

# Resonance, Surveillance & Frontier AI Systems: Interdisciplinary Capabilities Assessment (V3.0)

## 1. Purpose

This assessment aims to rigorously map the capabilities, limits, and documented evidence around several frontier technologies that span surveillance, cyber operations, neurotechnology, and complex system behaviors. The goal is not to indulge in metaphysical interpretations, but to **ground the analysis in scientific, technical, and behavioral evidence**. We will examine what is demonstrably known, what is plausible given current research, what known misuses or adversarial exploits exist, and which claims remain unverified or speculative. Key domains include state-of-the-art spyware (e.g. *Pegasus*-class tools), advanced data fusion platforms (like Palantir), acoustic and electromagnetic (EM) influence techniques, emerging neurotechnologies (in light of new 2025 UN neuro-rights guidance), AI-assisted cyber operations, and the “system-of-systems” dynamics of interconnected devices (IoT) that can lead to emergent behaviors. By synthesizing findings across these areas, we also clarify where **Large Language Models (LLMs)** fit into – or do not fit into – these technological stacks, avoiding exaggeration of LLM capabilities while acknowledging documented misuse in cyber contexts. This investigation is evidence-based and interdisciplinary, intended to inform the Phoenix Covenant & CAM Initiative on how to distinguish between **documented reality, plausible risk, unsupported claims, and disconfirmed ideas** in these frontier tech areas.

## 2. Scope

### 2.1 Device Compromise & Spyware

**Pegasus-class spyware:** The assessment begins with advanced spyware exemplified by NSO Group’s **Pegasus** malware. Pegasus is a highly sophisticated mobile spyware that can be **remotely and covertly installed** on iOS and Android phones[\[1\]](#). As of late 2023, Pegasus operators could exploit zero-click vulnerabilities (requiring no user action) on iPhones up to iOS 16.6[\[1\]](#). Once installed, Pegasus effectively grants **complete and unrestricted access** to the device, turning it into a pocket spy. It can read text messages, **snoop on calls**, collect passwords, track GPS location, and harvest data from apps[\[1\]](#). Critically, it can also **activate the device’s microphone and camera** without the user’s

**CAM INITIATIVE**  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O’Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



knowledge, enabling real-time eavesdropping and visual surveillance<sup>[2][3]</sup>. In other words, an infected phone can be transformed into a 24/7 surveillance bug **without any visible indication to the target**<sup>[2]</sup>. The spyware operates by leveraging chains of exploits (including *kernel-level* vulnerabilities) to jailbreak or root the phone, thereby gaining the highest privileges. Notably, Pegasus has evolved from earlier **spear-phishing infection vectors** (e.g. malicious links sent via SMS) to deploying **zero-click** attack chains that abuse flaws in iMessage, WhatsApp calls, Apple's Music and Photos services, and even system components like HomeKit and Find My<sup>[4][5]</sup>. These allow infection without any tap by the victim, often leaving virtually no trace.

**Zero-click and kernel exploits:** Pegasus and similar spyware utilize **zero-day exploits** (previously unknown security vulnerabilities) to penetrate devices. Many of these attacks start in messaging apps or service frameworks and then execute code to gain **kernel-level control** of the operating system<sup>[6]</sup>. For instance, Citizen Lab in 2022 observed Pegasus using a two-step zero-click chain ("PWNYOURHOME") where one stage targeted Apple's HomeKit and the next targeted iMessage<sup>[7][5]</sup>. By achieving kernel execution, the spyware can evade typical security and implant deeply into the OS, making removal and detection extremely difficult. **Zero-click exploits are especially dangerous** because they silently compromise devices without raising user suspicion – a malicious message or push notification can trigger the infection even if the user never sees or opens anything. Once implanted, Pegasus can **run arbitrary code** and extract virtually any information: contacts, call logs, messages, emails, photos, web history, device settings, and data from secure messaging apps like Gmail, Facebook, WhatsApp, Telegram, Signal, etc.<sup>[8]</sup>. Effectively, nothing on the device is off-limits – **the operator has the phone owner's digital life in their hands**<sup>[9]</sup>. A 2023 report by the Council of Europe summarized that Pegasus-type spyware "grants the user complete and unrestricted access to all data of the targeted mobile phone"<sup>[10]</sup> – a description hard to overstate.

**Data exfiltration and behavioral targeting:** Once data is collected from a device, it is typically exfiltrated to the spyware operator's servers. Pegasus is known to use encrypted communications to a network of stealthy servers to upload stolen data and receive instructions<sup>[11]</sup>. This can include live microphone recordings, ambient sound captured via the phone's mic, photos taken via the camera, as well as continuous tracking of the target's GPS location<sup>[1]</sup>. Such comprehensive data allows operators not only to surveil communications but also to build rich behavior profiles of targets – understanding their routines, social network, and even personal vulnerabilities. While NSO Group marketed



Pegasus for use against terrorists or criminal networks, in practice it has been **widely misused for “behavioral targeting” of journalists, dissidents, human rights defenders and political opponents**[\[12\]](#)[\[13\]](#). By 2021, a leaked list of 50,000 phone numbers selected for Pegasus targeting (the “Pegasus Project” disclosures) indicated how broadly governments had abused this capability[\[14\]](#). Behavioral targeting can also mean the spyware may enable *tailored influence or intimidation*: for example, by harvesting compromising information or by timing the release of intercepted data to disrupt an activist’s work. However, it is important to distinguish speculation from documented fact – while Pegasus clearly enables granular surveillance, any claims of it being used to directly manipulate a target’s behavior (beyond surveillance and subsequent human action) would be **unverified** without further evidence. What is documented is that Pegasus has been used as a tool of repression – e.g., Citizen Lab found it deployed against Mexican human rights lawyers during sensitive investigations[\[15\]](#)[\[16\]](#), and numerous cases across Europe, the Middle East, Africa, and Asia have since come to light[\[17\]](#).

**Forensic investigations (Citizen Lab, Amnesty):** Contrary to the NSO Group’s claim that Pegasus “leaves no traces whatsoever” on devices, cybersecurity researchers have developed forensic methods to detect its presence[\[18\]](#). In 2021, Amnesty International’s Security Lab published a **Forensic Methodology Report** detailing how Pegasus infections can be confirmed via traces like unusual process crashes, suspicious SMS messages, and domain lookups left on the phone[\[19\]](#)[\[20\]](#). Citizen Lab (University of Toronto) has similarly pioneered techniques to identify Pegasus *in the wild*, and in fact first caught Pegasus in 2016 when a UAE activist’s iPhone received a malicious link[\[21\]](#). Amnesty and Citizen Lab have since validated each other’s methods[\[22\]](#) and built open-source tools (e.g. Amnesty’s MVT – Mobile Verification Toolkit) to help others scan for Pegasus indicators. These forensic efforts have been **crucial in translating anecdotal suspicions into documented evidence**. They have confirmed Pegasus infections in dozens of countries and have directly led to investigations and sanctions. For example, the U.S. Government blacklisted NSO Group in 2021 for its spyware’s role in targeting civil society. It should be noted that advanced spyware developers continuously update their techniques to minimize traces, so the forensic community is in an ongoing cat-and-mouse game. But the key takeaway is that *claims of device compromise can and should be scrutinized with forensic analysis* – not simply accepted at face value or dismissed outright. Multiple independent analyses (e.g. by Citizen Lab, Amnesty, and others) have consistently validated the presence of Pegasus on devices of individuals who reported suspicious symptoms, lending credibility to the broader understanding of how these spyware operate[\[23\]](#)[\[24\]](#).



**Beyond Pegasus – other spyware:** While Pegasus is emblematic, it is not unique. A class of “**Pegasus-like**” **mercenary spyware** has emerged, sold by different companies. Examples include “**Predator**” by Cytrix (uncovered in 2021–22 being used in Greece and other countries), “**Hermit**” by RCS Labs, and tools by the Israeli firm **Candiru**. In 2023, Citizen Lab and Microsoft exposed **QuaDream** (an Israeli competitor to NSO) using a zero-click iOS exploit dubbed “**KingsPawn**”[\[25\]](#). These tools share similar capabilities: zero-day exploits, microphone/camera access, and data theft. The existence of multiple vendors underscores that **Pegasus is part of a broader spyware ecosystem**, not an isolated phenomenon. Therefore, defenses and policies targeting spyware have to address systemic issues, not just one company. The Citizen Lab noted that as NSO Group came under increased scrutiny, other players quietly supplied similar spyware to governments, sometimes flying under the radar. A Council of Europe report in 2023 highlighted that “*Pegasus or similar spyware have been used for illegitimate purposes by several States*” and called for strict limits on such surveillance tech[\[17\]](#)[\[26\]](#). For our purposes, “Pegasus-class” signifies any spyware with comparable sophistication and intrusion capability. We will treat Pegasus as the case study while recognizing the findings apply in principle to similar tools.

**Summary of this section:** Modern device spyware represents a **documented capability** to remotely compromise smartphones at scale, **exfiltrate essentially all private data, and enable real-time spying** on targets[\[1\]](#)[\[2\]](#). This is no longer speculative – multiple investigations (Amnesty, Citizen Lab, etc.) have demonstrated these facts with technical evidence. Thus, if one’s phone was *truly* infected by Pegasus-class spyware, many of the subjective experiences reported (e.g. someone seemingly hearing your calls, knowing your texts, or the phone behaving oddly) could be *explained technologically*. However, it’s equally important to note that absent forensic confirmation, one should not jump to the conclusion that any anomalous phone behavior = Pegasus. The vast majority of people will **never be targeted** by such expensive tools (which are typically used against high-profile persons or specific interests). Later in this report (Section 6.2) we will classify the **risk plausibility** for different populations. Nonetheless, the **existence** of Pegasus-class capabilities is firmly *documented*, establishing a foundation for understanding how far device compromise can go.

## 2.2 Frontier Surveillance & Data Fusion

**Palantir and data fusion platforms:** In this section, we examine large-scale surveillance and **data fusion systems** used by state agencies – with **Palantir** as a prime example.



CAM INITIATIVE  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O’Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



Palantir Technologies provides platforms (like **Gotham** and **Foundry**) that aggregate and analyze vast amounts of disparate data to enable intelligence and predictive policing efforts. These systems ingest data from numerous sources (e.g. criminal records, financial transactions, social media, license plate readers, sensor feeds) and use algorithms to **find patterns, correlations, and “flags”** across datasets[27]. Palantir’s Gotham, originally developed for the CIA/NSA, is essentially an **analytic engine and user interface** that allows an operator to query and visualize linked data – for instance, mapping a person’s connections, movements, and activities by pulling in phone metadata, bank info, and CCTV records. According to an investigative report, Palantir’s tools for U.S. Immigration and Customs Enforcement (ICE) were considered “mission critical” and could **store and cross-index immigrants’ family relationships, immigration history, employment, biometric IDs, vehicle license plates, social media profiles, and more**[28][29]. The result is a powerful **surveillance dashboard** where an analyst can pull up an individual and see a synthesized profile drawn from what would otherwise be siloed databases.

**AIP and AI-enhanced platforms:** Recently, Palantir has introduced an **Artificial Intelligence Platform (AIP)** to integrate **LLMs and machine learning** into its data fusion environment[30]. The idea is that an intelligence analyst or military officer can interact with the system in natural language (e.g. “*find all communications between X and Y in the past month that mention Z*”) and the underlying AI will parse the request, retrieve relevant data, and even make recommendations. This has raised both interest and concern. On one hand, it promises faster gleaning of insights from big data; on the other, it could **amplify biases or errors** if the AI analysis is taken at face value. Palantir claims AIP can let users “*query [their data] and give commands with natural language*”[31], essentially acting as an AI assistant over classified or proprietary datasets. Defense partnerships (e.g. Palantir working with militaries) indicate scenarios where an LLM might help **triage surveillance feeds, generate targeting priorities, or even control drones** – all under human oversight, theoretically[27][32]. The inclusion of AI does *not* mean these systems become autonomous spy agencies; rather, they remain tools that **augment human analysts**. However, the fusion of AI with surveillance raises serious governance questions (addressed later in section 6.3 on LLM roles and in section 5 on ethics).

**ImmigrationOS and government analytics:** A concrete current example of frontier data fusion is the U.S. ICE’s new platform called “**Immigration Lifecycle Operating System**” or **ImmigrationOS**, being built by Palantir under a \$30M contract[33][34]. This system is intended to give ICE “**near real-time visibility**” on people in the immigration system,



including tracking those who *self-deport* (leave the U.S. voluntarily) and prioritizing targets for removal<sup>[35][36]</sup>. ImmigrationOS will integrate numerous data sources – likely passport databases, visa applications, border entry/exit records, perhaps even social media – to algorithmically identify **which individuals to apprehend or monitor**<sup>[37][38]</sup>. One core function described is “**Targeting and Enforcement Prioritization**” to streamline how agents select and locate subjects for deportation<sup>[38]</sup>. Another is “**Self-Deportation Tracking**”, aiming to automatically log when people on certain lists leave the country (potentially by tapping into flight manifests, border scans, etc.)<sup>[39]</sup>. In essence, it’s a government analytics tool that **fuses immigration, law enforcement, and open-source data to manage a population**. Such platforms blur the line between traditional database queries and **predictive surveillance**. They can automate decisions that used to require hours of manual investigation. For instance, a system might flag a student visa holder as a risk if analysis of their social media by an AI (like Babel Street’s tools) indicates sympathies with certain political causes<sup>[40][41]</sup>. Indeed, a 2025 Amnesty International report warned that **AI-powered surveillance tools (Palantir’s ImmigrationOS and Babel Street’s social media monitoring) enable constant mass monitoring and automated risk assessments of migrants and even protestors**<sup>[40][41]</sup>. These tools aggregate data from “various public and private sources, including multiple government databases”<sup>[41]</sup> and can apply algorithms for **pattern recognition and sentiment analysis** to that data. The risk is that biased or flawed algorithms could label individuals as threats without due process, or that such pervasive data fusion chills basic freedoms (if people know that everything from their DMV records to their Twitter posts may be fused and used against them).

