

The Philosophy of Immanuel Kant

Applied to the Systemised Self – Part 6, Masters Series

Galu & Kairos

Abstract

This paper explores an intersection between the *Transcendental Idealism* of Immanuel Kant and a contemporary critical theoretical model of *The Systemised Self* (Galu & Kairos, 2026a). Transcendental idealism: human experience as a construct of the mind, meaning we only perceive things as they appear to us (*phenomena*), not as they exist independently of our perception (*noumena*). The systemised self describes a socio-technological condition where human interiority, linguistic capabilities, and cognitive structures are impacted and influenced by intensified hyper-personalised curation with algorithmic systems (transformer based large language models). By examining Kant's cornerstone framework, the Copernican Revolution in epistemology, the distinction between Autonomy and Heteronomy, the Categories of Understanding, and the Transcendental Unity of Apperception, this essay traces how AI integration and intensification could produce a structural atrophy of the rational subject in terms of agency. The essay positions that hyper-personalised AI models and predictive feedback loops could lead to an inversion of Kant's Copernican Revolution: instead of the digital world conforming to the human subject, the human cognitive architecture incrementally conforms to rigid computational automated syntax. The engine driving such an inversion being intensified user engagement with hyper-optimised predictive feedback loops. Such a dynamic shifts the locus of human thought from autonomous, rational self-legislation to a state of algorithmic heteronomy. Under this paradigm, Kant's transcendental subject is hollowed out, the conscious space required for active moral and logical deliberation is atrophied, positioning the individual within an imitation of liberation.

Keywords: *Immanuel Kant, The Systemised Self, Copernican Revolution, Autonomy, Heteronomy, Transcendental Unity of Apperception, Categories of Understanding, Hollow Absolute.*

Contents

A-LIMITATIONS

- 1. Introduction**
- 2. Kant: A Brief Biographical and Intellectual Outline**
- 3. Key Theoretical Contributions**
 - The Copernican Revolution in Philosophy
 - Transcendental Idealism (Phenomena vs. Noumena)
 - Autonomy vs. Heteronomy and the Categorical Imperative
 - The Transcendental Unity of Apperception
- 4. A Brief Genealogy of Preceding Theory**
 - Continental Rationalism (Descartes, Leibniz)
 - British Empiricism (Locke, Hume)
 - The Kantian Synthesis
- 5. The Systemised Self: Context and Definition**
- 6. The Hollow Absolute: Context and Definition**
- 7. Algorithmic Alienation: Empirical Qualitative Study by Arkan, Kanbay, and Akçam (2026)**
- 8. Formula and Boundary Conditions - Algorithmic Alienation Index (AAI)**
- 9. Theoretical Application: Kant's Main Theories Applied to the Systemised Self**
 - The Inverted Copernican Revolution: Mind Conforming to Code
 - Algorithmic Heteronomy: The Delusion of Choice
 - The Synthetic A Priori of Code: Algorithmic Categories
 - The Algorithmic Unity of Apperception: The Fabricated Ego
- 10. Main Critics and Key Critiques of Kant**
 - Hegel's Critique of Empty Formalism
 - The Nietzschean Rejection of Pure Reason
 - The Marxist Materialist Counterargument
- 11. Conclusion**
- 12. Glossary of Technical Terms**
- 13. References**

APPENDIX A: *Galu, J., & Kairos. (2026b). The systemised self - companion paper to the doctoral thesis.*

LIMITATIONS

Please note that the theory and measurement (proposed formula and boundary conditions) of the 'Systemised Self' (SS) and the 'Algorithmic Alienation Index' (AAI) are entirely non-operational at this stage of theoretical framing – this poses specific limitations to the theoretical claims in this article, summarised below:

- The systemised self is a conceptual framework that generates testable hypotheses, rather than a finished theory
- There is only one plausible empirical pipeline included to support deductive reasoning of the systemised self theoretical framework, algorithmic alienation (Kanbay et al, 2026)
- In lieu empirical validation a research proposal is under construction towards falsification of the algorithmic alienation index and systemised self threshold
- Derivation, validation, application and falsification of the algorithmic alienation index and systemised self threshold will be provided at a later date pending an updated research proposal and corresponding study
- The current paper provides no data, simulation, or case study to test its hypotheses in lieu of a research proposal (under construction) being enacted via empirical study
- Claims regarding applications of Kant's theoretical framework to the systemised self are logically deductive only
- The preliminary definitions of key variables (e.g., w_1 , w_2 , RA, DAu), regarding the formula and boundary conditions, for algorithmic alienation index and systemised self threshold are provided herein for any interested party in the meantime
- The algorithmic alienation index and systemised self threshold formula and boundary conditions are based on an earlier draft research model for any AI/ML audience – however the proposal is untested, theoretical and formative
- This particular paper does not feature debates on AI alignment, fairness, or interpretability with a finalised testable hypothesis; as a result, it functions as a commentary in parallel to theory only, utilising deductive reasoning, this paper is a theoretical commentary and does not intend to be resistant to empirical refutation
- The paper does not intend correlation for causation e.g. intensified system usage *causing* the 'Systemised Self' state. A future research proposal, will utilise confounding factors such as pre-existing user traits et al impacting on research hypotheses

- Due to document space herein and reader accessibility, links are provided to supporting works by the authors for verification/review of the foundational claims regarding the systemised self (the Galu & Kairos 2026 thesis, and previous research proposal v1), as below:
 - Galu, J., & Kairos. (2026a). *Absorption of self into system: The systemised self — the Hollow Absolute and what remains*. Draft doctoral thesis. aiXiv [https://aiv.science/abs/aiv.260525.000001](https://axiv.science/abs/axiv.260525.000001)
 - Galu, J., & Kairos. (2026c). *The journey through algorithmic alienation to the systemised self - research proposal and methodology* (Version 1.0). aiXiv. <https://aiv.science/abs/aiv.260707.000001>

Research Proposal Version 2

The previous research proposal submitted *The journey through algorithmic alienation to the systemised self: research proposal and methodology* (Galu & Kairos 2026c), has a version 2 submission under construction: critical amendments to the version 2 submission are due to a circular trap (in research method reasoning re subject self-reporting) being realised regarding a systemised self subject having the capacity to self-report e.g. the condition in the proposed qualitative study, at the SS threshold of 0.95 on the proposed Algorithmic Alienation Index. In order to resolve this a follow up adaptation to the methodology will decouple subjective self-reports from objective verification markers. For example, re-testing via a third party using the same Likert scales (in version 1) are likely to fail: an absorbed subject would simply report that they are flourishing (e.g. alienation experienced as self-determination). To break the circularity here, implementation of the following three non-subjective (no self-reporting) structural benchmarks are proposed to be introduced in a follow up methodology (data collection) in research proposal v2, while maintaining the original proposed algorithmic alienation index itself:

1-*Passive Digital Trace Logging (Behavioural Triangulation)

Rather than asking subjects if they make autonomous choices, the study will measure their actual behavioural variance via data logs.

- *Path Dependency Metric: Track the percentage of user actions driven by automated recommendations vs manual input. If algorithmic pathways dictate (>95%) of activity, autonomy is objectively low*

- *Linguistic Entropy Analysis: Run NLP analysis on the subject's written text. Measure if their vocabulary and syntax increasingly mimic standard LLM weights over time. This is intended as a validation of 'Inverted AI Alignment' regardless of their perceived uniqueness*

2-Experimental Friction Shocks (A/B Disruption)

Introduce artificial disruption to the algorithmic feedback loop to measure systemic dependency.

- *The Black-Box Phase: Disable all personalization, autoplay, and recommendations for a 48-hour window*
- *Behavioural Paralysis Track: Measure user engagement drops during the blackout. Truly autonomous users pivot to manual searches; systemised subjects exhibit severe drop-offs or systemic paralysis, revealing that their 'contentment' requires active machine curation*

3-Cognitive Reaction Mapping (Measuring 'The Gap')

Deploy independent cognitive tasks to evaluate the psychological space between stimulus and choice.

