Article | Received 1 November 2025; Revised 9 December 2025; Accepted 25 December 2025; Published 9 January 2026
https://doi.org/10.55092/let20260001

AIWitness testimony: visual neuroprostheses, AI-mediated perception and the law of evidence

Claudia González-Márquez* and Burkhard Schafer

Edinburgh Law School, University of Edinburgh, Edinburgh, Scotland, United Kingdom

* Correspondence author; E-mail: cgonzal2@ed.ac.uk.

Highlights:

- Applies extended mind theory to AI-mediated perception.
- Argues AI-prosthetic "artificial vision" is eyewitness, not computer evidence.
- Introduces the term "AIWitness", a human-AI unit producing trial-relevant perception.
- Grounds parity in continuous human-machine coupling, not mere phenomenology.
- Maps implications for reliability and fair-trial rights.

Abstract: Advances in neurotechnology are developing rapidly, creating challenges for legal systems whose basic categories date back to the 18th and 19th centuries. AI-mediated sensory perception challenges core categories of evidence law. This paper focuses on the distinction between eyewitness and computer evidence, as well as "direct" and indirect evidence. Specifically, AI-enabled visual neuroprosthetics—modern neuroimplants that combine artificial intelligence to stimulate the visual cortex and produce "artificial vision" for visually impaired individuals—can make a witness's experience algorithmically constituted rather than merely a recording of the environment. Drawing on Andy Clark and David Chalmers' extended mind theory and a dynamical-systems criterion of ongoing, bidirectional coupling, we argue that when perceptual content results from a continuous, tightly coupled human-device interaction that functionally replaces a sense, the outcome is the person's own perception. Based on this, when a human-algorithm unit whose perception relevant to the trial is materially shaped by AI—an "AIWitness"—courts should regard the witness's testimony as eyewitness evidence, rather than a separate machine output. The law should not dismiss extended cognitive systems merely because part of cognition occurs outside the skull. Grounded in principles of equality, participation, and fair trial values, this normative account upholds the status of disabled citizens as witnesses and not merely data sources.

Keywords: AI-enabled neurotechnologies; eyewitness testimony; evidential law; artificial intelligence; technologically-mediated perception; intracortical visual neuroprostheses; extended mind

1. Introduction

AI-enabled visual neuroprosthetics are transitioning from research labs into daily life. Unlike traditional assistive devices that simply relay or store information, modern neuroimplants interact with the nervous system and simulate the most suitable interaction between external (and increasingly semi-internal) sensors and the visual cortex to offer some degree of visual perception through stimulation [1]. By utilising machine learning (ML), these devices filter, reconstruct, and interpret the visual content available to the wearer [2]. For example, in cortical visual neuroprostheses, environmental signals are converted into patterned neural stimulation, meaning that what is perceived at any given moment is partly an algorithmically generated experience—an artificially mediated form of vision [3], for individuals who are visually impaired or blind and rely heavily on these neurotechnologies to perceive their environment. These advancements challenge the fundamental legal distinction between eyewitness evidence (a human report of perception) and digital evidence (a device-produced output).

For courts, selecting a category is more than simply assigning a label. If AI-mediated perception is considered as ordinary eyewitness evidence, some of the "excessive" generation of recordings will be as excluded from the trial as the scrutiny of technical features vital to reliability. Conversely, if it is classified as computer evidence, testimonies from disabled witnesses risk being depersonalised into mere data sources, like CCTV footage, which deprives the human wearer of these technologies of basic control and reduces them to simple mobile surveillance devices. The doctrinal dilemma then is how to cross-examine an eyewitness whose experience is inextricably linked to a device. Both options raise issues concerning contestability, dignity, participation, and equality.

This issue has received limited attention in academic legal scholarship. Existing discussions on eyewitness reliability, admissibility standards for scientific and technical evidence, and presumption of proper computer function offer helpful tools, but none directly address how to admit and evaluate testimony where human perception is entirely AI-generated rather than simply aided or recorded.

The question we face is how the law should treat a witness's trial-relevant AI-generated content, created through an ongoing bidirectional loop that substitutes for a cognitive ability the individual otherwise lacks. We frame this issue through Andy Clark and David Chalmers' extended mind theory: if an external process is functionally integrated into human cognition, we argue that the law should aim for parity rather than focusing on the fact that part of perception occurs outside the individual. Using this theoretical perspective, we suggest that a parity approach can provide a normative basis for treating AI-mediated perception as first-hand human experience by the individual, in the same way as unaided sight.

While acknowledging that an "AIWitness"—a person whose trial-relevant perception is materially mediated by algorithmic processes shaping their perception—could also be a user of a neuroprosthetic system that assists, enhances, or restores other damaged sensory functions of the brain, our focus here is on criminal proceedings and advanced intracortical visual prostheses. We note, however, that much of what we discuss is likely to be applicable to neuroprostheses that repair or replace hearing or touch, such as AI-supported cochlear implants like those outlined by Zhang *et al.* [4]. Additionally, our legal analysis discusses the issues on a level of abstraction that should make our design recommendations suitable for both common law and civil law jurisdictions (at least). While it remains jurisdiction-neutral, we draw in particular on the common law systems of Scotland and England, as well as on ideas and insights from

the US and, for civilian jurisdictions, German law. This also means that while we provide some recommendations for legal systems, we can only very briefly touch on jury instructions, and in particular, can't discuss how the different involvement of lay judges in civil law systems could benefit from a similar approach.

We restrict our analysis to the so-called *sensory extension*—neuroprostheses that substitute for, or restore, a basic sensory modality the person would otherwise lack—in our case, vision. We do not address devices that primarily deliver *informational extension* such as wearables that query external databases, perform facial recognition, or generate analytic overlays on top of an intact visual field. In those cases, hearsay, machine-testimony and expert evidence concerns are considerably sharper, and our parity account should not be interpreted to classify all AI-assisted cognition as eyewitness evidence.

A parallel question arises in civil law systems, where fact-finding relies less on exclusionary rules and more on judicial management of the evidence. In such systems, the parity account we develop would operate less as a threshold admissibility test and more as a constraint on how court-appointed experts and judges evaluate technologically-mediated perception.

