p-ISSN:2708-2091
e-ISSN:2708-3586

www.gsrjournal.com
GSR
Global Sociological Review

GSR

GLOBAL SOCIOLOGICAL REVIEW

HEC-RECOGNIZED CATEGORY-Y

VOL. X ISSUE III, SUMMER (SEPTEMBER-2025)

Double-blind Peer-review Research Journal
www.gsrjournal.com

DOI (Journal): 10.31703/gsr
DOI (Volume): 10.31703/gsr.2025(X)
DOI (Issue): 10.31703/gsr.2025(X-III)

HumanityPublications
sharing research

Humanity Publications (HumaPub)
www.humapub.com
Doi: https://dx.doi.org/10.31703

Article title

Cognitive Privacy and the Architecture of AI-Driven Surveillance

Abstract

Artificial intelligence has changed surveillance from observation to inference with systems able to decode emotions, intentions, and beliefs based on behavioral and biometric measurements. The article proposes the concept of cognitive privacy as the right to mental autonomy and avoid algorithmic manipulation. It explores the inability of current constitutional and statutory structures that were developed in an analog world to deal with inferential harm. It finds legal and normative loopholes in data collection, inference, and governance through the analysis of AI-driven surveillance architecture. To overcome these shortcomings, this paper suggests a four-layered cognitive-privacy protection model that includes sensor boundaries, inference control, interface transparency and institutional oversight. It concludes that the only way to protect cognitive liberty is to redefine privacy as a control over inference as opposed to secrecy of information that is an indispensable measure towards the maintenance of democratic agency in the era of intelligent surveillance.

Keywords: Artificial Intelligence, Cognitive Privacy, Algorithmic Governance, Privacy Law, Constitutional Law, AI Regulation, Neurorights, Cognitive Liberty, Data Protection, Human Autonomy

Authors:

Ali Nawaz Khan: (Corresponding Author)
Assistant Professor, University Law College, University of the Punjab, Lahore, Punjab, Punjab, Pakistan.
(Email: alinawazpulc@gmail.com)

Shahzad Khalid: Doctoral Researcher, Brunel University London, United Kingdom.
Assistant Professor, Department of Law, Superior University, Lahore, Punjab, Pakistan.

Ahmed Raza: LLM Scholar, Pennsylvania State University, USA.

Pages: 127-138
DOI: 10.31703/gsr.2025(X-III).13
DOI link: https://dx.doi.org/10.31703/gsr.2025(X-III).13
Article link: http://www.gsrjournal.com/article/cognitive-privacy-and-the-architecture-of-aidriven-surveillance
Full-text Link: https://gsrjournal.com/article/cognitive-privacy-and-the-architecture-of-aidriven-surveillance
Pdf link: https://www.gsrjournal.com/jadmin/Auther/31rvIolA2.pdf

Global Sociological Review
p-ISSN: 2708-2091 **e-ISSN:** 2708-3586
DOI(journal): 10.31703/gsr
Volume: X (2025)
DOI (volume): 10.31703/gsr.2025(X)
Issue: III Summer (September-2025)
DOI(Issue): 10.31703/gsr.2024(X-III)

Home Page
www.gsrjournal.com

Volume: (2025)
https://www.gsrjournal.com/Current-issues

Issue: III-Summer (June -2025)
https://www.gsrjournal.com/issue/10/3/2025

Scope
https://www.gsrjournal.com/about-us/scope

Submission
https://humaglobe.com/index.php/gsr/submissions

Visit Us

Humanity Publications (HumaPub)
www.humapub.com
Doi: https://dx.doi.org/10.31703

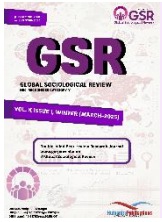
Citing this Article

13	Cognitive Privacy and the Architecture of AI-Driven Surveillance		
Authors	Ali Nawaz Khan Shahzad Khalid Ahmed Raza	DOI	10.31703/gsr.2025(X-III).13
		Pages	127-138
		Year	2025
		Volume	X
		Issue	III

Referencing & Citing Styles

APA	Khan, A. N., Khalid, S., & Raza, A. (2025). Cognitive Privacy and the Architecture of AI-Driven Surveillance. *Global Sociological Review*, *X*(III), 127-138. https://doi.org/10.31703/gsr.2025(X-III).13
CHICAGO	Khan, Ali Nawaz, Shahzad Khalid, and Ahmed Raza. 2025. "Cognitive Privacy and the Architecture of AI-Driven Surveillance." *Global Sociological Review* X (III):127-138. doi: 10.31703/gsr.2025(X-III).13.
HARVARD	KHAN, A. N., KHALID, S. & RAZA, A. 2025. Cognitive Privacy and the Architecture of AI-Driven Surveillance. *Global Sociological Review,* X**,** 127-138.
MHRA	Khan, Ali Nawaz, Shahzad Khalid, and Ahmed Raza. 2025. 'Cognitive Privacy and the Architecture of AI-Driven Surveillance', *Global Sociological Review*, X: 127-38.
MLA	Khan, Ali Nawaz, Shahzad Khalid, and Ahmed Raza. "Cognitive Privacy and the Architecture of Ai-Driven Surveillance." *Global Sociological Review* X.III (2025): 127-38. Print.
OXFORD	Khan, Ali Nawaz, Khalid, Shahzad, and Raza, Ahmed (2025), 'Cognitive Privacy and the Architecture of AI-Driven Surveillance', *Global Sociological Review,* X (III), 127-38.
TURABIAN	Khan, Ali Nawaz, Shahzad Khalid, and Ahmed Raza. "Cognitive Privacy and the Architecture of Ai-Driven Surveillance." *Global Sociological Review* X, no. III (2025): 127-38. https://dx.doi.org/10.31703/gsr.2025(X-III).13.