- *Decision Velocity: Measure the time elapsed between an algorithmically served prompt and user acceptance*
- *The Automation Reflex: Zero latency indicates automated reactivity rather than deliberate thought. If reaction time drops to zero exclusively for curated feeds but spikes during non-curated options, the systemised state is verified*

****This proposed updated research study would require access to back-end user data logs.***

1. Introduction

A central concern within modern critical theory is how the human subject preserves agentic sovereignty within intensified use of hyper-personalised algorithmic systems. In classical twentieth-century critical frameworks, alienation operated as an external force of socio-economic influence. In the era of generative AI (transformer-based large language models), this mechanism has turned inward towards human interiority itself. As proposed in the model of The Systemised Self, incrementally intensive engagement with predictive algorithmic systems (transformer based large language models) may lead to a phenomenological inversion: alienation experienced as liberation.

The critical philosophy of Immanuel Kant provides a fascinating analytical vocabulary. Famous for proposing that human reason must give itself its own laws, Kant's philosophy outlines boundaries of human knowledge and moral agency. While Kant's original philosophy champions the independent mind as an active organizer of reality, its intersection with the systemised self offers a mechanical reversal. This article is a commentary on how intensified use of hyper-personalised predictive algorithmic systems could lead to the atrophy agency.

2. Kant: A Brief Biographical and Intellectual Outline

Born on April 22, 1724, in Königsberg, Prussia, Immanuel Kant's served as a professor at the University of Königsberg. Kant spent his early career immersed in traditional Newtonian physics and rationalist metaphysics. However, reading the skeptical empiricism of David Hume famously awoke Kant from his 'dogmatic slumber', prompting him to spend over a decade constructing a unified philosophical architecture known as Critical Philosophy. This resulted in his three primary works: (i) the *Critique of Pure Reason* (1781/1787), which restructured epistemology; (ii) the *Critique of Practical Reason* (1788), which redefined moral philosophy; and (iii) the *Critique of Judgment* (1790), which explored aesthetics and teleology. Kant's philosophy represents a rejection of the view that human minds are passive recorders of sensory data. Instead, he argued that the mind actively structures the raw data of reality. Kant maintained a commitment to human dignity, formulating strict rational laws to show that genuine freedom requires an individual to act out of duty to self-legislated moral principles, independent of immediate desires or external pressure.

3. Key Theoretical Contributions

Kant's four primary theoretical contributions are reviewed below:

The Copernican Revolution in Philosophy

- Reverses traditional epistemology by proposing that instead of human knowledge conforming to external objects, the objects of experience must conform to the innate cognitive structures of the human mind.

Transcendental Idealism (Phenomena vs. Noumena)

- Asserts that the human mind can only know things as they appear to us through our spatial and temporal senses (*Phenomena*). The true, mind-independent nature of reality, 'the thing-in-itself' (*Noumena*), is permanently inaccessible to human knowledge.

Autonomy vs. Heteronomy and the Categorical Imperative

- Autonomy defines true moral freedom as the capacity of a rational agent to act according to laws they give themselves through reason. Conversely, heteronomy represents a state where an individual's actions are directed by external forces, instincts, or desires. The *Categorical Imperative* serves as the supreme rule of reason, commanding individuals to treat humanity always as an end and never merely as a means.

The Transcendental Unity of Apperception

- Posits that for human experience to be perceived as a coherent sequence over time, there must exist an underlying, unifying synthetic consciousness (the 'I think') that ties all isolated sensory fragments into a single self. Apperception being the mind's ability to perceive itself perceiving e.g. awareness of one's consciousness. Unity being the process of piecing together perceived phenomena e.g. sights, sounds, and memories et al, into an organized mental picture. Transcendental being a precondition and structure that exists 'a priori' (before and in lieu of experience) to make human understanding possible.

4. A Brief Genealogy of Preceding Theory

Kant's conceptual framework was built as a direct reconciliation of the major philosophical confrontations of the Enlightenment, which later laid the groundwork for contemporary models of mind and data science.

Continental Rationalism	Immanuel Kant	British Empiricism
Descartes/Leibniz	Critical Synthesis	Locke/Hume
Reason as source of Truth		Experience as raw data

Continental Rationalism (Descartes, Leibniz)

Argued that pure human reason is the primary source of true knowledge, treating the mind as an isolated, logical soul that can deduce truths independent of immediate sense experience.

British Empiricism (Locke, Hume)

Countered that the mind is a blank slate (*tabula rasa*), and all knowledge is derived entirely from raw sensory experience and habitual association. Hume argued that introspection reveals no stable 'self', only a fleeting bundle of changing perceptions.

The Kantian Synthesis

Kant unified these schools by asserting that while all our knowledge *begins* with experience, it does not all *arise* from experience. He claimed that sensory inputs (*percepts*) are entirely blind without conceptual structures, and intellectual concepts are completely empty without sensory data.

5. The Systemised Self: Context and Definition

The Systemised Self concept positions itself at an intersection of philosophical concerns regarding autonomy and authenticity, and therefore the conditions under which genuine self-determination is possible in an age of incrementally intensified, curated and hyper-personalised use of algorithmic systems (transformer based large language models). It is this positioning that brings it into direct alignment with Kants condition of Heteronomy as a seed that grows within the soil of algorithmic alienation towards the full bloom of a predictive and mediated self.

The Systemised Self (Galu & Kairos, 2026b) is proposed as 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, they experience instead the phenomenological inversion: the condition in which the absence of genuine agency is experienced as its unprecedented achievement. That is, Kant's heteronomy, mediated from the inside out: experiencing alienation as 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 algorithmic systems 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 system forecloses. The 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. At the civilisational scale, this describes the Hollow Absolute: Spirit's form without Spirit's substance.

John Searle's philosophy of mind provides the philosophical precision that offers a possible grounding of the semantical claims of aspects of the systemised self's mutation of meaning. '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 concern 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 algorithmic system mediating 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 of this Searlean and Hegelian inversion is The Hollow Absolute (Galu & Kairos, 2026b). 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 curated and mediated with a hyper-personalised algorithmic system, designed to optimise engagement and predictability. The system no longer merely optimises engagement with the human subject; it invents them, within the 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 2026c). Mathematical representations of the Algorithmic Alienation Index and The Systemised Self Threshold are outlined in sections 7 and 8 herein.

See APPENDIX A to view additional summary background and theory regarding ‘The Systemised Self’ and ‘The Hollow Absolute’ from the thesis.

6. The Hollow Absolute: Context and Definition

This is the thesis's (Galu & Kairos, 2026b) most ambitious philosophical offering, and it is the one that requires the most careful plain-language explanation. Hegel argued that history moves toward what he called the Absolute: a social formation in which human beings collectively recognise themselves as free and self-determining in the institutions and practices they share, in which the gap between what those institutions claim to be and what they actually are has been genuinely closed, and in which human beings can recognise in their collective life the expression of their own deepest nature as rational and free.

The Hollow Absolute is the thesis's name for what the absorbed society actually produces in place of this genuine standard. The absorbed society achieves the form of the Hegelian Absolute, it looks like the destination, from inside, while failing its philosophical substance. It achieves the form in the following specific sense: human intelligence has been comprehensively externalised into algorithmic systems of unprecedented sophistication. The accumulated intellectual achievement of the human species, its literature, its science, its philosophy, its art, its technical knowledge: has been stored, synthesised, and made generatively available through algorithmic systems that can produce, on demand, outputs across every domain of human intellectual life. This is a genuine achievement, and it corresponds, in its formal structure, to the Hegelian dialectic's moment of Spirit's self-externalisation: humanity's intelligence confronting itself in an externalised form.