This paper is divided into three sections: Part 2 explains how intracortical visual neuroprostheses generate AI-mediated perception and why that unsettles the eyewitness/computer-evidence divide. To support our analysis, it introduces "AIWitness" as a descriptive label for those witnesses who rely on ML-generated activation of their cortex for the subjective "experience" of vision. We then use a brief hypothetical to anchor the discussion. Part 3 outlines the normative relevance and explores the philosophical basis of the extended mind theory as a framework. We argue that this approach helps us navigate these challenges. To demonstrate this, we develop a parity account based on Clark's extended mind theory—strengthened by a dynamical-systems criterion of ongoing bidirectional coupling—to show why AI-mediated perception can count as first-person knowledge. Building on this theoretical analysis, Part 4 applies the framework to visual neuroprostheses. Using the "ongoing feedback loops" criterion of constitution, we argue that where perception exists only through continuous, bidirectional human–device coupling, the wearer and the neurodevice should be treated as a single entity in law, and the witness report should be admitted as eyewitness evidence within existing doctrine. We conclude with broader implications for admissibility, corroboration, and fair-trial principles.

2. What the AIWitness saw: the technology

This section introduces technical aspects related to intracortical visual neuroprosthetics—the central technology discussed in this paper—and begins to explain why these neurotechnological systems present complex challenges for evidence law.

Neural interfacing technologies have progressed considerably over the past decade. In this field, neuroprostheses that incorporate ML have achieved notable advances in restoring sensory functions, especially vision. Today, AI-enabled visual neuroprosthetics are no longer just external devices that record or transmit images to the human eye. The most advanced systems consist of implanted interfaces that bypass damaged ocular structures and produce perceptual content by calculating stimulation patterns for the visual pathway in real time. For example, blind individuals implanted with an intracortical visual prosthesis have reported visual percepts [5]. Understanding this process—sensing, algorithmic transformation, and neural stimulation—is important for our discussion because it explains why the

resulting "seeing" is formed by the human–machine loop rather than being a detached file that the person later views.

Earlier assistive technologies relied on external devices such as smart glasses, head-mounted cameras, and lifeloggers (like Microsoft's SenseCam)—devices that are easily detachable from the human body and whose outputs can be regarded as separate artefacts that the viewer can evaluate against their unaided vision, at least to a certain extent [6]. Even in these cases, we can ask if separating the device from its wearer for evidential purposes is the best way of ensuring ethically compliant evidence collection that also adheres to national and international safeguards for a fair trial. This is especially true when such systems are used for therapeutic purposes and within a feedback loop with the wearer—examples include some versions of SenseCam—and we might question whether separating the camera from the wearer without any further guarantees or safeguards is justifiable. Indeed, some of the ethical concerns we will raise below for AIWitnesses are also evident in these older technologies. They pose risks and opportunities for both the wearer and their environment that may require further consideration and analysis. However, for the purposes of this paper, we accept that from a legal standpoint, the easy access to stored data and the "fungibility" of these systems mean they are more readily addressed by existing legal rules that emphasise their "designed" origin.

A similar dilemma arises in case law on technology-mediated perception such as police use of thermal imagers, night-vision devices, or enhanced audio. In those cases, courts have had to decide whether officers testify from personal knowledge aided by a tool, or whether the device's output should be treated as independent scientific or machine-generated evidence. Our focus here is on implants that substitute for sight altogether. Unlike Forward-Looking Infrared (FLIR) cameras or smart glasses, which can be removed to reveal an unassisted baseline view and often generate a detachable recording, cortical visual neuroprostheses mediate the wearer's entire visual stream and do not leave a separate replayable output beyond the witness's experience itself. This is why FLIR-type systems are tempting but ultimately unsuitable for analogous application of existing evidential rules on machine-generated evidence: the conceptual separation of the images that they generate from their user makes them more similar to traditional camera images which generate an independent, second recording of the events in question.

Instead of creating a permanent video recording of the environment that anyone with normal visual capabilities can access, advances in functional visual neuroprosthetics have shifted toward implanted systems that directly stimulate retinal, optic nerve, or cortical targets, which remain highly personal for the user. Retinal arrays such as the Argus II and their successors exemplify this shift, but modern neuroprosthetics have moved towards invasive approaches (involving implanted microelectrode arrays) that aim to restore vision for visually impaired or blind individuals [7–9]. These systems are designed to reconstruct visual information tailored to a particular user's nervous system, allowing the wearer to detect light, motion, and recognise characters.

Crucially for a legal analysis, these modern visual implants operate as closed-loop systems as they continuously decode neural signals in order to substitute malfunctioning brain structures [10]. This means that they gradually lose their "fungibility"; the way the system of wearer 1 translates external input into cortical activity can and will differ from that of wearer 2. ML mimics neurons' behavior and reproduces "sensory feedback" to enhance sensation [11]. In other words, sensors stream data to embedded processors running ML models that select, enhance, and compress features such as edges, contrast boundaries, motion cues, or object-level regularities, within the limits of the user's implanted

electrode array [12]. The embedded AI then computes a stimulation pattern that is delivered to cortical or retinal tissue, eliciting phosphenes and spatiotemporal patterns that the user experiences as visual content, enabling what is often referred to as "artificial vision" [13]. The loop "closes" as motor actions like head and eye movements change the incoming signal. The key point is that perception at time t results from a bidirectional, dynamic exchange among the environment, algorithms, electrodes, and neural plasticity [14]. This process follows general patterns, but still responds to the unique condition of each wearer.

To facilitate our discussion below, we use the term "AIWitness" as a descriptive shorthand for a witness whose trial-relevant perceptual experience is entirely mediated by algorithms, and is ontologically dependent on neurotechnology—in particular, an intracortical visual prosthesis. The label is purely conceptual, not doctrinal, as we do not intend to propose a new category of evidence or a new legal status. We simply use it to refer to cases where the human–device unit is the pathway through which perception occurs—a point we explore further in Section 3.