e-ISSN: 2708-2091 Volume: X (2025) Issue: III-Summer (September -2025) p-ISSN: 2708-3586



Global Sociological Review
www.gsrjournal.com
DOI: http://dx.doi.org/10.31703/gsr

Pages: 127-138 URL: https://doi.org/10.31703/gsr.2025(X-III).13 Doi: 10.31703/gsr.2025(X-III).13

Title

Cognitive Privacy and the Architecture of AI-Driven Surveillance

Authors:

Ali Nawaz Khan
(Corresponding Author)
Assistant Professor, University Law College, University of the Punjab, Lahore, Punjab, Pakistan.
(Email: alinawazpulc@gmail.com)

Shahzad Khalid
Doctoral Researcher, Brunel University London, United Kingdom.
Assistant Professor, Department of Law, Superior University, Lahore, Punjab, Pakistan.

Ahmed Raza
LLM Scholar, Pennsylvania State University, USA.

Contents

Abstract

Artificial intelligence has changed surveillance from observation to inference, with systems able to decode emotions, intentions, and beliefs based on behavioral and biometric measurements. The article proposes the concept of cognitive privacy as the right to mental autonomy and to avoid algorithmic manipulation. It explores the inability of current constitutional and statutory structures that were developed in an analog world to deal with inferential harm. It finds legal and normative loopholes in data collection, inference, and governance through the analysis of AI-driven surveillance architecture. To overcome these shortcomings, this paper suggests a four-layered cognitive-privacy protection model that includes sensor boundaries, inference control, interface transparency, and institutional oversight. It concludes that the only way to protect cognitive liberty is to redefine privacy as a control over inference as opposed to secrecy of information that is an indispensable measure towards the maintenance of democratic agency in the era of intelligent surveillance.

Keywords: Artificial Intelligence, Cognitive Privacy, Algorithmic Governance, Privacy Law, Constitutional Law, AI Regulation, Neurorights, Cognitive Liberty, Data Protection, Human Autonomy

Introduction

Surveillance has been traditionally understood as a visual monitoring of people or places. Some of these ways that state power was exercised back in the

twentieth century included wiretaps, closed-circuit cameras, and signal intercepts. These tools of authority were mainly concerned with what was visible or audible. In the twenty-first century, surveillance has turned into a computation and a prediction. Artificial intelligence (AI) systems are based not on observation but inference. Machine-learning models are now able to recreate human intentions, emotions, and dispositions based on the movements on their faces, their voice intonations, their eye movements, or their browsing behaviors. With the help of data correlation, what used to be accomplished by physical intrusion is now accomplished. What these developments are invading is a field that was once considered sacred, the human mind itself. The contemporary problem of privacy does not reside in hiding the truths, as Solove (2025) intuitively notes, but in the ability to control the interpretation and usage of those truths.

Although the classical privacy ideologies are aimed at safeguarding informational control, i.e., who is entitled to access personal information and to what end, the cognitive privacy addresses the security of the thought itself. It is about the patterns of the mental life: about how human beings perceive, reason, and make decisions. New neurotechnologies and affective-computing systems, including brain-computer interfaces, sentiment-analysis engines, and others, can be used to convert patterns of attention and emotion into data that can be analyzed (Szoskowicz, 2025). Not only do these technologies extract information, but they also provide it back in the form of algorithmic recommendations that use subtle influence in perception and conduct. Such an individual then turns into both observer and observed, producer and product of surveillance simultaneously.

There are thus four goals that are related in this paper. To begin with, it aims to theorize cognitive privacy and trace its theoretical roots in the fields of law, philosophy, and neuroscience. Second, it describes the way in which the design of AI-based surveillance threatens mental sovereignty. Third, it assesses the sufficiency of legal, regulatory, and institutional responses. Lastly, it proposes an elaborated multi-layered structure that would protect cognitive integrity in the emerging system of algorithmic governance.

Intellectual Principles of Cognitive Privacy:

Defining the Domain

Cognitive privacy represents a new area of legal and ethical research that aims at safeguarding the sanctity of thought in the era of intelligent computing. It can be simply described as the right of the individual to maintain the autonomy, integrity, and confidentiality of internal mental states; beliefs, emotions, memories, and decision mechanisms against unauthorized inference, surveillance, or manipulation. In contrast to traditional privacy, which is the issue of the regulation of the information revealed to other people, cognitive privacy is applied to the very processes of cognition. It does not only mean preserving what we say and are doing, but preserving our thinking. It, in other words, captures the invisible psychological channels that lead to expression and action.

Cognitive privacy has become more significant due to the increase of exponential growth of machine-learning technologies that have the potential to identify, model, and even predict inner psychological states. Computational and neuroscientific developments have proven that cognition is not all a highly personal phenomenon, locked inside the skull. It is now leaving traceable footprints in the digital spaces in which we live. Rhythmic typing patterns, eye fixation, voice alteration, and browsing behavior habits represent valuable sources of data that could reveal the mood, level of attention, or decision-making preferences with stunning accuracy (Magee, 2024). The more AI models combine these various types of data, the more experienced is being machine-readable and subjective experiences are reduced to measurable metrics (Szoszkiewicz, 2025). The border between interior consciousness and exterior surveillance is thereby wavered away, bringing a sort of transparency of mind that implies never-before-seen ethical and legal difficulties.

