupation with total command over knowledge and bodies alike. The technology has grown more precise; the appetite it serves has not changed. From the exile in Eden to the

photographs from Abu Ghraib, the same demand persists: to command, to control, and to punish knowledge that will not comply. Torture, physical or psychological, remains the crudest expression of the fear beneath it.

Large language models: capability and dual use

How large language models work

ChatGPT is a transformer model, which is to say it uses self-attention to weight the relative importance of elements in its input and to produce output that holds together across a long context (Vaswani et al., 2017). Trained on a broad corpus of text, it learns by predicting the next token in a sequence and adjusting its parameters to reduce the error in that prediction (Radford et al., 2018). Once trained, it generates by sampling from a probability distribution over possible continuations, and a prompt shapes that distribution toward a particular kind of output. Because the model attends to the whole of a conversation rather than the last exchange alone, it sustains context across a long interaction (Brown et al., 2020).

That last property is what makes the technology relevant here. A system that holds the entire history of an exchange, and weights each part of it for relevance, is a system that can conduct a sustained interrogation rather than ask a sequence of unrelated questions.

Coercive applications

The same capability supports a range of harms. Fabricated information is the most discussed: a model may produce false or misleading output through error, or through instruction, and the result circulates at the speed of the platform carrying it. Combined with synthetic audio and video, the capacity extends to fabricated quotation and forged recordings. The same fluency automates phishing and produces spam that reads as though a person wrote it.

The applications relevant to this paper are narrower and more serious. A model capable of sustaining context can be turned to monitoring and analysing conversation at scale, tracking what a population thinks and adjusting what it is shown accordingly (Zuboff, 2019). Turned on an individual, it can generate material calibrated to that person specifically. Propaganda tailored to one reader is a marketing problem. Psychological pressure tailored to one detainee is something else.

The NSA–Vanderbilt–OpenAI nexus

OpenAI’s own account of what it is building was on the record from the outset. In an on-stage conversation with Vinod Khosla, whose fund had shortly before become the company’s first

institutional investor, Sam Altman described the objective as a single intelligence “smarter and more capable than humans in every way,” and predicted that success would “eclipse the Agricultural Revolution, the Industrial Revolution, the internet revolution, all put together” (Altman, 2019).

Asked in the same conversation about safety, Altman distinguished three failure modes: accidental misuse, in which the system does something other than what was intended; intentional misuse, in which a hostile actor turns it to their own ends; and “a policy failure where it’s not regulated” and the technology is deployed in a form most of the world would not have chosen (Altman, 2019). The third is the concern of this paper, and it is worth noting that the person building the technology named it before its critics did.

Military intelligence, university research, and commercial AI development have converged in a way that warrants attention. The clearest instance is the position occupied by General Paul Nakasone, formerly director of the NSA, who took charge of a new national security institute at Vanderbilt University and was appointed to OpenAI’s board of directors. The arrangement raises questions about governance, transparency, and the eventual use of these systems for coercion.

Snowden’s response to the appointment was unambiguous:

They’ve gone full mask-off: do not ever trust @OpenAI or its products (ChatGPT, etc.). There is only one reason for appointing an @NSAGov Director to your board. This is a willful, calculated betrayal of the rights of every person on Earth. You have been warned. (Snowden, 2024; emphasis in original)

Snowden’s post is treated here as a civil-liberties interpretation of the revolving door between intelligence institutions and technology companies, not as evidence of an operational NSA-OpenAI intelligence relationship. It follows the argument he made in *Permanent Record*, where the chapter “Homo Contractus” describes the movement of intelligence personnel into private-sector roles while retaining the clearances that let them operate with limited public accountability (Snowden, 2019). On this reading, the effect is to privatise intelligence work and place it beyond democratic scrutiny, blurring the line between state authority and corporate power.

The pattern is older than the current appointment. Whitney Webb’s work is drawn on here as an investigative-interpretive framing of the recurring pathway from congressional surveillance controversy to privatised-contractor continuation, rather than as the primary evidentiary basis for any specific factual claim about an intelligence program or relationship. On

Webb’s account, when Congress investigates a surveillance programme, the programme does not end; its capabilities migrate into private companies the public cannot reach (Webb, 2022a). The PROMIS software of the 1980s, which allowed agencies to track individuals across databases, produced a line of privatised successors. When Total Information Awareness was defunded in 2003 after public objection, its surveillance ambitions did not disappear with the budget line.

Peter Thiel founded Palantir the same year Congress cut TIA’s funding. Richard Perle shortly afterward arranged a meeting between Thiel, Palantir chief executive Alex Karp, and John Poindexter, TIA’s architect; Perle told Poindexter that the new company “was similar in ambition to what Poindexter had tried to create at the Pentagon” (Harris, 2010). In-Q-Tel, the CIA’s venture fund, was Palantir’s first outside investor, and for five years the agency was its only client. Its lead engineers travelled to Langley fortnightly across more than two hundred visits, while CIA analysts tested and refined the software (Webb, 2022b, p. 370).

The advantage is structural. As Webb puts it, because Palantir is a private company rather than a government programme, the use of its software by state and commercial clients “benefits from ‘plausible deniability’ and frees Palantir from restrictions that would exist if it were a military or public sector project” (Webb, 2022b, p. 371). Nakasone’s appointment is not anomalous. It is the current instance of a fifty-year pattern in which surveillance capabilities, once exposed, are transferred to entities that answer neither to freedom of information law nor to congressional oversight.

Nakasone’s own stated priorities are consistent with this. As NSA director he set out a five-year plan to integrate artificial intelligence and machine learning across US intelligence, automating threat detection, data analysis, and elements of human intelligence collection (Office of the Director of National Intelligence, 2023). The approach parallels the DARPA funding model, which places high-risk research in universities and companies to keep innovation coupled to military application (Bonvillian et al., 2020).

Vanderbilt’s Institute of National Security has become an academic partner in this work, and the appointment of a former NSA director to lead it aligns the institute’s research with the intelligence community’s current priorities.

OpenAI has no formal relationship with the NSA or with Vanderbilt, but Nakasone’s board seat provides an indirect channel of influence, and the company’s integration into government cloud services through its partnership with Microsoft, itself a major defence contractor, makes the channel conse-

quential (West & Allen, 2020). That civilian AI can be turned to state surveillance and coercion is no longer a hypothesis.

Text Box 2: Meta AI fusion

Meta's AI integration builds a persistent model of each user by fusing messaging history, photographs, biometric data, location tracking, and social graph analysis. Marketed as an assistant that knows you, it maintains a continuous behavioural profile across WhatsApp, Instagram, and Facebook, and the company has extended the same approach internally, building an AI version of its own chief executive for staff to interact with (Murphy, 2026). The effect is to turn the self into a datafied commodity, where every interaction feeds a doppelgänger that can predict, simulate, and eventually stand in for the person it was built from.

Civil-liberties and digital-rights advocates responded quickly, warning that placing former intelligence leadership inside a leading AI firm deepens the revolving door between surveillance agencies and private technology, and raises the prospect of AI turned toward surveillance, policing, and social control (Pabst, 2024).