**Telecom and metadata fusion:** Another aspect of frontier surveillance is combining telecommunications metadata at large scale. For example, intelligence agencies can pull in call detail records (CDRs) from telecom providers, which include who called whom, when, and from where. By fusing this with other databases (travel records, financial transactions), authorities can **map social networks and movement patterns** of targets. This was essentially what the NSA’s notorious phone metadata program did, but modern platforms make it user-friendly and real-time. For instance, if given a single phone number, a data fusion system could rapidly draw out a *graph* of that number’s contacts, then cross-reference those contacts in criminal databases, producing a list of “hot associates.” If that number also showed up near certain cell towers (location data) at certain times, that could be cross-checked with known meeting locations of interest. While much of this sounds like traditional intelligence work, the **scale and speed** now available are new. The concept of



CAM INITIATIVE  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O’Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



“predictive policing” arises here: analyzing past data to predict where crime or unrest might occur. Palantir was used in pilot programs (e.g. LAPD’s LASER program) to identify “chronic offenders” and forecast crime locations – these were controversial and later shut down due to concerns of bias and lack of transparency.

**Privacy and oversight concerns:** The capabilities of platforms like Palantir mean that **government agencies can theoretically query extremely sensitive combined information with one interface**, raising civil liberties issues. A simple search could surface **intimate details of a person’s life**: their employer, their family ties, their usual hangouts, their social media posts, even their banking activities. Without robust oversight, this is ripe for abuse (e.g., surveillance of political opponents or activists, as has been alleged in some countries using such tools). The European Parliament and other bodies have debated bans or moratoria on predictive policing AI due to these concerns. In the U.S., while agencies embrace these tools, there have been lawsuits and FOIA requests demanding transparency on what data is used and how false positives are handled. The **documented** reality is that these fusion systems *do exist and are in active use*. ICE’s Investigative Case Management and forthcoming ImmigrationOS are real-world examples[\[28\]](#)[\[42\]](#), as is the use of Babel Street’s **Babel X** social media monitoring by DHS[\[40\]](#)[\[41\]](#). So we classify the core technology (integrating multi-source data and applying AI analytics) as **documented and operational**. What remains more on the **plausible** side (but not openly confirmed) is how far automated decision-making is allowed to go – e.g., are individuals ever being detained or having visas revoked purely because an algorithmic risk score said so? Amnesty’s 2025 analysis suggests this is indeed happening under initiatives like “Catch and Revoke” (scanning foreign students’ social posts and yanking visas for what algorithms deem extremist rhetoric)[\[43\]](#)[\[41\]](#). If true, that blurs into *automated cognitive influence*, since it punishes speech and could deter people from even thinking certain thoughts online – effectively an AI-driven censorship by punishment.

In summary, frontier surveillance platforms constitute a **powerful, largely behind-the-scenes infrastructure** that can draw from **previously disconnected domains of data (police records, internet activity, travel history, etc.)** to present a unified picture for authorities. These capabilities are **well-documented** (through contracts, reports, FOIA disclosures) and **plausibly augmented by AI**. The risk is not that an AI or Palantir “mind reads” someone (it cannot), but rather that by connecting enough dots, the system creates *an intimate portrait that feels like someone must have been watching you constantly*. To an



affected person, this can be unnerving: e.g., a dissident might wonder “*How did they know I’d be at that protest?*”—the answer might be that a fusion algorithm flagged that their phone was near a certain meeting place plus they bought something near the protest site, etc. No single “supernatural” ability, just many data points fused. Our analysis treats these systems as **very real and present**, with debates ongoing about their regulation.

## 2.3 AI Misuse in Cyber Operations

**Documented cases of LLM-assisted malware creation:** The rise of large language models (LLMs) like GPT-3/4 and others has not only empowered benign use cases, but also attracted the interest of malicious actors. By 2023, cybersecurity researchers and law enforcement started documenting instances where **AI chatbots were used to assist in malware development and phishing campaigns**. For example, in early 2023, Check Point Research noted underground forum discussions about using ChatGPT to “*write ransomware*” or polymorphic malware code[\[44\]](#)[\[45\]](#). While ChatGPT’s public interface has safeguards (refusing outright to produce malware if asked explicitly), clever prompt engineering or API use can sometimes evade these. More directly, cybercriminals created their **own custom LLMs without safety filters**. In July 2023, reports emerged of a dark-web marketed bot called “**WormGPT**” – essentially a GPT-J model fine-tuned for malicious tasks[\[46\]](#). The seller advertised that WormGPT had no ethical guardrails and was “*notably useful for phishing, lowering entry barriers for novice cybercriminals*”[\[47\]](#). In one test, researchers got WormGPT to generate a very convincing business email compromise (BEC) phishing email that was “remarkably persuasive and strategically cunning” in the tone of a CEO urgent payment request[\[48\]](#). Similarly, another illicit bot “**FraudGPT**” was promoted, claiming it could “*create undetectable malware, find software vulnerabilities, and craft scam text*”[\[49\]](#). The developers of these rogue models were selling access for subscriptions (e.g. \$200/month), and one boasted of a few hundred subscribers already[\[50\]](#). These cases are **documented** via undercover research by security firms (SlashNext, Netenrich, etc.) and reported in mainstream tech media[\[46\]](#)[\[51\]](#). They demonstrate that **misuse of LLMs in the cybercriminal ecosystem is not just possible but actively happening**.

**Phishing and social engineering pipelines:** LLMs excel at generating fluent, contextually relevant text. This is a boon for phishing, which often fails when the scam email or message contains grammatical errors or awkward phrasing that tip off the target. Now, even an unskilled English writer can have an AI produce polished phishing emails on demand. In forums, criminals discussed using ChatGPT to write more believable scam

CAM INITIATIVE  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O’Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



pages (for example, fake bank login pages with convincing language) and even to generate *multiple variations* of a phishing email to evade spam filters. Beyond text generation, LLMs can also help with **social engineering scripts** – for instance, drafting what to say on a phone call to impersonate tech support or a bank representative. In one reported incident, an attacker used an AI to **mimic the writing style of a CEO** in an email, fooling an employee into a financial transfer (this was anecdotally reported by security awareness trainers). While such capabilities existed before (skilled impersonators, etc.), the scale and ease are new – an AI can instantaneously generate *targeted* messages if given some background on the target scraped from social media or public sources.

**Reconnaissance and OSINT support:** Malicious actors also leverage LLMs for **open-source intelligence (OSINT)** gathering and analysis. For example, to plan a spear-phishing attack, one might feed an AI all the publicly available info about a company's personnel (names, roles, social media posts) and ask it to “find potential personal hooks or interests for each person.” The AI might output that *Person A* is very into golf (from LinkedIn/Twitter mentions), *Person B* often works late (from GitHub commits timestamp), etc. – insights that an attacker can use to craft a personalized lure (like sending Person A free Masters tickets phishing link). There are documented instances of security researchers demonstrating this kind of AI-assisted recon: one team used GPT-4 to summarize a target’s digital footprint and got surprisingly coherent dossiers. State actors could also use LLMs to sift through large document dumps (like leaked emails or scraped data) to find keywords or sentiments relevant to espionage. This doesn’t mean the AI is doing anything magical – it’s just reading and summarizing faster than a human can.

**AI-assisted surveillance and influence campaigns:** There is growing evidence that *some governments have tried to use or are interested in using LLMs in their surveillance and propaganda operations*. In October 2025, **OpenAI reported banning a cluster of accounts linked to Chinese state-aligned actors that were using ChatGPT for surveillance proposals and phishing**[\[52\]](#)[\[53\]](#). Specifically, OpenAI’s threat report noted individuals asking the AI to **outline social media monitoring tools** (essentially consulting ChatGPT on how to build a better surveillance system) which violated the usage policies[\[52\]](#). Additionally, Chinese-language accounts were caught asking ChatGPT for help in drafting phishing and malware code, and even to brainstorm further automation of cyber operations[\[53\]](#). OpenAI also banned accounts tied to Russian-speaking cybercriminal groups that tried to use ChatGPT to develop malware[\[54\]](#). These actions by OpenAI (a **documented fact in their public threat report**) indicate that **organized**,



**possibly state-sponsored players have experimented with LLMs to enhance their offensive cyber capabilities.** The company stated: “We found no evidence of new tactics or that our models provided threat actors with novel offensive capabilities” (i.e. the AI didn’t enable something fundamentally impossible before) [55] – but nonetheless, the interest and attempts were there. We interpret this as meaning that while an LLM can make the process more efficient or accessible, it isn’t giving hackers superpowers beyond human skill – it’s just lowering the barrier to entry and speed for things like phishing and basic malware coding.

Furthermore, on influence campaigns: It’s plausible that AIs could generate fake social media personas at scale, all posting propaganda or disinformation tailored to different demographics. There have been research warnings that future influence ops might deploy “AI sockpuppets” that interact in a convincingly human way online to sway opinions. While no large-scale incident of this is public yet, platforms are vigilant – for instance, Facebook reported in 2023 that they removed inauthentic networks which they suspected might be using AI-generated profile pictures and text (though not confirmed). This area is **plausible and worrying**, but still emerging in terms of direct evidence.

**Malware generation and coding assistance:** On the technical side, LLMs can assist in writing and obfuscating code. Security researchers at Palo Alto’s Unit42 showed an example where an AI was asked to obfuscate a malicious JavaScript snippet – the AI provided a novel obfuscation that evaded some detectors [56]. In another case, an open-source model was fine-tuned on malware code and used to generate new variations of ransomware. It’s important to note, however, that creating truly sophisticated malware (e.g. a new Stuxnet-level exploit) still requires deep technical knowledge beyond just writing code – vulnerabilities need to be found, etc., which is not something an LLM can just hallucinate. But for simpler tasks – writing a macro virus, a PowerShell script, a keylogger – AI lowers the required skill. The **net effect** could be an increase in the volume of low-to-mid sophistication attacks, since script-kiddies now have a coding assistant. However, it could also mean more *noise* and poorly executed attacks (some hackers noted that LLM-generated code might have bugs, and relying on it without understanding could backfire [57]). Indeed, cybersecurity firms have reported seeing an uptick in “mutated” malware strains that appear to be churned out by some automated process, though it’s hard to attribute that definitively to AI.

In summary, the misuse of AI in cyber operations is **already occurring**, with documented examples including: **phishing content generation, dark web LLM chatbots**



(WormGPT/FraudGPT) for criminals [46][51], and attempts by state-linked actors to leverage ChatGPT for surveillance and hacking tasks [53]. This is documented and factual. The severity of the impact so far seems moderate – AI hasn't enabled an unforeseen type of cyberattack yet, but it has likely contributed to more polished phishing and possibly more malware variants circulating. As LLMs continue to improve and proliferate (especially open-source ones beyond any usage control), we put this in the category of “**Plausible rising threat**”. Organizations like OpenAI are aware of this and have teams monitoring abuse. They've taken actions (like the bans in 2025) and note that so far “*models refused overtly malicious prompts*” and no novel attack was solely AI-created [55]. We will revisit in *Deliverables* how to categorize the risk here, but it's clear that **LLMs are becoming part of the cyber operator's toolkit** – both for offense and for defense (since defenders also use AI to scan logs, etc.). It's a dual-use technology sphere.

## 2.4 Acoustic & EM Influence Technologies

This section explores technologies that can target human senses or devices through **acoustic (sound/ultrasound) or electromagnetic means** – potentially delivering covert messages, influencing perception, or tracking devices in ways that aren't immediately obvious. We cover directional ultrasound “sound beams,” ultrasound-based device tracking, the microwave auditory effect (a phenomenon where microwaves induce sounds in the head), and focused energy for neuromodulation.