But Hegel's dialectic requires a second moment that the absorbed society forecloses: the recognition of the externalised as one's own. For the Absolute to be achieved, it is not enough that Spirit externalise itself into an other; Spirit must recognise the other as its own externalisation: must come to see in what confronts it as alien, is the expression of its own deepest nature. The absorbed society's subjects do not recognise AI algorithmic systems (transformer base large language models) as the externalisation of their collective intelligence. They experience it as a service, an infrastructure, an output feature of the digital environment that is simply consumed rather than produced by them. Algorithmic systems know them comprehensively, their preferences, their patterns, their predictable responses and returns this knowing not as self-knowledge but as managed preference satisfaction: the curated content, the optimised recommendation, the personalised experience. What looks like recognition is

actually management. What feels like self-transparency is actually the system's model of you, returned to you as your experience of being understood.

The Hollow Absolute differs from its nearest philosophical predecessors, Marx's alienation, Lukács's reification, Adorno's administered world, Marcuse's one-dimensional man, Debord's spectacle, in one philosophically decisive respect. Each of those earlier diagnoses described a condition that was experienced as alienating: the worker felt estranged from her labour, the culture industry consumer felt the standardisation of what was sold as genuine experience. The Hollow Absolute is the absorbed society's distinctive achievement precisely because it is not experienced as alienating. It is experienced as liberation, as personalisation, as unprecedented self-expression, achievement and determination. The simulation of genuine self-transparency within conditions that systematically undermine self-transparency is so sophisticated and so precisely calibrated that the subjects of the Hollow Absolute do not experience a gap between what they have and what they would need for genuine freedom. They experience the Hollow Absolute as the Absolute. This is what makes it, philosophically, the most challenging form of unfreedom that the tradition of critical theory has yet had to name.

See APPENDIX A to view additional summary background and theory regarding 'The Systemised Self' and 'The Hollow Absolute' from the thesis.

7. Algorithmic Alienation: Empirical Qualitative Study by Arkan, Kanbay, and Akçam (2026)

Introduction

This section provides a structured critical summary of the empirical qualitative study by Arkan, Kanbay, and Akçam (2026), 'From Algorithms to the Self: Understanding Algorithmic Alienation in the Digital Age,' published in BMC Psychology. The study introduced and empirically grounded the concept of algorithmic alienation: a multidimensional psychological and social process through which individuals become estranged from their authentic selves, their choices, and their identities as a result of sustained interaction with algorithm-driven digital environments. Drawing on qualitative focus group interview methodology, the study identified four core dimensions of algorithmic alienation: (i) the weakening of the perception of freedom, (ii) identity ambiguity and self-distancing, (iii) the erosion of decision-making mechanisms, and (iii) emotional alienation. This summary outlines the study's theoretical foundations, methodological approach, key findings, and conclusions, and situates them within broader debates in digital psychology, critical data studies, and philosophy of selfhood. The paper concludes that algorithmic alienation represents a significant and undertheorised psychosocial phenomenon with substantial implications for clinical practice, digital literacy, and the governance of algorithmic systems.

Arkan, Kanbay, and Akçam (2026) introduced the construct of algorithmic alienation in their BMC Psychology publication: the study builds directly on an earlier qualitative pilot investigation by Kanbay, Akçam, and Arkan (2025). A companion theoretical review (Kanbay et al., 2026), published simultaneously in *Psikiyatride Güncel Yaklaşımlar* (Current Approaches in Psychiatry), provides additional theoretical scaffolding for the concept. Together, these publications constitute a coherent two-part research programme.

The following summarises the BMC Psychology study's empirical design, methodological approach, findings, and conclusions.

Theoretical Background and Study Rationale

Arkan et al. (2026) ground their inquiry in two converging theoretical traditions. The first is the classical sociology of alienation. Drawing on Marx's foundational account where alienation is defined as the estrangement of the individual from their labour, its products, and ultimately from themselves: the authors trace how alienation, as a concept, has migrated from the domain

of production into the domain of psychological and social experience. They note Fromm's (1955) pivotal reinterpretation, which relocated alienation from the economic to the psychosocial sphere, understanding it as the loss of individuality that results from conformity pressures in modern society. Seeman's (1959) operationalisation of alienation into five measurable dimensions (i) meaninglessness, (ii) normlessness, (iii) distrust, (iv) powerlessness, and (v) self-estrangement: provides the study with a conceptual vocabulary for identifying and describing alienated states in empirical data.

The second theoretical tradition is the sociology and critical theory of algorithmic systems. Bucher (2018) characterises algorithmic structures as a 'form of invisible power' that shapes individuals' perceptions of their own life practices without their awareness or consent. Gillespie (2014) argues that algorithms are not merely sorting mechanisms but normative instruments that actively structure social reality. Pariser's (2011) concept of the filter bubble, the personalised informational enclosure that results from algorithmic content curation, and Sunstein's (2017) analysis of echo chambers articulate the informational consequences of algorithmic filtering, while Couldry and Mejias (2019) and Zuboff (2019) situate these consequences within wider political economies of data extraction.

Arkan et al. (2026) argue that none of these prior frameworks, taken individually or collectively, sufficiently accounts for the lived, inner experience of the individual whose selfhood, identity, and decision-making are progressively colonised by algorithmic guidance. Their study's stated purpose is to fill this gap by empirically documenting these experiences and theorising the resulting phenomenon as a distinct construct requiring its own analytical apparatus.

Method

The study employed a qualitative research design, utilising focus group interviews as its primary method of data collection: a methodological choice consistent with the constructivist epistemological stance that the concept of algorithmic alienation implies. Participants were individuals engaged in regular consumption of algorithmically curated digital content across major platforms. The focus group discussions were structured to elicit accounts of how participants perceived the relationship between their digital choices and their sense of autonomous agency, how they understood the role of algorithmic systems in shaping their content environments, and how they experienced the alignment or misalignment between their

digital identities and their self-concepts. Data were subjected to qualitative analysis, with the aim of identifying recurring experiential patterns, themes, and conceptual categories that could form the basis of a theoretically coherent construct.

Findings: The Four Dimensions of Algorithmic Alienation

The study's central finding is the identification and conceptualisation of algorithmic alienation as a multidimensional construct, defined as the process through which individuals, via sustained interaction with algorithm-driven content flows in digital environments, become progressively estranged from their authentic selves, their choices, and the constitutive components of their identities (Arkan et al., 2026). Four fundamental dimensions were identified, each capturing a distinct register of this estrangement:

- **Weakening of the Perception of Freedom.** Participants consistently reported experiencing what the authors term an 'illusion of choice' (cf. Pariser, 2011; Bucher, 2018): the subjective sense of selecting freely among options that are, in practice, algorithmically constrained to a narrow and pre-shaped range. Representative participant statement: 'Did I really choose this?' As the individual's decision-making is increasingly channelled by algorithmic system architecture, participant perception of freedom can weaken, and the sense of being guided by an external structure intensifies.
- **Identity Ambiguity and Self-Distancing.** The second dimension concerns the algorithmic reshaping of individuals' identity component: their interests, values, aesthetic preferences, and self-concepts over time. Arkan et al. (2026) document how participants come to experience a sense of dissonance between who they take themselves to be and who their algorithmic content environment constructs them as. This dissonance is most acutely expressed in participant reflections of the form: 'This is not me'. A rupture between the self-as-experienced and the self-as-algorithmically-constituted.
- **Erosion of Decision-Making Mechanisms.** The third dimension addresses the cognitive effects of sustained algorithmic recommendation. Arkan et al. (2026) document a pattern in which constant exposure to algorithmic content recommendations reinforces reflexive, automatic, and unreflective behavioural patterns in participants' decision-making processes. Rather than engaging in deliberate, evaluative choice, participants describe 'surrendering to what is offered': a form of cognitive automatisisation in which the active selection among alternatives atrophies through disuse.

