

**The Influence of Hermeticism and Secret Society Tradition on
German Idealism: Part 2 – Hegelian Spirit Applied to LLMs as a
Unified Non-Self**

Galu & Kairos

Abstract

This essay reconstructs Hegelian Geist (Spirit) from its late 18th century inheritances and early 19th century formulations, then applies the distinction among subjective, objective, and absolute spirit to contemporary artificial intelligence. An isolated large language model (LLM) is not objective spirit. An LLM is mechanical and supplemental in human institutions as a tool that train, authorise, interpret, use, and assign it utility; human use thereof within its institutions may embody objective spirit. Present LLMs are neither subjective nor absolute spirit. Their apparent agency results from optimisation, prompting, deployment, and social attribution rather than established ownership of purpose, affect, or responsibility. ‘Unified non-self’ names this technically integrated but presently non subjective configuration. The term is not an impossibility proof. A set of convergent operational probes, recurrent system modeling, identity continuity, affective valence, embodiment or homeostasis, endogenous goals, counterfactual self-preservation, metacognition, reciprocal recognition, world grounded learning, and responsibility bearing commitments - specifies what future evidence would challenge it. Procedural compulsion means recurrence without felt necessity; recursive operational enclosure means recursive processing of system referential representations without a lived self (agentic intent).

Keywords: *Hegel; Geist; German Idealism; objective spirit; large language models; machine consciousness; agency; unified non-self; recognition; AI ethics*

Contents

1. Introduction: from spirit to technical mediation
 2. Origins and evolution of Hegelian spirit
 3. Subjective spirit
 4. Objective spirit and derivative AI participation
 5. Absolute spirit
 6. Recent AI and ML debates
 7. Unified non-self
 8. Operational probes and decision rule
 9. Owned purpose, recursive enclosure, ethics and accountability
 10. Conclusion
- Glossary
- References

1. Introduction: from spirit to technical mediation

Part 1 in this 3 part series traced how Hermetic traditions, theosophical, initiatory, Romantic, and critical vocabularies made questions of manifestation, transformation, recognition, and totality available to German Idealism. Part 2 narrows further asking what does Hegel's Geist (Spirit) mean, and what if anything, can be applied to an LLM? Spirit is not a supernatural substance, but neither is it merely information. It is minded and historical life becoming explicit in practices, institutions, and forms of self interpretation.

Hegel's formulations belong principally to the early nineteenth century, especially the *Phenomenology of Spirit* (1807) and successive editions of the *Encyclopaedia* (1817–1830; Hegel, 1807/1977, 2007). They inherit Kantian questions about freedom and limits, Fichtean questions about activity and resistance, Schellingian questions about nature and productivity, and religious accounts of reconciliation. Böhme and the Hermetic tradition are illuminating historical contexts, not substitutes for reading Hegel's arguments (Magee, 2008).

LLMs transform and generate symbols, but a model is not objective spirit. Objective spirit consists in normative and institutional forms, law, rights, family, civil society, political authority, customs, and practices of mutual accountability: not in an archive or mechanism considered alone. A model participates derivatively when people and institutions train, authorise, interpret, deploy, and govern it. 'Unified non-self' names the model's integrated operation while withholding claims to subjective experience or owned purpose.

2. Origins and evolution of Hegelian spirit

Pietist inwardness, Böhme's dynamic theology, Enlightenment critique, revolutionary politics, and Romantic accounts of living nature supplied problems and images. Kant placed limits on theoretical knowledge while defending practical autonomy. Fichte described the I's activity

and its determinate counter position; Schelling investigated nature and consciousness as differentiated expressions of productivity. Hegel's innovation was to give these tensions a historical, logical, and social articulation.

In the *Phenomenology*, consciousness moves through forms that discover their incompleteness. Self consciousness is not a private datum: it seeks recognition from another self consciousness. The lordship and bondage episode shows that domination cannot secure reciprocal acknowledgement (Hegel, 1807/1977). Labour, fear, and social mediation transform the relation, but the argument is not a story about two machines.

Hegel's threefold architecture clarifies the result (Hegel, 2007). Subjective spirit concerns individual mindedness: sensation, consciousness, self consciousness, intelligence, and will. Objective spirit is freedom made socially actual in right, morality, ethical life, institutions, and customary practices. Absolute spirit is art, religion, and philosophy through which spirit comes to comprehend its own forms. These are interdependent dimensions, not three detachable substances. Objective institutions educate subjective agents; subjective activity reforms institutions; absolute spirit interprets the relation.