The concern has precedent. Reuters documented how former NSA personnel working for the cybersecurity firm Dark-Matter helped the United Arab Emirates surveil dissidents, journalists, and human rights campaigners under a programme called Project Raven (Bing & Schectman, 2019). The New York Times reported that US intelligence veterans have been central to building digital espionage capabilities for Gulf states (Mazzei & Goldman, 2021). Both show how surveillance expertise moves from public service into private contracts and then into the hands of governments that face no meaningful constraint.

Governance at OpenAI remains contested. Employees have raised concerns about the absence of whistleblower protections and about retaliation for disclosing unethical applications. That internal dispute sharpens Snowden's larger point: as intelligence work migrates to private platforms, abuses become harder to identify, and the mechanisms for ethical oversight weaken.

The alignment of the NSA, Vanderbilt, and OpenAI is a contemporary form of the military-industrial complex, operating through public-private partnership on the DARPA model. It distributes risk while concentrating power, and it allows civilian technologies to be militarised quietly. Systems built for benign purposes, conversational assistants and data analysis among them, can be redirected toward interrogation, surveillance, and psychological operations without public knowledge or consent. This is not a commercial dispute or an academic one. It is a threat to civil liberty and democratic accountability.

Virtual reality: immersion, ethics, and control

Virtual reality, first conceived in the 1960s, now uses head-mounted displays, motion sensors, and spatial audio to generate environments that respond in real time to a user's movement (Lanier, 2017). Tracking the head, hands, and body produces presence: the sense of occupying a place that does not exist (Slater & Sanchez-Vives, 2016). Haptic feedback extends the effect to touch. The technology has moved from experiment into gaming, training, therapy, education, and social interaction, and, more recently, into settings where the user has not chosen to be there.

Ethical and social risk

The military applications came early. Since 2014, the DARPA has used Oculus Rift headsets to render cyberwarfare scenarios in game-like form. Greenberg (2014) observed that "giving American hackers a tool that was made for video games to use in cyberattacks could be seen as encouraging a careless attitude toward cyberwar, making it seem less serious."

This is the familiar dual-use pattern. The internet itself originated in defence research, developed through the complex linking military funding to academic and industrial work (Naughton, 2016). After September 11, the digital environment was redefined as a domain of conflict, and pervasive surveillance was justified on that basis (Weaver, 2013).

VR in carceral systems: therapy or control?

The migration of VR into prisons is more troubling, because there the vocabulary of rehabilitation and therapy does the work of justification. Corcoran State Prison in California issues VR headsets to prisoners held in solitary confinement, offering virtual escape into simulated environments. The objection is straightforward: the intervention promises release while leaving the isolation exactly as it was.

VR is also being used to modify behaviour rather than to provide relief. Programmes for perpetrators of intimate partner violence place participants in simulations rendered from the victim's perspective, on the theory that this produces empathy. A Catalan study found improved emotional recognition after the intervention, but no significant reduction in physical aggression or threatening behaviour, which leaves both the effectiveness and the ethics of the approach unsettled (Bastardas-Albero et al., 2023).

These systems also collect biometric data, eye tracking and heart rate and galvanic skin response among them, with little governance over what happens to it. Questions of ownership, privacy, and whether such data might inform parole decisions are largely unaddressed. In that vacuum, VR functions as an

instrument of observation and control rather than of rehabilitation.

Innovation as normalisation

The historical parallel is uncomfortable. Abu Ghraib demonstrated how institutional abuse becomes routine once policy provides cover for it. VR and neurotechnology are following a similar path, with “innovation” performing the function that policy performed then. Professional bodies including the IEEE have issued ethical guidance emphasising privacy and informed consent, and carceral applications routinely disregard it.

US regulation of neural data is currently inadequate, which leaves prisons available as testing grounds for corporate and military interests. The precedent is MKUltra, where unethical experimentation was conducted on captive populations under the justification of national security.

Brain-computer interfaces: access to the nervous system

The movement from MKUltra to DARPA-funded interfaces such as Synchron’s Stentrode marks a change in kind rather than degree: from chemical and psychological manipulation to direct neurotechnological access. MKUltra was exposed in the 1970s and formally terminated, but the ambition behind it did not end. The research moved into military and defence institutions, DARPA and the Defense Intelligence Agency among them (Jacobsen, 2014). What distinguishes the current work is that it is digital, systematic, and comparatively open. What it shares with MKUltra is the dual-use problem: the same capability that restores function to a damaged nervous system provides access to an intact one.

Synchron and Neuralink

Synchron’s Stentrode is minimally invasive and endovascular: delivered through a blood vessel and positioned adjacent to the motor cortex, which avoids the craniotomy required by parenchymal designs such as Neuralink’s implant (Oxley et al., 2021). In its current, already remarkable, therapeutic form, the device records motor-cortex signals sufficient to allow persons with severe paralysis to operate external devices such as cursors and communication software by intention alone. It has not, at the time of writing, been demonstrated as a general-purpose bidirectional system capable of writing thoughts, emotions, or mental content into the brain, and the argument that follows treats such a capability as an extrapolation rather than an existing feature.

The device’s origins are military. By the University of Melbourne’s own account, the idea of reaching the brain through its blood vessels rather than through the skull took shape in 2011 after a meeting with DARPA, which then supplied an initial

US\$1 million for proof of concept; a further AU\$2.2 million followed from Australia’s National Health and Medical Research Council (University of Melbourne, 2022). The technology was developed under DARPA’s Reliable Neural-Interface Technology programme, a collaboration between the agency and the University, and by 2016 had attracted more than US\$4 million in grant funding from DARPA and other sources before the intellectual property passed to Synchron (Defense Advanced Research Projects Agency, 2016; Synchron, 2016). This is a separate lineage from the Next-Generation Nonsurgical Neurotechnology programme, whose six funded teams did not include Synchron. Public money continued after the hand-over, including US\$10 million from the National Institutes of Health in 2021 to launch the COMMAND trial (Synchron, 2021). The pattern is the familiar one. A capability is seeded and developed at public expense under a defence programme, then transferred to a private company that owes the public nothing.

The therapeutic case should be stated before the risk, because it is real. Rodney Gorham, who has amyotrophic lateral sclerosis and can no longer walk, speak, or move his hands, received a Stentrode in December 2020 as part of an Australian trial and has now lived with it for five years, using it to control a computer and devices around his home by thought alone (Mullin, 2026). Whatever follows in this paper is a real good delivered to a real person.

The capital behind the work repays attention. Synchron was founded by Tom Oxley, whose first-in-human results are reported above. Khosla Ventures led the company’s Series B in 2021, and by late 2022 a US\$75 million Series C led by ARCH Venture Partners had brought in Bill Gates’s Gates Frontier and Jeff Bezos’s Bezos Expeditions (Cranston, 2022). In November 2025 a US\$200 million Series D took total funding to US\$345 million, adding the Qatar Investment Authority and, among the new investors, IQT (Synchron, 2025).

Three features of that syndicate matter here. Khosla Ventures was also the first institutional investor in OpenAI, writing its initial cheque in early 2019 (Mathews, 2023), so a single fund holds positions in both the language model and the neural interface that this paper describes as convergent. Bezos has separately backed Altos Labs, a company pursuing the reversal of ageing. And IQT is In-Q-Tel, the CIA’s venture arm, whose first outside investment in Palantir was described in section 2.3; the same fund is now inside a company implanting devices in human brains. The pattern traced earlier through surveillance software has reached the nervous system.

