

From Derivation to Integration: A MANHAJ Framework for Independence of Thinking in AI

Scientific Research Paper — Paper 5

Author: Faouzi Messeoud — فوزي مسعود

Methodology: MANHAJ — Intellectual Modeling Methodology

Version: 1.0 — October 2026

Status: Scientific Research Paper

About this Paper

This paper applies the Law of the Overall Duality of Mental Movement (Law 5) of the Intellectual Modeling Methodology (MANHAJ) to artificial intelligence, together with the Law of Dual Direction of Examination (Law 2), the Law of Rightness (Law 3), the Law of Derivation of the Cognitive Constant (Law 7), and the Theory of Doctrinal Centrality. It distinguishes the derivational movement of knowledge from the integrative movement of thought, gives the integrative movement a formal criterion, and defines the conditions under which an AI system can pass from cognitive thinking under an implicit framing to independence of thinking under a declared one.

The paper is part of the MANHAJ research program. It continues Paper 02, which formalized derivation and its closure, and Paper 04, which established that every learned representation operates under an embedded doctrinal centrality and made that centrality measurable from outside the system.

Relationship to the MANHAJ Reference Document

This paper develops a specific research dimension of the methodology presented in: Messeoud, Faouzi. *Intellectual Modeling Methodology (MANHAJ) — Reference Document — Version 1.0*. Zenodo, 15 September 2026. DOI: 10.5281/zenodo.22769627. The present paper is an independent scientific study and should not be considered a replacement for the reference document.

Citation

Messeoud, Faouzi. *From Derivation to Integration: A MANHAJ Framework for Independence of Thinking in AI*. Scientific Research Paper, Paper 05, Version 1.0, 2026.

The scientific paper begins on the following page.

Abstract

Artificial intelligence systems are increasingly described as reasoning, and progress in their reasoning is measured by depth, length, and correctness. The Intellectual Modeling Methodology (MANHAJ) applies a different criterion. Under its Law of the Overall Duality of Mental Movement, knowledge is a derivational movement from an accepted reference toward details, and thought is an integrative movement from details back to the foundations that produced them. The criterion that separates the two is the position of the movement relative to its framing, not the intensity of the reasoning within it.

Applied to AI, this criterion yields the paper's founding distinction. A system that operates a posteriori, under a framing it has accepted without examining it, performs cognitive thinking: a derivational movement that produces meaning but remains under its framing and descends along the meaning variable toward closure. Current approaches — extended reasoning, self-critique, abduction, and the extraction of principles from judgments — operate at this a posteriori level. In MANHAJ terms they are knowledge, however sophisticated, and not thought.

The paper gives the integrative movement a formal criterion. An operation is integrative when its output lies at a higher level of the foundational order than its inputs, when its inputs are derivable from that output, and when the output becomes an object of examination rather than the reference by which it is itself evaluated. Integration so defined is an inverse problem whose solution is not unique: it determines a class of framings, and the selection within that class requires a reference external to the data. Two propositions follow. A process is derivational with respect to any reference it does not examine, whatever its internal complexity; this includes the processes of current AI, which are derivational with respect to their embedded framing. And, under explicit conditions, an ascent whose selection draws only on the data and the learned model does not satisfy the condition of independent examination, and so cannot examine the implicit framing, because its criterion of examination is drawn from the very framing under examination, which is a methodological circularity and a begging of the question. At the summit of the ascent, the Law of Rightness applies as a binary judgment on the reached foundation, distinct from the quantitative correctness of its fit to the data. The judgment is relative to the declared centrality of the examining field; it is not an absolute verdict.

On this basis the paper defines independence of thinking as a temporal process: the repeated capacity of a system to ascend from its own outputs to the framing under which they were produced, to confront that framing with a declared and recognized centrality, to judge its rightness, and to re-derive under the admitted frame. Independence of thinking is independence from unexamined framing, not exit from all framing. An appendix presents designs for testing the resulting capabilities.

Keywords: MANHAJ; artificial intelligence; knowledge and thought; derivation; integration; cognitive thinking; independence of thinking; doctrinal centrality; a priori examination; Law of Rightness; inverse problems.

1. Introduction

1.1. The founding distinction

The approaches through which artificial intelligence reasons operate at the a posteriori level. They proceed from a framing already in place — carried by training data, by the selection of that data, by the design of the model, and by the definition of its objectives — and they evaluate each of their steps by

criteria drawn from that same framing. In the terms of MANHAJ, such processes are knowledge, not thought. This holds however deep, long, or correct the reasoning becomes, because what decides the classification is not the exercise of reason but the framing under which reason is exercised.

The distinction comes from the Law of the Overall Duality of Mental Movement (Law 5). Knowledge is a derivational movement — in the sense of mathematical derivation along the meaning variable, not in the loose linguistic sense of the word — from a known origin or reference toward particulars, details, and dependent relations; thought is an integrative movement — in the sense of mathematical integration along the meaning variable — that gathers scattered domains and details and reconnects them toward the overall foundation (Reference Document, Sections 5.5 and 9.2). The a priori movement is the movement of thought, and the a posteriori movement is the movement of knowledge (Reference Document, Section 8.2). Derivation is not empty: it produces new meaning at every step before closure. The Reference Document calls this cognitive thought (Section 8.2.1); this paper calls it cognitive thinking — meaning-producing derivation within a framed field. Cognitive thinking is a form of the derivational movement, and it is not thought in the integrative sense.