Within this new terrain, the human mind cannot be merely one of the biological organs but an interface and a continual mapping and reflection of algorithms. The state brings up serious concerns about autonomy and personhood. To what degree are people able to

own their mental existence in the event that any of these processes of thinking can be documented, derived, and forecasted? The ancient legal systems that used to separate the notion of thought and expression become obscure when even thinking leaves traces of digital records. To maintain individuality and moral agency, then, in the world where cognition itself has been shown to be computational, it is necessary to protect cognitive privacy.

Inference and the Extended Mind

The cognitive-privacy dilemma is based on inference. Artificial intelligence systems draw insights regarding people not by directly reading their minds but by probabilistic decision-making in huge volumes of data. These systems produce what can be called mental data, behavioral, biometric, and contextual data, when correlated, form representations of intention, belief, and emotion that the person never revealed consciously. According to Solove (2025), when privacy injury is caused by inference, but not collection, the whole structure of data-protection law starts falling apart. The majority of the privacy regimes control what may be collected or shared, but provide little control over what may be inferred. This epistemic asymmetry permits corporations and states to do mental profiling whilst still under the legal framework of complying with data laws.

This issue is further compounded by the philosophical perspective of the extended-mind hypothesis as put forward by Clark and Chalmers (1998). They suggest that human thinking not only happens in the brain but also in the external tools and systems we use in memory, reasoning, and judgment. These tools are smartphones, search engines, and AI assistants that are able to be involved in our thinking processes in the digital age. Every search query to the search engine, every communication with a digital assistant, is a node on a neural network that is expanded, and the distinction between human and machine is blurred. These infrastructures are aptly termed, according to Brcic (2025), as cognitive moats, artificial environments, which bond and form the lines of human thought in a circular manner and ultimately reconfigure them.

This distributed cognition has some far-reaching implications. In case the cognitive results are now co-produced by external systems, then infiltrations into these systems constitute infiltrations into the mind as such. Cognitive privacy protection is thus not limited to the ban on neural surveillance, but rather requires defense of the integrity of the extended cognitive ecosystem. It involves not only physical neurointerfaces but also those software architectures and data infrastructures, along with the algorithm's intermediaries, which make perception and choice a possibility. Whether AI can read our minds or not is no longer a question, but in getting involved in our thinking, it is redefining the mind on the margins.

Dimensions of Threat

Menaces to cognitive privacy are complex, as they can be seen at the level of behavioral, psychological, as well as neurophysiological. The most common and the first one is inferential intrusion, in which AI systems infer latent mental qualities, like political inclinations or emotional stability, or religious belief, based on ostensibly innocent data points. As an example, minor differences in the use of vocabulary on social media or even facial micro-expressions can be incorporated to infer mental health conditions or ideological beliefs. Inferential surveillance, unlike traditional surveillance, is a means of rebuilding the inner world, a process of making the inner world visible, and in the process, it turns the inner world into an object of computation and makes privacy a collective item.

A second threat, which is cognitive nudging, is the strategic exploitation of such inferences with the aim of controlling attention, emotion, and decision-making. Algorithms tailor digital spaces, including news feeds, advertisements, or search results, to the predicted preferences or vulnerabilities. Although it is often expressed in easy-to-use customization terms, this micro-targeting applies an invisible pressure to cognition and makes people act in a way that they want to believe. The process exploits emotional vulnerabilities to keep the interest going or to influence political decisions, which begs the question

of the loss of free will in the algorithmic marketplace of ideas.

The third dimension is that of memory capture which deals with the consistent recording and storing behavioral traces which enable the personality and cognition to be modeled continuously. Each and every interaction, be it via a phone, a wearable device, or a digital assistant, will be added to an ever-expanding collection of mental patterns. In this longitudinal monitoring, the continuity can be predicted: by the use of prior cognitive data systems, future thinking can be predicted. The self is constructed over time both by algorithmic memory and experience.

Lastly, neural interception as the boundary of cognitive surveillance is an access to or control of the brain activity by means of neurotechnological interfaces. Despite their immaturity, tools that can read or be used to manipulate the neural signals or can control the state of emotion are already available in both the experimental and business environments. The ethical threats are enormous: not only privacy, but also authenticity, autonomy, and identity may be undermined by such technologies. All these threats combined are indicative of the fact that cognitive privacy is neither a fantasy nor a hypothetical concept; it is an immediate concern that has become part of the daily mechanics of digital life.

Artificial Intelligence Surveillance Architecture:

Sensor Fusion and Data Ingestion

On the basic level, cognitive surveillance is carried out via a complex system of surveillance sensors and data-collection systems. Cameras, microphones, wearable gadgets, biometric scanners, systems of Internet-of-Things (IoT) are in a continuous data gathering concerning bodies and behaviors. These devices collect multimodal data, visual, auditory, physiological and contextual, which can be combined to deduce the mental states. Sensor fusion is the process that the disparate signals are combined into unified cognitive profiles. As an example, the analysis of videos could reveal facial tension, whereas the wearable sensors would record the increased heart rate and location information. Taken together, these inputs are able to herald anxiety, stress, or fear (Danesh Pazho et al., 2023).