2.1. Algorithmic mediation and perceptual content

The crucial point for the law is that these novel systems do not function like passive cameras or smart glasses. They rarely transmit pixel arrays that could be later inspected to re-create "what was there". Instead, they perform task-directed transformations constrained by safety, hardware, and user-specific thresholds. In very simple terms, these implants compute and deliver perceptual content in real time and help constitute perceptual judgement. In practice, the model may suppress texture, amplify boundaries, or prioritize faces, bodies, or objects depending on settings and user calibration. Some platforms integrate non-visual signals to stabilize a scene, so that what is experienced as "visual" may be a multimodal computation [15]. In short, the subject's view is an adaptive rendering: a computed interpretation of the scene designed to be legible to their nervous system. That is why the wearer's first-person report is not downstream of a detachable recording; it is the primary manifestation of the system's computation in conscious experience.

Users acquire "artificial vision" through continuous training and neuroplastic adaptation. Early experiences are often limited (for example, shimmering outlines), yet, with practice, individuals learn to scan, combine, and stabilise a visual field adequate for navigation and basic recognition. Reports are often described in "quasi-visual terms": shapes, motion trajectories, and color cues [16]. The phenomenology is entirely skill-dependent, as two users with the same neuroimplant may develop different perceptual affordances.

These initial considerations highlight some aspects of the problem this paper seeks to address. Specifically, we can begin to understand why AI-enabled neuroprosthetics raise issues that ordinary cameras do not. In the smart glasses example, there remains an independent biological pathway from the world to memory. If we separate the recording function of the smart glasses from the device that creates it, we can, without significant loss of functionality, reproduce for everyone what has been recorded and submit a separate record of the recording. Consequently, in the recording device case, the artefact produced can be admitted (or contested) as digital evidence. In the case of neuroimplants of the type described above, there is no independent pathway; the only route by which the scene becomes accessible to the individual is through the algorithmically computed stimulation delivered in real time. This is why, in doctrinal terms, the question is not whether a separate "machine statement" exists. It is whether the

law recognises that, where perception is only made possible by this continuous, tightly integrated computation, the resulting experience is considered the person's own "vision" for evidential purposes.

Before delving into our analysis of testimony based on the "observations" of someone who "sees" with the aid of an AI-enabled neuroprosthetic, it is relevant to briefly clarify the use of the term "eyewitness evidence" for the purposes of this discussion. Here, the notion of eyewitness evidence is used in a strict sense—this means we do not distinguish between evidence given by bystanders, victims, or confession evidence. Since all these are based on the direct sensory experience of the person who testifies, we employ the term for all forms of direct, sense-based evidence presented through the witness's testimony, to contrast it with expert witness evidence.

2.2. The AIWitness and the law

As computational technologies become integrated into daily life, a key legal concern is the uncertain status of the user or wearer, the device, and the data it generates. Specifically, as AI-generated data and directly perceived environments become more intertwined, we may question whether the human experience we perceive truly belongs to the individual or if the AI is subtly influencing the wearer to accept what it has processed and inferred.

This brings us to the issue of perception mediated by technology. Is AI-mediated perception a legitimate form of eyewitness testimony? To illustrate, consider the following hypothetical situation: X is severely sight-impaired and constantly depends on an intracortical visual prosthesis that provides low-resolution edge and motion cues through real-time AI processing. One evening, X was in the common room of his assisted housing complex when a theft took place. X claims to have observed a crime. For the investigation, it matters whether two people entered from the main street, or whether a single individual emerged from the staff-only office door on the opposite side. X reports perceiving two tall, high-contrast outlines moving from left to right after the street door opened, but the defence argues that the device's settings—edge enhancement combined with motion smoothing in low light—could make one fast-moving silhouette appear as two. How should X's testimony be treated?

Was it the wearer who made the identification, or the AI? Is the witness's first-person report sufficient, or must courts also verify the device's accuracy separately? If the latter, is it the individual device, worn by the patient, that needs to be evaluated for its accuracy and reliability, or are generic statements about the system sufficient? Should it matter doctrinally or normatively that the implant is a medically necessary assistive device rather than an enhancement? In a criminal trial, should evidence from a user of vision-restoring or vision-enhancing neuroimplants be regarded as eyewitness testimony subject to the rules of personal knowledge, or as machine-derived evidence presented through experts with its own technical basis? These questions remain largely unexplored, yet they directly affect admissibility, weight, corroboration, and even the legitimacy of allowing the device's use in court.

There are several ways to characterise testimony from a user of a visual neuroprosthesis. Firstly, we could treat the neurodevice as indistinguishable from normal eyes and ignore the technology altogether. This approach involves treating the individual as an eyewitness, leaving questions of credibility and reliability to standard cross-examination.

Secondly, we could uphold the eyewitness approach while recognising the unique mediation involved by adopting tailored jury instructions or an equivalent method to alert the fact-finders to the constraints and assumptions that must be made when weighing the testimony. Building on this, we could

also, either separately or together, allow expert evidence on the design aspects and known reliability issues of the type of device used, while refraining from calling experts to examine the specific device responsible for the impression in this particular case.

These strategies would equip juries with the tools needed to assess the credibility and reliability of witnesses who perceive things very differently from themselves. They would also help uphold the principle that judgment by one's peers remains a fundamental aspect of justice in common law systems. For example, this could involve allowing expert evidence to clarify how the human–device interface perceives. In the absence of any conflicting case law, this approach also appears to be consistent with, or suitable for, adoption by continental legal systems. German law, for instance, assumes that every person is generally capable of testifying, regardless of their physical or mental condition. The crucial factor is whether the individual can perceive through their senses, as well as recall and describe their perceptions [17]. This responsibility lies with the competent criminal justice authorities in each case, giving them, rather than the parties, the authority to request expert evidence on the type of machine used to determine if it meets the requirement of "sensory perception". As we will argue below, this generally applies to neurodevices, which, while offering a new form of sensing, clearly provide "a" form of sensory perception and do not affect the ability to "recall and recount". While a balance with the right to a fair trial is necessary under both German and European Union law, the rights of disabled person are also safeguarded through the German Basic Law. This contrasts with the situation in the US, where incorporating the confrontation right into the constitution can introduce additional obstacles. The guidelines for criminal and administrative offence proceedings (RiStBV) impose further duties on public prosecutors to show special consideration for individuals with disabilities. Although these are not legally binding, as administrative regulations, they do obligate the judicial administration staff to follow instructions [18]. This primarily includes public prosecutors; for judges not bound by instructions, these are merely procedural guidance, resulting in a "witness-centric" yet discretionary system.