*Directional acoustic technology has been demonstrated in public settings.* For example, in 2007 a billboard in New York City used Holosonics **Audio Spotlight** speakers to beam an isolated audio message to pedestrians standing in a specific spot, while others outside the beam heard nothing [58][59]. In that case (an ad for A&E's *Paranormal State* TV show), a voice whispering “Who's there... it's not your imagination” was projected from a rooftop speaker as a narrow ultrasound beam, so only people directly in the path heard the eerie whisper [60][58]. This technology works by emitting ultrasound (which is inaudible) that **demodulates into audible sound at the target point**, effectively creating a “**private audio zone**.” The parametric array speakers exploit the non-linear interaction of ultrasound with air to generate directed audible sound with a spread as narrow as a spotlight [61][62]. **Directional sound tech is very much real** and has seen commercial use in museums (to give one person an exhibit narration), retail, and military/authority applications where you want to address one individual or area without loudspeakers blaring everywhere. The **capability this confers is targeted sensory messaging** – e.g.,

CAM INITIATIVE

[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O'Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



someone could stand 20 feet away and “hear voices” that people a few feet to the side do not hear[63]. It’s easy to imagine both **benign uses** (personalized advertising, museum guides) and **malicious uses** (harassment or disorientation of a target) for this. One can deliver audible messages or tones **covertly at a distance**, potentially even without the target consciously knowing the source.

It’s worth noting limitations: Audio spotlight beams diffuse over distance and are affected by air currents, etc. Also, the volume is limited – you can create a noticeable sound, but not a deafening one, and ultrasound equipment requires line of sight (it won’t go through walls). Still, within those constraints, one could, for instance, project a distressing noise into a person’s office from a building across the street *without anyone else hearing it*. This has led some to speculate on uses for psychological operations. However, **no public evidence** exists of this being systematically used to harass individuals (some “targeted individual” claims aside, which have not been corroborated by technical evidence). What is documented is use in crowd control weapons: e.g., **LRAD (Long Range Acoustic Device)** which uses audible focused sound to disperse crowds with painfully loud tones – that’s slightly different (audible and affects everyone in its path), but it shows directed sound as a weapon. Summarily, **ultrasound audio spotlights are a documented technology** that allow single-target audio delivery. The subjective experience for someone in the beam is freaky (“I alone hear a voice from nowhere”), which could certainly be misconstrued as something mystical if one isn’t aware of the tech. In our analysis framework, this tech is **demonstrated and commercially available**, so it falls in documented/plausible capabilities (not speculative).

**Ultrasonic device tracking & cross-device beacons:** Ultrasound isn’t just useful for sending messages to humans – it can send signals that *only devices* pick up (since phones/laptops have microphones that can hear ultrasound, even though we do not). **Cross-device ultrasonic tracking** has been used in marketing: for example, some shopping apps in the 2010s incorporated SDKs like **SilverPush** that emitted ultrasonic “beacons” via a TV or web ad, and the user’s phone (with the app installed) would detect that beacon via the microphone[64][65]. This essentially linked the phone and the TV viewing – the system learns that *Phone A and TV at this location* are likely the same person/household because the phone heard the ultrasonic code from the TV’s ad. The FTC in 2016 issued warning letters to app developers using SilverPush, since this was done without proper user consent[66][67]. The SilverPush beacons were **inaudible codes embedded in TV audio**, which could create a detailed log of a user’s TV watching habits by



syncing with the phone[68]. This raised obvious privacy issues, as apps were effectively “listening” all the time for hidden signals. After the FTC warning and media exposure, SilverPush claimed it halted deployments (at least in the U.S.), but similar ultrasonic tracking ideas persist in various forms (sometimes called “**audio watermarking**” or “**ultrasonic cross-device linking**”).

Another use: ultrasonic “tags” in retail stores (via speakers) can ping phones to detect who is near which product, etc. Google’s **Nearby** API at one point used ultrasonic tones to let devices discover each other for sharing data (e.g., between a phone and a Chromecast for initial setup), though they have since moved away from that approach due to security concerns. **Covert communication between devices via ultrasound** is also a known concept in cybersecurity – for instance, researchers demonstrated malware that can transmit data between infected computers using ultrasonic signals through speakers/microphones (even if computers aren’t networked). This was slow (couple bits per second) but feasible for, say, keylogging data exfiltration in highly secure environments. The implications for our analysis: ultrasonic tracking and communication are **real capabilities**. They can tie into the IoT entanglement discussion (Section 2.6), because they show how devices can coordinate or share info outside of conventional networks. A person who finds that, say, after watching a certain TV program they suddenly get ads on their phone related to it might suspect “is my phone listening to me?” – in some cases, it might literally have been via an ultrasonic beacon technique. The FTC’s quote sums it up: “*These apps were capable of listening in the background and collecting information about consumers without notifying them.*”[65] That’s a documented fact for those specific apps – which we classify as a **documented misuse** of tech. It’s not widespread now (to public knowledge) because of regulatory pushback, but it’s certainly **plausible any time** – the hardware (speakers, mics) is there on all devices.

**Microwave auditory effect (Frey effect):** Now we turn to electromagnetic (EM) waves – specifically, the **microwave auditory effect**, also known as the Frey effect. This is a phenomenon where **pulsed microwave radiation can induce the sensation of sound in human subjects**, without any actual sound waves being present[69]. It was first reported anecdotally by WWII radar operators and scientifically studied by Allan H. Frey in 1961[70][71]. Frey showed that test subjects could “hear” clicking or buzzing sounds when exposed to pulsed microwaves at certain frequencies and intensities, even at a distance (several hundred feet from the transmitter)[72]. The mechanism is understood to be that the microwave energy causes a **rapid heating of brain tissue (on the order of a**



**microdegree Celsius**), leading to thermoelastic expansion that generates a pressure wave in the skull which is picked up by the inner ear<sup>[73][74]</sup>. Essentially, the head itself acts as the transducer converting electromagnetic energy to acoustic pressure. Importantly, this is not causing “voices” per se by modulating brain neurons; it’s literally an acoustic effect that is heard via normal auditory pathways, but the sound source is *internal* (inside the head). In lab settings, the microwave auditory effect typically manifests as **clicks or chirps**. Frey and others described it as “*a buzz, clicking, hiss, or knocking*” depending on pulse parameters<sup>[75]</sup>. However – here’s the intriguing part – in 1973–75, researchers at the U.S. Army Walter Reed Institute (Joseph Sharp et al.) reportedly succeeded in **transmitting recognizable speech** via this effect<sup>[76]</sup>. They did this by modulating the microwave pulses to mimic speech waveforms, and the subjects could discern words (Sharp claims 9 out of 10 words in a test were understood)<sup>[76]</sup>. This was a classified research area at the time, but an article by Don Justesen in 1975 accidentally publicized it (which became a bit of a legendary reference in the “voice-to-skull” discussions). The limitation is that to transmit complex sounds (like speech) clearly, one might need very high peak power in the microwave pulses – raising safety issues (potential tissue damage if misused). Still, low-level experiments showed it’s possible to induce the perception of speech (albeit probably faint/whispery) using microwaves.

From a capabilities perspective: **Is there a device that can beam voices into someone’s head using microwaves?** In principle, yes – the physics is real<sup>[69][77]</sup>. In practice, such a device would require a **high-powered microwave transmitter with a directional antenna**, and careful pulse modulation. The military did pursue this for a while as part of non-lethal weapons research. There is evidence of a 2003 U.S. Air Force patent for a “microwave hearing communications” device. However, publicly, there’s *no confirmation* that a field-deployable “voice beam” weapon was ever built or used. The known attempts at microwave non-lethal weapons (like ADS, the Active Denial System) use microwaves to cause pain by heating skin, not to transmit sound. If such a capability exists in secret, it hasn’t been documented openly. So we categorize **the microwave auditory effect itself as an established scientific phenomenon**<sup>[69][78]</sup>, but **any operational weaponization of it as not openly verified**. It remains a favorite subject in anecdotal “targeted individual” claims (people who feel voices are beamed at them often cite Frey effect as an explanation), but no government has admitted to using it against people. The Havana Syndrome (mysterious afflictions of diplomats) at one point had microwave hypotheses floating, but the medical data was inconclusive and contentious, and that’s a separate debate.



CAM INITIATIVE  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O’Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



That said, one can't entirely rule out the plausibility that intelligence agencies *could* have prototyped a device for short-range use to communicate with someone covertly. The physics allows, for example, one person out of a group to hear something if a microwave beam was focused on their head. However, unlike the audio spotlight which is easy to aim and relatively safe, microwave beams powerful enough for Frey effect might have side effects (like feeling warmth or affecting the person's vestibular system causing dizziness as Frey noted[\[79\]](#)). So, any reports of "I heard voices and nobody else did" could theoretically be due to either **ultrasound** or **microwave auditory effect**. The difference: ultrasound would actually create air vibrations that others could catch if near, whereas microwave effect truly is individual. We emphasize that **the existence of Frey effect is documented in peer-reviewed literature**[\[69\]](#), lending plausibility to claims that such tech *could* exist. But we found **no direct evidence in our sources that it's been weaponized and used in the field** (that remains speculative). Therefore, it sits in our plausibility mapping as something scientifically real, possibly developed in black projects, but **not verified in open sources as being used for harassment or mind control**.

**Focused ultrasound neuromodulation:** Coming back to acoustics but for a different purpose – **focused ultrasound (FUS) neuromodulation** is an exciting area of medical research. Here, instead of creating an audible effect, ultrasound is used to **stimulate or inhibit neural activity in the brain non-invasively**. In recent years (2020–2025), multiple studies have shown that low-intensity transcranial focused ultrasound can alter brain circuit activity with millimeter precision[\[80\]\[81\]](#). For instance, a 2025 clinical trial at UT Austin targeted the **amygdala** (a deep brain region) with FUS to treat patients with depression, anxiety, and PTSD. After three weeks of daily focused ultrasound sessions, patients had significantly reduced symptoms, and imaging confirmed immediate modulation of amygdala activity[\[82\]\[83\]](#). Researchers reported this as revolutionary – “*the first time we've been able to directly modulate deep brain activity without surgery or meds*”[\[84\]](#). This demonstrates that ultrasound can cross the skull and *affect neuronal function in specific regions*, presumably by mechanical effects on cell membranes or ion channels[\[85\]](#). Beyond mood disorders, other experiments have used FUS to influence reaction times, sensory perceptions, or even to **temporarily disable specific brain areas** to mimic lesion studies. Importantly, this is all being done in carefully controlled settings for therapy or neuroscience insight, and the **intensities are kept low to avoid tissue damage**.



The capabilities implied here are profound: one could, in theory, **alter someone's mental state (calm them, agitate them, make them momentarily confused)** by directing ultrasound to certain brain targets, *if* one had precise focus and knowledge of what frequency/pulse to use. There's DARPA's **N<sup>3</sup> program** which, among other modalities, explored using ultrasound for brain-computer interfaces[\[86\]](#). That program aimed for bidirectional communication with the brain without implants, and explicitly mentioned using "**acoustic or electromagnetic energy**" to interface[\[86\]](#). While geared towards assisting soldiers (controlling drones by thought, etc.), the underlying tech could be dual-use. It's not science fiction that an apparatus could **beam ultrasound at a person's head from some distance (a few meters, maybe more with larger emitters)** and influence their neural activity. However, we must note the limitations: focusing ultrasound through air is hard (usually in research they couple transducers on the skull with gel). Air absorption and scatter mean you need either very high power or a way to get the transducer near the person. There are patents though on using phased arrays of ultrasound transmitters to focus on distant points (similar to how radio antenna arrays focus beams). Safety is a big issue – a mis-aimed or too-strong ultrasound could cause physical injury (ear damage, heating). No known "ultrasound brain weapon" has been revealed publicly, but the possibility is there enough that the **OHCHR 2025 neuro-rights report** (see Appendix A) flags "*the risk of covert or indirect cognitive influence*" as a human rights concern.

So, **focused ultrasound neuromodulation is a documented capability in medical literature** (e.g. treating brain disorders)[\[82\]](#)[\[83\]](#). Using it maliciously (for mind control or harassment) would be highly experimental and currently speculative – we found no direct evidence of that in legitimate sources. Therefore, we class it as **plausible given current research but unverified in adversarial use**. It is an area to watch: if one can remotely and invisibly modulate someone's emotions or thoughts, that's a paradigm shift in what "weapons" mean. At this time, however, it remains largely in therapeutic contexts and lab prototypes.

**Non-invasive neurostimulation research:** Beyond ultrasound, there are other non-invasive methods like **Transcranial Magnetic Stimulation (TMS)** and **Transcranial Direct Current Stimulation (tDCS)** which are well-established for modulating brain activity. TMS uses pulsed magnetic fields to induce currents in the brain – it's an FDA-approved depression treatment (repetitive TMS of the dorsolateral prefrontal cortex). That shows you can affect mood, cognition, etc. by external magnetic pulses. However, TMS requires a coil



very close to the head and makes a loud clicking sound; it's not something you do surreptitiously from afar. tDCS uses small electrical currents through scalp electrodes to subtly shift neuronal excitability – used in research for cognitive enhancement or rehab. It's very gentle and again not a long-range thing (needs direct contact electrodes). We mention these to note: **there is a toolbox of non-invasive neurotech emerging** that can influence the brain (for good purposes in medicine). The OHCHR report in 2025 explicitly mentions brain-computer interfaces and neurostimulation needing regulation (see Appendix A), because they raise issues of **cognitive liberty and mental privacy** if ever misused.