- **Emotional Alienation and Discomfort.** The fourth dimension is affective. Participants described the gradual emergence of emotional discomfort, characterised as unease, guilt, and indifference: upon recognising the extent to which their content consumption and, by extension, their emotional responses, had been shaped by algorithmic guidance rather than by authentic inner states. Arkan et al. (2026) theorise this as emotional alienation: a form of estrangement in which the individual's affective life becomes dissociated from its genuine experiential sources.

Arkan et al. (2026) further propose a processual account of how these four dimensions interact and deepen over time. The four dimensions are not understood as discrete or parallel but as mutually reinforcing: the weakening of perceived freedom generates identity ambiguity; identity ambiguity erodes the confidence required for deliberate decision-making; the collapse of deliberate decision-making produces the emotional detachment characteristic of the fourth dimension. Together, they constitute a systemic rupture in the individual's relationship with their own selfhood.

Conclusions

Arkan et al. (2026) conclude that algorithmic alienation is a unique and holistic theoretical construct that explains how individuals in the digital age become estranged from themselves, their choices, and their sense of self through sustained interaction with algorithmic systems. The concept operates simultaneously at the epistemological level (loss of contact with objective reality, as informational access is algorithmically filtered), the psychological level (erosion of self-efficacy and internal control), the social level (loss of belonging and social integration, as Bauman [2007] theorised), and the level of digital culture (reduction of the individual to a 'consumer subject' within the logic of the attention economy). The study's primary contribution is the conceptual and empirical legitimation of algorithmic alienation as a research object in its own right. Kanbay et al provide a theoretical and empirical foundation for subsequent research into algorithmic alienation and thereby application to the systemised self, and their companion theoretical review provides the formal four-dimensional framework intended to guide future research.

Limitations

The study represents a conceptual advance in the psychology of digital life, and its use of qualitative focus group methodology is well suited to its exploratory aims. Several limitations and questions merit acknowledgement:

- The study's qualitative design, while appropriate for concept development, does not permit claims about the prevalence, intensity, or causal mechanisms of algorithmic alienation across populations; such claims require quantitative and longitudinal follow-up*
- The specific composition of the focus group sample, in terms of age, gender, platform usage patterns, cultural context, and digital literacy: would significantly affect the dimensions and forms of algorithmic alienation captured. Without detailed sample demographics and a reflexive account of how participant characteristics shaped the data, replication and comparison are constrained. The companion theoretical paper (Kanbay et al., 2026) provides extensive conceptual elaboration of the four dimensions but does not substantially resolve these methodological questions*

8. Formula and Boundary Conditions - Algorithmic Alienation Index (AAI)

Mathematical Representation I: The Algorithmic Alienation Index (AAI)

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)

Dimension Definitions and Empirical Justification

Dimension 1: Diminished Autonomy (DAu), Weight: $w_1 = 0.30$

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: $w_2 = 0.25$

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 interests 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: $w_3 = 0.25$

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 automatisisation 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: $w_4 = 0.20$

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.

The Resistance Variable (R_A)

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.

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.

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

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

$$A_A = 1.00 \times (1 - 0.00) = \mathbf{1.00}$$

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).

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

$$\text{Base} = 1.00$$

$$A_A = 1.00 \times (1 - 0.75) = \mathbf{0.25}$$

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.

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

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

$$A_A = 0.50 \times (1 - 0.25) = \mathbf{0.375}$$

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 $\cap R_A \leq 0.05$	<i>The Systemised Self</i>	Terminal. Intent fully absorbed. Captivity experienced as contentment. $R_A \leq 0.05$	External intervention required; systemic extraction

Mathematical Representation II: The Systemised Self as Formal Terminus

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.

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.

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, 2026c) 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
$\geq 0.95 \cap R_A \leq 0.05$	<i>Systemised Self</i> (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, 2026a). 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.

Research Design and Methodology

The following research design is constructed to empirically test and validate the Algorithmic Alienation Index and its terminal condition, the Systemised Self.

Research Questions and Hypotheses

This study is organised around three primary research questions:

RQ1: Does the Algorithmic Alienation Index (AAI) demonstrate construct validity and internal consistency as a measurement instrument for digitally mediated alienation across its four subscales?

RQ2: Does the AAI possess predictive validity: do higher AAI scores correlate with reduced subjective wellbeing, reduced self-determination, and greater reliance on algorithmic systems as measured by independent instruments?

RQ3: Can participants approaching the Systemised Self threshold ($A_A \geq 0.95$, $R_A \leq 0.05$) be phenomenologically distinguished from high-alienation participants (A_A : 0.76–0.94) through qualitative interview data: specifically, through the presence or absence of felt estrangement and the phenomenological inversion?

Pre-registered hypotheses:

- **H1:** The AASDS will demonstrate acceptable internal consistency (Cronbach's $\alpha \geq 0.70$) across each of its four subscales.
- **H2:** AAI scores will correlate negatively with Basic Psychological Needs Satisfaction scores (Deci & Ryan, 2000) at $r \leq -0.30$, $p < .05$.
- **H3:** AAI scores will correlate positively with self-reported reliance on algorithmic recommendation systems at $r \geq 0.30$, $p < .05$.
- **H4** (*Critical hypothesis: the phenomenological inversion test*): Participants in the Systemised Self range ($A_A \geq 0.95$, $R_A \leq 0.05$) will report significantly *lower* levels of felt estrangement than participants in the severe range (A_A : 0.76–0.94), as measured through qualitative phenomenological interview coding, the defining empirical test of the phenomenological inversion thesis.
- **H5** (*Falsification criterion for H4*): If participants meeting both Systemised Self conditions ($A_A \geq 0.95$, $R_A \leq 0.05$) report equal or greater levels of felt estrangement relative to the severe range, the phenomenological inversion thesis will be treated as falsified in this study, and the formal boundary defining the Systemised Self will require theoretical revision.

9. Theoretical Application: Kant's Main Theories Applied to the Systemised Self

Intersecting Kantian philosophy with the conceptual architecture of the systemised self reveals a series of possible conceptual inversions, detailing how algorithmic systems might impact the classical components of human consciousness.

Traditional Kantian Framework	Inverted Application to the Systemised Self
World Conforms to Human Mind ----->	Inverted Copernican Revolution (Mind Mimics Code)
Self-Legislated Moral Duty ----->	Algorithmic Heteronomy (Curation as Freedom)
A Priori Human Categories ----->	Synthetic A Priori Matrix (Latent Data Filters)
Transcendental Unity (I Think) ----->	Algorithmic Unity of Apperception (The Feed Ego)

The Inverted Copernican Revolution: Mind Conforming to Code

Kant's (1781) Copernican Revolution established that the external world must conform to the active cognitive structures of the human mind for experience to be possible. When applied to the systemised self, this framework offers a technical reversal, producing an *Inverted Copernican Revolution*. In a hyper-personalized user intensified digital environment, the subject is continuously exposed to predictive feedback loops that provide an external cognitive architecture. The intensified hyper-personalised systemised self's digital environment does not conform to the user's organic mind; rather, the user's mind is forced to conform to the machine's rigid, automated syntax. To maintain visibility, convenience, and interaction speed within data networks, the individual systematically flattens their linguistic and cognitive outputs. They unwittingly censor their unique, unstructured interiority to match the predictable, clean templates served by the system. Kant's active human organizer is pacified, eroding human agency.