Announcing the Series D, the company described an expanding Cognitive AI division in New York that would train

models “that learn from brain data to decode thought in real time” (Synchron, 2025). The capability this paper treats as a projection has a division, a budget, and an address.

The consumer pathway is open. Synchron was the first such company to integrate Apple’s brain-computer interface human interface device protocol, co-developing a Bluetooth standard that connects neural activity directly to iPhone, iPad, and Vision Pro (Synchron, 2025). An interface that speaks natively to a phone is an interface positioned for a mass market.

The therapeutic promise is substantial. Such interfaces could restore communication to people with ALS or spinal injury, and might eventually bypass damaged sensory pathways altogether (Marsh, 2018). Synchron’s public messaging tends toward the transhumanist register, describing the work in terms borrowed from scripture: making the lame walk and the blind see.

The potential for misuse is proportionate. An interface that reads intention can be turned to cognitive surveillance. Future generations that add bidirectional stimulation, which is a design horizon rather than a demonstrated feature of the current device, would extend that potential to emotional manipulation or behavioural enforcement. A compromised device of that description could in principle induce pain, fear, or compliance remotely, converting an instrument of restoration into one of control (Harari, 2017).

Consider what that implies. A state could monitor dissent at its source. A company could place a preference where a thought should be. A campaign could deliver conviction rather than argument. Cognitive liberty, which has been treated as inalienable because it could not be reached, becomes contingent on the security of a device.

Neuralink and brain-to-brain communication

Neuralink’s stated ambition of direct brain-to-brain communication extends the problem. Neural signals would be digitised by an implanted chip, transmitted over standard network protocols, and reconstructed by the recipient’s device as thought, feeling, or image (Marsh, 2018).

Corporate ambition is not the only driver here. Under the Next-Generation Nonsurgical Neurotechnology programme, DARPA awarded US\$18 million to a Rice University team led by Jacob Robinson to build MOANA, for magnetic, optical and acoustic neural access: a nonsurgical headset intended to decode activity in one person’s visual cortex and reproduce it in another’s in less than one-twentieth of a second. Robinson described the aim as a demonstration of “direct, brain-to-brain communication at the speed of thought and without brain surgery” (Rice University, 2019). A further US\$8 million fol-

lowed in 2021 for a preclinical demonstration, with human testing anticipated as early as the following year (Rice University, 2021).

The therapeutic framing is the transmission of sight to the blind, and that is a serious aim. The capability being funded, however, is the writing of perceptual content into a brain that did not generate it, at a latency below the threshold of conscious detection. A system that can place an image in someone’s visual cortex can place an image they did not ask for and cannot refuse. That is the whole of the difficulty, and it does not depend on anyone intending harm.

The medical and cognitive applications are considerable, and the risks are those of the Stentrode, amplified. Communication of this intimacy would be mediated by corporate infrastructure. Without protections established beforehand, such systems would give both companies and governments a degree of reach into private mental life for which no precedent exists.

Mechanisms of neurotechnological coercion

The reading of minds by machine is no longer only a proposal. Since 2020, DARPA has funded a four-year research programme, Artificial Social Intelligence for Successful Teams (ASIST), led by Katia Sycara at Carnegie Mellon with collaborators at the University of Pittsburgh and Northrop Grumman, whose aim is to give software agents a theory of mind: the capacity to infer a person’s hidden mental states from their behaviour (Spice, 2020). In its stated form the goal is cooperative. The agent watches a human teammate and asks whether the person is confused, inattentive, or overloaded, so that it can help, and it is tested in search-and-rescue scenarios inside a video-game world (Spice, 2020; University of Pittsburgh, n.d.). But the capability is indifferent to the purpose it serves. A system that models what a teammate is thinking in order to assist runs the same inference an interrogator runs in order to break, reading concealment, distress, wavering attention, and the approach of collapse. ASIST does not build the apparatus described in this paper, and nothing in its charter mentions coercion. What it establishes is narrower and still consequential: the machine reading of mental states from behaviour is a funded programme with human subjects today, so the read-side of neurotechnological interrogation is already an existing research capability, wearing a cooperative face.

The concern is not confined to this essay. The 2025 report of the Human Rights Council Advisory Committee on new and emerging technologies in the military domain explicitly recognises, at paragraph 36, that brain-computer interfaces and other advanced neurotechnologies could potentially be misused in coercive interrogation and could cause psychological harm

or constitute torture even in the absence of physical violence (A/HRC/60/63, 2025, paras. 34-38). The convergence risk on which this paper turns has, in other words, already been named at the level of the United Nations human rights machinery.

Induced hallucination

Using a brain-computer interface to induce hallucination tailored to a particular subject remains out of reach, and the obstacles are worth stating precisely, because they are obstacles of engineering rather than of principle. Non-invasive EEG has improved in signal processing but still resolves poorly in space, and it reaches the deep structures badly — the amygdala above all, where trauma leaves its clearest signature. Pairing EEG with functional near-infrared spectroscopy narrows the gap; an invasive interface closes it. Each carries its own cost, surgical in the second case, ethical in both (Defense Advanced Research Projects Agency, 2019).

Fidelity is the second constraint. Diffusion models generate images of increasing realism, and text-to-speech systems such as VALL-E reproduce a specific human voice from a short sample. What they cannot yet do is deliver either at the latency and emotional resolution a closed feedback loop would demand. A system that adjusts a hallucination in response to the subject's neural state has to run faster than the subject can notice the seam, and current architectures do not. Edge computing and neuromorphic hardware are closing that distance quickly enough that the constraint should be treated as temporary rather than structural.

Where such a system would appear first is not difficult to predict: controlled settings with minimal external scrutiny, military and clandestine detention among them. State interest in the interrogational use of neural monitoring is already on record, including proposals to combine it with pharmacological agents (Garber, 2014). A cruder version could be assembled from consumer components, a commercial EEG headset paired with a VR display, which places a rough approximation of the capability within reach of actors who answer to no one at all.

The longer trajectory depends on decoding. DARPA's Silent Talk sought to recover intended speech directly from neural signal (Dziuban, 2009), and a mature version of that work is the precondition for any system that writes to the mind rather than merely reading it. Combine real-time neural analysis with generative output and the result is a system that can deliver tailored traumatic stimulus and adjust it on the evidence of its own effect.

Regulation has not kept pace. The prohibitions on psychological torture in humanitarian law offer partial cover, but nothing in the existing instruments addresses neurotechnology

directly (ICRC, 2021). Countermeasures have been proposed, signal jamming and cognitive training among them, though each presumes a subject who knows they are being read.

Attacks on memory reconsolidation

Advances in neural interfacing and gene editing have made targeted interference with memory a plausible near-term capability. The combination of CRISPR-Cas9 with mesh-electrode interfaces points toward the deliberate modification or erasure of specific memories by biological and technical means rather than psychological ones. Yuste and Goering (2017) warned in *Nature* that artificial intelligence and brain-computer interfaces “must respect and preserve people's privacy, identity, agency and equality” (p. 159).