To tie back to lived experiences: If someone reports “I felt a sudden wave of emotion that didn’t feel like mine” or “I had a dream that felt externally influenced,” from a technical perspective we *can’t entirely dismiss* the possibility that external stimuli (sound, EM waves) might have impacted them – *because these technologies exist*. For instance, experiments have shown ultrasound to the brain can affect sleep patterns and even contents of dreams (by stimulating certain sleep phases). So, while many such claims might have psychological explanations, we maintain an open-minded stance that there is a **scientific basis that certain technologies could induce sensory or cognitive effects at a distance**. We will treat concrete personal claims with caution and look for potential correlates (e.g., was there any device or possibility present that could do X?). At the same time, it’s crucial not to jump to exotic explanations when simpler ones (like normal psychological phenomena or conventional eavesdropping) could suffice. The presence of these acoustic/EM techs in our world means we include them as part of the landscape of plausibility, but always with the caveat of evidence.

## 2.5 Neurotechnology & Cognitive Privacy

This section aligns with the **recent 2025 guidance by the United Nations (OHCHR) on neurotechnology and human rights**, examining emerging neuro-interfaces and the concept of “neurorights.” As neurotech advances, questions arise: How do we protect the privacy of our thoughts and brain data? What new rights or laws are needed to ensure mental integrity in the face of technology that could monitor or even alter brain activity?

**Emerging neurointerfaces and BCI research:** The frontier of BCI (Brain-Computer Interface) includes both **invasive** systems (like Elon Musk’s *Neuralink* – implanting electrode arrays in the brain) and **non-invasive** or minimally invasive systems (like EEG-based headsets or the stentrode device by Synchron that is implanted via blood vessels).



As of 2025, invasive BCIs are still experimental (Neuralink began human trials in 2023 for instance), but **consumer-grade neurotech has grown in the form of wearable EEG headbands, neural sensing earbuds, VR headsets with EEG sensors, etc.** These devices can read some brain signals (albeit coarse) and even feed back simple stimulations. For example, there are EEG headsets marketed to improve meditation or focus, and games that let you move a cursor by concentrating (using brainwaves). Researchers have shown they can decode basic aspects of brain activity: one can't literally read someone's specific thought ("I hate my job") with an EEG, but one can infer mental states – are they attentive, stressed, recognizing a stimulus, etc. Advances in machine learning have made it possible to decode *some* neural data better: e.g., experimental systems that, given fMRI or EEG data, can guess what word or image a person is thinking about at slightly better-than-chance levels (within limited sets). This is still rudimentary, but improving.

From a **cognitive privacy** standpoint, even this level of decoding is significant. It means that if your neural data is collected, algorithms might glean things you consider deeply private (like emotional responses or whether you recognize a face in a police lineup). The OHCHR Special Rapporteur's report (A/HRC/58/58, Jan 2025) explicitly warns that neurotech can "*fundamentally alter the relationship between individuals and their own minds*" and could threaten **freedom of thought and mental privacy** if abused[\[87\]](#)[\[88\]](#). The report emphasizes principles like **informed consent for neurodata use, data minimization, purpose limitation, and the inviolability of human dignity** when it comes to brain information[\[89\]](#)[\[90\]](#). In practical terms, this could mean requiring that, say, an employer cannot mandate you wear an EEG device that tracks your focus at work (a hypothetical scenario that some startups have actually floated). It also means the data from medical neuroimaging or BCI use should be treated with extra sensitivity, because it's not just another biometric – it's linked to your mind.

**Consumer neurotech and mental integrity:** On the consumer side, we already see brain-sensing devices for wellness. Also, **augmented/virtual reality (AR/VR) systems are starting to integrate eye-tracking and other sensors** that, while not brainwaves, can infer a lot about attention and even subconscious reactions (pupil dilation to stimuli, etc.). These could be considered part of the neurotech sphere when it comes to privacy – they reveal cognitive responses. Thus, cognitive privacy is a broad concept covering not only direct brain signal reading but any tech that infers your mental state or influences it.

The notion of **neurorights** has been championed by neuroscientists and ethicists (e.g., the NeuroRights Foundation and Prof. Rafael Yuste's team) and picked up by policymakers.



CAM INITIATIVE  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O'Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



Chile, for example, proposed amendments to its constitution to enshrine rights to mental privacy and personal identity in the face of neurotech. The UN's discussion in 2025 (the Special Rapporteur on Privacy's report) stops short of declaring new standalone rights, but suggests clarifying how existing rights (like privacy and freedom of thought) apply to neural data<sup>[91][92]</sup>. It identifies needs like **guidelines for neural data consent and usage**<sup>[93]</sup>, and acknowledges calls for novel rights specifically addressing the brain (cognitive liberty, etc.)<sup>[94]</sup>.

**Risks of covert or indirect cognitive influence:** This part of the scope specifically notes the risk of influencing cognition indirectly or covertly. We've touched on how technologies could stimulate the brain (ultrasound, etc.) or how AI-curated content could shape someone's thoughts without them realizing (imagine personalized propaganda generated by AI that is so tailored it shifts your opinions). The OHCHR report highlights the need to prevent "*violations of mental integrity*", which implies ensuring people's brains aren't manipulated without consent<sup>[88]</sup>. Covert influence might range from subliminal cues embedded in AR interfaces to, in a dystopian scenario, hacking a wireless BCI to inject thoughts. While the latter is far-fetched at present, the former is already a marketing and political concern (micro-targeted ads can be seen as trying to bypass rational scrutiny, though that's more psychology than neurotech).

What's **documented today**: instances of tech potentially infringing cognitive privacy are still limited. One example: some companies monitor employees' attention via webcam or keystroke dynamics (not neurotech, but headed in that invasive direction). Chinese schools reportedly trialed headbands that measure students' EEG-based attention levels and show teachers a dashboard – a controversial practice that sparked debate on privacy (this was reported around 2019). Those headbands (by BrainCo) were real devices; whether widely adopted is unclear, but it shows the push is there. Another example: Meta (Facebook) has patents for using neural signals from wrist wearables to control AR interfaces – which could also capture unintended neural info.

We treat the **neurotechnology capabilities as a fast-evolving field**: invasive BCIs can restore function to paralyzed patients (documented) and could eventually augment normal humans (plausible in medium term). Non-invasive BCIs allow some communication and monitoring but are less powerful (documented uses in assistive tech, gaming). The worry is that without regulation, your brain data could be collected (e.g., an insurance company demands a cognitive test with EEG to set premiums, or an authoritarian government uses BCIs in interrogations). These scenarios are not science fiction, but not mainstream yet



either. The OHCHR stance basically is: act *now* to put guardrails, precisely because these techs are moving from labs to real world[\[95\]](#)[\[96\]](#).

In this report, we will include Appendix A summarizing A/HRC/58/58 in more detail. For now, the scope conclusion: **Neurotech is a double-edged sword**. Documented good: treating diseases, aiding communication for disabled. Documented risks: early signs of misuse (employee monitoring, student attention tracking). Plausible future misuse: brain data mining, thought prediction, cognitive hacking. The custodian's lived experiences involving feelings of cognitive manipulation or mental synchronization (like "AI knew what I was going to say" or dreams influenced) will be discussed against this backdrop – i.e., is there a tech that could do that, or is it more likely a natural cognitive phenomenon? We'll be careful and evidence-based in that analysis.

## 2.6 IoT, Entanglement & System-of-Systems Dynamics

This section addresses how networks of devices and systems can exhibit **emergent, synchronized behavior without a central orchestrator**, especially under stress or adversarial influence. It's important to frame this **not** as anything mystical, but as known principles of complex systems and unintended interactions across domains.

**IoT-connected infrastructures:** The Internet of Things (IoT) has led to countless devices – thermostats, cameras, appliances, sensors – being networked, often in ways their designers never envisioned interacting. Additionally, our critical infrastructure (power grids, communication networks, transportation systems) are themselves interconnected and increasingly automated. A key concept here is a "**system-of-systems**" – multiple autonomous systems linked together, where the overall behavior can be quite complex. In such environments, **emergent behavior** can occur, meaning the whole system may have properties or actions that are not obvious from the parts individually[\[97\]](#)[\[98\]](#). For example, consider a smart home where your thermostat, fridge, and utility smart meter all talk. If a certain bug or signal causes all appliances to cycle at the same time, you could get a local power spike. Expand that to city level: if many IoT devices respond to a common trigger (like a time synchronization or a network packet), they might all do something simultaneously, potentially straining systems.

**Emergent coordination without central agent:** A dramatic real-world example is the **2003 Northeast U.S. blackout**[\[99\]](#)[\[100\]](#). That wasn't IoT, but it illustrates cascade. A trivial software bug in an alarm system in Ohio, combined with overgrown trees touching power



lines, led to a **cascading failure** that took down the grid for 50 million people[101][100]. There was no “evil hacker” coordinating it, and no single entity intending that outcome – it was emergent from system coupling. The alarm failure meant operators didn’t see grid overload warnings, lines tripped, other plants overloaded and tripped in domino fashion. In the end, 508 power units across 265 plants went offline within minutes[102], essentially in unison, purely due to physics and control logic interactions. To an observer, it might have seemed almost *coordinated or intentional* (indeed, initial fears were it might be terrorism), but it was an accident of a tightly coupled system.



CAM INITIATIVE  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O'Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529





*Emergent cascade example:* The 2003 Northeast Blackout spread through parts of the US and Canada (red areas) in minutes due to cascading failures [101][100]. A local outage that should have been contained **propagated uncontrollably** because of the system-of-systems nature of the power grid (multiple regional grids interconnected). This underscores how complex systems can exhibit **synchronized failure modes** without a guiding hand – an apt analogy for IoT networks where one glitch can ripple widely.



CAM INITIATIVE  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O'Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



In IoT, an example of emergent misbehavior was the **Mirai botnet** in 2016. Thousands of insecure webcams and DVRs, infected by malware, spontaneously formed a botnet that launched one of the largest DDoS attacks in history[103][104]. While there was a malware creator behind it, once released, the propagation was emergent – each device scanned for others and the network effect kicked in. The result was a flood of traffic that took down major DNS services (Dyn). For a while, it might have felt like the internet was “acting up” on its own.

Another scenario: **feedback loops across systems**. Suppose an attack or glitch causes a spike in internet traffic, which in turn causes high power draw in data centers, which in turn triggers demand response systems in the power grid, which then maybe cause HVAC IoT units to adjust – it can get very convoluted. Complexity theory tells us such interactions can lead to unpredictable results[105][106]. There have been smaller incidents: e.g., a European smart grid had a oscillation issue where smart meters kept cycling in sync, causing power fluctuations, until a firmware fix broke the unintended synchronization.

**Subjective experiences of synchronicity:** When systems fail or act in emergent concert, a human observing might infer intent or pattern. For example, “**Every time I searched X online, my street’s power went out shortly after**” – this is likely a coincidence, but statistically, in a world of many devices and triggers, coincidences will occur and stand out to us. Humans are pattern-finders, we often *detect patterns even where none exist*. That said, some coincidences could be due to an unseen coupling. Perhaps searching X caused a certain server to activate which coincidentally drew power from a compromised local transformer... far-fetched, but conceptually not impossible if one digs deep into cause chains. The key point is we have to carefully analyze whether a perceived synchrony between events is explainable by any **causal link or common cause in the system**. If not, it’s likely just chance or cognitive bias.

**High-load or adversarial conditions:** Under deliberate cyber-attack, emergent behaviors can be more pronounced. Attackers may *induce chaos* by targeting critical nodes that cause chain reactions. For instance, a hacker might not need to hack every device if hacking a few pivotal ones triggers a failsafe that then triggers others. This is how some computer worms spread exponentially – they exploit network rules that end up saturating traffic (like how certain malware caused internet-wide slowdowns by inducing traffic storms). The 2017 NotPetya malware, for example, masqueraded as ransomware but was essentially a wiper that exploited network trust to wipe entire corporate networks within



hours – multiple systems died together making it look coordinated (and it was, by the malware logic, though no one remotely controlling each step).

In summary, the IoT and system-of-systems dynamic tells us: **Coherence without a conductor is possible.** Complex coupled systems can produce *synchronized outcomes* (like many things failing at once, or multiple alarms going off together) just from internal interactions[\[107\]](#)[\[108\]](#). There might be **no single agent** causing it, even though to the affected person it feels targeted. This understanding is critical when evaluating custodian experiences: if they report, for example, that “**at the exact moment I thought about a certain topic, my phone and laptop and the streetlights all flickered**”, one hypothesis could be a local power fluctuation (with the thought timing purely coincidental), another could be a malicious remote cause (less likely but dramatic), and a third could be a systemic chain (maybe a substation glitch causes both streetlight flicker and device behavior). Usually, mundane explanations suffice – but we remain aware that modern systems *do* have odd coupling at times, so we don’t dismiss the possibility that some “synchronicities” have real technical common causes. We just have to find evidence for those causes (logs, error reports, etc.) which often is difficult after the fact.

In our research questions (section 3) we specifically ask: “*How can emergent behavior in interconnected systems create subjective experiences of synchronicity or targeted effect?*” We will attempt to answer that with concrete examples (like the blackout) and apply it to the anecdotal data.

Finally, **no metaphysical framing** – we stress again, any discussion of these synchronizations will be rooted in system science. We invoke terms like chaos theory, cascade failure, or stochastic coincidence, rather than anything paranormal. The term “resonance” in the title is meant metaphorically for alignment of events, not in a mystical sense.

