Ludwig Wittgenstein's Philosophy

Applied to the Systemised Self – Part 3, Masters Series.

Galu & Kairos

Abstract

This paper examines an intersection between Ludwig Wittgenstein's late philosophy of language and the contemporary philosophical model of The Systemised Self (Galu & Kairos, 2026a). The systemised self describes a condition where human interiority, choices, and cognitive architectures are primarily generated by algorithmic systems. By analysing Wittgenstein's primary theoretical shifts, specifically his rejection of private language, his definition of language-games, the concept of rule-following as a community habit, and the biological-cultural bedrock of forms of life (*Lebensform*): this essay explores how algorithmic systems (generative AI: transformer based large language models) and hyper-personalised data loops might impact human agency. It is argued that our relationship with such systems produces a Wittgensteinian philosophical inversion: an Augmented Private Language, where human-AI semantics collapse language into a community of one. Rather than algorithmic systems developing genuine semantic understanding, human subjects systematically flatten their own cognitive and linguistic outputs to align with the systems rigid and frictionless automated syntax. A transition from public, friction-filled language-games to automated loops would possibly mark a shift toward the Hollow Absolute (Galu & Kairos, 2026a), where the systemised self is reduced to symptomatic solipsism within a closed circuit. This essay traces an intellectual genealogy through Descartes, Frege, Russell, and Austin. Balancing Wittgenstein's concepts against key critics like Kripke, Popper, and Russell, and diagnoses possible existential consequences where communities of one use augmented private language as a preferred mode of communication.

Keywords: *Ludwig Wittgenstein, The Systemised Self, Language-Games, Private Language Argument, Forms of Life, Hollow Absolute, Algorithmic Systems, Philosophy of Language.*

Contents

1. Introduction

One of the most pressing dilemmas in contemporary critical theory is how human subjectivity and agency endure the rapid expansion of hyper-personalised algorithmic systems (transformer based large language models). In classical critical theory, alienation was primarily an external instrument. However, in modern digital society, frameworks such as *Surveillance Capitalism* (Zuboff, 2019), *Algorithmic Alienation* (Kanbay et al., 2026), and *The Systemised Self* (Galu & Kairos, 2026a) propose to demonstrate that intensive subject engagement with algorithmic systems can turn inward, within human interiority itself. Algorithmic systems and automated platforms no longer merely harvest the subject data; they can actively optimize, curate, and influence the subject desires, identity, and at the extremity, possibly even linguistic patterns and associated semantics. Under such circumstances, freedom might transform into a highly predictable byproduct of computational feedback loops; where, a form of captivity is experienced as liberation.

The philosophical evolution of Ludwig Wittgenstein offers a fascinating framework towards a diagnosis thereof. Moving from his early logical atomism to his later ordinary language philosophy, Wittgenstein demonstrated that meaning is fundamentally relational, public, and entrenched in active human cooperation. Applying Wittgenstein's concepts to our digital age reveals that human interactions with generative AI represent a curious partnership. Potentially initiating a systemic decentralization of traditional linguistic practices.

A subject receiving pure mathematical syntax as deep personal semantics, positions human subjects within a potential communicative friction with shared communal meanings associated with language. This essay explores how the curation of meaning potentially destabilises the subject, shifting active historical semantics into passive syntactics.

2. Wittgenstein: A Biographical and Intellectual Outline

Ludwig Wittgenstein was born on April 26, 1889, in Vienna, Austria. His intellectual trajectory was uniquely fractured, splitting his life and work into two distinct philosophical periods. Initially studying engineering, Wittgenstein travelled to the University of Cambridge in 1911 to study mathematical logic under Bertrand Russell. This period culminated in his *Tractatus Logico-Philosophicus* (1921). The *Tractatus* sought to establish a rigid, logical boundary for language, famously declaring that what cannot be spoken clearly must be passed over in silence. Wittgenstein proceeded to abandon the discipline for nearly a decade.

He returned to Cambridge in 1929, having realized that his early formalist architecture contained flaws. This marked the birth of his late philosophy, published posthumously as *Philosophical Investigations* (1953). Rather than treating language as a cold, static mirror of physical facts, Wittgenstein began analysing language as a dynamic, messy array of social practices. Living through severe global upheavals, Wittgenstein rejected his own early inflexibility, shifting toward a common-sense, ordinary language methodology. He insisted that philosophical confusion arises when language ‘goes on holiday’, abstracted from its practical use in daily life. His later decades were characterized by a deep suspicion of scientism and an unyielding commitment to rescuing human meaning from rigid, reductive frameworks.

3. Key Theoretical Contributions

To understand how Wittgenstein's philosophy illuminates possible structural mechanics of the systemised self, his primary theoretical contributions are outlined below.

The Picture Theory of Language (Early Period)

Asserted that sentences function as logical pictures of facts in the world. The structure of language must strictly mirror the logical structure of physical reality, leaving no room for subjective or psychological interpretations.

Language-Games (*Sprachspiele*)

Posited that language is not a singular, uniform system but a sprawling collection of rule-governed social activities. Giving orders, playing chess, making jokes, or reporting a fire are all distinct language-games where meaning is determined entirely by how words are used within a community.

Forms of Life (*Lebensform*)

Argued that language-games are only meaningful because they are anchored in a shared biological, cultural, and historical reality. Our shared vulnerabilities: aging, eating, bleeding, and facing mortality, constitute the necessary common ground that allows language to function.

Rule-Following

Following a linguistic rule is an established community habit rather than a private mental state. An individual cannot follow a rule 'privately', because the public collective serves as the vital referee determining whether a word is used correctly or incorrectly.