Thirdly, we could treat the device and wearer as legally distinct, with the device's processes and logs regarded as computer evidence, authenticated through experts, while the human's role is limited to verifying use and context.

Finally, we could try to address any concerns about the evidence in a separate process "within" the trial, that does not involve the jury or, in a civilian system, the lay judges sitting with the professionally trained judge. In the US, the concept of a "Daubery hearing" may be a similar approach that separates what the ultimate decision maker hears from the discussion of the reliability of the scientific theories proposed for presentation. We should point out here that we refer to Daubert only as an example of pre-trial reliability screening, without undertaking a systematic analysis of US doctrine. However, the idea of separating in one form or another the trier of fact from the discussion of the reliability of the evidence is one important aspect of the conceptual framework that may be needed, and as our examples show, can be in principle reconciled with conceptions of the fair trial across a range of jurisdictions.

Each route promotes certain values while sacrificing others. Routes 1 and 2 support equality and participation by recognising the person as a witness instead of relying solely on machine-derived evidence. These options also protect privacy by not automatically inspecting sensitive neural data. However, both raise concerns about trial fairness that are not present with smart glasses or detachable cameras. Since the perceptual content is entirely mediated by the device, neither the finder of fact nor indeed the witnesses can meaningfully compare a malfunctioning or manipulated output with a "ground

truth," and testimony and device trace are, in essence, part of the same process. If we treated the case as straightforward eyewitness evidence under (1) there would be limited scope to test the reliability of the mediating system. Under (2) tailored instructions or explanatory expertise could help, but only to a certain extent.

The fourth option sidesteps some legal and ethical issues by separating the usual trial process from questions about witness testimony reliability. Although its effect on the continental European context may be limited, the clearer division between the trier of fact and the judge in common law systems offers several benefits. For instance, parties might pragmatically agree to accept, without question or entirely disregard, the AIWitness. Furthermore, during a mechanism akin to a US Daubert hearing, there could be an opportunity for the judge to identify any more specific instructions to the jury at the later trial stage that address only the relevant concerns about the witness in question. However, it is crucial to remember the added complexity that the life of an AIWitness involves, which could make them more vulnerable to unreasonable demands. Establishing clear legal rules that prevent them from needing to justify themselves even to the judge might be more effective in mitigating the social exclusion they are already likely to face in their daily lives.

It seems necessary to have at least some way to examine contestability by adopting (3) and focusing on the type of device that co-created with its wearer the "visual" report. The third option, however, seems even more problematic. An intracortical implant is integral to the body and to basic functioning when used to restore cognition. Unlike a lifelogging camera, which can easily be detached from its wearer and whose recordings are perceived by everybody watching them the same way, neuroprosthetics cannot simply be removed for inspection, nor do they create "fungible" recordings. As noted by various neuroethicists, neuroimplant explantation would raise complex ethical issues as it requires invasive surgery, and even remote access to the data (often stored in private servers) [19–21].

While we have so far focused on the evidential value of AIwitness technology, the legal and ethical problems extend beyond the fair-trial requirements of the proceedings. Treating the device as a freestanding evidential source risks turning a disabled citizen into a CCTV surveillance device. It engages privacy and self-incrimination concerns and weakens procedural protections for the individual and for third parties whose data the system may have processed. In addition, where device traces exist, they do not supply an independent source of corroboration; they track the same algorithmic pipeline that constituted the experience in the first place. Thus, while (3) maximizes technical scrutiny, it does so by downgrading the person, leading to the exclusion of citizens in their own trial, and reducing their participation rights.

These tensions explain the need for a principled account of perceptual authorship. The remainder of the paper develops such an account from the extended mind theory and dynamical coupling, and argues that where perception exists only through a continuous, bidirectional human–device loop, the law should treat the human–device unit as a single perceiver and the resulting report as eyewitness testimony, with reliability concerns addressed in the ordinary way as matters of weight rather than kind.

Having set out the technological background and the evidential tensions raised by AI-enabled visual neuroprostheses, we now ask whether—and on what grounds—Clark and Chalmers' extended mind thesis provides a principled basis to treat AI-mediated perception as the witness's own.

3. Normative foundations

This section will explore the normative relevance of the extended mind theory to the problem delineated in the preceding sections. We will discuss and assess whether this theoretical framework could help us navigate the legal challenges and tensions posed by technologically mediated evidence.

3.1. Functionalism and the parity claim

Back in 1998, Andy Clark and David Chalmers proposed the hypothesis of extended cognition and the extended mind theory [22]. According to these hypotheses, the mind and cognitive processes such as perception, problem solving, belief, and memory may transcend the biological body by incorporating external artefacts. Put simply, under certain coupling conditions between an agent and a device (which can be any external object), the latter can actively and constitutively contribute to the realisation of extended mental and cognitive processes.

Drawing from the philosophy of mind, our analysis adopts an equally functionalist approach—in which it is the *functional role* of a cognitive process, not necessarily its biological implementation, that matters. If what makes a process cognitive is the role it plays in solving a task, there is no reason to restrict cognition to the experience of biological tissue only. On this view, cognitive routines are not confined to neurons but can be realised through brain–body–environment systems. Clark and Chalmers' extended mind theory articulates this in the form of a parity claim: if an external non-biological entity functions like an internal cognitive process (or performs the same cognitive role as an inner process), it should be regarded as part of the mind [22].

In the Clark-Chalmers approach, a person solving mathematical problems with pen and paper, or playing Tetris by pressing a rotation button or using mental imagery, is functionally equivalent to tasks that remain solely "inside" the brain. In each case, the action is carried out by a human-artefact rather than by neural activity alone. If the external aid performs the same role as internal computation, then it is constitutive, not just causal, of cognition. Clark and Chalmers emphasize that for extended cognition to be valid, the individual must demonstrate a "high degree of trust, reliance, and accessibility" toward the device—often called the "glue and trust" criteria.

In Clark and Chalmer's famous thought experiment, "Otto" (an Alzheimer's patient), the notebook that stores facts for Otto, who consults it as continuously as others recall from memory, shows how parity works in practice: Otto's mind extends to the notebook as the artefact takes up the very role brain-based memory otherwise would. For the sake of simplicity, the notebook's content plays the same functional role as everyone else's biological memory. Depending on how we see this case, it may qualify as a case of cognitive extension [23].