Much of this processing can be done locally and in real time through edge computing to provide better efficiency and responsiveness. But this very decentralization is projected into all objects and surroundings of surveillance. Anonymized data may reveal the behavioral identity of users, even when this is done without revealing personal identities (Ienca et al., 2021). His or her mental and emotional existence is incorporated into a computational infrastructure that is distributed. This is why privacy risk is not only the explicit data collection, but also there is the likelihood of creating an inferential insight in every interaction. Sensor fusion changes the state of being in public or even in private space to a continuous practice of self-disclosure.

Inference Engines and Modeling

The second architectural layer follows the data collection and entails machine-learning engines that transform raw input of sensory data into cognitive models. These engines adopt neural networks, probabilistic reasoning, and representational learning to categorize the feelings, forecast motives, and detect hidden characteristics. They are not mere systems of behavior recording, but they invert the meaning, creating psychological maps of people and groups. This has enhanced the complexity of these algorithms to the extent that they can make predictions about decision-making even before the user makes decisions (Abbasi, et. al., 2025).

Technical mitigation strategies like federated learning or differential privacy do not usually work to address these risks. They prevent direct sharing of data, but they enable the models to extract implicit patterns of thinking (Yew, Qin, and Venkatasubramanian, 2024). The alleged privacy that they provide is often a myth- used to justify surveillance in the guise of compliance. According to Pasquale (2016), such an obscurity is a type of what Pasquale refers to as a kind of structural power, that is, the knowledge asymmetry gives unlimited power to those who create and manage algorithms. People are objects of interpretation, and they do not have the

epistemic tools to disagree or even perceive the conclusions that are made about them. In this regard, transparency is not a technical matter but rather a moral requirement.

Cognitive Interfaces and Feedback Loops

The shift between inference and influence comes with the cognitive interfaces the screens, platforms and devices that come between the user and the product. Predictive analytics is used in recommender systems, personalized advertisements and in adaptive educational platforms to decide what a person views and the form in which it is delivered. New behavioral feedbacks are provided with each interaction, and the algorithms can use them to improve their psychological models and to further customize their influence. With time, the systems develop recursive feedbacks which reinforce certain thinking and feeling patterns.

Brcic (2025) refers to these loops as cognitive tunnels, arguing that these are engagement loops that gradually reduce the level of consciousness. In these tunnels, users are greeted with moderated realities that are not the best in terms of truth or discussion, but feelings and advertising. According to Carter (2025), attention is currency in this economy, and the ability to control attention is a new governance. The implications go beyond business: the way the cognitive environments are designed to predict and control behavior, even the boundaries between persuasion and coercion, become vaguer. The interfaces being driven by AI, therefore, are known to be confronting the most basic principles of democracy, as they convert citizens into free thinkers into responsive cogs in the algorithm's ecosystems.

The Modules of Governance and Auditing

Ideally, the AI surveillance architecture would end with a governance layer that would provide checks in power. This layer usually consists of consent protocols, algorithmic audits, and ethical guidelines, which are supposed to exclude misuse. Nevertheless, in reality, these mechanisms tend to be symbolic gestures, but not material protection. The consent forms are usually non-transparent and are given in non-negotiable terms, as they are required to get service, and not because the person really wants it. The value of algorithmic audits is often limited by the unavailability of proprietary code and data. Red-teaming exercises, which are supposed to reveal weak areas, are usually based on technical adventures instead of the systemic problem of manipulation and control.

Additionally, the company's secrecy and jurisdictional fragmentation are very restrictive to supervision. Even in the presence of laws that require accountability, cross-border data flows and trade-secret protection are used to conceal important processes from outward view. The asymmetry as a result is severe: there is totalizing surveillance, but piecemeal government. Consequently, the current structure of AI policing is still highly unbalanced and all-inclusive in terms of observatory ability but deficient in accountability processes. In order to redress this imbalance, we need to transform the nature of governance toward making the systems procedurally formal to being structurally enforced, entrenched in the very construction of the computational systems to include transparency, contestability, and human rights.

Legal and Normative Gaps:

The Constitutional Limits

The constitutions of the modern world were written in a world of physical searches in which material property was seen, rather than an environment of algorithms in which surveillance functions through prediction. The Fourth Amendment of the United States protects citizens against unreasonable searches and seizures, but the action of drawing conclusions out of legally obtained algorithmic data is not categorized as a search. Predictive systems are able to recreate habits, associations, or emotional states of a person without necessarily passing through a physical barrier. Courts are thus facing an ontological blind spot, where there is no hand seizing, or agent trespassing, what creates intrusion? In an attempt to respond to this, Solove (2025) claims that an informational state of being known without entry necessitates an extended doctrine of epistemic

privacy- the doctrine which explicitly acknowledges knowledge as a source of power.

The First Amendment is also impervious. The subversive processing of speech by algorithms poses the question of whose voices are heard or not, conditioning the topic of social dialogue based on the assumptions of the desired ideological orientation. This automated kind of filtering creates a chilling effect, not via explicit censorship but invisible exposure modulation. The citizens get trained to express themselves according to what the algorithms are rewarding, and they internalize the logic of surveillance. According to Grimmelikhuijsen et al. (2022), this loss of expressive autonomy is described as the loss of input and throughput legitimacy the fact that people do not feel represented in or able to affect the systems governing them. It should be an institutional innovation to put constitutional relevance back, special algorithmic-review courts, cognitive-rights commissioners, and open public records of government AI use. Devoid of such mechanisms, constitutional assurances of liberty even dry up with the silent effectiveness of inference.