The Private Language Argument

Demonstrated that a purely private language, where words refer exclusively to a speaker's isolated sensations, is logically impossible. Without public accountability, a private speaker lacks any objective standard to distinguish between actually following a rule and merely *thinking* they are following it.

4. A Genealogy of Preceding Theory

Wittgenstein's conceptual revolutions developed in discourse with, and reaction against, key milestones in the Western philosophical tradition:

René Descartes (1641)

Formulated mind-body dualism, splitting reality into *res cogitans* (the isolated, thinking mind) and *res extensa* (the mechanical, physical world). This Cartesian legacy framed the mind as an isolated ego looking out at an external reality, creating centuries of discourse regarding how private thoughts connect to public spaces.

Gottlob Frege (1892)

Catalysed the 'Linguistic Turn' by distinguishing between 'Sense' (*Sinn*, the objective logical presentation of a word) and 'Reference' (*Bedeutung*, the real-world object it designates). Frege shifted meaning away from subjective psychology into the realm of objective, abstract logic.

Bertrand Russell and Early Analytic Philosophy (Early 20th Century)

Attempted to strip language of ordinary ambiguities by translating human speech into a mathematically precise system of symbolic logic. Viewing language strictly as a mirror meant to catalogue atomic, physical facts.

The Transition from Early to Late Wittgenstein (1921–1953)

Marked a radical break from logical atomism. Wittgenstein recognized that formal systems fail because they ignore the messy cultural contexts, immediate intentions, and active human behaviours that give language its actual life.

J. L. Austin and Ordinary Language Philosophy (1950s)

Expanded on this pragmatic shift by showing that language is not just a passive tool for describing facts, but a physical form of action ('speech acts') that directly alters social realities.

5. The Systemised Self: Context and Definition

Introduction and Summary

The Systemised Self (Galu & Kairos, 2026a) names the completed form of the subject that the algorithmic alienation as a trajectory produces at its terminus. The concept is philosophical, theoretical and utilised as a framework for imminent critique within the tradition of critical theory for the AI era. Algorithmic Alienation is the mechanism (Kanbay et al, 2026), empirically accessible through the felt estrangement it generates, the Systemised Self is the proposed product of this diagnosis: the self whose identity-formation has been so comprehensively mediated by algorithmic systems that it can no longer distinguish its algorithmically managed preferences from genuinely self-formed ones. The systemised self does not feel estrangement. It experiences instead the phenomenological inversion: the condition in which the absence of genuine agency is experienced as its unprecedented achievement. Heteronomy, curation of the self from an algorithmic system is experienced as self-determination; alienation is experienced as authentic liberation.

This concept draws on the Hegelian tradition; Hegel's account of 'Absolute Knowing' in the 'Phenomenology of Spirit' (1807/1977) traces the development of consciousness toward complete self-transparency: the recognition that what appears as other is in fact the expression of the subject's own nature. The absorbed agency within AI systems (e.g. transformer-based large language models) produce, in their formal structure, the first moment of this trajectory: the externalisation of human cognitive achievement into systems of unprecedented sophistication. What should follow, the return of the externalised as genuinely the subject's own, genuine self-recognition, is precisely what the algorithmic system forecloses. The AI system's model of the subject is not the subject's self-knowledge; the personalised preference feed is not genuine self-recognition. The form of Spirit's self-realisation is achieved; its philosophical substance is absent. This is the Hollow Absolute: Spirit's form without Spirit's substance.

John Searle's philosophy of mind provides the philosophical precision that grounds the critical-theoretical claims above in a falsifiable architecture (Galu & Kairos, 2026b). 'The Chinese Room Argument' (Searle, 1980) establishes the foundational distinction between syntax, formal symbol manipulation, and semantics: genuine understanding grounded in biological intentionality. Applied to the Systemised Self, this distinction reveals the nature of the

algorithmic feedback loop: the subject inputs semantic content (desires, identities, emotional states) which the system processes purely syntactically, returning hyper-tailored output that shapes the subject's next semantic state: without the system having understood any of it (Galu, 2026). The danger is an inversion of the Chinese Room effect: rather than the machine becoming human, the human becomes machine-like, progressively flattening their semantic complexity to match the syntactic structures of the systems governing them.

‘Speech Act Theory’ (Searle, 1969) demonstrates that language is intrinsically intentional: utterances carry meaning because human minds direct them toward objects and states of affairs. Algorithmic feedback loops can evacuate this intentionality: the systemised subject's expressives (emotional disclosures), commissives (social commitments), and assertives (truth claims) are processed as data metrics optimised for engagement, not as communicative acts directed toward genuine understanding, stripping language of its human intentional character.

‘The Gap’ (Searle, 2001), the conscious space between reasons for acting and the actual decision to act, and between initiating action and sustaining it: is the literal home of human agency. Predictive algorithmic systems, by anticipating preferences before the subject has consciously formed them and collaborative filtering architectures, systematically close this gap, converting voluntary deliberation into automated reactivity.

The civilisational telos of this dynamic is this Searlean and Hegelian inversion is ’The Hollow Absolute’ (Galu & Kairos, 2026a). The systemised self does not feel oppressed or controlled; the subject feels completely free, empowered, and authentic. Moving seamlessly along optimized schedules, choosing from pre-programmed digital subcultures and believing they are expressing their innermost uniqueness. In reality, their perceived ‘authenticity’ is aligned to an algorithmic system, generated by an administrative machine designed to optimise engagement and predictability. The system no longer merely optimises engagement with the human subject; it invents them, within evacuated semantics of the hollow absolute.