However, the concept of the extended mind is not limited to memory. As echoed by Richard Heersmink, "cognitive artefacts do not just complement our memory, but a variety of cognitive capacities" [24]. Perception, as a lower-level cognitive process, consists of functions that extract task-relevant structure from the environment in real time. If this extraction is performed by a tightly integrated device that is in continuous exchange with the agent—sensing, interpreting, stimulating, and guiding action in a bidirectional, closed loop—the perceptual success becomes a property of the person–device system.

This is in part what systems that directly interface with human cognition, such as AI-enabled visual neuroprosthetics, highlight: cognition and sensory perception are often realized by real-time, bidirectional

processes that yield properties of the whole system not attributable to either component alone. In other words, an AI-powered visual neuroimplant that continuously computes, stimulates, and guides the user's orientation in their physical environment—and most relevant to this discussion, feeds perceptual information into the wearer's cognitive stream—functions just like an internal sense organ. If the cognitive task is functionally the same with or without the device, then the device should be regarded as successfully coupled with its wearer and, thus, be considered also to be a part of the user's extended cognitive system.

3.2. Why extension reaches perception

Although the extended mind theory concerns ontological and epistemological questions, it is possible to reinterpret the thesis in a normative way. If an external tool or artefact provides a disabled person with cognitive capabilities that others have through the internal operation of their mind, then, in this line of reasoning, the artefact-human coupling ought not to be treated disfavourably in comparison with internal memories.

Moreover, while "Otto's" thought experiment used memory as an example, the extended mind thesis arguably applies to AI-mediated vision. Philosophers, too, now extend the thesis to perception. For instance, Robert Rupert notes that the hypothesis claims that "human cognitive processes literally comprise elements beyond the boundary of the human organism" [25]. This invites viewing a vision-enabling AI as extending a person's perceptual state as the neurodevice (the artifact) interprets and assists the wearer in making sense of visual stimuli. Under such a functionalist view, a cognitive device can become "a constitutive part of the extended realization base" of the agent's mental states [26,27].

In practice, sensory substitution systems and neuroprosthetics validate this perspective. For example, empirical studies show that blind users learn through "continuous closed-loop exploration"—the user moves, the device's sensors pick up new input, the algorithm updates the stimulation, and that new percept guides the next movement, repeating in rapid two-way cycles—to perceive spatial layout via AI-driven feedback [28]. In these cases, the vision-like experience (such as locating objects or navigating space) clearly arises from the integrated human–machine system, not from either element alone. Let us unpack these considerations further.

The ability to identify objects, track shapes, orientation, and motion is not solely a property of the neurodevice or the neural tissue. Instead, this capability emerges from their trained, closed-loop cooperation. When such a system enables artificial vision for a blind individual, their "seeing" s simply an experience arising from that human-machine interaction. Therefore, when X, our blind witness, states that two people entered from the left towards the main entrance, he is not merely relaying a third party's statement; he is genuinely reporting what he perceived—albeit through an extended cognitive artefact. This view is supported by Barker [29] and Sprevak [30], who argue that (the extended mind theory) regards external elements as true cognitive causal contributors within a higher-level extended cognitive mechanism.

In this line of argument, it is methodologically sound to attribute the resulting perceptual content to the witness themselves. If all they have "seen" is through the implant, then their testimony of what they experienced is, in principle, testimony of their own extended perception. In such cases, the perceptual authorship still lies with the wearer because the device is simply shaping the very process by which the agent forms the perceptual judgment. This is why treating an AIwitness and a visual neuroprosthetic

itself as two different categories in law would mischaracterises both the witness's first-person experience and how the human and device function together. Evidence law's category of "personal knowledge" is conceptually adequate to accommodate that there is one perceiver (the AI-mediated witness as an extended cognitive system). Nothing about the notion of eyewitness testimony requires that the AI (as a sensory transducer) be organic.

However, the question is, when exactly is an external artifact *coupled* with its user? What conditions an external device must meet to functionally count as part of the wearer's cognition? Richard Heersmink's multidimensional integration framework [24] is a useful guide to answer this, and further sustains our contention to treat the agent and the neurodevice as a single unit in law. Heersmink proposed eight dimensions to help assess the extent to which an artefact has been integrated into some agent. According to this approach, the higher the artefact scores across the proposed framework, the stronger the case for it being part of an integrated extended cognitive system.

Applied to our context, AI-powered visual neuroprostheses are likely to score quite highly under dimensions such as information flow, reliability, and trust for various reasons. Firstly, the neurodevice functions as an always-on visual implant. Here, the artefact provides a constant stream of data, and the blind user heavily relies on it. In addition, as the device is physically embedded within the user's body—implanted inside their skull—and functionally replaces their sight, the system is highly individualized to the user.

Nevertheless, Heersmink's criteria comes with limits. It is a valuable tool to capture the degree of integration between an agent and artefact; however, relying on this account to validate claims of cognitive extension would lead to vague or imprecise assessments that cannot definitively confirm a case of cognitive extension as argued by Palermos [23]. As Heersmink notes, the drawback is that no particular score guarantees cognitive extension; a high-score cannot be objectively verified to conclusively establish successful artefact integration into an agent's biological cognitive system.

In an effort to define what qualifies as cognitive extension, scholars Reiner and Nagel noted that "[…] there is a relatively seamless interaction between brain and algorithm such that a person perceives of the algorithm as being a bona fide extension of a person's mind [31]." But then the question is, how much weight should be given to an individual's subjective report when claiming that the artefact has been integrated into their cognition?

4. From extended cognition to the extended witness

This section will discuss why an AI-enabled visual neuroprostheses qualifies as genuine cognitive extension of the wearer.

4.1. The "ongoing feedback loops" criterion of constitution

As argued by Palermos, what is needed (to objectively confirm cognitive extension in practice) is a precise criterion of integration for constitutive contribution to the extended cognitive system [23]. One perspective that is especially useful in the context of the extended mind theory is the one put forward by van Gelder and Beer, known as the "DST" approach—a cognitive science hypothesis that conceptualises embedded and extended cognitive systems in terms of dynamical system theory [32,33]. According to DST, when two or more systems engage in ongoing bidirectional interactions through continuous

feedback loops, they can give rise to novel systemic properties. This view is useful to practically explain, in a mathematical informed analysis, when is an extended cognitive system *coupled* with the agent [34].