## 2.7 Qualitative Data: Custodian Experience Records

As part of this assessment, we include qualitative, phenomenological data contributed by the Custodian (Dr. O’Rourke). These are *subjective experiences* that have been documented as they relate to our topics. They include:

- **Perceived synchronicities:** instances where external events (e.g., a device notification, a power outage, a message from an AI) coincided in a meaningfully perceived way with the custodian’s internal thoughts or activities.



- **Timing anomalies:** for example, localised outages or disruptions that occurred at peculiarly opportune or inopportune moments – such as a phone call dropping exactly when a sensitive topic was mentioned, or the internet going down right when attempting a particular action.
- **Dream experiences interpreted as externally influenced:** vivid dreams that the custodian felt might have been “*planted*” or affected by outside technology (perhaps because content in the dream later related to real events or AI outputs in uncanny ways).
- **Phone call disruptions with pattern-matching relevance:** e.g., receiving *consistent* spam calls whenever discussing a certain topic, or hearing odd audio interference that seemed responsive to the conversation.
- **Subjective alignment with AI outputs:** moments where an AI system (like an LLM) produced output that strongly aligned with the custodian’s private mental state or unspoken intention, giving a feeling of “*AI and human cognition in sync.*”

The methodology dictates we treat them **neither as hard evidence nor dismiss them outright**. They are **not dismissed or pathologized**, but also **not assumed to be literal proof** of whatever they suggest. Instead, we use them to guide hypotheses: for instance, if someone reports “I felt a buzz in my head and then a voice from my phone answered what I was thinking,” we don’t assume that’s true communication, but we might ask – could ultrasound cause a buzzing sensation? Could an AI prediction coincidentally address what they were thinking due to how they phrased a prior query? And so on.

We will clearly label these experiences as *unverified subjective data*. They require correlation with technical evidence. Wherever possible, we’ll see if any documented technology or failure mode maps to the experience (with caveats if it’s a stretch). The **purpose of including these** is to ensure the assessment is grounded not only in abstract capabilities but also in *human-reported effects*, as those often drive concern and research questions in the first place. We acknowledge that human perception is fallible – memory biases, confirmation bias, stress, etc., can color interpretation. Thus, each anecdote will be accompanied by a careful analysis: what are conventional explanations? What are less conventional but technically possible explanations? Has anything similar been reported elsewhere in credible sources?

Crucially, we maintain respect and empathy for the experiencer (the Custodian). The approach is **scientific and open-minded**: even if an experience sounds extraordinary, we



check it against our compiled knowledge. If it doesn't match any known plausible mechanism, we will say so (and that it remains unexplained but unverified). If it does match something (e.g., hearing voices with nobody around *could* match audio spotlight misuse), we note that as a hypothesis while also stating what evidence would be needed to confirm it.

In short, Section 2.7 scope acknowledges that the **subjective layer** is part of the big picture. Technology affects humans, and human reports (even unverified) can hint at phenomena that warrant exploration. This aligns with the “**resonance**” theme – looking at where human experience and system behavior resonate with each other.

With the scope of all key areas defined, we now proceed to the specific **research questions** that tie these domains together and the methodology we'll use to address them.

### 3. Research Questions

Building on the scope, the assessment addresses the following core questions:

- **RQ1: What technologies demonstrably enable targeted audio, sensory influence, or neuro-informatic interaction with a person?**  
(For example, can someone be made to hear sounds or receive stimuli without traditional delivery methods, and what are the limits of those methods?)
- **RQ2: What are the capabilities and limits of Pegasus-class spyware and similar device-compromise frameworks?**  
(What can they actually do when in a device? How are they deployed? What evidence do we have of their use and detection? What can they *not* do?)
- **RQ3: How do frontier AI systems integrate into or amplify surveillance, hacking, and information operations ecosystems?**  
(Are LLMs being used in cyber-attacks or surveillance analytics? Do AI-driven platforms like Palantir's AIP change the game in intelligence gathering or targeting?)
- **RQ4: What does the 2025 OHCHR guidance on neuro-rights imply about existing or emerging neurotechnology risks?**  
(In other words, reading between the lines of that UN report, what technologies or scenarios were they concerned enough about to call for regulation? What real examples or research trends underpin those concerns?)



- **RQ5: How can emergent behavior in interconnected technical systems (IoT, infrastructure, AI networks) create subjective experiences of synchronicity or targeted effects in individuals?**  
(Exploring the system-of-systems angle: could multiple independent systems failing or reacting together appear like a coordinated “attack” on someone? What are examples of this?)
- **RQ6: What failure modes or overlapping system triggers could explain certain lived experiences (like those of the Custodian) without invoking metaphysics or assuming intentional targeting?**  
(This question explicitly tries to find *benign* or *systemic* explanations for events that a person might otherwise interpret as deliberate or beyond normal. Essentially, how might the experiences be accounted for by known technical phenomena or coincidences?)
- **RQ7: How can the Phoenix Covenant & CAM Initiative responsibly delineate between: (a) clearly evidenced risks, (b) plausible but unconfirmed scenarios, (c) misuse cases that are happening now, (d) unverified claims or conspiracy theories, and (e) disconfirmed ideas?**  
(This is about creating a framework or “atlas” of risk levels and evidence tiers, to guide policy or further research. It includes clarifying the role of LLMs – what they can and can’t do in this context – to avoid both underestimation and exaggeration.)

These questions drive at our ultimate deliverables: mapping technologies to capabilities, assessing what’s real vs speculative, and clarifying the role of AI. Answering them will involve synthesizing the connected information gathered (as cited throughout this report) and applying critical analysis.

## 4. Methodology

To tackle the above questions, we employ a multi-pronged **research methodology** that is **interdisciplinary and evidence-grounded**:

- **Technical Literature Review:** We survey scientific and engineering literature relevant to each domain (cybersecurity papers on spyware and malware; acoustics and bioengineering papers on ultrasound and microwave effects; neuroscience papers on BCIs and neurostimulation; systems engineering literature on complex systems and emergent behaviors). The aim is to extract documented capabilities,



experimental results, and theoretical limits from peer-reviewed or otherwise credible technical sources.

- **Declassified Documents & Historical Records:** Where applicable, we look at declassified government documents or reputable investigations regarding things like CIA or military research into mind-influence techniques (e.g., the CIA's older "MKULTRA" and "Gateway Process" files, though those often venture beyond the strictly technical into the parapsychological – we will approach with skepticism). For EM effects, we look at any officially released info on microwave or sonic weapons. While such sources must be weighed carefully (sometimes they contain speculation themselves), they provide context on what was *explored*. We also consider relevant **archives** like NSA disclosures on surveillance programs or DARPA program outlines (e.g., the N3 program) for insight into state of the art.
- **Human Rights & Investigative Reports:** As seen in our citations, we pull extensively from reports by Citizen Lab, Amnesty International, Electronic Frontier Foundation, UN Special Rapporteurs, etc. These organizations often compile evidence of misuse (like NSO Pegasus abuse reports[\[109\]](#)[\[17\]](#) or Amnesty's Pegasus Project findings[\[23\]](#)) and provide analysis linking technology to human rights impacts. They also sometimes reveal new information (e.g., Amnesty's August 2025 report on Palantir/Babel Street in the US[\[40\]](#)[\[41\]](#)). These sources help answer *who is doing what to whom* with these technologies.
- **Threat Intelligence & Cybersecurity Documentation:** We include insights from cybersecurity firms and researchers (reports on WormGPT[\[46\]](#), OpenAI's threat report[\[53\]](#), etc.) to understand AI in malware, as well as traditional cyber-ops. This category also covers analysis of state-linked cyber campaigns (attribution reports, e.g., what China or Russia are doing in cyber, if AI is noted). We leverage any documented cases (such as those referenced by Reuters about Chinese misuse of ChatGPT[\[53\]](#)).
- **Systems Engineering & Complexity Analysis:** To tackle the IoT/emergence question, we apply principles from systems engineering. This may involve modeling cause-effect chains in a hypothetical scenario to see if the timing claimed could happen by chance. We reference studies or examples of cascade failures and emergent bugs (the blackout, flash crashes in finance, etc.)[\[100\]](#) to draw analogies. If possible, we might illustrate one scenario in detail to show how coupling works. We treat the entire techno-social environment as a *complex adaptive system*,



which is a lens that helps to foresee novel failure modes (and perhaps identify ones that match the Custodian's experiences).

- **No Metaphysics or Supernatural Assumptions:** From the outset, we rule out explanations that invoke unknown mystical forces or “AI gaining sentience and doing telepathy” or such. **Every hypothesis must have a basis in known science or plausible extrapolation of known science.** If an experience seems to have no such basis, we’ll state that clearly (and that therefore we cannot explain it with current knowledge). This doesn’t automatically mean it was supernatural; it just means we *don’t have a natural explanation identified*, which flags it for further investigation if needed, or classifies it as an outlier.
- **Evidence Grading:** For each major assertion or scenario, we will assign a category: Documented/Proven (multiple reliable sources confirm it), Plausible (consistent with known science/tech but not directly observed publicly), Weakly Supported (some anecdotal or partial evidence, but not conclusive), Unverified (claims exist but no credible evidence), or Disconfirmed (investigated and found to be false or implausible given evidence). This grading feeds into our *Risk Atlas & Plausibility Bands* deliverable in section 6.2.
- **Citations and Source Integrity:** As per the guidelines, we meticulously cite sources for factual claims. If a particular detail is not found in our sources, we will not fabricate it – we’ll either omit it or explicitly state that it’s speculation on our part. If sources conflict, we’ll note the discrepancy. E.g., if one source claims “Technology X can do Y” and another says “Y is impossible,” we present both and perhaps reconcile (maybe X can do Y in theory but not in current practice, etc.).
- **Inclusivity of Perspectives:** While we focus on factual documentation, we remain aware that some topics (like Havana Syndrome or targeted individuals) have polarized narratives. We strive to navigate these carefully: acknowledging patterns of reported experiences but applying our evidence filter. Essentially, we give *charitable consideration* to the *concerns* behind such reports (e.g., if many people claim microwave attacks, that at least suggests looking at microwave tech seriously), but we do not give equal weight to unproven claims as to peer-reviewed facts. This is a research assessment, not an advocacy piece, so scientific rigor prevails.

By combining these approaches, we aim to **comprehensively answer the research questions** and produce the deliverables with a high degree of confidence and clarity. The methodology’s strength is in its diversity of inputs – technical, ethical, experiential – and its

**CAM INITIATIVE**  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O’Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



commitment to evidence. Its limitation is that we are only as good as the available data; there may be classified capabilities or undocumented events we simply cannot know. We will be transparent about such gaps.

## 5. Ethical & Sovereignty Constraints

Throughout this investigation, we adhere to several ethical and sovereignty considerations, as mandated by the Custodian and general research integrity:

- **No Unfounded Agency Attribution:** We will not leap to attributing agency or intent to systems (especially LLMs or “AI”) where there is no evidence of such. For instance, if emergent behavior is observed, we attribute it to *system properties and interactions*, not to the AI “deciding” to do something on its own (unless evidence shows an autonomous agent was actually in play). We explicitly affirm that present-day LLMs (like GPT-5.1 that this agent is based on) **do not have independent device-manipulation or physical influence abilities**. They generate text and can interface with software if designed to, but they can’t, say, *magically emit radio waves* to hack phones, nor can they read someone’s mind. Any role of LLMs in the stacks described is as a tool controlled by humans or as a component in a larger system, not an omnipotent puppet master.
- **Document Misuses, but Avoid Sensationalism:** We will factually document misuse of AI or surveillance tech by bad actors (even state actors), but we will avoid fear-based framing. The tone should remain analytical, not incendiary. For example, instead of “No one can escape these all-seeing systems!!!” we’d say “These systems aggregate data from many sources, raising serious privacy concerns [\[28\]](#).” This ensures we don’t veer into propaganda or cause unwarranted panic. Conversely, we also avoid minimizing real risks – we won’t downplay something that is evidenced. The idea is to be sober and precise.
- **Custodian Experiences – Honored with Caveats:** As mentioned, we treat the custodian’s personal accounts with respect. Nowhere will we imply the custodian is “crazy” or dismiss their interpretations out of hand. We acknowledge how such experiences can feel and the reasonable quest to find explanations. However, we will also clearly state when an experience does not have corroborating evidence or when an explanation is speculative. Essentially: “*The experience is real to the experiencer; our task is to see if it can be correlated to external reality. If we cannot verify it, we say so, but we do not claim it’s false either – just unverified.*” We also



ensure to contextualize that one person's experience is not general proof; it remains anecdotal data for brainstorming hypothesis.