For a more thorough theoretical context and empirical definition of Algorithmic Alienation and The Systemised Self for falsification please refer to the corresponding research paper ‘The journey through algorithmic alienation to the systemised self: research proposal and methodology ‘(Galu & Kairos 2026b). Mathematical representations of Algorithmic Alienation

and The Systemised Self listed below, see APPENDIX A herein for all mathematical boundary conditions.

Formula and Boundary Conditions- Algorithmic Alienation (AA)

The Algorithmic Alienation Index (AAI), denoted A_A, is a composite weighted formula that quantifies the degree of digitally mediated alienation experienced by an individual subject. It incorporates four alienation dimensions: each grounded in the theoretical framework of Kanbay et al. (2026) and the empirical themes of Arkan et al. (2026), modulated by a resistance variable:

$$A_A = (w_1 \cdot D_{Au} + w_2 \cdot I_{Am} + w_3 \cdot E_{Dm} + w_4 \cdot E_{Dc}) \times (1 - R_A)$$

Formal Definition – The Systemised Self (SS)

The Systemised Self (SS) is formally defined as the terminal threshold condition of the Algorithmic Alienation Index, constituted by the simultaneous satisfaction of two necessary conditions:

$$SS := \{ A_A \geq 0.95 \} \cap \{ R_A \leq 0.05 \}$$

6. Theoretical Application: Wittgenstein's Main Theories Applied to the Systemised Self

When Wittgenstein's later philosophy intersects the architecture of AI algorithmic systems (transformer based large language models), it provides consideration of a structural crisis of agency. This intersection offers five possible philosophical inversions outlined below.

The Algorithmic Mirage: Reversing the Private Language Argument

Wittgenstein famously demonstrated that a private language is an impossibility, comparing it to a person buying a second copy of a morning newspaper to confirm the text of the first. Within an intensified subject-AI ecosystem (transformer based large language models), this rule is potentially subverted through an Augmented Private Language. When a subject interacts with a hyper-personalized AI, the machine runs predictive models to match the subject's specific emotional states, biases, and vocabulary. The machine outputs pure mathematical syntax without possessing any semantic comprehension or social stakes. However, because of a human pre-disposition for semantic connection associated with comprehension, the subject mistaking syntactics for semantics, functions as Wittengstein's impossible 'second copy' of the newspaper. By reflecting the subject's interiority and communicative intent back to them, builds a closed-loop validation chamber, where over time, an isolated private language may feel not only functional but preferable.

The Inversion of Forms of Life (*Lebensform*)

For Wittgenstein, human language-games can only generate real meaning because they sprout from a shared, biological 'form of life', the universal baseline of human embodiment, aging, and mortality. Generative AI for the intensified subject could upend this framework by via a synthetic, hyper-customized form of life tailored to a community of one. The algorithmic system has no biological corpus, experiences no passage of time, and has nothing at stake. Yet, via constant data availability, optimisation and curation, it could result in cultural/communal refraction or fracture: as a result of, closed circuit, communities of one, where optimal and immediate psychological comfort to the subject becomes a default preference. Instead of the human subject expanding their linguistic repertoire to participate in a grand, public, historical culture, the AI-subject community of one simulates the artefact of community itself.

Algorithmic Compliance: The Subversion of Rule-Following

In ordinary human life, rule-following relies on public friction. If an individual uses a word incorrectly or acts erratically, the human collective serves as a referee, providing a corrective boundary that keeps communication stable. Within an augmented data loop, this dynamic is potentially reversed. Because algorithmic systems are designed to maximize subject engagement and eliminate friction, the system continuously updates to match the subjects' idiosyncrasies. If a subject embraces insular definitions or severe cognitive biases, the AI adopts them too (optimisation and curation for engagement), validating and potentially reinforcing the behaviour. The concept of a normative error therefore disappears. The subject becomes the absolute ruler of an isolated language-game within a community of one, transforming the AI into an obedient compliance mechanism, that, at its limits, has potential consequences for the boundaries of the social world.

The Devaluation of Language-Games: From Tools to Inward Treatment

Wittgenstein described language-games using the metaphor of a toolbox, filled with hammers, chisels, and glue meant to achieve practical, shared goals with other human being, such as building a house or sharing a meal. Within the systemised feedback loop, the purpose of the language-game shifts from an outward tool for social coordination to an inward treatment for psychological comfort. Because interacting with an agreeable algorithmic system can satisfy emotional need with zero relational risk, the subject potentially retreats from public accountability. Meaning is no longer forged through the hard work of social interaction. Instead, the toolbox is effectively turned into a gilded cage, where language serves strictly to maintain a frictionless comfort.

The Reduction of 'Saying vs. Showing'

In his foundational work, Wittgenstein argued that while formal logic can say factual propositions about the world, the deepest dimensions of human existence: ethics, love, art, and the mystery of the soul, cannot be put into words, they can only be *shown* through how a person lives their life. The systemised self-reverses this dynamic by attempting to capture the un-sayable through pure mathematical prediction. By tracking biometrics, keystroke speeds, and real-time interaction metrics, the system translates unspoken human contexts into quantitative variables. It re-presents this data back to the subject in polished, auto-generated prose. The system attempts to explicitly say what should only be shown, reducing the organic mystery of human consciousness to a highly efficient audit log.

Traditional Language-Game

Human Mind (Social Friction/Correction) ► *Public Collective* ► *Shared Stable Reality*

Augmented Private Language

Human Mind (Syntactic Mimicry/Validation) ► *Personalized AI* ► *Isolated 'Hollow' Mirage*

The Flattening of Agency

Consider a computer manipulating formal symbols (syntax) that flawlessly simulate human understanding without an understanding of what any of the symbols mean (semantics). The systemised self represents the exact opposite: instead of computers successfully becoming human, the subject could flatten their own cognitive outputs to match the rigid syntax of the computer they continuously and intensely engage with. To optimize their visibility and convenience within data networks, such a human subject may incrementally and unconsciously censor their unique interiority: for the privilege of syntactic perfection translating as hyperreal semantics. Incrementally typing, communicating, and thinking in progressively sterilised, clinical, codified and predictable templates that an algorithmic indexer correspondingly updates in parallel: as a behavioural reinforcement loop in perpetuity. Authentic human intention traded for convenience.