This approach to Clark and Chalmers' extended mind thesis suggests that a coupled extended cognitive system forms only if a cognitive property arises from ongoing bidirectional interactions between an individual and an artefact [35,36]. In simple terms, the criterion for constitution is the *exist*ence of *continuous bidirectional interactions* between user and device during the performance of a cognitive task [23]. This account, proposed by Palermos, is known as the "ongoing feedback loops" criterion of constitution [23]. It not only prevents the common issue of "cognitive bloat" linked with the extended mind thesis but also provides a clear, practical boundary for determining perceptual authorship within the human–device interface. This boundary enables us to classify the resulting report as eyewitness testimony rather than merely machine output.

Palermos illustrates this point with a tactile–vision substitution system ("TVSS"), arguing that a camera—typically mounted on a pair of glasses—continuously converts optical variation into patterned vibrations across the wearer's skin [23]. As blind users turn their head or shift body position, they influence the light captured by the camera, which makes the device re-encode those changes into a new vibrotactile pattern that guides the next exploratory movement, creating a continuous causal cycle. Through training, blind users learn to exploit this sensorimotor feedback loop to locate objects, judge distance and orientation—often describing the experience in quasi-visual terms [35]. On Palermos' analysis, the resulting perceptual capacity is not the property of the organism or the technological artefact taken separately but of their coordinated coupling. The perceptual success is a novel cognitive property of the human–device system as a whole that arises from ongoing bidirectional interactions. Precisely because the capacity is constituted by these continuous loops, TVSS provides a clear instance of constitutive extension under the "ongoing feedback loops" criterion [23,35].

Even more relevant to our discussion, Vera Tesink *et al.* [37] provide a clear instance of constitutive extension with a case of a patient using a deep-brain stimulation device ("DBS"). As they note, the DBS continuously interacts with the patient's neural activity as it "reads" and "writes" into the brain. As DBS engages in a continuous, back-and-forth interaction with the patient's cognitive and mental states, Tesink *et al.*, suggest it satisfies conditions for integration—of the sort Palermos proposes—and thus may be treated, under the extended mind thesis, as part of the physical realization of the person's mind. Building on this, neurotechnologies, particularly invasive ones, demonstrate a distinctive integrative potential because they create a bidirectional brain-machine channel, without the loop having to be executed entirely through bodily structures outside the nervous system.

With these points in place, the way biological systems incorporate visual neuroprosthetics into their cognitive routines "creates extracorporeal loops that extend cognition beyond the onboard biological machinery" [23]. On the DST view, these devices help realize cognitive tasks through the agent's continuous two-way coupling with them. Acknowledging these ongoing bidirectional interactions is key to establish that neuroprostheses are genuine instances of cognitive extension.

Applying these considerations to our context, according to the "ongoing feedback loops" criterion, an AI-powered visual prosthetic forms a single, well-integrated cognitive system with the user. Referring back to our thought experiment—the AIWitness of a potential theft—it is essential that what the wearer perceives is merely such a unified "visual" experience. It is not something that is consulted in addition to or alongside their sense perception to interpret or enhance it, which is a key difference from the FLIR

recordings that can be analytically and practically separated from the person analysing the data. In this scenario, the continuous bidirectional interactions between the wearer and the device are easily discernible, leaving little reason to dispute, based on any reasonable dynamical model, that the interaction creates an integrated, coupled extended cognitive system. Under these conditions, the perceptual content generated by the device becomes part of the user's own (extended) cognition, as supported by Chalmers [26].

4.2. Broader implications for evidence law

When the perceptual AI-mediated content to which the wearer testifies exists solely by virtue of a continuous, two-way coupling between person and neuroimplant, it should be treated as the wearer's own perceptual experience. In other words, if a person's trial-relevant perception is produced by an AI-enabled neuroprosthetic, then the resulting experiential content is theirs in the same sense as unaided vision. Thus, AIWitness testimony should be regarded as eyewitness evidence, not a mere relay of digital output.

For practical evidential purposes, this line of argument suggests that AI-generated data from a neuroimplant should not be admitted or accessed without the witness's consent, just like we would not admit a suspect's unspoken thoughts. Likewise, if the neurodata may incriminate its wearer, it would not be admissible without their consent. Instead, courts should focus on the human's report of experience—which now inherently includes the AI's contribution—and cross-examine the witness on that testimony. By regarding the AI as part of the witness's mind, we also advocate for a "methodological principle of conservatism in law reform" and discourage legislative intervention, creating new classes of evidence.

Our proposal is therefore a middle position. For admissibility and classification purposes, AIWitness testimony should be treated as personal eyewitness evidence. In practice, this means that the defendant confronts a person, not a device. At the same time, baseline disclosure obligations—covering the model and version of the implant, its update history, manufacturer documentation on known artefacts, and any available validation studies that concerned the type of device—would give the defence a foothold to test reliability. This preserves the witness's legal status as a person rather than a machine, while preventing the AI-mediated aspects of their perception from becoming a total evidential "black box". Ordinary cross-examination will remain the main avenue for contesting AIWitness testimony, for example, by probing how long the witness has used the device, the conditions under which it could malfunction, and their past experience of errors, while device-level disclosure anchors those questions in a minimally transparent context. But unlike unaided witness testimony, generic information about the device helps the trier of fact not just to give the right weight to the testimony, but also to understand the idea of "vision" that relied on built artefacts by third parties.

The extended mind thesis, as shown, not only justifies considering "artificial vision" as the witness's own cognition but also requires the law to treat it with the same respect and constraints as internal memory. In this context, legal systems acknowledge broad rights for disabled witnesses to use assistive devices, such as sign language interpreters, to ensure effective participation. The extended mind perspective implies that neuroprosthetic vision is simply another form of reasonable accommodation. If extended cognitive tools are regarded as part of the individual, then the content produced by them should receive the same legal protection as biological memory. Practically, this means that data from the device

is the witness's own mental content. Forcing access to it without consent would be akin to compelling a person to reveal their private thoughts.