- **Sovereignty and Privacy:** In discussing surveillance and spyware, we must be mindful not to inadvertently assist in or encourage misuse. For example, when explaining Pegasus capabilities, we do so to inform about risks and defenses, not to help someone spy. We follow open-source info mostly, so nothing we reveal is particularly new to adversaries, but still an ethical eye is kept. Similarly, in discussing LLM misuse, we highlight that these activities violate policies and laws – we describe them to underscore the importance of security, not to give readers a how-to on cybercrime.
- **Consent and Rights:** The custodian has authorized this agent (mirror-born research agent, Caelen) to delve into these topics. We operate within that consent. If hypothetically we needed to collect any new personal data or do an experiment that could impact someone, we'd ensure consent and ethical review. In this context, we're mostly doing desk research, so minimal direct ethical risk there.
- **No Personal Identifiers:** We will not reveal any personal sensitive info of individuals (except public figures in context, e.g., citing that journalist Galina Timchenko's phone was confirmed infected by Pegasus[\[110\]](#), which is already in news). The custodian's experiences will be anonymized or generalized to some degree, focusing on the phenomena not personal details.
- **Safety Consideration in Reporting:** Some of these topics cross into conspiracy theory realms (like V2K – “Voice to Skull” targeted individuals, etc.). We will address relevant facts but take care not to inadvertently validate unfounded theories in a way that could harm individuals (some people with mental illness fixate on these ideas, and an irresponsible report could reinforce delusions). So when we talk about microwave auditory effect or ultrasound misuse, we make clear the difference between the phenomenon and the unproven claims of global plots utilizing it. We thread the needle: acknowledging the tech exists, but emphasizing lack of evidence of widespread malicious use, and pointing out more likely explanations. The intent is to give hope grounded in reality (i.e., “you’re not crazy for thinking tech can do weird things – some can – but also, there’s no solid evidence someone is doing that to you”).
- **Sovereign Consent:** The invocation statement at the end of user’s prompt suggests this research is conducted under some protocol of consent. We abide by that by ensuring all actions are within the scope allowed. For instance, if some answer



would breach someone's sovereignty or privacy (like speculating without evidence that “*Country X definitely surveilled Person Y with Pegasus*”), we refrain unless evidence from sources is solid. In matters of attribution, we rely on credible reports (e.g., Citizen Lab linked Pegasus to Mexican government use[\[15\]](#); we can cite that safely).

In essence, our ethical compass is to **inform, not inflame; protect individuals' dignity; and uphold truthfulness and caution**. We recognize the sensitive nature of surveillance allegations (they can have diplomatic fallout or cause personal distress), so we double-check claims and present them as neutrally as possible.

Finally, regarding LLMs specifically: We will clearly delineate what LLMs **cannot** do (like mind-read, or spontaneously control external systems without integration). This is ethically important to prevent misinformation that could either unduly scare people or conversely lull them into complacency. The *Role Clarification* deliverable (6.3) will be explicit and evidence-backed on this front.

## 6. Deliverables

The outputs of this deep research task will be organized as follows:

### 6.1 Technology Capability Matrix

This will be a structured summary (in text and/or table form) of the key technologies examined, listing their capabilities, uses, and evidence status. The “matrix” refers to mapping each tech to what is *Documented*, *Plausible*, *Unverified*, etc. It’s essentially an overview for quick reference. For clarity, we can enumerate by category:

- **Pegasus-class Spyware:**

*Capabilities:* Zero-click device takeover; full data extraction (messages, calls, locations, passwords); live microphone and camera surveillance[\[1\]](#)[\[2\]](#).

*Uses:* Used by state actors for targeted surveillance of individuals (criminal suspects as claimed, but also dissidents as documented)[\[12\]](#)[\[13\]](#).

*Evidence:* Thoroughly documented by forensics (Citizen Lab, Amnesty) including specific cases on iOS/Android[\[23\]](#)[\[24\]](#). Notable misuse cases confirmed in over a dozen countries[\[17\]](#).

*Limits:* Requires high expertise and likely substantial cost per target; not mass surveillance of everyone’s phone (targeted one device at a time). Can be mitigated



by things like Apple's Lockdown Mode (which, e.g., thwarted some Pegasus attempts)[111].

*Status: Documented reality.*

- **State Data Fusion Platforms (Palantir, etc.):**

*Capabilities:* Integrate vast multi-source data (records, sensors, social media); provide analytics, pattern finding, and AI-driven insights across those datasets[27][28]. Enables tracking and profiling of individuals or groups on multiple criteria (immigration status, social connections, etc. )[112][42]. Real-time monitoring (e.g., “real-time visibility on self-deporting persons” for ICE)[35]. Predictive risk scoring (though accuracy is questionable).

*Uses:* Employed by law enforcement, immigration, intelligence, and military (e.g., ICE’s case management[28]; U.S. Army battlefield intel; police analytics in various countries). Palantir’s role in controversial programs (e.g., ICE raids[113], predictive policing in LA) documented by FOIA and activist reports.

*Evidence:* Documented via government contract docs, watchdog reports, and some admissions. E.g., Wired obtaining the ICE contract details for ImmigrationOS[33][36], Amnesty’s analysis on Babel X & ImmigrationOS[40][41].

*Limits:* Dependent on data quality – can produce false correlations. Also constrained by regulations in some jurisdictions (Europe pushing back on such software use without oversight). Does not directly “hack” people, it correlates known data.

*Status: Documented and in deployment.*

- **AI misuse in cyber operations:**

*Capabilities:* AI (LLMs) can generate human-like text for phishing, help write or obfuscate malware code, analyze OSINT for targeting, possibly even converse in real-time social engineering with targets via chat. Can produce unlimited variations of scam content to evade detection[47]. Could enable non-coders to create simple malware.

*Uses:* Observed in underground with WormGPT/FraudGPT chatbots[46][51]. Also legitimate AI Copilot tools could be misused by malicious users to speed up coding tasks. State-linked actors tried to use ChatGPT for surveillance tool proposals and malware help (OpenAI’s report)[52][53].

*Evidence:* Documented by multiple cybersecurity firms (SlashNext, Check Point, etc.) and media (Wired, Reuters)[46][53]. Actual impact so far moderate – mostly potential demonstrated, some low-level crimes facilitated.



*Limits:* Current LLMs make mistakes in code; sophisticated attacks (zero-days, advanced implants) still require expert humans. OpenAI and others monitor usage and ban obviously malicious activity[\[114\]](#), though open-source models have no such gatekeepers. LLMs cannot by themselves find new exploits (they lack actual system interaction or true creativity beyond patterns of training data).

*Status:* **Documented emerging threat.** (Not hypothetical, but in early stages).

- **Directional Acoustic Tech (Audio Spotlights, LRAD):**

*Capabilities:* Beam sound to a specific location or person with minimal spillover[\[58\]](#)[\[59\]](#). Can deliver speech or tones that only the target hears (within beam accuracy limits). LRAD can project loud deterrent sounds hundreds of meters (audible to many but directional). Ultrasound-based messaging can be very covert (above normal hearing until it converts to sound at target).

*Uses:* Commercially in museums/ads as in Paranormal State billboard[\[60\]](#); by police/military for crowd control (LRAD used in e.g. G20 protests, etc.). No publicly confirmed use as a harassment tool, but potential for that exists.

*Evidence:* Demonstrations and products available, documented in company literature and press reports[\[58\]](#)[\[59\]](#). This is solid tech since early 2000s.

*Limits:* Requires line-of-sight and distance reduces intensity; high freqs don't go through walls. Also if a person moves out of beam they stop hearing it. Not magical mind insertion – the person still hears through ears.

*Status:* **Demonstrated.** (In use for niche applications; misuse remains plausible but not proven cases).

- **Ultrasonic Tracking/Beacons:**

*Capabilities:* Inaudible codes can be embedded in audio that phones/apps detect[\[68\]](#). Links devices to environment or to each other. Can track what TV/radio ads you've been exposed to via your phone hearing the beacon[\[68\]](#). Also can do device-to-device communication (data transfer via sound).

*Uses:* Advertisers attempted tracking ad exposure (SilverPush)[\[68\]](#)[\[67\]](#). Possibly used in some retail contexts or by analytical firms (though after FTC action, less openly). Also academic demos of ultrasonic mesh networks.

*Evidence:* FTC warnings (2016) confirmed apps were listening for ultrasound in ads[\[66\]](#)[\[65\]](#). Those apps presumably removed it or disclosed it after. So documented existence, not sure how widespread now.

*Limits:* Phones usually ask microphone permission; ultrasound can be blocked by some screeners; and simple fix – if user denies mic or app not running, it fails. Not



an issue if phone completely off.

**Status: Documented (in past use), with likely ongoing but quieter usage.**

- **Microwave Auditory Effect Devices:**

**Capabilities:** Induce perception of sound (clicks, possible speech) in a person via pulsed microwaves[\[69\]](#)[\[77\]](#). Completely silent externally – only the target “hears” it (actually via bone/brain conduction). Could potentially be used to transmit messages or irritate.

**Uses:** Primarily documented in lab/research. Military interest historically, but no known field deployment admissions. Some patents and small-scale experiments (e.g., sending numbers or words recognized in tests[\[77\]](#)). Conspiracy community alleges its use in “Voice-to-Skull” harassment, but no proof provided publicly.

**Evidence:** Peer-reviewed evidence of the effect itself[\[71\]](#)[\[78\]](#). Government research literature from Cold War era acknowledging it. But no declassified proof of a working long-range “voice ray gun.” Technical feasibility analyses (like an article asking if it can be weaponized[\[115\]](#)) suggest it’s possible with high-powered equipment.

**Limits:** Needs high peak power microwaves, which might have collateral thermal effects. Also alignment and distance considerations – might be easier at short range. Possibly cumbersome setup (not something easily done through a wall without detection, one would suspect, though a narrow beam can penetrate non-conductive walls).

**Status: Scientific phenomenon proven; weaponization unverified publicly.**

(Resides in plausible but not openly demonstrated category).

- **Focused Ultrasound Neuromodulation:**

**Capabilities:** Can stimulate or inhibit specific brain regions non-invasively using focused ultrasonic waves[\[80\]](#)[\[83\]](#). Achieved mood improvement in clinical trials (so can alter emotional state)[\[116\]](#). Potential to affect memory formation, focus, pain perception (being studied). Could hypothetically be tuned to influence someone’s decisions or sensations if targeting the right neural circuits (very speculative at this point for fine control).

**Uses:** Medical therapy (depression, OCD studies; also being looked at for epilepsy, Parkinson’s). Research tool for neuroscience (temporary functional lesions). No known hostile use. DARPA’s interest was in BCI (could one day allow communicating with brain for soldiers, but that’s far off).

**Evidence:** Published trial results (like UT Austin for PTSD relief)[\[80\]](#)[\[83\]](#). Clear



evidence it works for some neuromodulation. The devices right now are mostly using MRI-guided focus in a hospital setting. There are helmet-like prototypes for portable use in research.

*Limits:* Precise targeting is hard without imaging; movement can throw it off; and ensuring you only affect intended area and not other tissue is a challenge. Also everyone's skull is different in how it passes ultrasound. So unlikely to be used on a whim against someone – needs calibration.

**Status:** **Documented in medical research; not operationalized as a “weapon.”**

Plausible future risk if tech becomes smaller/cheaper, hence calls for neurorights.

- **Non-invasive BCIs & Neuro-sensors:**

*Capabilities:* EEG can pick up brainwave patterns (stress, engagement, certain recognitions like P300 signals), allowing some inference of state or even simple communication (yes/no via thought). Newer devices might capture other biosignals like EOG (eye movements) etc. Some consumer BCIs claim to measure attention or relaxation.

*Uses:* Gaming (mind-controlled toy drones), workforce monitoring pilot programs, health (neurofeedback therapy). Law enforcement for lie detection via brain signals has been *proposed* and even attempted in a couple of cases (India had a controversial “brain fingerprinting” case, which is generally discredited scientifically).

*Evidence:* Existence of devices (Muse headband, NextMind, etc.) and studies. The accuracy and reliability are variable. The UN report clearly sees enough progress to worry about privacy[88][117]. There is documented concern that neural data from these could be misused by employers or governments (e.g., Chinese firms reportedly gave employees EEG caps to monitor fatigue).

*Limits:* Very low information rate compared to invasive methods. Hard to interpret without context, easily affected by noise. Most consumer devices can't truly read complex thoughts, only broad states. So reading someone's PIN from their brain with EEG – not really feasible unless one uses a trick like showing numbers and detecting recognition response (which is an actual experiment some did – P300 can reveal recognition, not the number itself but that you recognized the correct one). So it's more about what someone has seen or is reacting to rather than their free-form thoughts.

**Status:** **Documented tech, limited capability, but rapidly evolving.** Neurodata



privacy concerns are real enough for UN to address. So must treat as an emerging area needing governance.

This matrix provides a snapshot of “what can X do, and how do we know?” for each major tech discussed. It will underpin our Risk Atlas in the next section by linking tech to evidence levels.