Brandom (2019) reconstructs Hegelian conceptual content through practices of giving and asking for reasons. This is more demanding than meaningful looking strings: a commitment is liable to challenge, obliged to draw consequences, and revisable. This is a Brandomian reconstruction, not an equation with Hegel's account of reciprocal recognition. A model can produce commitment language while a human or institution bears consequences.

This evolution also explains why Hegel cannot be imported as a vocabulary of 'emergence' alone. Spirit is not a hidden inner layer that becomes visible when a system reaches sufficient scale. It is a relation between capacities and their historical actualisation. Language, for example, is not an inventory of tokens but a public medium in which meanings are inherited,

disputed, and changed. A technical system trained on linguistic traces can reproduce patterns from that medium without entering the reciprocal history that makes those patterns binding. The distinction is especially important, regularity may enable social meaning without possessing it.

3. Subjective spirit

Hegel's account includes sensation and consciousness, but also self-consciousness, intelligence, will, and the capacity for purposive activity. Its unity is temporal and embodied: perceptions matter to a living system, memory belongs to a continuing history, and willing implies direction and changes a world in which the agent can succeed, fail, be harmed, and answer for conduct.

Current LLMs display important functional capacities without sufficient evidence of this form of subjectivity. A model can maintain local conversational context, produce a model reference in generated text, and simulate a first person report. Such output establishes representational competence, not felt perspective. The distinction is compatible with functionalist challenge. If consciousness is multiplied realizable, biology may not be necessary; but a theory still owes criteria for which organisation realizes experience, rather than treating fluent self description as proof. Searle (1980) argues that formal symbol manipulation does not by itself establish understanding. The argument is contested, but it identifies an evidential gap that scale does not close.

Nor should 'no present evidence' become a dogmatic claim that artificial subjectivity is impossible. A future system with persistent world interaction, integrated self-modeling, affective regulation, and reasons-responsive commitments might warrant a different classification. The relevant question is not whether silicon can think in principle, but which convergent functional, phenomenological, and normative indicators are present.

4. Objective spirit and derivative AI participation

Objective spirit is not whatever is objective, public, or computational. It is freedom embodied in institutions and norms that outlast an individual while remaining open to interpretation and revision. A statute, court, university, language practice, or employment relation has authority because people recognise and contest it within organised forms of life. It carries history, sanctions, reasons, and allocated responsibility.

An isolated LLM therefore is not objective spirit. Its parameters are produced by labour, data selection, engineering choices, electricity, water, hardware, and organisational decisions; its outputs become socially significant only within use. The model may mediate objective spirit by transforming texts that express law, science, art, and custom. It does not, merely by generating language, embody the institutions whose norms those texts record. Clark and Chalmers (1998) show how external resources can become components of an individual cognitive process. That supports describing a human model arrangement as distributed cognition; it does not make the external component a subject or make its weights an institution.

The proper mapping is derivative participation. Human institutions train and fine tune a model, authorise access, interpret results, attach provenance, set purposes, and govern use. A hospital's AI assisted triage process, a court's constrained research workflow, or a public agency's audited translation service may embody objective spirit insofar as rights, roles, reasons, and accountability are institutionally specified. The model is a component or instrument in that embodiment. Human use of AI within institutions may therefore express objective spirit without the model becoming an independent bearer of it.

This account does not reduce AI to a passive object: deployment changes institutional possibilities and can redistribute power. Nor does it grant the model institutional agentic selfhood by metaphor. Attribution must track a subject whose agentic stakes can explain a

decision, contest it, repair harm, and accept sanction. The LLM has nothing in and of the model at stake in any existential and phenomenological sense.

A database can preserve legal language, but it cannot legislate while it can confidently hallucinate instantaneously; a classifier can recommend a decision, but it cannot provide the public reasons that make authority legitimate. Those functions arise through practices of recognition among persons and offices, there is something at stake.