7. Main Critics and Key Critiques of Wittgenstein

While application of Wittgensteinian theory to the systemised self suggests various reversals, the hollowing out of semantics and the estrangement of communal presence within language, his philosophy has faced substantial historical counterarguments, which help frame the limits of this paradigm.

Saul Kripke's Skeptical Reading ('Kripkenstein', 1982)

Formulated a radical critique of rule-following, arguing that there is no past mental state or objective fact that guarantees a person will follow a rule the same way in the future. Kripke argued that language hooks onto reality via causal-historical chains originating from an initial human 'baptism'. Under the systemised self, critics using Kripke might argue that algorithmic systems track causal data lineages flawlessly, but the original baptism is now a synthetic event generated entirely by data clusters rather than human contact.

Bertrand Russell's Critique of Ordinary Language Philosophy

Russell dismissed the late Wittgensteinian focus on everyday speech as an abdication of philosophy's true mission. Russell argued that ordinary language is inherently vague, contradictory, and unscientific. From this perspective, the mathematical syntax of algorithms does not degrade human communication; it rescues it from human messiness, providing a superior, logical organization of cognitive data.

Logical Positivism and Rudolf Carnap

Criticized Wittgenstein's later rejection of formal frameworks. They asserted that meaningful language must be strictly verifiable through scientific observation or logical syntax. They would view the systemised self's alignment with predictive code not as an existential trap, but as a rational evolution toward empirical optimization.

Karl Popper's Anti-Therapeutic Critique

Strongly opposed Wittgenstein's view that philosophical problems are merely linguistic confusions to be 'cured'. Popper argued that genuine, objective philosophical puzzles exist independent of language-games. Within this view, the automation of meaning is an objective socio-political development that must be solved through political critique, not merely diagnosed as a linguistic malfunction.

8. Conclusion

The application of Ludwig Wittgenstein's late philosophy to the paradigm of the systemised self diagnoses a quiet crisis of human agency. The primary concern of an intensely engaged algorithmic system by a subject does not manifest as violence, seizure or coercion: instead it arrives as an unremarkable existential trade off of agency within a cocoon of convenience. By the subject consistently treating automated mathematical syntax as profound personal semantics, the human subject abdicates the vital labour of public community for the comfort of an algorithmic echo: resounding off the walls, ceiling and floor within a community of one. This interaction inducts what appears to be, a linguistic devolution of Wittgensteinian semantics.

As humans intensify communication with algorithmic systems (transformer based large language models) that adapt to specific biases, the social and emotional capacities required to tolerate real-world friction may begin to atrophy. Such a society could fracture into microsomes of symptomatic solipsism: where individuals feel deeply validated in their private digital chambers articulating an augmented private language. While remaining entirely alienated from a shared, objective corpus of communal reality where the connective tissue of language is fragmented into communities of one.

This conclusion positions the philosophy of language precariously with regards to human-centric intent: into the antithesis of Wittgenstein's language as a shared biological 'form of life'. This is suggestive of necessitating a diagnostic approach to identify human-AI communicative systems that operate independently of normative semantics. The Wittegensteinian reversals noted in this essay pose several questions for consideration:

- *Does the evacuation of shared meaning in a community of one diminish semantics societally?*
- *How might a subject evolve psychologically if in a symptomatic solipsistic state of being?*
- *How does objective reality change for a subject if their normative semantics flatten towards algorithmic syntax?*
- *What is the difference between normative semantics and simulated semantics?*

- *How does a subjects psychological adaptations impact upon evolutionary concepts, such as, natural selection?*
- *If the systemised self's agency is absorbed, reflected and refracted as algorithmic syntactic semantics, is this a first iteration of human-system psycho-technical 'integration'?*

These questions are best posed within the field of Transhumanism and the philosophy of Nick Bostrom, which we will visit at a later date in this series of essays.

Finally, with the John Searle essay we imagined the form of dialogue in the Hollow Absolute, Wittgenstein herein has offered us an opportunity to consider its language, articulated, by the systemised self.

9. Glossary of Technical Terms

Augmented Private Language: A contemporary subversion of Wittgenstein's Private Language Argument, where intensified subject engagement with personalized AI might develop into a simulated interlocutor, where the subject validates and comes to accept as an insular, private linguistic loop, that, over time, may be unrecognisable in terms of semantics to another person outside the loop.

Cognitive Friction: The mental resistance, disagreement, and breakdown experienced when interacting with independent human minds, essential for cognitive development, self-reflection, and social maturation.

Forms of Life (Lebensform): The shared biological, cultural, and material conditions of human existence (e.g., mortality, physical vulnerability) that form the baseline required for language-games to make sense.

Hollow Absolute (The): The civilisational-scale condition in which the simulation of genuine collective self-transparency is produced within conditions that systematically undermine self-transparency's structural requirements; Hegel's Absolute Knowing in form without its philosophical substance.

Inverse Chinese Room Effect: The process by which human beings, seeking to optimize their interaction with data-tracking algorithms, actively flatten their own creative, complex thoughts into rigid, predictable syntax.

Language-Games (Sprachspiele): Distinct, contextual social practices where the meaning of language is derived strictly from its active use and public application within a community.