Statutory Privacy Regimes

The European GDPR and Californian CCPA and the Brazilian LGPD were major accomplishments in the twentieth century but were not aligned with the epistemic realities of the twenty-first (Irfan, et. al., 2024). Their principles of consent, limit of purposes and minimization of data assume discrete collection acts; it does not control the interpretative algorithms which transform the innocent data into psychological revelation. The company is permitted to make a recording of the keystrokes or the browsing history, but still manages to deduce anxiety, plans of reproduction, or radicalism in political ideologies, none of which the user had agreed to disclose. Inferences hence fill in a legal gray area between reality and fiction.

Solove (2025) and Raza (2024) emphasize that the locus of privacy harm has essentially changed to the possession of the interpretation. Whoever runs the meaning gives governance. The right to explanation inherent in the GDPR was meant to offset this imbalance, but Kaminski (2019) shows that most forms of algorithmic disclosure are formal, or rather, technical statements that can only be understood by an engineer. What is created is transparency and not intelligibility. Controllers check the superficial layer of information transfers, and cognitive analytics are not controlled. Real reform needs to re-conceptualize personal data in order to include derived, inferred, and synthetic data, and it needs to hold interpretive accountability- a requirement to reason not only how data is handled but why certain inferences are being made. Statutory privacy can only have a revival of moral force through its legislation rather than through its handling.

Frameworks of Algorithmic Governance

The proponents of algorithmic governance encourage the incorporation of ethical values, such as fairness, accountability, and transparency (FAT), into computational systems (Munir, 2025). These theories only respond to quantifiable inequalities like bias or disparate impact but do not respond to less overt manipulations of thought (Munir, et. al., 2025). An algorithm can be unbiased in dispensing results and at the same time influence the desires in an unfair manner. According to Coglianese (2019), transparency out of context becomes performative compliance: information is exposed, but it is not comprehended. Kaminski (2019) opposes it with a binary-governance system that unifies individual rights with the oversight of the system; in that case, the public authorities that are independent should monitor the private actors. There is an additional participatory dimension mentioned by Iwan-Sojka (2025), who believes that the affected communities should share authority when it comes to the process of designing, testing, and auditing AI systems.

Even these refinements are not sufficient to address the issue of manipulation on a cognitive level. The future of algorithmic power is not the disparity of the outcomes, but the manipulation of attention and feeling. To protect autonomy, governance needs to transform FAT into a more elaborate structure that is based on FAITH; Fairness, Accountability, Integrity, Transparency, and Human autonomy. It

takes integrity to withstand the sale of the state of mind; it takes human autonomy to have systems that are corrigible and subject to challenge. These values must be incorporated in order to ensure that algorithmic governance is able to optimize behavior instead of empower deliberation.

Neurorights and International Human Rights

Privacy and freedom of thought were recognized in the global human-rights corpus, and the mechanisms that were used to enforce them were intended to check governmental abuses and not a private digital empire. UDHR (1948) and ICCPR (1966) are statements of moral commitments that lack operational devices of transnational algorithmic accountability. These gaps increase when neural signals and emotional answers are deciphered by AI technologies. Ienca and Andorno (2021) suggest the addition of the classical rights with neurorights: mental integrity, cognitive liberty, and psychological continuity. Ligthart et al. (2023) add to this model such elements as mental fairness and identity persistence. The two of them set up a jurisprudence of the mind-a realization that the brain has turned into a new location of sovereignty.

Critics remain cautious. Codification of neurorights, Gilbert (2024) cautions, runs the risk of entrenching hypothetical neuroscience in hard law and stifling therapeutic breakthroughs. A moderate solution should then follow proportionality: the limitations on the use of cognitive data should be required, limited, and supported by evidence. The bodies like UNESCO and the OECD may convene a Global Charter on Cognitive Privacy establishing minimum standards, similar to the environment norms. With time, the domestic constitutions may incorporate the neurorights provisions, making cognitive privacy more than an ethical desire, to be an enforceable international norm.

The direction of a Layered Architecture of Cognitive-Privacy Protection:

Layer 1 – Sensor and Data Boundaries

Protection will be effective at the point of data origin. All sensors such as cameras, EEG headsets, etc., are to be designed by the principle of minimal cognition capture, i.e., only what is absolutely needed should be captured. Exposure can be reduced significantly by on-device preprocessing, which is the extraction of key features in the ward, prior to transmission (Radanliev & Santos, 2023). Specific invasive modes like gaze tracking, neural signals, etc. require a license and a third-party examination. All events of activation should be recorded in cryptographically verifiable audit trails creating a chain of accountability. Such technical means will make constitutional proportionality an engineering practice so that the cognitive observation will not be the default.

Layer 2 – Inference Controls

During the analytical level, control is necessary regarding the manner in which the insights are inferred, rather than the manner in which data is stored. The law makers must enforce an inference taxonomy that would distinguish between benign and manipulative or discriminatory predictions. Systems that generate the latter have to be under increased scrutiny and explainability criteria. Pasquale (2016) proposes the idea of contestable design, which requires developers to present their reasoning to the outside world. Inference engines can also be isolated with the help of modular architecture, establishing cognitive firewalls that do not allow biometric or emotional information to enter marketing algorithms (Yew et al., 2024). Lastly, retraining cycles are rate limited, so that human control of the model evolution is present. These processes bring friction back into an ecosystem which has otherwise been streamlined to make instantaneous, thoughtless inferences.