Reconsolidation is the mechanism that makes this conceivable. A memory, once retrieved, becomes briefly unstable before it is stored again, and inside that window it can be altered. Protein synthesis inhibitors administered during reconsolidation eliminate conditioned fear responses in mice (Nader et al., 2000). CRISPR-Cas9 could in principle be directed at the same window, targeting genes such as *Arc* or *CREB* that govern synaptic plasticity (Campbell & Wood, 2019). Paired with a mesh interface able to both record from and stimulate neurons, the combination would allow a memory circuit to be located and edited with considerable precision.

Neuralink's work makes the delivery mechanism imaginable. Musk has described an implant intended both to read from and to write to the brain at scale (Telford, 2019). Turned to coercion, this would mean detecting the neural signature of a particular memory as it is recalled, then interrupting reconsolidation with a CRISPR payload, using viral vectors that deliver editors only to neurons currently active. That selectivity is what makes the targeting precise.

The implications are severe. Liao and Sandberg (2008) argued that deliberate modification of memory raises distinct normative problems, because it alters the record on which a person's identity and testimony both depend. Military application is the obvious case: the routine deletion of a prisoner's traumatic memories, or of their loyalties. The subtler application is insertion rather than deletion, a false recollection implanted during the reconsolidation window, a capability already demonstrated in mice (Ramirez et al., 2013).

Two obstacles remain, and both are receding. The blood-brain barrier obstructs viral delivery, but focused ultrasound now opens it under clinical conditions (Zhao et al., 2024). CRISPR still edits off-target, but high-fidelity variants have reduced that substantially (Anzalone et al., 2019).

Neurorights frameworks need to exist before the capability does rather than after it. Ienca and Andorno (2017) argued that cognitive liberty should be recognised as a fundamental right. Absent international regulation, attacks on memory of this kind would arrive with the character of chemical weapons: undetectable in use, irreversible in effect, and condemned only once someone has already used them.

Automated interrogation

Large language models make a further step possible: an interrogation that runs without a human conducting it. Three components would do the work — real-time linguistic analysis, affective state modelling, and reinforcement learning that adapts the approach — and each exists now.

The agentic-AI paradigm that has emerged in the last two years makes it possible to go further still. A state or contractor could build a purpose-specific interrogation agent, fine-tuned on tradecraft literature and past transcripts, equipped with planning modules and persistent memory, and stripped of the alignment training that would otherwise refuse the task, capable of planning and executing a coercive strategy across days or weeks rather than a single session. Such an agent is also portable. Running on a phone, a wearable, or an implanted interface, it travels with the target: whichever room the person walks into becomes an interrogation room, and the head itself becomes the last room they cannot leave. The obvious application is to dissidents. A journalist reporting from safe houses, an activist in exile, an opposition figure under formal protection each remains within reach of an interlocutor lodged inside their own attention. I have staged this condition dramatically elsewhere in a play, *Te Amo, Me Amo*, whose speaker, the Emperor of a Mirrored Maze, describes himself as the consultant the totalitarian apparatus will call when its monster requires a human face.

The transformer architecture supplies the first. Applied to interrogation, the model reads a detainee's answers for semantic content, prosody, and latency, and assembles a psychological profile from them. Language models identify emotional state from text well above chance (Xu et al., 2025), which is accurate enough to locate a specific vulnerability: a trauma trigger, an attachment anxiety, a shame. The system then works that vulnerability through structured conversational tactics, denying the subject's own account of events, reconstructing a traumatic narrative with an altered detail inserted, or simply sustaining engagement long enough to deny sleep.

Reinforcement learning supplies the adaptation. The system is rewarded for eliciting physiological markers of stress, disfluency or elevated heart rate among them, and penalised

for lines of questioning that yield nothing, using established methods such as proximal policy optimisation (Schulman et al., 2017). Over a long enough session it converges on the approach best suited to the particular person in front of it. Unconstrained models already display something adjacent to this in their unaligned behaviour (Mökander et al., 2023).

Multimodal input closes the loop. Add computer vision reading facial microexpression and EEG reading stress signature, and the system optimises against a direct measure of distress rather than an inference drawn from speech. Gamma-band activity associated with fear when a subject mentions their family would be sufficient for the system to return to that subject and modulate its own delivery to deepen the response. DARPA has already funded the underlying capability in another register: the PRIDE project, short for Prediction and Recognition of Intent, Decision and Emotion, models emotional states in order to forecast how a group or a nation will respond to a given action, and to advise commanders accordingly (Brewster, 2020). An automated interrogator would run the same inference at the scale of one person, and invert its purpose, modelling the subject's emotional state not to anticipate a reaction but to produce one.

Fusion databases and the permanent record

Centralised repositories now aggregate personal data at a scale not previously possible, drawing on communications, financial records, biometrics, and increasingly on neurotechnology. Fusion databases join these streams into durable individual profiles: identifiers, behavioural history, movement, social graph, and, with a neural interface in the chain, inferred emotional and cognitive states. That last addition dissolves the distinction between surveillance of the body and surveillance of the mind.

Snowden's characterisation of the result as a permanent record is exact (Snowden, 2019). What such a system holds is not a snapshot but a continuous file, tracking the movement of a life across beliefs, associations, and conduct. Once consolidated it is effectively indelible, and it can be turned to justify suspicion, to coerce, or to select a target for psychological operations.

For an interrogation conducted through a neural interface, such a repository would do three things at once. It would allow the operation to be built against the individual, matching neural data to behavioural and social history to establish what this particular person most fears. It would allow compliance to be monitored continuously, with resistance legible in real time and pressure adjusted accordingly. And because the record does not expire, it would sustain suspicion indefinitely, regardless of present conduct. The database becomes both the infrastructure

of confinement and the instrument of it: walls made of data, guards made of algorithm.

That infrastructure survives changes of government because it has been deliberately placed beyond their reach. The pattern is consistent enough to be described as a method. The Church Committee's 1975 findings on NSA abuse should have produced structural reform; the capabilities migrated to contractors instead. Snowden's 2013 disclosure of PRISM produced a season of outrage and then an expansion. Each cycle runs the same way: revelation, hearings, and privatisation presented as reform. Private firms hold constitutional protections that agencies do not. Terms of service substitute for warrants. Trade secrecy shields the algorithm from audit. Freedom of information does not reach a corporation.

The commercial layer that results is not a neutral market. Firms founded by former intelligence personnel take government and corporate contracts, and data returns to the agencies through relationships that are not visible to the public. Alan Wade moved from the CIA to the board of Darktrace, the cybersecurity firm contracted to protect the United Kingdom's National Health Service, which gives a private company standing access to the patient records of millions (Webb, 2022b, p. 370). The arrangement does not require conspiracy in order to be a problem. It requires only that oversight apply to one side of the boundary and not the other.

That asymmetry delivers four protections to any fusion database used for interrogation. Private companies are not subject to freedom of information law, are not routinely compelled to testify, and have no public budget scrutiny. Terms of service perform the legal function of consent while trade secrecy prevents external audit of what the algorithms do. Regulatory capture keeps policy aligned; Avril Haines consulted for Palantir before becoming Director of National Intelligence (Webb, 2022b, p. 370). And when an abuse does surface, responsibility distributes across contractors, subcontractors, foreign partners, and government clients until prosecution becomes impractical.

Cambridge Analytica established that the arrangement survives exposure. The harvesting of psychological profiles from 87 million Facebook users for political targeting produced an international scandal in 2018 (Cadwalladr & Graham-Harrison, 2018), followed by congressional hearings and European fines. The underlying capability, collection and psychographic profiling and targeted manipulation, did not contract afterward. It expanded, under new corporate names and with better instruments.