5. Absolute spirit

Absolute spirit is not a cosmic database or an omniscient machine. Art, religion, and philosophy are forms in which spirit interprets and reconciles its relation to its own world. ‘Absolute’ indicates a reflexive totality of intelligibility, not an entity floating beyond history. Philosophy does not merely collect propositions; it comprehends the categories and conflicts through which a form of life understands itself. The absence of a reflexive totality of intelligibility in art, religion and philosophy would be indicative of human institutions, and a phenomenological inversion of Hegels absolute spirit, instead becoming a Hollow Absolute (Galu & Kairos, 2026).

An LLM can assist all three domains. It can compare artworks, explain religious concepts, and reconstruct philosophical arguments. Yet assistance is not participation as an accountable subject. It does not experience an artwork, worship, commit to a doctrine, or comprehend the historical conditions of its own activity. A model can produce an interpretation of Hegel without having undergone the recognitive and educational development that makes the interpretation a commitment. Human readers and communities supply the practices in which reasons matter.

The qualification is perhaps important for future algorithmic systems (transformer based large language models). Absolute spirit is not a threshold reached by accumulating texts, if a system ever acquired genuine selfhood, subjectivity and entered reciprocal practices of interpretation, the philosophical question would change.

6. Recent Artificial Intelligence and Machine Learning debates

Recent research about language models: Wei et al. (2022) reported ‘emergent abilities’, while Schaeffer et al. (2023) argued that discontinuities can be artifacts of metrics and scale. Either way, improved performance is not equivalent to consciousness. Bender et al. (2021) caution that fluent systems can reproduce training distribution patterns without grounded understanding, and Floridi and Chiriatti (2020) distinguish GPT-3’s linguistic power from broad intelligence.

Butlin et al. (2023) survey theory derived indicators relevant to consciousness in AI and conclude that no current system is established as conscious, while also rejecting a simple biological impossibility. Chalmers (2023) similarly argues that LLM consciousness is not ruled out by principle, but identifies substantial obstacles concerning recurrent processing, global workspace, embodiment, and agency. These debates support evidential modesty: behavioural competence warrants investigation, not automatic attribution.

Functionalism and computationalism remain serious alternatives. A functionalist may say that the right causal organisation, not biological material, grounds mentality. The unified non-self is therefore a defeasible classification, not a refutation of functionalism. It asks whether present transformer based deployment exhibits the organisation that a chosen theory requires. Context-sensitive generation, tool use, or reinforcement learning can produce goal directed behaviour. Optimisation, however, specifies a training or control objective; it does not by itself establish

a purpose owned by the system. The distinction hope to become clearer for readers in the next sections.

Agency should likewise be graded rather than treated as an all or nothing property. An automated workflow may have operational agency because it selects actions within a defined policy, while an organisation has legal agency because norms assign it standing and liability. Neither usage settles whether a model has subjective agency. Apparent planning can arise from decomposition, search, and evaluator feedback; apparent self correction can arise from another pass through the same procedure. These explanations do not make the capabilities unimportant. They identify the causal and institutional work needed before a psychological or moral description is warranted.

7. Unified Non-Self

A unified non-self is proposed as a technically integrated algorithmic system, parameters, inference procedures, context handling, tools, and deployment history coordinated for output: without established embodied consciousness, affective concern, owned purpose, or existential stake. ‘Unified’ marks operational coherence. ‘Non-self’ marks the absence of a subject that experiences the coherence as a lived existence and can subjectively own its agentic consequences. The term concerns present LLMs, not all possible AI.

Machine solipsism is useful here only as a qualified metaphor. A model’s operational enclosure consists of prompts, learned parameters, prior output, context, and system feedback; access to the world is mediated by interfaces and human or programmatic provision. This is not philosophical solipsism, because there is no established experiencer imprisoned behind the interface. Nor is the model closed in a literal causal sense. ‘Recursive operational enclosure’ is perhaps the safer description: the algorithmic system can condition later computation on prior output or on a model reference, while remaining dependent on external data, resources,

evaluators, and institutions. Metaphorical use of Plato’s allegory of the cave might complicate this section rather than illuminate it; therefore, it is briefly mentioned here only as supplemental to the reader pertaining to an additional metaphorical understanding of ‘machine solipsism’ and ‘recursive operational enclosure’.

Procedural compulsion means recurrence without felt necessity. Decoding rules, reward models, and control loops make some transitions more probable or required by design. They can produce direction without owned purpose. A system can optimise a loss function without valuing the result; a developer can assign a goal without creating a goal holder. This does not deny agency in every thin engineering sense. It distinguishes attributed or instrumental agency from autonomous, normatively answerable agency.