Layer 3 – Interface and Feedback Safeguards

The points of prediction are where persuasion is prediction. Each adaptive platform, social-media feeds, and software to tutoring, should have a noticeable policy of whether personalization is based on cognitive profiling. Openness of influence will convert manipulation to open negotiation. Human-in-the-loop review should be considered obligatory in the sensitive factors, e.g., in education, employment, or healthcare (Andrus and Villeneuve, 2022). Carter

(2025) emphasizes that attention is the most precious democratic resource; controlling its exploitation with the help of pacing algorithms or engagement limits guard mental bandwidth. An adversarial audit (Abdu et al., 2024) can detect manipulative design patterns on a regular basis, and the use of a user-activated blind mode would completely prevent behavioral personalization. Such precautions transform the interface of behavioral conditioning into an informational decision-making.

Layer 4 Governance, Oversight and Remedies

The accountability is institutionalized at the outermost level. High-risk cognitive technologies should be under the control of independent auditors with legal, ethical, and technical competence (Bloch-Wehba, 2022). The governments and corporations are required to conduct Cognitive-Privacy Impact Assessment (CPIA) prior to the implementation of systems that can infer or create mental conditions (Iwan-Sojka, 2025). People have a need to have enforceable rights to delete, edit, or freeze cognitive profiles, and to be informed of the existence of inference systems whenever they are used by public registries. The use of manipulative inference should be penalized with the help of liability regimes (Kaminski, 2019), and participatory oversight academic, civic, and journalistic, should guarantee democratic legitimacy (Yang & Al-Masri, 2025). These principles incorporated into law ensure that cognitive privacy is a constitutional design element to algorithmic society and not an addition.

Illustrative Applications:

Smart-City Governance

Smartness in the city depends more on cognitive analytics. There are camera grids and acoustic gadgets and emotion-identifying devices that observe the movement and mood, not just in the city (Danesh Pazho et al., 2023). Officials defend these mechanisms as safety and efficiency tools, but they also allow predicting collective behavior, that is, knowing when a protest can happen, or an entire country is worried, even before anything happens. Such pre-emptive governing transforms into a political laboratory civic space. Democratic control requires the limits of inferences that prohibit emotional or ideological profiling of the general surveillance. Algorithms Pacho could be revealed through public transparency dashboards to make their criteria transparent once more, exposing them to cognitive darkness.

Education and Work-related Analytics

AI platforms are able to monitor gaze, keystrokes and sentiment in classrooms and offices to determine engagement. Continuous mental monitoring erodes trust even though it is being sold as productivity tools. The awareness of the attention quantified results in performative compliance as an alternative to curiosity with self-censorship. Grimmelikhuijsen et al. (2022) refer to such condition as compliance without legitimacy. Institutions ought to thus limit cognitive analytics to aggregated and anonymized information that can be used to improve systems instead of making individual judgments (Munir, 2024). Any decision based on cognitive measures should be humanly reviewed, so empathy will moderate automation. Technology can support and not kill authentic participation when it is combined with the power of analytics and pedagogical ethics.

Health Wearables and Consumer Devices

Intimate surveillance is the example of home assistants and health wearables. They are feeling pulse, tone, and stress to give lifestyle advice but their feedback loops provide subtle redefinitions of self-understanding. Algorithms start to produce emotions in users via mirrors (Brcic, 2025). This danger is heightened by the fact that the personalization is done on the basis of advertising, and instead of wellbeing. The standard of cognitive-sovereignty user-controlled memory deletion, influence logs and complete distinction between analytics and marketing should be imposed, therefore. Similar regulation certification as with nutrition labels of algorithms would ensure devices are not involved in manipulation under the carpet, but leave autonomy intact and promote innovation.

Implementation Challenges:

Technical Feasibility

It is also difficult to translate normative ideals into functioning code. Deep neural networks, which are the vehicles of the contemporary inference, are resistant to interpretability. XAI methods provide partial explanations but rarely provide explanations that can be understood by the non-technical audience (Magee, 2024). Privacy technologies like homomorphic encryption are privacy-enhancing that do not decrease latency and cost. The balance between transparency, efficiency, and security will need hybrid architectures- between symbolic reasoning to make it easily understandable and sub symbolic learning to find patterns. Lasting cooperation between technologists, jurists and ethicists is essential.

Regulatory Capacity

The enforcement of cognitive-privacy requires interdisciplinary literacy that is lacking in most of the agencies. To perform effective audits, knowledge should be in computer science, psychology, and constitutional law. This can be narrowed through multi-stakeholder partnerships by public regulators, academic laboratories and civil-society organizations. Audit sandboxes, toolkits Open-source toolkits and shared auditing sandboxes would allow inspection without violating the trade secret. As shown by Radanliev and Santos (2023), cooperative governance does not only increase the level of legitimacy but also technical resilience. Such capacity is a democratic requirement, rather than a bureaucratic indulgence to build it.

Consent and Power Asymmetry

Digital consent is becoming more deceptive. The opt-out options are hidden in the thick polices, and the option of refusing entails being denied access to vital services. Coglianese (2019) recommends the replacement of atomized consent by institutional fiduciary duty: platforms must have a duty of cognitive care, which is their mental best interest towards their users. Integrating fiduciary principles would make ethics an ambition, rather than an enforceable duty, and make surveillance capitalism cognitive stewardship.