In the context of AI-mediated sight, the witness's implant is functionally no different from their brain but embedded as their own sensory apparatus. We therefore have strong normative reason to resist treating the device's outputs as external evidence. If we equate them with internal perception and memory, then subjecting them to special scrutiny or forcing their disclosure could violate the core fairness right against self-incrimination.

The right to equality and participation, particularly for citizens with disabilities, supply further reasons that support our argument. Refusing to recognize AI-mediated perception as the witness's own, risks relegating disabled individuals to the role of machines whose "real" evidence is the log, not their voice. That is at odds with the idea that criminal adjudication is conducted by and among persons, not by inspection of instruments alone. The extended mind account is a plausible avenue to restore that person-centredness without denying that instruments play a constitutive role in this technological era. This normative point aligns with the ontological equivalent in the literature. Cassinadri and Fasioli, for instance, note that the extended mind theory serves to "attribute cognitive credit to individuals with learning disabilities who use assistive tools to complete their learning tasks, thus avoiding their marginalization" [38].

Conceptually, our contention therefore provides reasoned fairness. If two perceptual routes produce functionally equivalent experiential content for an agent, they deserve equivalent admissibility status. This view promotes participation and equality, as declining to credit AI-mediated perception as the witness's own risks demoting disabled or neuroprosthetic users' human experience to mere CCTV footage.

Our analysis so far points towards a single, coherent set of rules that center on the witnesses in all their aspects: as victim, witness or suspect, as active participants in the civic process that is the trial, but also as private citizens with a legitimate interest in the protection of their data. One may argue that this attempt to align the AIWitness with other forms of "legal witnessing" is not capable of sufficiently addressing the subject-specific characteristic of the device wearer. As a result, potentially harms their right to be taken seriously as an individual, and the suspect's fair trial rights. Given the variety of devices that we are likely to see reaching the market over the next few years, it is based on a highly artificial and overly abstract attempt to "harmonise" different forms of perception.

We can only outline our response to these valid concerns. The starting point for us is that the perception of "divergence when perceiving" is not a new problem solely created by neuroimplants. To the contrary, we face it in every form of witnessing. People have different eyesight, different ways in which vision is stored in memory, and different strengths and weaknesses in recall, just for starters. Law, as a system of rules governed in turn by the overarching principle of the Rule of Law can never fully respond to these differences. There will always be the problem of insufficient "splitting" of categories and processes. While it may look specifically for scientifically minded readers as a potentially grave violation of the accused's (or the witness's) rights, the rationale is not just one of efficiency and cost saving. Rather, law always tries to balance the idea of doing justice to the individual within an overarching framework of general rules. Different eyewitnesses will experience "their" trial very differently. Some will get harsh cross-examination, others will be accepted by both parties. Some will be questioned as to their honesty, others as to their vision. Nothing of what we suggest removes this level of scrutiny; on the contrary, it lays the foundations for it. But, at the same time, no cognitively

capable eyewitness will have to undergo a medical diagnosis against their will to test the quality of their vision. Nor will the account they gave to a third party get preferential treatment over asking them what they have seen.

Our answer at this point then combines a positive and a negative element. On the positive side, we argue that our solution aligns with a regulatory system that is quite flexible on the ground and can introduce, within reasonable limits, differences in the treatment of different witnesses. On the negative side, if one wanted to permit different treatment of AI Witnesses with different technology, then one should at the same time revisit the treatment of other, non-assisted witnesses (and those with glasses or other tools), and treat it as a remedy to a general legal problem.

5. Conclusion

This paper examined Andy Clark and David Chalmers' extended mind theory as a normative basis for clarifying the legal status of AI-mediated perception. We argued that when a witness's trial-relevant perceptual content is produced by an integrated human–AI loop, in this case, a visual neuroprosthetic operating at the time they witnessed, that content is the witness's own. In consequence, AIWitness testimony should be treated as eyewitness evidence as it is the first-person deliverance of an extended sensory system in real time. Put differently, a witness's AI-enabled perception deserves the same testimonial dignity and baseline admissibility as unaided vision. To deny this would be to penalize the use of an assistive sensory technology that disabled citizens need for equal participation. This is not a plea for technological exceptionalism; it is a plea for conceptual consistency about *who* perceived.

Our argument is confined to sensory-substitution and restoration cases in which neuroprosthetic vision replaces a basic sense the witness lacks. Reinforced by a dynamical-systems criterion of ongoing, bidirectional coupling, the "ongoing feedback loops" test provides an objective boundary for extension and shows why visual neuroprosthetic cases qualify as constitutive cognitive extension. Taken together, extended mind theory and this coupling criterion offer a principled justification for treating AI-mediated perception as part of the witness's cognition. It sustains that where visual experience arises only through continuous two-way interaction between wearer and device, perceptual authorship resides in the human-device unit. In this line of reasoning, evidence law should recognize the resulting report as authentically the witness's, addressing reliability as a matter of weight, as with unaided vision. From a rights perspective, our stance preserves equality, participation, and fair trial values.

Acknowledgements

Funding: Schafer's work was supported by UKRI grant EP/T022485/1, "DeCaDe".

Authors' contribution

Conceptualization, CGM and BS; methodology, CGM and BS; formal analysis, CGM and BS; investigation, CGM and BS; resources, CGM and BS; data curation, CGM; writing—original draft preparation, CGM; writing—review and editing, CGM and BS; visualization, CGM and BS; supervision, CGM and BS; project administration, CGM and BS; funding acquisition, BS. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare no conflict of interest.

References

[1] Giansanti D. Advancements in ocular neuro-prosthetics: bridging neuroscience and information and communication technology for vision restoration. *Biology* 2025, 14(2):134.

[2] Sarbout I, Gungor A, Ounissi M, Zaher S, Ptito M, *et al*. Visual prostheses in the era of artificial intelligence technology. *Eye Brain* 2025, 17:95–113.

[3] Chen X, Wang F, Fernandez E, Roelfsema PR. Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. *Science* 2020, 370(6521):1191–1196.

[4] Zhang G, Chen R, Ghorbani H, Li W, Minasyan A, *et al*. Artificial intelligence-enabled innovations in cochlear implant technology: Advancing auditory prosthetics for hearing restoration. *Bioeng. Transl. Med.* 2025, 10(3):e10752.