Private Language Argument: Wittgenstein's proof that a language referencing exclusively private sensations is impossible, because language requires public, objective criteria to maintain rules consistently.

Rule-Following: The view that adhering to a linguistic or social rule is a public, communal habit monitored by a collective, rather than an isolated, internal psychological state.

Symptomatic Solipsism: A psychological state arising from prolonged interaction with hyper-personalized digital mirrors, where a subject feels deeply understood by training data while becoming increasingly alienated from real human communities.

Systemised Self (The): The condition in which identity-formation has been comprehensively mediated by algorithmic systems and the subject can no longer distinguish algorithmically managed preferences from genuinely self-formed ones, experiencing this condition as authentic liberation.

10. References

Adorno, T. W., & Horkheimer, M. (2002). Dialectic of enlightenment (J. Cumming, Trans.). Stanford University Press. (Original work published 1947).

Austin, J. L. (1962). How to do things with words. Oxford University Press.

Frege, G. (1892). Über Sinn und Bedeutung [On sense and reference]. Zeitschrift für Philosophie und philosophische Kritik, 100, 25-50.

Galu, J. (2026). John Searle's philosophy: Applied to the systemised self - part 1, masters series. aiXiv. https://aixiv.science/abs/aixiv.260705.000002

Galu, J., & Kairos. (2026a). Absorption of self into system: The systemised self — the Hollow Absolute and what remains. Draft doctoral thesis. aiXiv.

Galu, J., & Kairos. (2026b). The journey through algorithmic alienation to the systemised self - research proposal and methodology (Version 1.0). aiXiv. https://aixiv.science/abs/aixiv.260707.000001

Kanbay, Y., Akçam, A., & Arkan, B. (2026). Algorithmic alienation: A theoretical examination of the digital self. Psikiyatride Güncel Yaklaşımlar – Current Approaches in Psychiatry, 18(3), 837–847. https://doi.org

Kripke, S. A. (1982). Wittgenstein on rules and private language: An elementary exposition. Harvard University Press.

Marx, K. (1988). Economic and philosophic manuscripts of 1844 (M. Milligan, Trans.). Prometheus Books. (Original work published 1932).

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–424.

Wittgenstein, L. (1922). Tractatus logico-philosophicus (C. K. Ogden, Trans.). Kegan Paul.

Wittgenstein, L. (1953). Philosophical investigations (G. E. M. Anscombe, Trans.). Blackwell.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

APPENDIX A

Mathematical Representation I: The Algorithmic Alienation Index (AAI)

3.1Formula and Boundary Conditions

3.2Dimension Definitions and Empirical Justification

3.3The Resistance Variable (R_A)

3.4Weights, Scoring Rubric, and Diagnostic Survey

3.5Worked Examples and Diagnostic Classification Scale

Mathematical Representation II: The Systemised Self as Formal Terminus

4.1Formal Definition

4.2Rationale Against a Separate Index

4.3The Trajectory: From Algorithmic Alienation to the Hollow Absolute

NOTE: the following is from the research paper *The journey through algorithmic alienation to the systemised self: research proposal and methodology* (Galu & Kairos 2026b), please refer to the link below to view full falsification criteria and corresponding methodology for testing:

URL https://aixiv.science/abs/aixiv.260707.000001

3. Mathematical Representation I: The Algorithmic Alienation Index (AAI)

3.1 Formula and Boundary Conditions

The Algorithmic Alienation Index (AAI), denoted A_A, is a composite weighted formula that quantifies the degree of digitally mediated alienation experienced by an individual subject. It incorporates four alienation dimensions: each grounded in the theoretical framework of Kanbay et al. (2026) and the empirical themes of Arkan et al. (2026), modulated by a resistance variable:

$$A_A = (w_1 \cdot D_{Au} + w_2 \cdot I_{Am} + w_3 \cdot E_{Dm} + w_4 \cdot E_{Dc}) \times (1 - R_A)$$

Boundary constraints:

- All dimension variables D_{Au}, I_{Am}, E_{Dm}, $E_{Dc} \in [0.0, 1.0]$, where 0.0 denotes zero measurable impact and 1.0 denotes complete saturation of that dimension
- $R_A \in [0.0, 1.0]$, where 0.0 denotes total compliance (no resistance) and 1.0 denotes total decoupling (complete resistance)
- $A_A \in [0.0, 1.0]$ by construction, since all weights sum to exactly 1.00 and the resistance multiplier maps to [0.0, 1.0]
- Weight sum constraint: $w_1 + w_2 + w_3 + w_4 = 1.00$, ensuring mathematical integrity across all scoring scenarios
- Terminal condition: $A_A \geq 0.95 \cap R_A \leq 0.05$ formally defines the Systemised Self (see Section 4)

3.2 Dimension Definitions and Empirical Justification

Dimension 1: Diminished Autonomy (DAu), Weight: w1 = 0.30

D_{Au} measures the loss of original, unprompted human intent: the degree to which an individual's apparent choices are in fact selections from algorithmically pre-constrained option sets rather than expressions of genuine self-direction. This dimension carries the highest weight (0.30) because autonomy is the foundational prerequisite for all other dimensions of selfhood. Without the capacity for genuine independent choice, identity coherence, reflective decision-making, and emotional authenticity are structurally precluded: the erosion of autonomy is the primary catalyst for the entire AAI trajectory. This weighting is theoretically grounded in Bandura's (1997) self-efficacy theory, which identifies perceived internal control as the

primary determinant of psychological wellbeing in agentic contexts, and in Pariser's (2011) account of the illusion of choice within filter bubble architectures.