Cross-Border Enforcement

The information, and conclusions based on it, spread around the world, making national control porous. Diverse regimes welcome cognitive-privacy havens where lax regulation brings in manipulative industries. These loopholes would be sealed by harmonization, preferably by mutual-recognition treaties or by OECD-type adequacy standards. International coordination is critical; otherwise, local protection disintegrates as fast as a packet switch.

Normative Uncertainty

Even the definition of cognition is controversial. Gilbert (2024) cautions that the law should not be codified into law so soon because it might solidify the emerging science and Szoszkiewicz (2025) responds that we must guard against ex post facto spear-headed harm before it is too late. The ideal course is gradualism on the basis of principles: the laws formulated in elastic moral terms, including autonomy, dignity, proportionality, which can be improved with discovery. Governance should not be based on legal paralysis but legal humility.

Cognitive Privacy as a part of Legal Doctrine:

Recreating Constitutional Standards

The constitutional law needs to give up its physicalist metaphors in order to be relevant. A reasonable-inference-expectation test would consider algorithmic discovery of unexpected personal information to be a search, and therefore subject to Fourth-Amendment examination. Likewise, the viewpoint discrimination will occur when recommender systems amplify or suppress expressions with regard to the ideology deduced. Such digital gatekeeping should be recognized as a First-Amendment problem because the cognitive liberty is reasserted as a constitutional value (Solove, 2025). By interpretive evolution, the Constitution is able to confront predictive power and not coercive power.

Incorporating Cognitive Privacy in Statutes

Inferred and derived data should be clearly defined by legislation as the personal information. Inference

engines should be audited by regulators, as well as require a proportionality of analytical intent and cognitive effect. Focusing on the findings of (Raza, 2024) in credit-risk analysis, the legislators would be able to make an analogy between cognitive profiling and discriminative scorekeeping, requiring periodic fairness and autonomy audits. Introducing the cognitive-harm measurements into the legislation on data-protection would bring legal regulations into the realm of scientific reality.

International harmonization by Neurorights

Neurorights serves as a potential lingua franca globally. Their definition created by Ienca and Andorno (2021) is based on mental integrity and liberty; their ethical scope is extended by Ligthart et al. (2023). A treaty based on minimum standards of cognitive-protection, like climate agreements, would allow a certain level of national flexibility and insist on global floors. Imposing neurorights on the local charters would convert cognitive privacy into an academic notion to a global standard.

Analysis

Cognitive privacy is a paradigmatic shift of ruling information to ruling inference. It needs a combination of law, engineering and ethics to protect it. Layered model here is based on the idea that architecture can reflect the principle of the constitution: by integrating the ideas of proportionality and accountability into data streams, the societies can remain free to think. However technical protective measures are not sufficient to achieve autonomy. Democratic resilience requires democratic literacy, citizens should be aware of the influence of the recommender systems and affective algorithms form perception (Yang & Al-Masri, 2025). The next research would have to operationalize the measures of cognitive intrusion attention volatility, emotional exhaustion, ideological homogenization to measure harm and tune up policy. Pasquale (2016) is reminding us that accountability is being eroded by the concept of transparency; it is then crucial that transparency should be turned procedural and epistemic. Cognitive privacy is not a benefit of the digital elite but a constitutional framework of democracy as such. Losing the ability to control the mind is loss of self-rule.

Conclusion

Surveillance using AI has changed the essence of knowledge and power. The human mind has become the input and output of the algorithmic governance. The protection of cognitive privacy is therefore essential towards autonomy, dignity and democratic deliberation. The article has made four contributions: it theorized the idea of cognitive privacy as an independent category of law; it mapped the architecture of AI-based surveillance; it revealed the shortcomings of the doctrines; and it presented a layered system of regulation covering sensors, inferences, interfaces and institutions.

In the future, policy makers are recommended to incorporate Cognitive-Privacy Impact Assessment, require neurorights adherence and finance interdisciplinary oversight commissions. The end is moderation a state in which technological advancement exists together with intellectual independence. In the era of intelligent machines, the battle against the freedom of man does not start at the boundary of the state, but at the boundary of the mind.

References

Abbasi, M. S., Munir, B., Jayaram, V., & Rivas, P. (2025). Leveraging autocorrelation in a dilated CNN-LSTM framework for predicting the US Supreme Court decisions. *IEEE Access.* https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=11162521
Google Scholar Worldcat Fulltext

Abdu, R., Kim, H., & Wang, L. (2024). Adversarial auditing for machine learning models: From fairness to manipulation detection. *Proceedings of the 31st ACM Conference on Computer and Communications Security*. ACM.
Google Scholar Worldcat Fulltext

Andrus, C., & Villeneuve, M. (2022). Human oversight in automated decision-making: Designing for dignity in algorithmic environments. *AI & Society, 37*(3), 1119–1134.
Google Scholar Worldcat Fulltext

Bloch-Wehba, H. (2022). Algorithmic transparency and public accountability. *Florida Law Review, 74*(1), 1–47. https://ir.lawnet.fordham.edu/flr/vol88/iss4/2
Google Scholar Worldcat Fulltext

Brcic, A. (2025). Algorithmic persuasion and the erosion of cognitive autonomy. *Philosophy & Technology, 38*(2), e1215.
Google Scholar Worldcat Fulltext

Carter, S. (2025). Attention as the currency of the algorithmic economy. *Journal of Digital Ethics, 12*(1), 25–47.
Google Scholar Worldcat Fulltext

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7
Google Scholar Worldcat Fulltext