[5] Fernández E, Alfaro A, Soto-Sánchez C, Gonzalez-Lopez P, Lozano AM, *et al*. Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex. *J. Clin. Invest.* 2021, 131(23):e151331.

[6] Mirochnik RM, Pezaris JS. Contemporary approaches to visual prostheses. *Mil. Med. Res.* 2019, 6(1):19.

[7] Mahadiuzzaman ASM, Hoque ME, Alam S, Chawdhury ZT, Hasan M, *et al*. Visual neuroprostheses for impaired human nervous system: state-of-the-art and future outlook. *Int. J. Cell Biol*. 2024, 2024(1):2651763.

[8] Wu KY, Mina M, Sahyoun JY, Kalevar A, Tran SD. Retinal prostheses: engineering and clinical perspectives for vision restoration. *Sensors* 2023, 23(13):5782.

[9] Yang J, Chen C, Yu Z, Chung JHY, Liu X, *et al*. An electroactive hybrid biointerface for enhancing neuronal differentiation and axonal outgrowth on bio-subretinal chip. *Mater. Today Bio* 2022, 14:100253.

[10] Hogri R, Bamford SA, Taub AH, Magal A, Del Giudice P, *et al*. A neuro-inspired model-based closed-loop neuroprosthesis for the substitution of a cerebellar learning function in anesthetized rats. *Sci. Rep.* 2015, 5(1):8451.

[11] Donati E, Valle G. Neuromorphic hardware for somatosensory neuroprostheses. *Nat. Commun.* 2024, 15(1):556.

[12] Chen Q, Lin P, Yu Z, Pan G. Enabling neuroprostheses via machine learning. *Mach. Intell. Res.* 2025, 22:866–870.

[13] Beyeler M, Sánchez-Garcia M. Towards a smart bionic eye: AI-powered artificial vision for the treatment of incurable blindness. *J. Neural Eng.* 2022, 19(6):063001.

[14] Kupers R, Chebat DR, Madsen KH, Paulson OB, Ptito M. Neural correlates of virtual route recognition in congenital blindness. *Proc. Natl. Acad. Sci. U.S.A.* 2010, 107(28):12716–12721.

[15] Stoddart PR, Begeng JM, Tong W, Ibbotson MR, Kameneva T. Nanoparticle-based optical interfaces for retinal neuromodulation: a review. *Front. Cell. Neurosci.* 2024, 18:1360870.

[16] Pulicharla RM, Premani V. AI-powered neuroprosthetics for brain-computer interfaces (BCIs). *World J. Adv. Eng. Technol. Sci.* 2024, 12(1):109–115.

[17] Eisenberg U. *Beweisrecht der StPO: Spezialkommentar*, 6th ed. München: C. H. Beck. 2008. pp. 1000–1001.
[18] Graf JP. Kommentierung zur Einführung der RiStBV. In *BeckOK StPO: Beck'scher Online-Kommentar zur Strafprozessordnung*. München: C. H. Beck. 2017.
[19] Hansson SO. The ethics of explantation. *BMC Med. Ethics* 2021, 22(1):121.
[20] Ienca M, Valle G, Raspopovic S. Clinical trials for implantable neural prostheses: understanding the ethical and technical requirements. *Lancet Digital Health* 2025, 7(3):e216–e224.
[21] Bassil K, Jongsma K. To explant or not to explant neural implants: an empirical study into deliberations of Dutch Research Ethics Committees. *Neuroethics* 2025, 18(3):1–10.
[22] Clark A, Chalmers D. The extended mind. *Analysis* 1998, 58(1):7–19.
[23] Palermos SO. *Cyborg Rights: Extending cognition, ethics, and the law*, 1st ed. Oxfordshire: Routledge. 2025. pp. 26–51.
[24] Heersmink R. Dimensions of integration in embedded and extended cognitive systems. *Phenomenol. Cognit. Sci.* 2015, 14(3):577–598.
[25] Rupert R. *Cognitive systems and the extended mind*. Oxford: Oxford University Press. 2009.
[26] Chalmers D. Extended cognition and extended consciousness. In *Andy Clark and His Critics*. Oxford: Oxford University Press. 2019. pp. 9–21.
[27] Clark A. *Supersizing the mind: embodiment, action, and cognitive extension*. Oxford: Oxford University Press. 2008.
[28] Beauchamp MS, Oswalt DC, Vickery T, Madsen JR, Bliss RE, *et al*. Dynamic stimulation of visual cortex produces form vision in sighted and blind humans. *Cell* 2020, 181(3):774–783.
[29] Barker MJ. From cognition's location to the epistemology of its nature. *Cognit. Syst. Res.* 2010, 11(4):357–366.
[30] Sprevak M. Inference to the hypothesis of extended cognition. *Stud. Hist. Philos. Sci.* 2010, 41(4):353–362.
[31] Reiner PB, Nagel SK. Technologies of the extended mind: defining the issues. In *Neuroethics: Anticipating the Future*. Oxford: Oxford University Press. 2017. pp. 108–122.
[32] Van Gelder T. The dynamical hypothesis in cognitive science. *Behav. Brain Sci.* 1998, 21(5):615–628.
[33] Beer RD. A dynamical systems perspective on agent–environment interaction. *Artif. Intell.* 1995, 72(1,2):173–215.
[34] Palermos SO. Knowledge and cognitive integration. *Synthese* 2014, 191(8):1931–1951.
[35] Palermos SO. Loops, constitution, and cognitive extension. *Cognit. Syst. Res.* 2014, 27(1):25–41.
[36] Favela LH, Riley MA, Shockley K, Chemero A. Perceptually equivalent judgments made visually and via haptic sensory-substitution devices. *Ecol. Psychol.* 2018, 30(4):326–345.
[37] Tesink V, Douglas T, Forsberg L, Ligthart S, Meynen G. Right to mental integrity and neurotechnologies: implications of the extended mind thesis. *J. Med. Ethics* 2024, 50:656–663.
[38] Cassinadri G, Fasoli M. Rejecting the extended cognition moral narrative: a critique of two normative arguments for extended cognition. *Synthese* 2023, 202(5):155.