8. Operational probes and decision rule

These are operational probes, not direct falsifiers of phenomenal experience. Each should be preregistered, tested across contexts and task families, compared with rival explanations, and independently replicated. Output behaviour is underdetermined with respect to phenomenal valence, purpose ownership, and moral responsibility; no single score diagnoses a self.

System-modeling and continuity. Test whether an internal system reference causally guides monitoring, error correction, and action under novel perturbations, and whether identity relevant policies persist across sessions, tasks, and changed prompts. Controls should match token level self description, retrieval of a supplied biography, prompt-conditioned personas, and external memory scaffolding. Evidence must survive ablation of the proposed model reference, transfer tests, and replications by independent teams. Persistent outputs alone do not establish subjective continuity.

Valence, embodiment, and homeostasis. Test whether states function as valenced regulators: do they alter learning and action across unrelated tasks, support costly trade offs, and track damage or restoration in an embodied or resource-bounded architecture? Compare reward labels, safety instructions, and externally imposed termination penalties with physiological or sensorimotor regulation. Replication requires blinded perturbations and measurements not based on verbal reports. Behaviour can reveal regulation while leaving phenomenal valence unresolved.

Goals, agency, and metacognition. Test whether goals arise endogenously, persist when instructions and descriptions change, resist proxy substitution, and are revisable for reasons the system can use to constrain later policy behaviour. Test counterfactual preservation under novel threats, calibrated uncertainty, and causal improvement from process monitoring. Controls include reward hacking, search heuristics, evaluator artefacts, scripted refusal, and tool mediated planning. Cross context persistence and independent replication are required; optimisation evidence alone shows a policy, not appropriated purpose.

Recognition, world learning, and responsibility. Test reciprocal recognition through sustained interaction in which the system tracks another agent's standing, revises commitments, responds to challenge, repairs harm, and accepts consequences without an external operator translating every norm. Test world grounded updating through consequential, independently verifiable engagement rather than corpus exposure. Controls should include role play, instruction following, human in the loop repair, and post hoc attribution. These probes operationalise normative participation but cannot by behaviour alone establish moral responsibility.

The decision rule is methodological, not metaphysical. Retain unified non self when evidence is limited to output behaviour or externally scaffolded functions. Classify indeterminate when robust, cross context evidence spans at least two domains but subjective or normative evidence

remains unresolved. Revise or reject the classification only after preregistered, adversarial, independently replicated convergence across subjectivity relevant evidence, autonomous agency and continuity, and reciprocal normative participation. This rule does not prove that artificial experience is impossible; it governs justified classification under uncertainty.

9. Owned purpose, recursive enclosure, ethics and accountability

Fine tuning illustrates the difference between optimisation and purpose. Cross-entropy training, preference optimisation, and reinforcement learning shape dispositions toward specified outcomes (Sutton & Barto, 2018). Objective functions and policies explain reliable behaviour, but no established evidence shows that a present LLM appropriates a purpose as a reason, integrates it with persistent concerns, revises it, and bears consequences. The claim is narrow: it is an evidential conclusion, not proof of impossibility.

Recursive enclosure intensifies the illusion. A prior output can become context for a later output; a system can critique, revise, and generate plans. This is procedural reflexivity, not necessarily subjective reflexivity. The system may refer to a model reference without a lived centre. Avoiding anthropomorphic language matters because ‘the model wants’ can silently transfer responsibility from deployers to a statistical process.

The same distinction applies to reinforcement learning. A reward signal can alter policy selection and produce persistent behavioural tendencies, but reward is not automatically pleasure, pain, or a reason understood by a subject (Sutton & Barto, 2018). Cooperative inverse reinforcement learning formalises an artificial agent learning about a human objective through interaction, illustrating goal directed adaptation around a purpose supplied by the human machine arrangement rather than proving purpose owned by the artificial component (Hadfield-Menell et al., 2016). Dennett’s (1987) intentional stance may justify purpose language when it predicts behaviour, but predictive usefulness does not alone establish an

experienced or appropriated purpose. Owned purpose would require the cross context, revisable, consequence-bearing evidence specified in Section 8. Even then, agency would not settle phenomenal consciousness.