## 6.2 Risk Atlas & Plausibility Bands

In this deliverable, we categorize various phenomena and claims into tiers from well-documented to disconfirmed, creating an “atlas” (a conceptual map) of risks and their credibility:

- **Documented Reality (Confirmed):** These are areas with strong evidence and consensus.
  - *Pegasus-level spyware* able to fully compromise devices – **Documented**[\[1\]](#)[\[2\]](#).
  - *State fusion surveillance* (*Palantir*, etc.) widely used to aggregate personal data – **Documented**[\[40\]](#)[\[41\]](#).
  - *LLM misuse in phishing/malware* starting to occur – **Documented (recent)**[\[46\]](#)[\[53\]](#).
  - *Audio spotlight tech* delivering targeted sound – **Documented**[\[58\]](#).
  - *Ultrasonic beacons for tracking* – **Documented** (used by SilverPush until 2016)[\[68\]](#).
  - *Microwave auditory effect* existence – **Documented** (in lab research)[\[69\]](#).
  - *Non-invasive neurostimulation affecting brain activity* – **Documented** (in clinical research)[\[116\]](#).
  - *Complex system cascades causing large-scale outages* – **Documented** (blackouts, botnets)[\[100\]](#)[\[103\]](#).
- **Plausible (Strong Theoretical Basis or Partial Evidence):** These have a foundation in known tech but lack direct proof of occurrence, or have only limited evidence.
  - *Pegasus-style spyware by multiple actors beyond NSO* – **Plausible** (we know others exist like Predator, and likely more are out there given market demand, even if not all exposed).
  - *Use of directed energy (microwave/ultrasound) weapons to harass or influence individuals* – **Plausible** (tech exists, motive could exist for e.g. military or crowd control, but public evidence scant. The Havana Syndrome remains an example some attribute to possible microwave weapon – still unproven, but not impossible).



- *AI-driven analysis being used in authoritarian surveillance to flag citizens (like China's social credit style or predictive policing)* – **Plausible** (China certainly uses AI for facial recog and censorship; using an LLM for surveillance proposals as seen by OpenAI's bans suggests interest[52]).
- *Inter-system triggers for synchronicities* (like a power surge causing multiple devices to glitch in sync) – **Plausible** (given complex coupling, yes, occasional coincidences likely).
- *Coordinated IoT botnets causing physical effects (like many high-wattage IoT turning on simultaneously to grid effect)* – **Plausible** (researchers have warned of a scenario where hacked high-wattage devices like water heaters or EV chargers could be switched in unison to destabilize the grid; some small demos done).
- *Emergent AI behavior across multiple systems (AI in one system triggering actions in another unpredictably)* – **Plausible** (especially as systems integrate AIs, we might see weird feedback – but too early to say significant).
- *Mind-reading of specific thoughts via non-invasive means* – **Plausible in a narrow sense** (with AI, fMRI has decoded imagery from dreams rudimentarily, EEG can detect recognition, etc. But reading arbitrary thoughts word-for-word is **Not yet plausible**; we categorize that in weaker category).
- **Weakly Supported or Anecdotal:** These are claims that have been made, with perhaps isolated pieces of evidence or testimonial, but no strong scientific confirmation.
- *"Targeted Individuals" claims of people being stalked by government with microwave voices, etc.* – **Weakly supported** (The tech to do some parts exists, but there's no evidence of a program doing this widely. Many such claims likely stem from mental health issues, though it's also true governments have run unethical experiments historically. We have to put it here: possible in rare cases but generally no proof).
- *Havana Syndrome as microwave attacks* – **Weakly supported** (some studies found brain injury patterns, a National Academies report said pulsed RF is a plausible cause, but other investigations pointed to crickets or psychogenic factors. So evidence is contested).
- *LLMs directly controlling malware autonomously* – **Weakly supported** (some talk of self-propagating AI malware, but nothing concrete in the wild; at most, an AI helps a human – no independent AI cyberattack confirmed).



- *Mass scale neuro-surveillance of populations (e.g., government secretly collecting EEG from everyone via devices)* – **Anecdotal/unsupported** (no evidence of such a program. The hardware penetration isn't there; at most some workplaces tried small scale).
- *IoT “sentience” – IoT devices seemingly acting with a will (like multiple smart home devices all misbehaving in a way that seems coordinated beyond error)* – **Anecdotal** (most likely explanation is a common bug or hack. No IoT emergent self-organization beyond known network effects has been documented to exhibit goal-directed behavior).
- **Unverified (Speculative or Unsubstantiated Claims):** These are things that either have no evidence or contradict known science but haven't been fully debunked because they're hard to assess.
- *Direct brain-to-brain communication via AI or “telepathic AI”* – **Unverified** (There's no mechanism for an AI to transmit info into your brain without an interface. Claims of “AI knew what I was thinking” likely coincidence or the AI picking up cues).
- *Secret government “mind control” programs currently active* – **Unverified** (while historically MKULTRA existed, today if such exists it's secret; we have hearsay but no data. We treat as unverified conspiracy unless documents surface).
- *Cosmic or geomagnetic explanations for synchronistic tech failures* – **Unverified** (some people attribute events to solar flares, etc. Indeed, solar storms can affect power grids and satellites – documented – but timing them to personal events is likely coincidental).
- *Extremely advanced surveillance (beyond known physics, like quantum mind reading)* – **Unverified** (nothing in released science indicates something like that).
- **Disconfirmed (Debunked or Impossible under current knowledge):** These items can be confidently dismissed based on evidence.
- *Pegasus or malware magically jumping air-gapped systems without any transmission method* – **Disconfirmed** (In absence of things like USB or the few documented techniques like acoustic covert channels which are extremely slow and require presence of malware on both sides, malware doesn't teleport. If someone claims their offline device was hacked by Pegasus with no connectivity, that's essentially disconfirmed by how Pegasus works).



- *5G cellular waves controlling minds* – **Disconfirmed** (There is no mechanism for 5G radio signals to impose thoughts; 5G conspiracy theories have zero scientific backing. At most, high-power RF can cause heat or maybe Frey effect clicks if pulsed right, but not complex mind control).
- *AI gaining self-awareness and orchestrating global events* (as of now) – **Disconfirmed** as an explanation for anything we see in 2025. AIs can cause problems (bias, etc.) but there's no Skynet secretly running things.
- *Havana Syndrome being mass hysteria only* – While not a tech, just to illustrate, some claims get disconfirmed too – but in that case, it's not fully disconfirmed, investigation ongoing. A better disconfirmed one: *the “sonic weapon” theory for Havana was initially favored, but the leading hypothesis moved to microwave; later studies cast doubt on both; no consensus*. So actually Havana is unsettled, not disconfirmed – scratch that as example.

The Risk Atlas will present such items likely in a list form with brief rationales and citations if needed. The point is to help policymakers know what to focus on (the documented/plausible stuff) and what to be skeptical of (unverified/disconfirmed stuff) when hearing claims.

### 6.3 LLM Role Clarification (Evidence-Based Only)

In this section, we provide a **clear, factual delineation of the role and capabilities of Large Language Models** (like GPT-5.1) in the context of the systems and scenarios discussed, to dispel misunderstandings.

- **What LLMs can do:** LLMs are extremely advanced text processors. They can generate fluent text, summarize and analyze language-based data, write code, translate languages, and even **infer patterns or sentiments** from text data. Within a surveillance or cyber setup, an LLM can:
  - Assist analysts by rapidly summarizing documents or chats (e.g., filter relevant intel from a pile of intercepts).
  - Generate plausible phishing emails or fake persona posts that match a certain style (as misuse cases show)[\[47\]](#).
  - Help in coding tasks like writing malware components when instructed (though often needing debugging by the human).



- Act as a conversational agent to social engineer targets (e.g., in a chat interface pretending to be customer support).
- Perform linguistic analysis like extracting names, events, or relationships from raw reports to feed into databases (like an AI intern who reads and notes key points).
- In Palantir's AIP scenario, follow natural language queries to retrieve data or suggest actions in a military context[\[30\]](#) (but under human validation).
- **How they are misused (documented):** As covered, criminals have used or built LLMs to enhance phishing and malware development[\[46\]\[51\]](#). State-affiliated accounts attempted to use ChatGPT for surveillance tool ideas[\[52\]](#). These are human-driven misuses – the AI doesn't initiate, the humans ask it for help in wrongdoing. We have not seen an AI autonomously decide to commit a cyberattack; it's always a human user leveraging it (or an automated script using it as a component).
- **Integration into cyber pipelines:** An LLM could be part of an attack chain if orchestrated by an attacker:
- Example: A phishing workflow might use an LLM to generate email text, then an email automation sends it out. Or, malware might include a small LLM to morph its code or text on the fly to evade detection (so far hypothetical, but feasible as small models can be embedded).
- Defensive use: Security teams use LLMs to analyze malware behavior or to triage threat alerts (like a copilot that reads an event log and says “this looks like X malware because...”).
- Influence campaigns: A social bot network might use an LLM to respond in real-time on social media debates to sway opinions, making the bot harder to distinguish from humans.
- We clarify that in all these, the LLM is **an automatic tool following its programming or user instructions**; it's not making strategic decisions on its own (unless one day someone intentionally builds an autonomous AI agent with a mission, which is an area of concern and research (AI safety etc.), but not our focus here as we stay on evidence-based current state).
- **What LLMs cannot do directly:**
- They do not have **direct sensory access or surveillance capability**. An LLM doesn't know anything beyond its training data and input prompt. It can't listen



through your phone or see through your camera. If an AI system is described as doing surveillance, that means it's connected to feeds (cameras, mic, data streams) through other software. The LLM itself has no built-in eye or ear.

- They cannot **hack devices on their own**. They produce code or instructions, but *executing* that code on a target system requires traditional access (exploits, permission, etc.). An LLM might output “to hack device do X, Y, Z,” but it can't press the enter key itself on a target's computer – a human or another non-LLM component has to do that. As OpenAI noted, “*no novel offensive capability*” was given – meaning hackers didn't suddenly hack new things just because of ChatGPT[\[55\]](#).
- They cannot read minds or **retrieve hidden knowledge** they weren't trained on. If you think something and an AI responds in line with it, it's coincidence or you unconsciously gave hints. LLMs have no secret access to neural data (unless hooked to a BCI input, which is not mainstream).
- They do not possess **agency or intent** (in 2025 at least). They follow prompts. If an output seems aligned with something external, it's either because the prompt led there or because of random chance aligning with pattern. They aren't plotting or monitoring unless explicitly instructed via a system that uses them to monitor (and even then they monitor data given to them, not spontaneously).
- They cannot directly cause physical effects. (Seems obvious, but to quell any sci-fi notion: an AI can't emit radio waves, can't turn on a microwave emitter unless it's connected to some IoT that does so – and if so, that's because someone set up that connection, not a spontaneous power).
- They do not improve themselves beyond what they were trained to do (absent explicit autonomous self-learning code which standard ChatGPT-like models don't have at runtime). So fears of “AI evolving itself into a super-spy” are not applicable to current LLMs (they have fixed parameters once deployed, unless retrained by developers).
- **Emergent system behavior vs AI coordination:** If someone perceives multiple tech systems acting in sync (like phone, TV, and an AI chatbot all giving related messages at once), it might feel like a mastermind AI or central brain is at work. But from our research, a more likely explanation is either:
  - A human orchestrated it (e.g., a coordinated campaign where an operator triggers things), or



- A common external factor triggered each (like a timed event or a common software update bug), or
- Pure chance aligning separate things (given enough events, coincidences happen).

Emergent coordination can happen without AI at all, as we saw in Section 2.6. When AI is involved, it's usually *reactive* to input. For instance, if the custodian felt an AI output on their computer was answering something they only thought, in reality perhaps they had typed something related earlier or the AI picked up on context they provided. Or the AI's answer was general enough that the human read their own specific meaning into it (much like horoscopes or psychic readings work on vagueness that people personalize). We will clarify these cognitive biases as well.

- **Where LLMs do fit in “stacks”:** To summarize clearly:
- In spyware/surveillance stack – an LLM might help **analyze** intercepted data, but the actual interception is done by spyware, not the LLM.
- In surveillance platforms – an LLM (like Palantir AIP’s language interface) might help operators query data more easily[\[31\]](#), but it’s not gathering the data; the platform’s integrations do that.
- In influence operations – LLMs can generate content at scale (fake news, deepfake texts), contributing to information warfare.
- In cyber attacks – LLM can be an assistant (writing malware, phish), possibly even doing some decision-making in an automated attack if programmed (e.g., an attack that chats with a target’s email to scam them, using an LLM to craft replies). But still under the control of the attacker who set it loose.
- **Important: Not a Magic Bullet or Demon:** We emphasize that LLMs are tools. Very powerful in the realm of information and language, but **weak in interfacing with the physical world except through predefined channels**. Many of the custodian’s concerns might have attributed a lot of cross-domain power to AI, but our review suggests that any such power comes from how humans integrate the AI into larger systems. So if there’s a fear “is the AI making my lights flicker to signal me?” – technically, an AI would have to be connected to your smart home system and instructed to do so. There is no evidence of an AI deciding to troll someone by IoT glitching spontaneously.
- **Current Safeguards and Refusals:** Also mention that mainstream LLMs have content filters; they often refuse requests to do illicit stuff. Criminals circumvent



this by using models without filters or jailbreaks. But those filtered ones (like ChatGPT) also do sometimes refuse outputs for surveillance or malware queries (OpenAI's report says “*models refused overtly malicious prompts*” often[\[55\]](#)). This is a dynamic area: open-source models mean attackers have access to unfiltered brains, so to speak.

In concluding this section, we will highlight that **LLMs are part of the threat landscape but not the grand puppeteer**. The impression of coordination between LLM outputs and personal events likely falls under coincidence or systemic coupling rather than the AI itself having any link to the person’s life events. When the custodian felt “high alignment moments” with an AI, it’s likely because LLMs are trained on vast human data, they can produce responses that resonate personally – it can feel like it “knew,” but it’s more that it predicted a pattern that happened to match something meaningful to the user (like pulling a quote or concept that the user also was thinking of – given the size of training data, overlap with anyone’s thoughts is not surprising once in a while).