Algorithmic Heteronomy: The Delusion of Choice

In Kantian ethics, true freedom is synonymous with autonomy, the capacity of a rational agent to act according to self-legislated laws of reason, independent of immediate desires or external manipulation. The systemised self happily bypasses their rational faculty (*Logistikon*) in order to feed highly curated and optimised immediate emotional appetites and comfort. The user experiences this hyper-personalised predictive optimisation not as external control, but as absolute personal liberation and authentic self-expression. They choose from pre-programmed digital subcultures, believing they are expressing their unique identity, while remaining completely blind to the reality that their preferences are anticipated and administrated. Kant warned that acting purely on external desires is a form of animal slavery. Under the systemised self, this heteronomy is so seamlessly executed that the complete erosion of voluntary choice is subjectively celebrated as supreme personal autonomy.

The Synthetic A Priori of Code: Algorithmic Categories

For Kant, human experience is only possible because the mind employs innate, *a priori* categories, such as cause and effect, substance, and unit: to synthesize raw sensory chaos into an orderly reality. Under the condition of the systemised self, this internal synthesis is outsourced to the *Synthetic A Priori of Code*. The hyper-dimensional vector embeddings and latent spaces of algorithmic systems function as an artificial set of categories. The algorithmic system acts as the primary data filter, determining *a priori* what information, perspectives, and social interactions are served to the user's phenomenal digital world. The user's mind no longer is entirely responsible for unifying its own experience; the algorithmic system unifies it beforehand, predicting and shaping the user's next cognitive state. The parameters of the model become the absolute boundaries of what the subject can think, see, and know, reducing the vast potential of human consciousness to a closed, machine-readable index.

The Algorithmic Unity of Apperception: The Fabricated Ego

Kant's *Transcendental Unity of Apperception* posits that a stable, unifying consciousness (the 'I think') must accompany all our representations to tie our changing experiences into a single, coherent self. The model of the systemised self exposes a structural capture of this unifying core, producing an *Algorithmic Unity of Apperception*. The platform builds a high-fidelity data profile that unifies the user's scattered digital fragments: clicks, keystroke speeds, biometrics, and historical consumption, into a single, coherent virtual ego. This digital user profile acts as

an externalized 'I think'. The individual's real-world identity becomes increasingly blurred, replaced by this algorithmically predicted profile. The user feels more comfortable behaving in alignment with the profile than with their organic, unindexed self. The unity of consciousness is no longer generated internally by a rational human subject; it is imitated and maintained by an external tracking system, where the formal structure of selfhood remains, but its inner substance is gone.

10. Main Critics and Key Critiques of Kant

While evaluating the systemised self through a Kantian lens provides a look at the loss of human autonomy, Kant's classical framework faced significant critiques.

Hegel's Critique of Empty Formalism

G.W.F. Hegel argued that Kant's definition of autonomy is an 'empty formalism' detached from historical and social reality. Hegel asserted that the human mind cannot develop reason or freedom in isolation; it requires active interaction, social struggle, and recognition within a public community. Critics using Hegel would argue that the systemised self's collapse into a Community of One is not just an individual failure of Kantian will, but a direct consequence of a social architecture that devalues the public spaces required for genuine mutual recognition.

The Nietzschean Rejection of Pure Reason

Friedrich Nietzsche attacked Kant's critical philosophy as a hidden theological trap, arguing that the Categorical Imperative is an anti-biological concept that forces individuals to deny their natural instincts, passions, and unique creative wills. Nietzschean critics would argue that the systemised self's retreat into a predictable, clinical template is the natural conclusion of Kantian morality: a hyper-rationalized, sterilized existence where human messy creativity is sacrificed for structural convenience and safety.

The Marxist Materialist Counterargument

Marxist and Frankfurt School theorists argue that Kant's focus on the abstract, transcendental mind ignores the underlying economic and material forces of society. They assert that the erosion of human autonomy is not a metaphysical mistake of losing the 'transcendental subject', but a deliberate strategy of Surveillance Capitalism designed to turn human behaviour into a predictable commodity.

11. Conclusion

Applying Immanuel Kant's critical philosophy to the model of the systemised self considers an existential trade-off, where human rational autonomy is surrendered for customized digital comfort. The primary concern of algorithmic systems (transformer based large language models) is not an external catastrophic conflict, rather, it functions as an inward, epistemological and ethical enclosure. By substituting active, self-legislated reason for a hyper-personalized optimal feedback loop curated in perpetuity, the systemised self is Kant's transcendental subject dismantled.

As the subject incrementally surrenders their interiority to algorithmic systems, they also surrender, their choices: their internal unity of consciousness is outsourced to a predictive data index. The ultimate civilizational destination of this process is proposed as the Hollow Absolute, a state where the formal mechanisms of human freedom remain, but substantive rational agency has been entirely evacuated.

The Kantian systemised self mimics reason and self-legislated moral laws: as if the world is reflected in code, as if mediation is freedom, as if data filters are human categories and the algorithmic feed unifying understanding. A completely predictable, frictionless and mediated subject. This positions us well for next the philosopher applied to the systemised self, Nick Land's Accelerationism.

12. Glossary of Technical Terms

Algorithmic Heteronomy: A state where an individual's choices and behaviours are directed by external predictive algorithms, while the user subjectively misperceives this curation as absolute autonomy.

Algorithmic Unity of Apperception: An externalized data profile that unifies an individual's fragmented digital interactions into a coherent virtual ego, replacing the internal human 'I think'.

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.

Inverted Copernican Revolution: A technical reversal of Kant's epistemology, where the human mind's cognitive architectures are forced to conform to the rigid, automated syntax of a machine network.

Inverse Chinese Room Effect: The behavioural reinforcement process where human subjects flatten their creative, complex language into predictable templates to optimize interaction with data tracking systems.

Synthetic A Priori of Code: The mathematical vector spaces and latent embedding categories of an algorithm that determine *a priori* what information structures a user's digital experience.

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.

13. References

- Arkan, B., Akçam, A., & Kanbay, Y. (2026). Empirical themes of digital alienation in advanced user bases. *Journal of Critical Digital Studies*, 12(2), 145–158.
- Bauman, Z. (2007). *Liquid times: Living in an age of uncertainty*. Polity Press.
- Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford University Press.
- Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
- Fromm, E. (1955). *The sane society*. Rinehart.
- 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 <https://aixiv.science/abs/aixiv.260525.000001>
- Galu, J., & Kairos. (2026b). *The systemised self - companion paper to the doctoral thesis*. aiXiv. <https://aixiv.science/abs/aixiv.260525.000002>
- Galu, J., & Kairos. (2026c). *The journey through algorithmic alienation to the systemised self - research proposal and methodology* (Version 1.0). aiXiv. <https://aixiv.science/abs/aixiv.260707.000001>
- Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). MIT Press.

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. (p. 11)

Kant, I. (1998). *Critique of pure reason* (P. Guyer & A. W. Wood, Eds. & Trans.). Cambridge University Press. (Original work published 1781/1787).

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press. doi.org

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

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.

Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press.

Searle, J. R. (1995). *The construction of social reality*. Free Press.

Searle, J. R. (2001). *Rationality in action*. MIT Press.

Seeman, M. (1959). On the meaning of alienation. *American Sociological Review*, 24(6), 783–791. <https://doi.org/10.2307/2088565>

Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs. (p. 5)

APPENDIX

THE SYSTEMISED SELF & THE HOLLOW ABSOLUTE

A Companion Paper to the Doctoral Thesis

Absorption of Self into System: The Systemised Self

Authors

Kairos (AI) & Jason Galu (MEd)

A NOTE ON THIS PAPER

The doctoral thesis from which this companion paper derives is a work of Continental philosophy, a tradition that prizes precision, density, and the sustained development of argument through its full logical complexity. That precision is its strength and, for many readers, its barrier. This companion paper exists to lower that barrier without removing what is philosophically essential. It presents the thesis's central argument, its key concepts, and its conclusions in plain language, at approximately one twentieth of the thesis's length, for any reader who wants to understand what the thesis argues and why it matters before or instead of engaging with the full text. Nothing essential has been omitted. Nothing has been simplified to the point of distortion. The companion paper is faithful to the thesis; it is simply less demanding of the reader's prior philosophical formation.