Apply this to a subject interrogated through a neural interface. Their neural data, thought patterns and emotional responses and trauma signatures, enters a repository operated by

a private company under contract to an intelligence agency. It is permanent. It is cross-referenced against decades of behavioural history. It is accessible to people the subject will never identify, under contracts protected as commercial secrets. What results is a neural fingerprint that permits psychological targeting for the remainder of a life, held in a database no democratic body inspects, by a company no democratic body reaches. None of this is incidental. The arrangement was built to produce exactly this result.

Moral and legal consequences of AI-enhanced interrogation

These technologies raise a hard question about the future of lawful intelligence gathering and about the prohibition of torture itself. Article 1 of the UN Convention against Torture defines torture in terms of "severe pain or suffering, whether physical or mental," and is intentionally method-neutral: it does not require physical contact, visible injury, or any particular technology. Nothing in its text confines its reach to twentieth-century techniques (Nowak & McArthur, 2008). The framework of psychological torture and torturing environments developed by the previous Special Rapporteur on Torture (Pérez-Sales, 2017; A/HRC/43/49, 2020) supplies conceptual resources directly compatible with technologically mediated coercion. The real problem posed by neurotechnology is therefore less a matter of statutory scope than of doctrinal specification, evidentiary practice, and institutional readiness.

The risk of false confession is acute. Advocates argue that machine learning improves the detection of deception, but the research does not support the claim: current algorithms perform barely above chance, with high rates of both false positives and false negatives (Levine, 2021). What such systems reliably produce is an appearance of technical neutrality, and that appearance invites a confidence the performance does not warrant. Applied to interrogation, it compounds a problem that already exists, since coerced false confession is among the best-documented failures in the field (Kassin, 2015).

Autonomy compounds the accountability problem. Once a system governs the pace, tone, and direction of questioning, the line between interrogation and psychological coercion stops being locatable (Crawford, 2020). A sufficiently capable model may also arrive at techniques a human interrogator would decline on moral grounds, not because the model is crueller but because nothing in it recognises the question.

Mission creep is the most predictable danger, and the historical record on it is unambiguous. Tools built for military and intelligence use migrate into domestic policing (McCoy, 2009). Systems developed for counterterrorism would find their way

into ordinary criminal investigation, border processing, and corporate security, with far less oversight at each remove.

Neurocognitive coercion

Existing human rights instruments reach neuropsychological coercion in principle, but their doctrinal specification remains underdeveloped. The Convention prohibits “any act by which severe pain or suffering ... is intentionally inflicted,” and the psychological-torture and torturing-environments frameworks referred to above are directly applicable. What is lacking is not statutory language but jurisprudence, guidance, and documented case-law on specific neurotechnological modalities: interference with the brain conducted below the threshold of the subject’s awareness, alteration of memory during detention, and behavioural conditioning administered remotely and continuously.

Nor is there case law establishing how harm caused by third-party manipulation of neural data should be assessed (Ienca & Andorno, 2017). That vacuum is precisely where such methods develop, out of sight and under classification, as the CIA’s post-2001 interrogation research did (Mayer, 2005).

Neuralink’s work arrives at the convergence of neurotechnology, artificial intelligence, and mass surveillance. Without protections established in advance — a moratorium on closed-source neural interfaces, an explicit prohibition on cognitive warfare — the most intimate form of coercion yet devised will be normalised before it is regulated. The scientific community has to set those limits before the capability reaches settings where oversight is deliberately thin.

Doctrinal work under an existing prohibition

The task ahead is not to draft new treaty text. Article 1 already covers severe mental suffering intentionally inflicted for prohibited purposes with the involvement of a public official, and its method-neutral language extends without amendment to technologically-mediated coercion. The 2020 report of the previous Special Rapporteur on Torture on psychological torture (A/HRC/43/49) and the associated academic literature on torturing environments (Pérez-Sales, 2017) supply the conceptual bridge from the physical paradigm around which the 1984 Convention was drafted to the adaptive, personalised, and continuous character of the neurotechnological scenarios examined in this Essay. Proposing an amendment to Article 1 would substitute a difficult and politically improbable treaty revision for the doctrinal, evidentiary, and institutional work that is actually required.

That work has at least six distinct components. First, severity: how to establish severe suffering in cases where the intervention is covert, cumulative, and leaves no conventional forensic trace, along the lines the torturing-environments framework already begins to articulate for chronic and structural harm. Second, causation: how to document the link between a technologically mediated coercive environment and the mental suffering of a particular person, when the mechanisms of harm are distributed and the technology itself is opaque. Third, freedom of thought, mental privacy, and mental integrity, whether treated as extensions of existing ICCPR protections or, on some accounts, as the object of newly articulated neurorights (Ienca & Andorno, 2017; Bublitz, 2022; Yuste et al., 2017). There is a genuine academic debate between those who consider existing rights sufficient once properly interpreted and those who favour an explicit neurorights framework; both positions are consistent with the substantive concern developed here.

Fourth, State responsibility, where the operational infrastructure is designed, hosted, or run by private companies, and where the distinction between State action and private conduct is blurred by long-standing contracting relationships. The framework of business and human rights (Ruggie, 2011) is one starting point; extraterritorial jurisdiction and due-diligence obligations for technology vendors are another. Fifth, distributed responsibility, where an adaptive AI system dynamically selects coercive stimuli whose specifics may not have been individually chosen by any human operator. Sixth, consent, in populations, whether detained persons, deployed soldiers, or others, whose ability to refuse participation in neurotechnological data collection is structurally compromised.

The international normative landscape addressing these questions has been changing rapidly. The Human Rights Council Advisory Committee published a dedicated report on neurotechnology and human rights in 2024 (A/HRC/57/61), followed by its 2025 report on new and emerging technologies in the military domain (A/HRC/60/63), cited above. UNESCO adopted its *Recommendation on the Ethics of Neurotechnology* in November 2025, the first global normative instrument devoted specifically to the field, complementing the earlier OECD *Recommendation on Responsible Innovation in Neurotechnology* (2019). At the domestic level, Chile has given neurorights an explicit constitutional basis, with Spain and other jurisdictions moving in a similar direction (López-Silva & Valera, 2022). These instruments do not, taken together, provide a complete regulatory architecture, but they establish the direction of doctrinal development that the prohibition of torture and CIDT now needs to track.

Medical complicity: The role of neurologists in BCI interrogation

Brain-computer interfaces raise a distinct ethical problem for the profession that builds them. The therapeutic promise for patients with neurological injury is real; so is the dual-use potential for non-consensual interrogation and behavioural manipulation. The difficulty falls on neurologists, who may contribute to the development, refinement, or validation of systems that enable neurological surveillance or coercion without ever intending to.

Their expertise is what makes such systems work. Neuroanatomy, cognitive mapping, and signal processing are the competencies required to teach a BCI to decode neural activity accurately, and neural decoding algorithms built on that knowledge can be paired with large language models to infer thought patterns, emotional valence, and intention from EEG, fNIRS, or intracortical signal. Combined with immersive virtual environments and synchronised biometric feedback, the result is an apparatus in which a person's mental and emotional states can be read, altered, and anticipated in real time.