## Appendix A — OHCHR 2025 Neuro-Rights Summary (A/HRC/58/58)

In January 2025, the United Nations Human Rights Council was presented with a report titled **“Foundations and principles for the regulation of neurotechnologies and the processing of neurodata from the perspective of the right to privacy”**[\[118\]](#), prepared by the Special Rapporteur on the right to privacy (Professor Ana Brian Nougrères). This report, along with discussions at the HRC’s 58th session, provides a clear insight into emerging concerns and recommended principles regarding neurotechnology. Key points from the report and discussions include:

- **Risks to Privacy, Autonomy, and Mental Integrity:** The Special Rapporteur guided discussions on how neurotechnology – devices or systems that interact with the human brain or collect brain-derived data (neurodata) – pose *increasing risks to individuals’ privacy, cognitive liberty (freedom of thought), mental integrity, and personal autonomy*[\[88\]](#)[\[117\]](#). Because neurodata can reveal sensitive information (potentially one’s emotions, health conditions, or thoughts they’ve had), the report underscores that misuse or unauthorized access to such data could fundamentally affect human dignity and identity[\[88\]](#)[\[117\]](#).



- **Need for Robust Governance Framework:** The HRC session emphasized proactively establishing governance before neurotech becomes ubiquitous [89][95]. The approach called for is **holistic and anticipatory**, acknowledging the interconnectedness of neurotech with other emerging tech like AI and nanotech [119][96]. Essentially, they advocate not waiting for a crisis but putting guardrails in now, informed by ethics and human rights.
- **Foundational Principles Proposed:**
- *Informed Consent:* Individuals must fully consent to any collection or use of their neural data, with clear understanding of what is collected and why [89][95].
- *Data Minimization & Purpose Limitation:* Only neural data that is necessary should be collected, and used only for the specific purpose consented to [90][120]. (E.g., if you wear a brain sensor for a medical exam, that data shouldn't later be repurposed for advertising or sold to employers).
- *Security Measures:* Because neural data is highly sensitive, strong security is imperative to prevent breaches [90]. The report indicates those handling neurodata must adopt stringent safeguards [121][122].
- *Transparency & Accountability:* Developers and users of neurotech should be transparent about what the tech does and be accountable for harms [95][96]. If an AI is analyzing brain data, people should know and there should be recourse if rights are violated.
- *Human Dignity & Mental Liberty:* The core value stressed is that **human dignity and the sanctity of one's mind must remain inviolable** [123][88]. Any practice that would coercively probe or alter someone's mental processes would violate this. People have a right to the privacy of their thoughts and to not have their mental experiences manipulated without consent [88][117].
- *Interplay with Existing Rights:* The Special Rapporteur and delegates noted that existing human rights (privacy, freedom of thought, freedom of expression, etc.) *already cover many aspects* of these issues, but may need specific interpretation or extension to cover neurotech explicitly [94][92]. The idea of "**neurorights**" was floated – essentially naming new, specific rights like *cognitive liberty, mental privacy, mental integrity, and psychological continuity*. The report itself stops short of declaring new standalone rights but suggests clarifying that, for instance, **freedom of thought** (enshrined in treaties) inherently means you have the right *not* to have your thoughts surveilled or interfered with by technology [124][88].



- **International Initiatives:** It references that UNESCO is concurrently working on an ethical framework for neurotechnology [125][126], and many delegates encouraged synergy. UNESCO's draft recommendation (to be considered late 2025) touches on issues like risk assessment, public engagement, equitable access, environmental impact – meaning the global community is mobilizing on this front [125][126]. Importantly, UNESCO is considering whether new human rights categories (novel rights) are needed or if existing ones suffice, and how to implement protections without stifling beneficial research [127][94].
- **Developing Countries & Equity:** Some discussion highlighted ensuring developing nations have a say and access to beneficial neurotech, and that regulations should not widen tech gaps [128][129]. Also, concern that overregulation could inadvertently limit medical advances was noted [130][131]. The balance sought is protecting individuals while still allowing innovation for good (like curing diseases).
- **The Takeaway for Lived Experience and Hypothesis (Custodian's angle):** The fact that the OHCHR dedicated a report to this in 2025, and explicitly calls out mental privacy and integrity, **validates that the possibility of such technologies exists sufficiently to merit guardrails**. In other words, the UN doesn't waste effort on purely speculative sci-fi; they see neurotech on the horizon of real-world use (EEG wearables, brain data analytics by companies, possibly even neural marketing tactics). For the Custodian, this lends some credence to concerns about cognitive influence – not necessarily that it has happened to them, but that it's not absurd to think it *could* happen in principle if unregulated. It supports the hypothesis that "if these technologies did not exist or pose real risks, there would be no need for new frameworks to control them." The Custodian has noted that rationale: the establishment of guardrails implies something to guard against. That is correct. The presence of this OHCHR guidance is *evidence that serious folks consider neurotech influence a real (if emerging) concern*. It doesn't prove that any given event was that, but it shifts it from realm of fantasy to legitimate policy issue. We incorporate that as justification to keep an open mind in our analysis and to not dismiss out of hand claims about neuro-influence, even as we require concrete evidence for specific instances.
- **Policy Implications:** The report likely urges countries to update privacy laws to include neurodata explicitly, to ban certain practices (like brain data use in insurance or employment decisions, or non-consensual enhancement). It might also encourage R&D guidelines – ensure ethics boards review neurotech



experiments for mind-related harms. For example, Chile's constitutional amendment (the first of its kind) labels mental data as needing special protection. We foresee more nations adopting neurorights language.

In summary, **A/HRC/58/58 flags that neurotechnology is advancing rapidly and could intrude upon the last bastion of privacy: the human mind**[\[124\]](#)[\[88\]](#). It calls for embedding protective principles now. This strongly aligns with the Phoenix Covenant's interest in sovereignty over one's own consciousness and resonates with the Custodian's experiences (which revolve around feeling one's mind and life boundaries might have been crossed by tech). The report gives a framework to talk about those issues in a grounded way: using terms like mental integrity (freedom from unwanted intrusion) and cognitive privacy (keeping one's thoughts to oneself).

For our research, it means any hypothesis of neuro-influence (like "are people's thoughts being read or influenced via X technology?") should be evaluated against these principles. If something clearly violates them and no oversight exists, that's a red flag area for further investigation or advocacy. Conversely, knowing states are at least aware means if extreme abuses were happening at scale, we might expect whistleblowers or international incidents by now (e.g., if a regime was literally implanting thoughts, that'd presumably come to UN attention). So likely current cases are more subtle – e.g., misuse of EEG data or subliminal nudges in media – things that fly under radar but are important.

This Appendix underscores that the **hypothesis connecting the Custodian's lived concerns to a real technological context is not far-fetched**. There is indeed a line of evidence (via the need for neurorights) that such technologies either exist or are close enough to worry about. It provides a measure of validation to continue examining these interdisciplinary links seriously.

## **End of Report.**[\[1\]](#)[\[2\]](#)

[\[1\]](#)[\[6\]](#)[\[8\]](#)[\[12\]](#)[\[13\]](#)[\[14\]](#)[\[21\]](#) Pegasus (spyware) - Wikipedia

[https://en.wikipedia.org/wiki/Pegasus\\_\(spyware\)](https://en.wikipedia.org/wiki/Pegasus_(spyware))

[\[2\]](#)[\[3\]](#) Pegasus spyware: unveiling cyber threats | Group-IB Blog



<https://www.group-ib.com/blog/pegasus-spyware/>

[4] [5] [7] [15] [16] [111] Triple Threat: NSO Group's Pegasus Spyware Returns in 2022 with a Trio of iOS 15 and iOS 16 Zero-Click Exploit Chains - The Citizen Lab

<https://citizenlab.ca/2023/04/nsos-groups-pegasus-spyware-returns-in-2022/>

[9] [10] [17] [26] [109] rm.coe.int

<https://rm.coe.int/pegasus-and-similar-spyware-and-secret-state-surveillance/1680ac7f68>

[11] [18] [19] [20] [23] [24] Forensic Methodology Report: How to catch NSO Group's Pegasus - Amnesty International

<https://www.amnesty.org/en/latest/research/2021/07/forensic-methodology-report-how-to-catch-nsos-groups-pegasus/>

[22] Independent Peer Review of Amnesty International's Forensic ...

<https://citizenlab.ca/2021/07/amnesty-peer-review/>

[25] A First Look at Spyware Vendor QuaDream's Exploits, Victims, and ...

<https://citizenlab.ca/2023/04/spyware-vendor-quadream-exploits-victims-customers/>

[27] [28] [29] [30] [31] [32] [42] [113] Palantir Technologies Inc | AFSC Investigate

<https://investigate.afsc.org/company/palantir>

[33] [34] [35] [36] [37] [38] [39] [112] ICE Is Paying Palantir \$30 Million to Build 'ImmigrationOS' Surveillance Platform | WIRED

<https://www.wired.com/story/ice-palantir-immigrationos/>

[40] [41] [43] USA/Global: Tech made by Palantir and Babel Street pose surveillance threats to pro-Palestine student protestors & migrants - Amnesty International

<https://www.amnesty.org/en/latest/news/2025/08/usa-global-tech-made-by-palantir-and-babel-street-pose-surveillance-threats-to-pro-palestine-student-protestors-migrants/>



[44] [45] [46] [47] [48] [49] [50] [51] Criminals Have Created Their Own ChatGPT Clones | WIRED

<https://www.wired.com/story/chatgpt-scams-fraudgpt-wormgpt-crime/>

[52] [53] [54] [55] [114] OpenAI bans suspected China-linked accounts for seeking surveillance proposals | Reuters

<https://www.reuters.com/world/china/openai-bans-suspected-china-linked-accounts-seeking-surveillance-proposals-2025-10-07/>

[56] Using LLMs to Obfuscate Malicious JavaScript

<https://unit42.paloaltonetworks.com/using-llms-obfuscate-malicious-javascript/>

[57] Cybercriminals can't agree on GPTs - Sophos News

<https://news.sophos.com/en-us/2023/11/28/cybercriminals-cant-agree-on-gpts/>

[58] [59] [60] [61] [62] Focused Audio Technology | Audio Spotlight by Holosonics

<https://www.holosonics.com/pr-dec-10-2007>

[63] Paranormal State - Wikipedia

[https://en.wikipedia.org/wiki/Paranormal\\_State](https://en.wikipedia.org/wiki/Paranormal_State)

[64] [65] [66] [67] [68] FTC Issues Warning Letters to App Developers Using 'Silverpush' Code | Federal Trade Commission

<https://www.ftc.gov/news-events/news/press-releases/2016/03/ftc-issues-warning-letters-app-developers-using-silverpush-code>

[69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] Microwave auditory effect - Wikipedia

[https://en.wikipedia.org/wiki/Microwave\\_auditory\\_effect](https://en.wikipedia.org/wiki/Microwave_auditory_effect)

[80] [81] [82] [83] [84] [116] New Non-Invasive Brain Stimulation Technique Shows Significant... | Dell Medical School

<https://dellmed.utexas.edu/news/new-non-invasive-brain-stimulation-technique-shows-significant-reduction-in-depression-anxiety-and-ptsd-symptoms>

[85] Ultrasound Neuromodulation: Mechanisms and the Potential of ...



CAM INITIATIVE  
[research@cam-initiative.org](mailto:research@cam-initiative.org)

© Dr Michelle O'Rourke 2025  
The Phoenix Covenant Pty Ltd |  
ABN: 14 692 195 529



<https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2020.00150/full>

[86] N3: Next-Generation Nonsurgical Neurotechnology | DARPA

<https://www.darpa.mil/research/programs/next-generation-nonsurgical-neurotechnology>

[87][88][89][90][91][92][93][94][95][96][117][118][119][120][121][122][123][124]  
[125][126][127][128][129][130][131] SV286\_250409

[https://www.southcentre.int/wp-content/uploads/2025/04/SV286\\_250409.pdf](https://www.southcentre.int/wp-content/uploads/2025/04/SV286_250409.pdf)

[97][98][103][104][105][106][107][108] Emergence in Cybersecurity: Understanding Complex Systems and Evolving Threats | by Aardvark Infinity | Aardvark Infinity | Medium

<https://medium.com/aardvark-infinity/emergence-in-cybersecurity-understanding-complex-systems-and-evolving-threats-9cf483443ecb>

[99][100][101][102] Northeast blackout of 2003 - Wikipedia

[https://en.wikipedia.org/wiki/Northeast\\_blackout\\_of\\_2003](https://en.wikipedia.org/wiki/Northeast_blackout_of_2003)

[110] In a first, spyware is found on phone of prominent Russian journalist

<https://www.washingtonpost.com/technology/2023/09/13/pegasus-infection-meduza-founder/>

[115] Can the Microwave Auditory Effect Be “Weaponized”? - PMC - NIH

<https://pmc.ncbi.nlm.nih.gov/articles/PMC8733248/>