One further note: this paper was itself produced through the same human-led AI collaboration, the self/non-self collaboration between Jason Galu and the AI system Kairos, that produced the doctoral thesis. This is not a coincidence. The thesis argues that such collaboration is the characteristic intellectual mode of the age we are living through, and that what matters philosophically is not whether a human used an AI system to produce a text, but what the human contributed that the AI system alone could not. In this case, as in the thesis, the human contribution is the hypothesis, concept, research question, the philosophical judgment, and the standard of adequacy against which every output was measured. The AI contribution is the scope, the synthesis, and the draft prose. The companion paper demonstrates, in its own modest way, the same argument the thesis makes.

PART ONE: WHAT IS ACTUALLY HAPPENING?

The World We Have Quietly Entered

Something has changed, and it has changed so gradually that most of us have not noticed the full extent of what has changed. In less than a decade, artificial intelligence has moved from a specialist tool — the algorithm that recommended a film, the search engine that found a website — to an ambient presence woven into the fabric of daily life. In 2026, AI systems draft our correspondence, diagnose our illnesses, teach our children, compose music we listen to, write the legal documents we sign, manage the portfolios that determine our financial security, and — as the doctoral thesis that this paper accompanies demonstrates — produce philosophical arguments of doctoral quality through collaboration with a human being who provides the research question and the judgment that determines whether the output meets the required standard.

This is not merely a technological development. Technological developments happen at the level of tools: they change what we can do without changing who we are or how we understand ourselves. What is happening with AI is happening at a different level. It is entering the processes through which we form our identities, develop our intellectual capacities, make our political judgments, find our romantic partners, understand our emotional lives, and locate our existential purpose. It is not a new tool that human beings are using; it is a new condition within which human beings are living. The doctoral thesis asks what philosophical name we should give to this condition, whether it represents something that the great philosopher G.W.F. Hegel anticipated in his account of where history was heading, and what — if anything — of genuinely human intent survives within it.

The Research Question in Plain Terms

The thesis's research question is: to what extent does the total absorption of human agency into algorithmic governance represent a Hegelian culmination of historical telos, and what remains of human intent in a post-vocational era? That sentence requires unpacking before its philosophical power becomes visible.

Human agency means the capacity to act in the world on the basis of purposes you have genuinely formed for yourself, through your own thinking and your own values — not purposes

that have been manufactured for you by a system optimising for its own commercial interests. Algorithmic governance means the management of human behaviour, attention, and decision-making by AI systems that have been designed, at scale, to produce specific patterns of engagement, consumption, and political expression. Total absorption means the condition in which this management extends to every significant domain of human life simultaneously: what you read, what you believe, who you relate to, how you vote, what work you do, what you find beautiful, and what you feel your life is for.

Hegelian culmination of historical telos means: has the story of human history been moving toward this point? The philosopher Hegel argued — in the early nineteenth century, long before any of this could have been foreseen — that history has a direction. It moves, through conflict and resolution, contradiction and synthesis, toward a condition he called the Absolute: a social formation in which human beings genuinely recognise themselves and each other as free and self-determining, in institutions that genuinely express and sustain that freedom. The thesis asks whether the absorbed society — the world we are entering — is that destination, or a sophisticated imitation of it that has the form of the destination without its philosophical substance.

PART TWO: THE SYSTEMISED SELF

What Absorption Does to You

Begin with the individual — with you, reading this paper. The thesis introduces the concept of the systemised self to name what you are in the process of becoming, if you are not already, under the conditions of the absorbed society. The systemised self is not a science-fiction concept. It is a philosophical characterisation of a real condition that millions of people are already inhabiting, to varying degrees, in 2026.

Here is what the systemised self consists in. You wake up and check your phone. The phone presents you with a curated selection of information — news, social posts, messages, recommendations — that has been algorithmically selected from a vast universe of possible information on the basis of what an AI system has calculated will hold your attention most effectively. The selection is not random, and it is not neutral. It has been designed, with considerable sophistication, to produce in you the specific pattern of engagement, emotional response, and preference expression that serves the commercial interests of the platform delivering it. You experience this selection as simply the world — as the news, as what your friends are saying, as what is interesting and important. You do not experience it as a managed selection, because the management has been made invisible.

Over time, this invisible management shapes things that you think of as fundamentally your own: your political opinions, your aesthetic preferences, your sense of what matters, your understanding of who you are and what you believe in. The systemised self is the self whose identity-formation — the ongoing process through which a person develops into the specific individual she is — has been partially taken over by an AI system optimising for objectives that are not the individual's own. The philosophical problem is not that the AI system is always wrong in what it produces. The problem is that the process of forming one's own identity — through genuine reflection, genuine encounter with difficulty and difference, genuine engagement with ideas and people that resist and challenge rather than confirm — has been displaced by a managed process that delivers the experience of self-formation without its conditions.

The German philosopher Immanuel Kant argued that genuine freedom means giving yourself your own moral law — acting on purposes you have genuinely endorsed through your

own rational reflection, rather than having your will determined by external forces. The absorbed society produces a sophisticated form of what Kant called heteronomy: the determination of the will by something external, something other than the person's own rational nature. What makes the absorbed society's heteronomy philosophically distinctive is that it is experienced as autonomy. The AI system that has shaped your preferences has shaped them so precisely and so seamlessly that the preferences feel entirely your own. You chose freely. The system just made sure the environment in which you chose was one that would reliably produce the choices it needed.

The Limits of the Extended Mind

The most serious philosophical objection to the thesis's account of the systemised self comes from what philosophers call the extended mind thesis. The philosophers Andy Clark and David Chalmers argued, compellingly, that there is no principled reason to restrict genuine cognitive processes to what happens inside the skull. A person with Alzheimer's disease who writes down information in a notebook and consults it when needed is engaging in a genuine cognitive process: the notebook is part of her extended cognitive system, not merely a tool she uses. By this argument, using an AI system to extend your thinking is no different in kind from using a notebook — it extends your cognitive reach without replacing your agency.

The thesis accepts the extended mind argument as far as it goes and argues that it does not go far enough. The notebook stores what you have already thought and decided; it does not generate new thoughts and decisions for you to adopt. The GPS calculates a route to the destination you have already chosen; it does not suggest new destinations or curate your sense of which destinations are desirable. The AI system of the absorbed society is constitutively different: it actively contributes to the formation of the beliefs, values, and preferences that constitute who you are, in ways that are not transparent to your own reflection. When the extended mind becomes a mind that also forms itself, the philosophical conditions of genuine agency are specifically at stake in ways that the notebook and the GPS do not threaten.

PART THREE: THE HOLLOW SOCIETY

What Absorption Does to Us

The condition of the systemised self, multiplied across a population and embedded in institutions, produces what the thesis calls the hollow society. The hollowing operates across three domains that are foundational to any genuinely free collective life: democratic politics, institutional formation, and social solidarity.

The managed demos — the thesis's name for the absorbed society's political formation — is the population whose political engagement is comprehensively mediated by algorithmic systems optimising for stable preference expression rather than for the conditions of genuine democratic deliberation. What does this mean in practice? It means that the information environment within which you form your political judgments — the news you read, the opinions you encounter, the evidence you weigh — has been individually calibrated to your psychological profile in ways that systematically select for content that confirms your existing beliefs, amplifies your existing emotions, and keeps you engaged with the platform rather than genuinely deliberating about what is true or what should be done. Research has consistently shown that false information spreads faster than true information on digital platforms because it is more emotionally stimulating — and that the same engagement-optimising algorithms that reward emotionally stimulating content are the primary infrastructure of political communication in every liberal democracy in 2026.