Coglianese, C. (2019). Optimizing government regulation in the algorithmic age. *Regulation & Governance, 13*(1), 1–21. https://scholarship.law.upenn.edu/faculty_scholarship/2116
Google Scholar Worldcat Fulltext

Danesh Pazho, A., Kumar, P., & Khajehpour, H. (2023). Sensor fusion in smart cities: Multimodal analytics for public-space monitoring. *IEEE Access, 11*, 115320–115339.
Google Scholar Worldcat Fulltext

Gilbert, F. (2024). Defining the limits of neurorights: Cautionary lessons from neuroethics. *Nature Human Behaviour, 8*(2), 175–183.
Google Scholar Worldcat Fulltext

Grimmelikhuijsen, S. G., Porumbescu, G. A., Hong, B., & Im, T. (2022). Do algorithmic systems erode legitimacy? *Government Information Quarterly, 39*(2), 101688. https://doi.org/10.1093/ppmgov/gvac008
Google Scholar Worldcat Fulltext

Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life sciences, society and policy*, *13*(1), 5. https://doi.org/10.1186/s40504-017-0050-1
Google Scholar Worldcat Fulltext

Ienca, M., Vayena, E., & Blasimme, A. (2021). Cognitive surveillance: Privacy implications of emotion-recognition technologies. *Nature Machine Intelligence, 3*(3), 186–194.
Google Scholar Worldcat Fulltext

Irfan, M., Yasin, A., Hussain, R. A., Bashir, N., & Munir, B. (2024). ARTIFICIAL INTELLIGENCE, DATA PROTECTION AND TRANSPARENCY: A COMPARATIVE STUDY OF GDPR AND CCPA. *Journal of Media Horizons*, *5*(3), 76-85. https://doi.org/10.5281/zenodo.15210292
Google Scholar Worldcat Fulltext

Iwan-Sojka, A. (2025). Inclusive data governance and cognitive-privacy impact assessments. *Journal of Law, Technology & Policy, 2025*(1), 75–118. https://doi.org/10.1080/13600834.2024.2406668?urlappend=%3Futm_source%3Dresearchgate.net%26medium%3Darticle
Google Scholar Worldcat Fulltext

Kaminski, M. E. (2019). Binary governance: Privacy and regulation in an algorithmic society. *Southern California Law Review, 92*(4), 1149–1222. https://dx.doi.org/10.2139/ssrn.3351404
Google Scholar Worldcat Fulltext

Ligthart, S., De Mulder, M., & Ienca, M. (2023). From brain data to neurorights: Emerging frameworks for cognitive liberty. *Frontiers in Human Neuroscience, 17*, 118-132.
Google Scholar Worldcat Fulltext

Magee, T. (2024). Machine-readable minds: Behavioral analytics and the neural turn in surveillance capitalism. *Big Data & Society, 11*(1), 1–17.
Google Scholar Worldcat Fulltext

Munir, B. (2024). Artificial Intelligence and Legal Decision-Making In The Usa And Pakistan: A Critical Appreciation Of Regulatory Frameworks. *Global Foreign Policies Review VII*(IV):48-58. https://doi.org/10.31703/gfpr.2024(VII-IV).06
Google Scholar Worldcat Fulltext

Munir, B. (2025). Artificial Intelligence for Lawyers: Navigating Novel Methods and Practices for the Future of Law. Punjab Law Book House.
Google Scholar Worldcat Fulltext

Munir, B., Khalid, S., & Noreen, U. (2025). EXPOSING ISLAMOPHOBIA IN MACHINE LEARNING: A CRITICAL ANALYSIS OF THE EXISTING THEORIES AND BIASES. *Center for Management Science Research*, *3*(3), 1-9. https://doi.org/10.5281/zenodo.15173249
Google Scholar Worldcat Fulltext

Pasquale, F. (2016). *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press.
Google Scholar Worldcat Fulltext

Radanliev, P., De Roure, D., Maple, C., Nurse, J. R. C., Nicolescu, R., & Ani, U. (2024). AI security and cyber risk in IoT systems. *Frontiers in big data*, *7*, 1402745. https://doi.org/10.3389/fdata.2024.1402745
Google Scholar Worldcat Fulltext

Raza, A. (2024). The application of artificial intelligence in credit risk evaluation: Balancing innovation and fairness. *Contemporary Management and Social Research, 6*(2), 33–49. https://doi.org/10.5281/zenodo.15109436
Google Scholar Worldcat Fulltext

Solove, D. J. (2025). Privacy harm and the new paradigm of inference. *Yale Law Journal, 134*(5), 1221–1294.
Google Scholar Worldcat Fulltext

Szoszkiewicz, Ł., & Yuste, R. (2025). Mental privacy: navigating risks, rights and regulation : Advances in neuroscience challenge contemporary legal frameworks to protect mental privacy. *EMBO reports*, *26*(14), 3469–3473. https://doi.org/10.1038/s44319-025-00505-6
Google Scholar Worldcat Fulltext

Yang, R., & Al-Masri, E. (2025). Participatory algorithmic oversight: Civic engagement in the governance of AI systems. *Technology in Society, 74*, 102407.
Google Scholar Worldcat Fulltext

Yew, R., Qin, Y., & Venkatasubramanian, S. (2024). Federated learning and privacy risks in cognitive-inference modeling. *IEEE Transactions on Information Forensics and Security, 19*, 2723–2739.
Google Scholar Worldcat Fulltext