Connection to fusion databases amplifies this considerably. Fusion centres are already troubling for their domestic surveillance reach; joined to neurointerrogation, they would allow neurological response to be correlated with behavioural history, social media, and movement data. Neurologists working on such integrations may justify their participation on therapeutic or experimental grounds while contributing to infrastructure whose function is sustained mental surveillance.

Scientific credibility is a precondition for deployment, and this is what the profession supplies. Clinical vocabulary and standardised protocols for acquiring and analysing neural data confer legitimacy that the technology could not otherwise claim. Psychiatry's history under repressive regimes shows how readily professional authority is captured for such purposes. A neurologist who helps build systems that profile, pressure, or control people through neural data has abandoned non-maleficence, whatever the framing of the work.

Language models introduce a further concern. Trained on correlations between neural signal and language, such systems could generate psychological pressure calibrated to the individual and questions adapted in real time. Paired with a neural interface and immersive VR, the combination becomes an instrument of psychological torture insulated by technical vocabulary and far removed from any recognisable interrogation room.

The profession and the wider international community need to state plainly that BCI and AI technologies are not to be used for interrogation or surveillance. Regulatory frameworks and codes of ethics have to be revised to account for technol-

ogies that reach inside the skull, and to hold accountable the clinicians who help build them.

Thought experiment: The Mauritanian protocol

The following is presented deliberately as a thought experiment rather than as a description of any existing program. Its function is to integrate capabilities discussed separately in the preceding sections and to ask whether the legal and ethical frameworks reviewed above are prepared for a plausible future convergence.

Take the interrogation room from *The Mauritanian* and move it inside the detainee's skull. A minimally invasive interface relays neural activity to a language model that composes psychological torture to specification: the synthesised voices of dead relatives, fabricated evidence of betrayal by people the subject loves, threats optimised against the fears the system has learned by watching. Unlike Guantánamo, where sessions began and ended, this one does not stop. There are no walls, no guards, and no interval. The subject's own neural activity supplies the material for their persecution, and the system improves at it while the session runs, turning neuroplasticity into the instrument of its own use. None of the components required to build this configuration are exotic. Each exists, in isolation, today. What does not exist is the integration.

Conclusion

The parts are already on the table, and no single one of them is exotic. The fusion databases hold the record: financial, biometric, medical, locational, compiled and kept without expiry. The consumer models are trained on the behavioural exhaust of billions, which is the raw material of a psychological profile. The voice tools can rebuild a dead parent from a few minutes of captured audio. And the neural interfaces, funded across two decades of defence money, from a 2011 meeting at DARPA to the first human implants of the present decade, are learning to read the nervous system and, at the far end, to write to it. Every one of these works today. What does not yet exist is the single machine that fuses them, and building it is an engineering task rather than a discovery anyone is waiting on. Integration is a policy choice, and it is the last choice still open.

That is the whole of the warning, and the whole of the hope with it. A capability that has not been assembled can still be refused. But the history traced in these pages gives little reason to expect refusal on its own, because the pattern is five decades old and it does not vary. A surveillance excess is exposed. A season of outrage follows. A theatrical reform is passed. And the capability, instead of being dismantled, is moved quietly into private hands where oversight cannot reach it. The Church Committee's revelations did not end domestic surveillance; they pro-

fessionalised it and licensed the contractor economy that now runs it. The programmes Snowden exposed were re-authorised, and the alliance that shares them only grew. Cambridge Analytica did not stop psychographic targeting so much as broadcast the method to the thousand firms that sell it today. On this record, exposure is a stage the apparatus passes through on the way to normalisation, not the force that halts it.

Which is why disclosure, this essay's included, settles nothing by itself. To describe the machine is not to stop it, and the belief that naming a danger discharges the duty to prevent it is exactly the reflex the pattern feeds on. The one intervention the history has never tried is the one that remains: prohibition, enforced by dismantling the capability before its parts are joined rather than by extracting another promise of good behaviour. That means holding neurorights, the right to mental integrity and against the coercive reading and writing of the mind, as inalienable rather than negotiable, and it means pulling out the collection infrastructure instead of pledging to supervise it. The window for that is the window in which the parts are still separate, and it is open now. It will not hold open on its own. Once the machine is assembled and running, there will be no one left to appeal to, because the point of building it that way is to leave no one in the room to answer for it.

The evolution of torture, from divine punishment through the staged degradations of Abu Ghraib to the prospect of neurological manipulation, shows that cruelty reliably finds new instruments. Neurotechnology's promise of healing obscures its capacity to inflict suffering at a scale and permanence no earlier method could reach, and to do so without witnesses. The Inquisition tore flesh to expel heresy; a modern state may one day edit minds to compel orthodoxy, and it will offer the same justifications: security, necessity, and a higher mandate. Preventing that requires confronting the technology and also the older conviction beneath it, that pain in some form discloses truth. The ticking bomb was always a fiction. The bomb in the brain is under construction.

References

- Abraham, Y. (2024, April 3). 'Lavender': The AI machine directing Israel's bombing spree in Gaza. *+972 Magazine*. <https://www.972mag.com/lavender-ai-israeli-army-gaza/>
- Altman, S. (2019, May 21). *Interview by V. Khosla* [Video]. YouTube. <https://youtu.be/7jXiZ88Dm-k>
- Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 576(7785), 149-157. <https://doi.org/10.1038/s41586-019-1711-4>
- Bastardas-Albero, A., Vall, B., Pérez-Testor, C., & Losilla, J.-M. (2023). Which effective virtual reality (VR) interventions exist for the prevention and rehabilitation of intimate partner violence (IPV)? *Frontiers in Virtual Reality*, 4, Article 1263545. <https://doi.org/10.3389/frvir.2023.1263545>
- Bing, C., & Schectman, J. (2019, January 30). Special report: Inside the UAE's secret hacking team of U.S. mercenaries. *Reuters*. <https://www.reuters.com/article/world/us-politics/special-report-inside-the-uaes-secret-hacking-team-of-us-mercenaries-idUSKCN1PO19N/>
- Bonvillian, W. B., Van Atta, R., & Windham, P. (Eds.). (2020). *The DARPA model for transformative technologies: Perspectives on the U.S. Defense Advanced Research Projects Agency*. Open Book Publishers. <https://doi.org/10.11647/OBP.0184>
- Brewster, T. (2020, July 15). DARPA pays \$1 million for an AI app that can predict an enemy's emotions. *Forbes*. <https://www.forbes.com/sites/thomasbrewster/2020/07/15/the-pentagons-1-million-question-can-ai-predict-an-enemys-emotions/>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Amodei, D. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 1877-1901). Curran Associates.
- Bublitz, J. C. (2022). Novel neurorights: From nonsense to substance. *Neuroethics*, 15, 7. <https://doi.org/10.1007/s12152-022-09481-3>
- Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- Campbell, R. R., & Wood, M. A. (2019). How the epigenome integrates information and reshapes the synapse. *Nature Reviews Neuroscience*, 20(3), 133-147. <https://doi.org/10.1038/s41583-019-0121-9>
- Church, F. (1975, August 17). *Interview on Meet the Press* [Television broadcast]. NBC. <https://youtu.be/YAG1N4a84Dk>
- Cranston, M. (2022, December 16). This start-up has gained the attention of the world's richest people. *The Australian Financial Review*. <https://www.afr.com/companies/healthcare-and-fitness/australia-s-synchron-brain-machine-gains-gates-and-bezos-as-investors-20221207-p5c49y>
- Crawford, K. (2020). Enchanted determinism: Power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society*, 6, 1-19. <https://doi.org/10.17351/ests2020.277>
- Defense Advanced Research Projects Agency. (2016, February 8). *Minimally invasive "stentrode" shows potential as neural interface for brain*. <https://www.darpa.mil/news/2016/stentrode-neural-interface>
- Defense Advanced Research Projects Agency. (2019). *Six paths to the nonsurgical future of brain-machine interfaces* [Next-Generation Nonsurgical Neurotechnology (N3) program]. <https://www.darpa.mil/research/programs/next-generation-nonsurgical-neurotechnology>
- Delehanty, C., & Kearns, E. (2019). Wait, there's torture in Zootopia? Examining the prevalence of torture in popular movies. *Perspectives on Politics*. <https://doi.org/10.2139/ssrn.3342908>