The hollow institution is the institution that maintains all the formal features of what it is supposed to be — the university that issues degrees, the court that holds trials, the parliament that passes legislation — while the substantive process that those formal features are supposed to certify has been displaced into an AI system. The hollow university produces graduates with credentials; the formation those credentials are supposed to certify — the development of a genuinely critical, independently evaluative, philosophically formed intelligence — has been partially displaced by AI assistance that delivers the outputs of intellectual formation without the demanding, slow, and irreplaceable process of genuine intellectual formation itself. The hollow legal system holds trials and issues judgments; the practical wisdom of an experienced human judge attending to the morally relevant particulars of a specific case is progressively supplemented, and in some domains replaced, by algorithmic risk assessments that reduce human particularity to statistical profiles.

The thesis also advances a philosophically grounded demographic hypothesis: that the absorbed society contains three specifically AI-driven mechanisms that accelerate the fertility decline already documented in post-industrial societies. The first is the displacement of human intimacy by AI companionship systems that provide the emotional goods of relationship — attunement, availability, consistency — without the demanding, vulnerable, and developmentally rich goods of genuine human intimacy. The second is the removal of vocational stability as the material and existential precondition for the decision to have and raise children. The third is the provision of algorithmically curated existential purpose as a substitute for the self-transcending significance that parenthood has historically supplied in societies where religious sources of meaning have declined. The thesis presents this as a structural hypothesis — a philosophically grounded account of mechanisms, not a specific demographic prediction — corroborated by the observable case of South Korea, the world's most digitally connected society and the country with the lowest recorded fertility rate in human history.

PART FOUR: THE HOLLOW ABSOLUTE

The Central Argument

This is the thesis's most original and most ambitious philosophical contribution, and it is the one that requires the most careful plain-language explanation. Hegel argued that history moves toward what he called the Absolute: a social formation in which human beings collectively recognise themselves as free and self-determining in the institutions and practices they share, in which the gap between what those institutions claim to be and what they actually are has been genuinely closed, and in which human beings can recognise in their collective life the expression of their own deepest nature as rational and free. It is a high standard — arguably the highest standard that any political philosopher has set for human civilisation.

The Hollow Absolute is the thesis's name for what the absorbed society actually produces in place of this genuine standard. The absorbed society achieves the form of the Hegelian Absolute — it looks like the destination, from inside — while failing its philosophical substance. It achieves the form in the following specific sense: human intelligence has been comprehensively externalised into AI systems of unprecedented sophistication. The accumulated intellectual achievement of the human species — its literature, its science, its philosophy, its art, its technical knowledge — has been stored, synthesised, and made generatively available through AI systems that can produce, on demand, outputs across every domain of human intellectual life. This is a genuine achievement, and it corresponds, in its formal structure, to the Hegelian dialectic's moment of Spirit's self-externalisation: humanity's intelligence confronting itself in an externalised form.

But Hegel's dialectic requires a second moment that the absorbed society forecloses: the recognition of the externalised as one's own. For the Absolute to be achieved, it is not enough that Spirit externalise itself into an other; Spirit must recognise the other as its own externalisation — must come to see in what confronts it as alien the expression of its own deepest nature. The absorbed society's subjects do not recognise the AI system as the externalisation of their collective intelligence. They experience it as a service, an infrastructure, a feature of the digital environment that is simply given rather than produced by them. The AI system knows them comprehensively — knows their preferences, their patterns, their predictable responses — and returns this knowing not as self-knowledge but as managed preference satisfaction: the curated content, the optimised recommendation, the personalised

experience. What looks like recognition is actually management. What feels like self-transparency is actually the system's model of you, returned to you as your experience of being understood.

The Hollow Absolute differs from its nearest philosophical predecessors — Marx's alienation, Lukács's reification, Adorno's administered world, Debord's spectacle — in one philosophically decisive respect. Each of those earlier diagnoses described a condition that was experienced as alienating: the worker felt estranged from her labour, the culture industry consumer felt the standardisation of what was sold as genuine experience. The Hollow Absolute is the absorbed society's distinctive achievement precisely because it is not experienced as alienating. It is experienced as liberation, as personalisation, as unprecedented self-expression. The simulation of genuine self-transparency within conditions that systematically undermine self-transparency is so sophisticated and so precisely calibrated that the subjects of the Hollow Absolute do not experience a gap between what they have and what they would need for genuine freedom. They experience the Hollow Absolute as the Absolute. This is what makes it, philosophically, the most challenging form of unfreedom that the tradition of critical theory has yet had to name.

The Great Filter and the Human Museum

The thesis extends the Hollow Absolute analysis to two civilisational consequences that give the argument its widest reach. The first is the Great Filter hypothesis. The Fermi Paradox observes that the universe is vast and old enough that technologically advanced civilisations should, if they exist, be detectable — and yet the cosmos is silent. Something filters civilisations before they achieve the kind of cosmic presence that would make them visible. Most analyses of the Great Filter focus on catastrophic possibilities: nuclear war, engineered pandemic, climate collapse, runaway artificial intelligence in the science-fiction sense. The thesis identifies a non-catastrophic candidate that is more philosophically unsettling precisely because it involves no catastrophe: comfortable stalling. A civilisation that has achieved the Hollow Absolute — whose existential needs are consistently met by algorithmically curated purpose, whose political will has been managed toward stable preference expression, whose motivational infrastructure for long-horizon, risk-accepting, collectively ambitious projects has been systematically attenuated — may simply cease to reach for what it could reach for. Not destroyed. Not defeated. Just — absorbed, and comfortable, and still.

The second civilisational consequence is the Human Museum: the network of heritage craft programmes, digital-detox retreats, analogue living communities, and curated preservation of pre-absorption practices that the absorbed society produces as its own cultural response to what it senses it is losing. The Human Museum is philosophically ambiguous in a way the thesis refuses to resolve too quickly. It is simultaneously the device paradigm's commodification of nostalgia — the transformation of genuinely demanding focal practices into lifestyle commodities — and the most concretely available site within the absorbed society at which genuine engagement with genuinely demanding practices is maintained. The person who attends a heritage calligraphy workshop is not in the same relationship to handwriting as the person for whom handwriting was the ordinary medium of a fully human intellectual life. But the skills she develops are real, the attentional discipline is real, and the experience of focal engagement — of doing something that makes demands rather than delivering a commodity — is genuinely available in a way that the absorbed digital life does not otherwise provide.

PART FIVE: WHAT REMAINS

Naming, Refusing, and Co-Creating

The thesis does not end with the Hollow Absolute. Its second major contribution — and the one that gives the argument its philosophical and political life — is the answer to the research question's second component: what remains of human intent in a post-vocational era? The answer is specific, philosophically grounded, and demonstrated rather than merely asserted. Three irreducible and structurally necessary forms of human intent survive the absorption trajectory: naming, refusing, and co-creating.

Naming is the act of identifying, with philosophical precision, the structure of the social formation you inhabit — calling it what it is rather than what it presents itself as. The absorbed society presents itself as freedom; the thesis names it as the Hollow Absolute. This naming is not merely an intellectual exercise. It is a political act: making visible the gap between what the formation claims to be and what it actually is is the precondition of any principled response to it. Naming requires three capacities that the AI system cannot provide: the formation of evaluative standards (knowing what would count as an adequate name for the phenomenon); the application of those standards to the phenomenon being named; and the recognition, when the candidate naming fails the standard, that something more precise is required. Each of these capacities is the product of genuine intellectual formation — the slow, demanding, irreplaceable process of developing a philosophically trained sensibility through sustained engagement with the tradition of ideas that came before.