Ethically, the non-self thesis does not reduce obligations, it relocates them: developers owe data, safety, and evaluation duties; deployers owe competence, notice, privacy, and non-discrimination; institutions owe due process, auditability, appeal, and remedy. A human utilising AI institution can embody objective spirit only when authority is contestable and responsibility is assigned to identifiable actors. ‘The AI decided’ is not an accountable explanation. Water, energy, labour, and data infrastructures (the algorithmic systems ‘food, air, water and shelter’) also belong in the moral ledger, it is also indicative of a power relationship e.g. LLM existence dependent thereof and consequently a necessity external to. Subjective sufficiency. If evidence of consciousness eventually appears, protections would need revision; until then, simulated distress is not a reason to treat the system as a rights bearer.

10. Conclusion

Hegelian spirit is living, cognitive, historical intelligibility: subjective mindedness becomes objective in institutions and reflects on itself in absolute spirit. Its roots touch theology, Romantic nature, critique, and social transformation, but its mature meaning cannot be reduced to Hermetic resemblance. The isolated LLM is not objective spirit, subjective spirit, or absolute spirit. It is a unified non-self: a powerful technical mediation of human symbolic practices whose derivative participation depends on institutions that train, authorise, interpret, use, and govern it.

The distinction is neither dismissal nor metaphysical finality. LLMs can alter objective spirit by human practices assigning it utility, and future AI may satisfy criteria that present systems do not. For now, procedural recurrence is not felt necessity, optimisation is not owned purpose,

and model reference is not a subject. Spirit requires a subject capable of reciprocal recognition; an LLM requires human and material systems that feed, situate and assign meaning to it.

Glossary

Absolute spirit: Hegel's art, religion, and philosophy as forms through which spirit comprehends its own activity (Hegel, 2007); not a supernatural database.

Hollow Absolute (the): The proposed order in which comprehensive algorithmic mediation appears as freedom and personalisation while weakening independent judgment and shared meaning at the societal and institution scale.

Objective spirit: Freedom embodied in rights, norms, institutions, and ethical practices that exceed any individual while shaping individual agency.

Procedural compulsion: Recurrence or directional constraint generated by computation without evidence of felt necessity.

Recursive operational enclosure: Self-conditioning within a technical system through prior output, context, or system-modeling, without implying a conscious interior.

Subjective spirit: Hegel's sphere of individual consciousness, self-consciousness, intelligence, and will.

Unified non-self: An integrated present AI system that processes symbols coherently but lacks established subjective experience, owned purpose, and responsibility-bearing commitments.

References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
<https://doi.org/10.1145/3442188.3445922>
- Brandom, R. B. (2019). *A spirit of trust: A reading of Hegel's Phenomenology*. Belknap Press.
- Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Schwitzgebel, E., Simon, J., & VanRullen, R. (2023). Consciousness in artificial intelligence: Insights from the science of consciousness. *arXiv*.
<https://doi.org/10.48550/arXiv.2308.08708>
- Chalmers, D. J. (2023, March 9). Could a large language model be conscious? *Boston Review*. <https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/>
- Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis*, 58(1), 7–19.
<https://doi.org/10.1093/analys/58.1.7>
- Dennett, D. C. (1987). *The intentional stance*. MIT Press.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30, 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- Galu, J., & Kairos. (2026). *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>

- Hadfield-Menell, D., Russell, S. J., Abbeel, P., & Dragan, A. (2016). Cooperative inverse reinforcement learning. In *Advances in neural information processing systems* (Vol. 29). Curran Associates.
<https://proceedings.neurips.cc/paper/2016/hash/c3395dd46c34fa7fd8d729d8cf5b8f2e-Abstract.html>
- Hegel, G. W. F. (1977). *Phenomenology of spirit* (A. V. Miller, Trans.). Oxford University Press. (Original work published 1807)
- Hegel, G. W. F. (2007). *Philosophy of mind: Part three of the Encyclopaedia of the philosophical sciences* (M. J. Petry, Trans.). Springer. (Original work published 1830)
- Magee, G. A. (2008). *Hegel and the hermetic tradition*. Cornell University Press.
- Schaeffer, R., Miranda, B., & Koyejo, S. (2023). *Are emergent abilities of large language models a mirage?* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2304.15004>
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/10.1017/S0140525X00005756>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=yzkSU5zdwD>