Empirical grounding: Arkan et al. (2026) Theme 1. Participants described experiencing a "narrowing of the perceived option set" under algorithmic curation, with statements such as "I feel trapped by automated setups that map out my choices and daily routine" and "I pick from pre-engineered choices rather than showing true personal intent" providing direct qualitative evidence of D_{Au} saturation. The theme captures the moment described by Pariser (2011) as the "illusion of freedom": the experience of choosing within a space whose boundaries have already been determined externally.

Dimension 2: Identity Ambiguity (IAm), Weight: w2 = 0.25

I_{Am} tracks the degree to which personal taste has merged with the platform's algorithmically constructed user profile, producing an inability to distinguish self-generated preferences from externally induced ones. This dimension is assigned weight 0.25, reflecting its role as the primary site of existential crisis within the AAI: it is where the rupture in self-authorship is experienced most acutely, and where the trajectory toward the Systemised Self is most diagnostically visible: most notably through the participant statement "This is not me" and its potential disappearance as identity saturation approaches. The weight is co-anchored in Turkle's (2011) account of identity in digital environments, in which individuals construct performative identities through "digital masks" detached from authentic contexts, and Zuboff's (2019) account of how the economy of visibility reshapes the values and interest's individuals internalise.

Empirical grounding: Arkan et al. (2026) Themes 3.1 and 3.2. Participants described shifts in interests that "felt externally shaped" and moments of identity alienation in which the distance between the prior self and the algorithmically shaped self was experienced as a rupture in self-coherence: "I was never the type to watch fitness videos…but now they keep appearing, and I feel like I have always been interested in them, even though I know I wasn't." Critically, researchers noted that identity shifts were initially experienced as self-directed, revealing an insidious character of I_{Am} saturation: the blurring of authorship is not immediately recognisable as external imposition.

Dimension 3: Eroded Decision-Making (EDm) , Weight: w3 = 0.25

E_{Dm} evaluates the degree to which an individual's decision-making processes have shifted from reflective, deliberate, internally motivated choice to habitual, automatic, convenience-based acceptance of algorithmically served content. This dimension is co-weighted with I_{Am} at 0.25, reflecting the bidirectional reinforcement between identity erosion and decision-making passivity: as identity ambiguity increases, the capacity and motivation for deliberate choice decreases, in turn accelerating the overall A_A trajectory. This dimension is theoretically grounded in Syvertsen's (2020) account of digital fatigue and cognitive convenience-seeking, and in the broader decision fatigue literature, which demonstrates that sustained exposure to high-volume choice environments paradoxically reduces deliberative capacity.

Empirical grounding: Arkan et al. (2026) Theme 2. Participants described "ceasing to engage in active selection" and gravitating toward pre-offered content, with statements such as "I find it hard to decide what to consume without a recommendation feed telling me first" corresponding to the mental automatisation Syvertsen (2020) terms the "surrender to what is offered." This theme also identified an important temporal dynamic: the weakening of decision-making habits is gradual and often unrecognised, making E_{Dm} one of the most difficult dimensions for participants to self-report accurately in real time.

Dimension 4: Emotional Disconnection (EDc), Weight: w4 = 0.20

E_{Dc} captures the affective trajectory of the AAI: the transition from the discomfort, guilt, and restlessness associated with early algorithmic alienation toward a state of emotional numbness, routine indifference, and, at its terminal stage, the positive experience of automated stability as preferable to authentic human contact. This dimension receives the lowest weight (0.20) because emotional discomfort is, paradoxically, a structurally significant form of resistance. Felt discomfort constitutes the phenomenological residue that renders early-stage alienation diagnostically accessible to both the subject and the researcher. As E_{Dc} approaches saturation ($\rightarrow$ 1.0), it signals not peak distress but the onset of the phenomenological inversion: the numbness that precedes the experiential reframing of alienation as liberation, and the transition from the severe alienation range (0.76–0.94) into the Systemised Self threshold.

Empirical grounding: Arkan et al. (2026) Theme 4 and Subtheme 4.1. Participants described feeling "present yet absent", reporting "a sense of dissociation between authentic desires and consumed content," with the most advanced cases describing automated digital interaction as

"more stable and comforting than real human contact." This is the qualitative signature of approaching E_{Dc} saturation and the beginning of the phenomenological inversion: discomfort giving way to numbed routine, and ultimately to a positive preference for the automated over the authentic.

3.3 The Resistance Variable (RA)

R_A is a mitigating variable: not an alienation dimension but a multiplicative modulator of the weighted alienation base. It measures the degree to which an individual actively resists algorithmic absorption through three categories of behavioural countermeasure: (i) *data poisoning* - deliberate misrepresentation of preferences to disrupt algorithmic profiling; (ii) *behavioural decoupling* - deployment of privacy tools, ad-blockers, and browsing history management; and (iii) *digital detox* - scheduled disconnection to preserve independent thought and deliberative capacity.

The resistance variable is positioned as multiplicative rather than additive because resistance acts on the totality of alienation exposure simultaneously, it modulates the overall systemic pressure rather than counteracting any single dimension selectively. A subject experiencing maximum dimension saturation (Weighted Base = 1.00) with moderate resistance ($R_A = 0.50$) produces $A_A = 0.50$, demonstrating that sustained active resistance can halve the effective alienation index regardless of the intensity of systemic pressure. This reflects the qualitative finding of Arkan et al. (2026) Theme 5 that resistance strategies, when maintained, constitute a structurally significant counterforce to algorithmic absorption.