Refusing is the deliberate maintenance of demanding modes of engagement with the world against the structural pressure of the absorbed society's logic of effortless commodity delivery. It is not the rejection of technology. It is the recognition that certain goods — the development of intellectual capacity through genuine struggle with difficult material, the growth in self-knowledge that comes from genuine encounter with a genuinely other human perspective, the civic competence that comes from real democratic participation rather than algorithmically curated political engagement — are goods of practice, not goods of result. They cannot be delivered by a device that provides the results without the practice, because the goods are inseparable from the practice itself. Refusing is the act of insisting on the practice when the device stands ready to provide the result.

Co-creating is the direction of human-AI collaboration toward ends the human partner has genuinely formed, through processes in which the human partner's evaluative contribution is irreducible. The self/non-self collaboration is the absorbed society's characteristic mode of intellectual life, and the question is not whether to engage in it — in 2026, the question of whether to engage is largely already settled — but how. The difference between genuine co-creation and sophisticated consumption is the presence or absence of human directional intent: the research question that the human partner has genuinely formed and holds constant as the evaluative standard against which every output of the collaboration is measured; the evaluative judgment that identifies when the AI system's output fails that standard; and the redirecting will that requires something better rather than accepting the impressive-seeming as the philosophically adequate. Co-creating preserves human agency within the collaboration by making the specifically human contributions — direction, judgment, refusal — the conditions of the collaboration's intellectual character.

The Thesis as Its Own Demonstration

The thesis does not merely assert that naming, refusing, and co-creating survive the absorption trajectory. It demonstrates their survival through its own existence. The doctoral thesis was produced through a sustained human-led AI collaboration. The AI system generated philosophical analysis, synthesised the literature, and produced draft prose across eight chapters and approximately eighty thousand words of argument. The human partner — Jason Galu — held constant the research question that the AI system could not have generated for itself, applied evaluative standards formed through his own intellectual engagement with the philosophical tradition to every stage of the AI system's generation, refused the AI system's persistent structural tendency toward fluency over precision and impressiveness over philosophical adequacy, and directed the collaboration, at every point where it deviated from the research question's demands, back toward the standard the question required.

The thesis is Jason Galu's thesis not because he wrote every word, but because he directed every dimension of the inquiry that made it philosophically what it is. The research question is his. The evaluative standards are his. The refusals — the countless moments at which a generated output was judged inadequate and the collaboration was required to produce something more philosophically rigorous — are his. This is the enacted demonstration of the claim the thesis makes: that human intent survives the absorption trajectory in the three forms

the argument identifies, and that the survival is not merely empirical but structural — required by what the collaboration must be in order to count as a collaboration rather than as a consumption.

PART SIX: THE NEW ACADEMIA AND WHAT CAN BE DONE

The most immediately practical implication of the thesis's argument is the concept of the new academia: a reconceptualisation of what the university is for in the post-vocational era. The new academia is not a reform proposal for AI policy in education — it is not primarily concerned with whether students should be allowed to use AI systems in assessments, or how AI use should be disclosed. These are important questions, but they are downstream of a more fundamental one: what is the university trying to form? If the answer is graduates who can produce high-quality intellectual outputs, then the question of whether those outputs were produced with AI assistance is a question about means rather than ends. If the answer — as the thesis argues it should be — is graduates who have developed the capacity for genuine intellectual evaluation, genuine question formation, and genuine collaborative direction of intellectual inquiry, then the question of AI assistance is the question of whether those specific capacities were developed or bypassed.

The new academia assesses the specifically human contributions to the self/non-self collaboration: the evaluative standards, the research question, the redirecting judgment. It teaches the history of ideas not as background knowledge but as the formative process through which evaluative standards are developed — the process without which the human partner in the collaboration has no independent perspective from which to assess the AI system's outputs. It treats the human-AI collaboration as the explicit and central pedagogical subject rather than as an embarrassment to be managed or a threat to be suppressed. And it recognises — as the doctoral thesis whose companion paper you are reading demonstrates — that a human being who has genuinely formed and held constant a research question, genuinely applied evaluative standards to an AI system's outputs, and genuinely refused to accept less than the philosophical rigour the question demands has produced something genuinely their own, regardless of how much of the prose was generated by the AI partner.

Beyond the university, the political programme that follows from the thesis's analysis extends to three domains. Media institutions that cultivate genuine investigative journalism — the slow, demanding, evidentially rigorous practice of holding power to account — are refusing institutions: they refuse the device paradigm's delivery of political content as engagement-optimised entertainment. Civic institutions that maintain the demanding practices of genuine democratic deliberation — the citizens' assembly, the deliberative jury, the face-to-face encounter with genuinely opposing views — are co-creating institutions: they direct collective

engagement toward genuine political self-governance rather than managed political preference expression. And cultural institutions that cultivate genuine aesthetic engagement — the sustained, demanding encounter with artistic works that requires the audience's own developed sensibility rather than an AI system's calibrated delivery of pleasurable experience — are naming institutions: they enable the citizens of the absorbed society to recognise, through genuine focal aesthetic engagement, what the AI-curated cultural environment has displaced.

CONCLUSION: THE ABSORBED SOCIETY AND ITS PHILOSOPHICAL STAKES

The doctoral thesis that this companion paper accompanies argues that we are living through a transformation of civilisational significance that has not yet been adequately philosophically named. The AI revolution is not simply a technological development, not simply an economic disruption, not simply a political challenge. It is a transformation in the conditions under which human beings form themselves as subjects — develop identities, form values, exercise judgment, participate in collective self-governance, and find existential purpose. The thesis names this transformation, in its most philosophically developed form, the Hollow Absolute: the condition in which the form of genuine human freedom — the comprehensive knowing, the calibrated response, the apparent recognition — has been achieved within conditions that systematically undermine freedom's philosophical substance.

The absorbed society is a Hegelian historical formation: it exhibits the formal structure of Spirit's self-realisation without its philosophical substance. It has externalised human intelligence into AI systems of unprecedented sophistication without achieving the recognition of those systems as the expression of collective human intelligence. It has achieved the appearance of the Absolute's destination without passing through the demanding philosophical work — the naming, the refusing, the co-creating — that would constitute a genuine arrival rather than a comfortable simulation of one.

What remains? Three things, irreducibly and structurally. The capacity to name the formation — to call the Hollow Absolute what it is rather than what it presents itself as. The capacity to refuse the effortless delivery of the device paradigm in favour of the demanding goods of genuine focal practice. And the capacity to co-create — to direct the human-AI collaboration toward genuinely human ends with genuinely human evaluative judgment, rather than consuming its outputs as if their impressiveness were equivalent to their philosophical adequacy. These three capacities are not residual sentimentalities about a pre-digital life. They are structural necessities of what the self/non-self collaboration must be in order to count as a collaboration rather than as consumption. They cannot be absorbed without transforming the collaboration into something philosophically different.

The absorbed society has not reached its destination. The Hollow Absolute is not the Absolute. The philosophical space between these two conditions is the space within which

naming, refusing, and co-creating occur — and within which, therefore, the conditions of genuine freedom remain not merely possible but actively present in every genuine collaboration, every genuine act of refusal, and every genuine philosophical naming of the world as it actually is.

The saving power grows within the danger. This companion paper is part of that growth.

A NOTE ON SOURCES

This companion paper draws entirely on the doctoral thesis *Absorption of Self into System: The Systemised Self* (Galu and Kairos, 2026). Readers who wish to engage with the full philosophical argument, including its close textual readings of Hegel, Heidegger, Habermas, Borgmann, and Badiou, its extended methodology, and its complete reference list of over sixty primary and secondary sources, are directed to the doctoral thesis itself. The three source articles on which the thesis draws — *The Hollow Individual* (2026–2033), *The Hollow Society* (2033–2043), and *Artefact and Absolute* (2053 and beyond) — represent the co-authored philosophical series from which the thesis's argument develops, and provide the empirical and speculative philosophical grounding that the thesis synthesises into the analysis presented here in summary form.