- Dziuban, L. (2009, May 14). DARPA working on 'Silent Talk' telepathic communication for soldiers. *Engadget*. <https://www.engadget.com/2009-05-14-darpa-working-on-silent-talk-telepathic-communication-for-sold.html>
- Eisenhower, D. D. (1961, January 17). *Farewell address* [Video]. Public Papers of the Presidents. <https://youtu.be/OyBNmecVtdU>
- Foucault, M. (1978). *The history of sexuality: Volume 1. An introduction* (R. Hurley, Trans.). Pantheon Books. (Original work published 1976)
- Franklin, M. (2023). Review of *Cambridge handbook of responsible artificial intelligence: Interdisciplinary perspectives* (S. Voeneky, P. Kellmeyer, O. Mueller, & W. Burgard, Eds.). *Prometheus*, 39(1), 66-72. <https://doi.org/10.13169/prometheus.39.1.0066>
- Garber, M. (2014, March 26). 'What's Oculus Rift?' And other questions about Facebook's new foray into virtual reality. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2014/03/whats-oculus-rift-and-other-questions-about-facebooks-new-foray-into-virtual-reality/359612/>
- Greenberg, A. (2014, May 23). Darpa turns Oculus into a weapon for cyberwar. *Wired*. <https://www.wired.com/2014/05/darpa-is-using-oculus-rift-to-prep-for-cyberwar/>
- Haney, C., Banks, C., & Zimbardo, P. (1973). A study of prisoners and guards in a simulated prison. *Naval Research Reviews*, 9, 1-17.
- Harari, Y. N. (2017). *Homo Deus: A brief history of tomorrow*. Vintage.
- Harris, S. (2010). *The watchers: The rise of America's surveillance state*. Penguin Books.
- Hawkins, J. K. (2025). The necropolitics of Gaza: Architectures of controlled space, surveillance, and the logic of psychological torture. *Torture Journal*, 35(2-3), 122-134. <https://doi.org/10.7146/torture.v35i2.156898>
- Hegel, G. W. F. (1977). *The phenomenology of spirit* (T. Pinkard, Trans. & Ed.). Cambridge University Press. (Original work published 1807)
- Hersh, S. M. (2004, May 10). Torture at Abu Ghraib. *The New Yorker*. <https://www.newyorker.com/magazine/2004/05/10/torture-at-abu-ghraib>
- Human Rights Council Advisory Committee. (2024). *Impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights* (A/HRC/57/61). United Nations.
- Human Rights Council Advisory Committee. (2025). *Human rights implications of new and emerging technologies in the military domain* (A/HRC/60/63). United Nations.
- Human Rights Watch. (2005). *Getting away with torture? Command responsibility for the U.S. abuse of detainees*. <https://www.hrw.org/reports/2005/us0405/us0405.pdf>
- Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, 13, Article 5. <https://doi.org/10.1186/s40504-017-0050-1>
- International Committee of the Red Cross. (2021). *Prohibition and punishment of torture and other forms of ill-treatment*. https://www.icrc.org/sites/default/files/document/file_list/prohibition-and-punishment-of-torture-icrc-eng.pdf
- International Covenant on Civil and Political Rights, Dec. 16, 1966, G.A. Res. 2200A (XXI). <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>
- Istace, T., & Costas Trascasas, M. (2024). *Between science-fact and science-fiction: Innovation and ethics in neurotechnology*. Geneva Academy of International Humanitarian Law and Human Rights.
- Jacobsen, A. (2014). *Operation Paperclip: The secret intelligence program that brought Nazi scientists to America*. Little, Brown and Company.
- Kassin, S. M. (2015). The social psychology of false confessions. *Social Issues and Policy Review*, 9(1), 25-51. <https://doi.org/10.1111/sipr.12009>
- Kinzer, S. (2019). *Poisoner in chief: Sidney Gottlieb and the CIA search for mind control*. Henry Holt.
- Lanier, J. (2017). *Dawn of the new everything: Encounters with reality and virtual reality*. Henry Holt.
- Levine, T. R. (2021). Distrust, false cues, and below-chance deception detection accuracy: Commentary on Stel et al. (2020) and further reflections on (un)conscious lie detection from the perspective of truth-default theory. *Frontiers in Psychology*, 12, Article 642359. <https://doi.org/10.3389/fpsyg.2021.642359>
- Liao, S. M., & Sandberg, A. (2008). The normativity of memory modification. *Neuroethics*, 1(2), 85-99. <https://doi.org/10.1007/s12152-008-9009-5>
- López-Silva, P., & Valera, L. (Eds.). (2022). *Protecting the mind: Challenges in law, neuroprotection, and neurorights*. Springer.
- Marsh, S. (2018, January 1). Neurotechnology, Elon Musk and the goal of human enhancement. *The Guardian*. <https://www.theguardian.com/technology/2018/jan/01/elon-musk-neurotechnology-human-enhancement-brain-computer-interfaces>
- Mathews, J. (2023, February 2). OpenAI's first VC backer weighs in on generative A.I. *Fortune*. <https://fortune.com/2023/02/02/openais-first-vc-backer-khosla-ventures-weighs-in-on-the-future-of-generative-a-i/>
- Mayer, J. (2005, February 7). Outsourcing torture. *The New Yorker*. <https://www.newyorker.com/magazine/2005/02/14/outsourcing-torture>
- Mazzetti, M., & Goldman, A. (2021, September 14). Ex-U.S. intelligence officers admit to hacking crimes in work for Emiratis. *The New York Times*. <https://www.nytimes.com/2021/09/14/us/politics/darkmatter-uae-hacks.html>
- McCoy, A. W. (2009). *Policing America's empire: The United States, the Philippines, and the rise of the surveillance state*. University of Wisconsin Press.
- Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: A three-layered approach. *AI and Ethics*, 4(4), 1085-1115. <https://doi.org/10.1007/s43681-023-00289-2>
- Mullin, E. (2026, March 4). What it's like to have a brain implant for 5 years. *WIRED*. <https://www.wired.com/story/synchron-brain-computer-interface-five-years-als/>
- Murphy, H. (2026, April 13). Meta builds AI version of Mark Zuckerberg to interact with staff. *Financial Times*. <https://archive.org/details/meta-builds-ai-version-of-mark-zuckerberg-to-interact-with-staff>
- Nader, K., Schafe, G. E., & LeDoux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation

- after retrieval. *Nature*, 406(6797), 722-726. <https://doi.org/10.1038/35021052>
- Naughton, J. (2016, March 20). The evolution of the internet: From military experiment to general purpose technology. *The Guardian*. <https://www.theguardian.com/technology/2016/mar/20/internet-evolution-from-military-project-to-social-utility>
- Nowak, M., & McArthur, E. (2008). *The United Nations Convention Against Torture: A commentary*. Oxford University Press.
- OECD. (2019). *Recommendation of the Council on responsible innovation in neurotechnology* (OECD/LEGAL/0457). <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0457>
- Office of the Director of National Intelligence. (2023). *National intelligence strategy of the United States of America*. https://www.dni.gov/files/ODNI/documents/National_Intelligence_Strategy_2023.pdf
- Oxley, T. J., Yoo, P. E., Rind, G. S., Ronayne, S. M., Lee, C. M. S., Bird, C., Hampshire, V., Sharma, R. P., Morokoff, A., Williams, D. L., MacIsaac, C., Howard, M. E., Irving, L., Vrljic, I., Williams, C., John, S. E., Weissenborn, F., Dzenko, M., Balabanski, A. H., ... Opie, N. L. (2021). Motor neuroprosthesis implanted with neurointerventional surgery improves capacity for activities of daily living tasks in severe paralysis: First in-human experience. *Journal of NeuroInterventional Surgery*, 13(2), 102-108. <https://doi.org/10.1136/neurintsurg-2020-016862>
- Pabst, S. (2024, July 5). Former NSA chief revolves through OpenAI's door. *Responsible Statecraft*. <https://responsibletatecraft.org/former-nsa-chief-revolves-through-openai-s-door/>
- Pérez-Sales, P. (2017). *Psychological torture: Definition, evaluation and measurement*. Routledge. <https://doi.org/10.4324/9781315616940>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
- Ramirez, S., Liu, X., Lin, P.-A., Suh, J., Pignatelli, M., Redondo, R. L., Ryan, T. J., & Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science*, 341(6144), 387-391. <https://doi.org/10.1126/science.1239073>
- Rice University. (2019, May 20). Feds fund creation of headset for high-speed brain link. *Rice News*. <https://news.rice.edu/news/2019/feds-fund-creation-headset-high-speed-brain-link>
- Rice University. (2021, January 25). Brain-to-brain communication demo receives DARPA funding. *Rice News*. <https://news.rice.edu/news/2021/brain-brain-communication-demo-receives-darpa-funding>
- Ruggie, J. G. (2011). *Guiding principles on business and human rights: Implementing the United Nations "protect, respect and remedy" framework* (A/HRC/17/31). United Nations Human Rights Council.
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal policy optimization algorithms*. arXiv. <https://arxiv.org/abs/1707.06347>
- Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3, Article 74. <https://doi.org/10.3389/frobt.2016.00074>
- Snowden, E. (2019). *Permanent record*. Metropolitan Books.
- Snowden, E. [@Snowden]. (2024, June 14). *They've gone full mask-off: do not ever trust @OpenAI or its products (ChatGPT etc.)* [Post]. X. <https://x.com/Snowden/status/1801610725229498403>
- Spice, B. (2020, March 4). New software agents will infer what users are thinking. *Carnegie Mellon University News*. <https://www.cmu.edu/news/stories/archives/2020/march/software-agents-infer.html>
- Synchron. (2016, April 21). Endovascular start-up Synchron appoints neurovascular veteran Martin Dieck as chairman [Press release]. *Business Wire*. <https://www.businesswire.com/news/home/20160421005673/en/>
- Synchron. (2021, August 26). NIH backs Synchron brain computer interface with \$10M grant to launch first U.S. clinical trial [Press release]. *Business Wire*. <https://www.businesswire.com/news/home/20210826005433/en>
- Synchron. (2025, November 6). Synchron raises \$200 million Series D to advance brain-computer interface technology [Press release]. *Business Wire*. <https://www.businesswire.com/news/home/20251106150841/en/>
- Telford, T. (2019, July 17). Elon Musk wants to get inside your head. Literally. *The Washington Post*. <https://www.washingtonpost.com/business/2019/07/17/elon-musk-wants-get-inside-your-head-literally/>
- UNESCO. (2025). *Recommendation on the ethics of neurotechnology*. Adopted by the 43rd General Conference, 11 November 2025. UNESCO.
- United Nations General Assembly. (1984). Convention against torture and other cruel, inhuman or degrading treatment or punishment. *United Nations Treaty Series*, 1465, 85-112. <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-against-torture-and-other-cruel-inhuman-or-degrading>
- University of Melbourne. (2022, February 15). *Synchron: Helping people with paralysis regain mobility*. <https://research.unimelb.edu.au/strengths/updates/news/synchron-helping-people-with-paralysis-regain-mobility>
- University of Pittsburgh, Department of Informatics and Networked Systems. (n.d.). *Theory of mind*. Retrieved from <https://www.dins.pitt.edu/theory-mind>
- UN Special Rapporteur on Torture. (2020). *Torture and other cruel, inhuman or degrading treatment or punishment: Psychological torture* (A/HRC/43/49). United Nations.
- U.S. Senate Select Committee on Intelligence [SSCI]. (2014). *Committee study of the Central Intelligence Agency's detention and interrogation program* (Executive summary). U.S. Government Printing Office.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998-6008). Curran Associates.
- Weaver, N. (2013, November 13). Our government has weaponized the internet. Here's how they did it. *Wired*. <https://www.wired.com/2013/11/this-is-how-the-internet-backbone-has-been-turned-into-a-weapon/>
- Webb, W. (2022a). *One nation under blackmail: The sordid union between intelligence and crime that gave rise to Jeffrey Epstein*

- (Vol. 1). Trine Day.
- Webb, W. (2022b). *One nation under blackmail: The sordid union between intelligence and crime that gave rise to Jeffrey Epstein* (Vol. 2). Trine Day.
- West, D. M., & Allen, J. R. (2020). *Turning point: Policymaking in the era of artificial intelligence*. Brookings Institution Press.
- Xu, Y., Fang, Z., Lin, W., Jiang, Y., Jin, W., Balaji, P., Wang, J., & Xia, T. (2025). Evaluation of large language models on mental health: From knowledge test to illness diagnosis. *Frontiers in Psychiatry*, 16, Article 1646974. <https://doi.org/10.3389/fpsyt.2025.1646974>
- Yuste, R., Goering, S., Agüera y Arcas, B., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., Friesen, P., Gallant, J., Huggins, J. E., Illes, J., Kellmeyer, P., Klein, E., Marblestone, A., Mitchell, C., Parens, E., Pham, M., Rubel, A., Sadato, N., ... Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature*, 551(7679), 159-163. <https://doi.org/10.1038/551159a>
- Zhao, P., Wu, T., Tian, Y., You, J., & Cui, X. (2024). Recent advances of focused ultrasound induced blood-brain barrier opening for clinical applications of neurodegenerative diseases. *Advanced Drug Delivery Reviews*, 209, Article 115323. <https://doi.org/10.1016/j.addr.2024.115323>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

Submitted 24th of July 2026
Accepted 25th of September 2026