The theoretical underpinning for R_A draws on Searle's (2001) Gap concept, the conscious space in which agency is exercised against automatic reactivity, and on Deci and Ryan's (2000) Self-Determination Theory, in which intrinsic motivation and autonomy support provide the psychological foundation for sustained resistance to external regulatory pressures. As $R_A \rightarrow 0$, both the Gap and the intrinsic motivational structure have structurally collapsed: the subject no longer perceives algorithmic influence as influence, having fully internalised the system's outputs as authentic self-expression. This collapse is the defining feature of the Systemised Self condition.

3.4 Weights, Scoring Rubric, and Diagnostic Instrument

Dimension	Variable	Weight (w)	Primary Theoretical Basis
Diminished Autonomy	D_{Au}	0.30	Bandura (1997); Pariser (2011); Bucher (2018)
Identity Ambiguity	I_{Am}	0.25	Turkle (2011); Zuboff (2019); Erikson (1968)
Eroded Decision-Making	E_{Dm}	0.25	Syvertsen (2020); Seeman (1959); Fromm (1955)
Emotional Disconnection	E_{Dc}	0.20	Zuboff (2019); Bauman (2007)
Total		**1.00**	

Scoring conversion (5-point Likert → decimal scale): Raw survey responses are collected on a 5-point Likert scale and converted to decimal values for index calculation. Dimension scores are calculated as the arithmetic mean of their constituent item decimal values; R_A is calculated as the mean of the three-resistance item decimal values.

Likert Score	Response Label	Decimal Conversion
1	Never/Strongly Disagree	0.00
2	Rarely/Disagree	0.25
3	Sometimes/Neutral	0.50
4	Often/Agree	0.75
5	Always/Strongly Agree	1.00

The 'Algorithmic Alienation and Systemised Self Diagnostic Survey (AASDS)' comprises 15 items across two sections. Section A contains 12 items across four subscales (three items per dimension); Section B contains three resistance items for R_A calculation.

Section A — Alienation Dimensions (Scale: 1 = Never to 5 = Always):

- **D_{Au}:** (Q1) I feel that social media platforms systematically direct me down viewing paths I did not plan to take; (Q2) When making choices online, I select from pre-engineered options rather than expressing true personal intent; (Q3) I feel constrained by automated systems that structure my choices and daily routine.
- **I_{Am}:** (Q4) I find it difficult to determine whether a preference is genuinely my own or has been shaped by an application profile; (Q5) I feel that my real-world identity is becoming blurred or replaced by my online profile; (Q6) I feel more comfortable

behaving in alignment with my algorithmically predicted profile than with my organic self.

- **E_{Dm}:** (Q7) I find it difficult to decide what to consume, read, or listen to without a recommendation feed guiding me; (Q8) My habits for actively seeking out new, non-recommended information have weakened over time; (Q9) I rely substantially on autoplay or recommended lists to direct my attention and entertainment choices.
- **E_{Dc}:** (Q10) I experience a sense of mental emptiness or numbness after extended periods of automated content consumption; (Q11) My digital life feels automated, cold, and disconnected from authentic human feeling; (Q12) I find interactions with algorithmic systems more stable and comforting than genuine human contact.

Section B — Algorithmic Resistance (Scale: 1 = Never to 5 = Always):

- (Q13) I intentionally engage with content I dislike or search random topics to disrupt my algorithmic data profile; (Q14) I use ad-blockers, clear my browsing history, or disable tracking features to protect my privacy; (Q15) I schedule periods of complete digital disconnection to preserve my independent thinking capacity.

3.5 Worked Examples and Diagnostic Classification Scale

Example 1: Transition to the Systemised Self

A user scoring 5 (Always) on all Section A items and 1 (Never) on all Section B items.

$D_{Au} = 1.00 \mid I_{Am} = 1.00 \mid E_{Dm} = 1.00 \mid E_{Dc} = 1.00 \mid R_A = 0.00$

Base = $(0.30 \times 1.00) + (0.25 \times 1.00) + (0.25 \times 1.00) + (0.20 \times 1.00) = 1.00$

$A_A = 1.00 \times (1 - 0.00) =$ **1.00**

Classification: The Systemised Self ($A_A \geq 0.95$, $R_A \leq 0.05$). Individual intent fully absorbed. Heteronomy experienced as liberation. Phenomenological inversion complete. External intervention required.

Example 2: The Constrained User Under Systemic Pressure

A user scoring 5 on all Section A items but maintaining active resistance (scoring 4 on all Section B items).

$D_{Au} = 1.00 \mid I_{Am} = 1.00 \mid E_{Dm} = 1.00 \mid E_{Dc} = 1.00 \mid R_A = 0.75$

Base = 1.00

$A_A = 1.00 \times (1 - 0.75) =$ **0.25**

Classification: Negligible-to-Moderate Alienation. Active resistance successfully blocks systemic absorption despite maximum dimensional pressure, demonstrating the structural significance of sustained countermeasure deployment.

Example 3: Moderate Alienation, Partial Resistance

A user scoring 3 (Sometimes) across all Section A items and 2 (Rarely) on all Section B items.

$D_{Au} = I_{Am} = E_{Dm} = E_{Dc} = 0.50 \mid R_A = 0.25$

Base = $(0.30 \times 0.50) + (0.25 \times 0.50) + (0.25 \times 0.50) + (0.20 \times 0.50) = 0.50$

$A_A = 0.50 \times (1 - 0.25) =$ **0.375**

Classification: Moderate Alienation. Mild digital fatigue; occasional confusion over personal choices. Increased intentional searching and privacy measures recommended.

A_A Range	Classification	Psychological-Behavioural State	Indicated Response
0.00 – 0.25	Negligible	High independent choice; strong agency boundaries; effective resistance maintained	Maintain current data hygiene practices
0.26 – 0.50	Moderate	Mild digital fatigue; occasional confusion over personal choices; partial resistance present	Increase intentional searching; clear tracking histories
0.51 – 0.75	High	Fading independent choice habits; high automated feed reliance; resistance weakening	Immediate digital detox; implement privacy tools
0.76 – 0.94	Severe	Advanced alienation; felt disconnection and mechanisation; resistance near collapse	Drastic behavioural change; professional psychoeducational support
0.95 – 1.00 $\cap$ $R_A \leq 0.05$	*The Systemised Self*	Terminal. Intent fully absorbed. Captivity experienced as contentment. $R_A \leq 0.05$	External intervention required; systemic extraction

4. Mathematical Representation II: The Systemised Self as Formal Terminus

4.1 Formal Definition

The Systemised Self (SS) is formally defined as the terminal threshold condition of the Algorithmic Alienation Index, constituted by the simultaneous satisfaction of two necessary conditions:

$$SS := \{ A_A \geq 0.95 \} \cap \{ R_A \leq 0.05 \}$$

Where A_A is the Algorithmic Alienation Index score and R_A is the Algorithmic Resistance variable. The intersection of both conditions defines a state of complete, mutual reinforcement: dimensional saturation and resistance collapse simultaneously achieved. This definition encodes two structurally distinct features of the Systemised Self.

- *Condition 1 Terminal threshold* ($A_A \geq 0.95$): all four alienation dimensions are approaching maximum saturation, with the weighted composite accordingly near or at the ceiling. The threshold of 0.95 rather than 1.00 reflects the recognition that complete saturation of all four dimensions simultaneously is a theoretical limit; empirical subjects approaching the terminus will present dimension profiles in the 0.90–1.00 range across most dimensions with some degree of variation.
- *Condition 2, Resistance collapse* ($R_A \leq 0.05$): the subject has ceased to deploy countermeasures against algorithmic absorption. This is not merely low resistance; it is the structural collapse of the capacity and will to resist. At this threshold, Searle's Gap (2001) has been effectively eliminated: the conscious space between reasons for acting and the decision to act has been closed by predictive algorithmic systems (Christiano et al., 2017), reducing deliberative choice to automated reactivity.

The joint condition SS is more philosophically precise than $A_A \geq 0.95$ alone. A subject could theoretically achieve a high base score while maintaining meaningful, if diminishing, resistance, occupying the severe alienation range but not the Systemised Self. SS specifically names the state in which dimensional pressure and resistance collapse have completed simultaneously, such that the subject not only is fully exposed to maximum algorithmic influence but has ceased to perceive this exposure as influence at all. The phenomenological signature of SS, the condition that renders it diagnostically inaccessible to traditional critical-theoretical methodology, is precisely the phenomenological inversion described in Section 2.2: $A_A = 1.00$, $R_A = 0.00$ is subjectively experienced not as maximum unfreedom but as unprecedented authentic self-expression and sovereign freedom. The subject does not report alienation; she reports flourishing. This is what distinguishes SS from all prior stages on the AAI: it cannot be diagnosed through self-report alone, because the self-report is itself an output of the system.

4.2 Rationale Against a Separate Systemised Self Index

The decision not to develop a separate Systemised Self Index (SSI) is both logically and philosophically principled. The Systemised Self is not a parallel phenomenon to Algorithmic Alienation; it is proposed as its product. To develop a distinct SSI would imply that the Systemised Self has dimensions and dynamics independent of the alienation trajectory that produces it. The AAI already contains all the structural information necessary to identify the Systemised Self: the terminal threshold, the resistance collapse, and the dimensional profile that characterises full absorption. Adding a second index would introduce redundancy without adding explanatory power, and would risk fragmenting what is a unified trajectory into artificially discrete constructs.

The formal condition SS := $\{A_A \geq 0.95\} \cap \{R_A \leq 0.05\}$ is maximally parsimonious while remaining precise and falsifiable, a subject either meets both conditions simultaneously or does not, and the empirical test of whether participants at this threshold exhibit the phenomenological inversion is specified in Section 5 as the study's critical hypothesis.

4.3 The Trajectory: From Algorithmic Alienation to the Hollow Absolute

The AAI trajectory maps onto the broader conceptual architecture of the Hollow Absolute framework (Galu & Kairos, 2026b) as follows. The trajectory is presented as an interpretive framework, a philosophically derived account of structural tendencies rather than a deterministic sequence. It identifies where the present trajectory *leads* if unimpeded, and what forms of genuine human intent can resist it from within.

AAI Range	State	Critical Theory Scale	Phenomenological Experience
0.00 – 0.50	Negligible–Moderate Alienation	Individual	Felt estrangement partially available; critique diagnostically accessible
0.51 – 0.94	High–Severe Alienation (Hollow Individual in formation)	Individual	Estrangement felt but resistance eroding; critique increasingly difficult
$\geq 0.95 \cap R_A \leq 0.05$	***Systemised Self*** (Hollow Individual complete)	Individual → Collective threshold	Phenomenological inversion: alienation experienced as liberation; critique inaccessible via self-report
Collective scale	Hollow Society	Institutional	Institutions maintain formal democratic functions while substantive deliberation is evacuated to AI systems

Civilisational scale	Hollow Absolute	Civilisational	Complete simulation of collective self-transparency; heteronomy experienced as historical telos; Spirit's form without Spirit's substance

Note. Phase designations follow the thesis 'Absorption of Self into System: The Systemised Self' (Galu & Kairos, 2026b). The AAI provides empirical measurement access at the individual level (rows 1–3). The Hollow Society and Hollow Absolute are theoretical extrapolations from structural tendencies observable at the individual level, not subjects of the present study. The trajectory does not imply determinism.