This paper is built on that distinction. It states it at the outset because the rest of the argument depends on it: a system can produce new, correct, and complex results and still remain entirely within the movement of knowledge.

1.2. The problem in AI terms

AI research measures progress in reasoning by the depth and length of reasoning chains, by search over candidate solutions, by self-critique and revision, and by accuracy on benchmarks. Each of these improves the derivational movement. None of them, by itself, changes the position of the system relative to its framing. A system that reasons for longer under an unexamined framing performs more cognitive thinking under the same framing.

The consequence is structural. A system that operates a posteriori accepts its framing implicitly and works under its direction. Its outputs may be novel, but their novelty is a novelty of derivation: each output is admissible relative to a reference that the system has never examined. The question of whether that reference itself is admissible does not arise within the system's operation.

1.3. The gap left by Papers 02 and 04

Paper 02 formalized derivation as a meaning-producing movement and detected its closure through the Law of Derivation of the Cognitive Constant, $d/dx(C) = 0$. It treated the descending movement. Paper 04 founded the measurement frame of meaning a priori and established, as a methodological consequence of MANHAJ, that every learned representation operates under an embedded doctrinal centrality, Γ_L . It made the orientation of Γ_L measurable — but by an examiner standing outside the system.

Two things remain open. The integrative movement has not been formalized: the program has a model of descent but not of ascent. And the system itself has not been given the capacity to examine the framing under which it operates: Γ_L has been measured, not reached from within.

1.4. Central question and thesis

The central question of the paper is:

What distinguishes the integrative movement from the derivational movement along the meaning variable, and under which conditions can an AI system perform it — ascending to the framing under which it operates and judging its rightness?

The thesis is that an a posteriori system performs cognitive thinking under an implicit, unexamined framing, and that the integrative movement, governed by Laws 5, 2, and 3 and anchored in a priori founding, moves it to independence of thinking: a declared a priori foundation to which the system itself can ascend and which it can examine.

Independence of thinking, so defined, is independence from unexamined framing. It is not independence from all reference, which no movement of thought can reach, and it does not include the production of a new centrality, which the paper leaves open.

1.5. Contributions

- **C1 — The founding distinction applied to AI.** A criterion, derived from Law 5, that classifies an AI process as knowledge or thought by its position relative to its framing, and not by the depth or correctness of its reasoning (Proposition 1).
- **C2 — A formal criterion of integration.** Three jointly necessary conditions — level, generation, and objecthood — that separate integration from summarization, abduction, and principle extraction.
- **C3 — Non-uniqueness and the role of the a priori.** A treatment of integration as an inverse problem whose solution is a class of framings, and a conditional result that an ascent whose selection draws only on the data and the learned model does not satisfy the condition of independent examination (Proposition 2).
- **C4 — Rightness at the summit.** The application of Law 3 to the foundation reached by the ascent, distinguishing a binary judgment of rightness from the quantitative correctness of fit.
- **C5 — Independence of thinking.** A definition of independence of thinking as a temporal process, and its integration with the derivational model of Paper 02 and the founded frame of Paper 04.
- **C6 — Test designs (Appendix A).** Four designs that specify how the resulting capabilities can be tested in subsequent applied work.

1.6. Scope

The paper concerns the movement of an AI system relative to its framing. It does not claim that any existing system performs the integrative movement, and it does not claim that the designs presented in its appendix have been carried out. The centrality relative to which the system's framing is judged is recognized and declared by the examining field, as in Paper 04; the paper does not treat the production of a founding doctrinal centrality by the system, which it leaves as an open theoretical question (Section 9).

2. Related Work

The works reviewed here are strong within their own aims. They are discussed for one question only: where each stands relative to its framing. In every case the reference that evaluates the process is drawn from the same data, objectives, or preferences that formed the system, so the process remains a posteriori in the sense of Law 2.

2.1. Depth and length of reasoning

Chain-of-thought prompting elicits intermediate steps before an answer (Wei et al., 2022). Self-consistency samples several chains and selects the most frequent answer (Wang et al., 2023). Tree-of-thoughts searches over branching partial solutions (Yao et al., 2023). Scaling test-time computation allocates more inference to harder problems (Snell et al., 2024). Each extends the derivational path or

searches among derivational paths. The criterion of selection — agreement, a learned value, or a verifier — belongs to the framing under which the paths are generated. These methods increase cognitive thinking; they do not change its position.

2.2. Self-critique and principle-guided revision

Self-Refine and Reflexion have a model criticize and revise its own outputs (Madaan et al., 2023; Shinn et al., 2023). Constitutional AI revises outputs against a written list of principles (Bai et al., 2022). The constitutional approach comes closest to a declared reference, since its principles are written down. But the principles are applied as a fixed parameter of revision: they evaluate outputs and are not themselves brought under examination. Self-critique without such a list draws its criteria from the model itself, and it has been shown to be unreliable as a corrector of reasoning (Huang et al., 2024). Explanations produced by a model about its own reasoning may also fail to reflect the factors that actually drove its outputs (Turpin et al., 2023).

2.3. Abduction and the extraction of principles

Abductive reasoning seeks the most plausible explanation of given observations (Bhagavatula et al., 2020). Inverse Constitutional AI compresses a set of preference judgments into a short list of principles that reconstruct them (Findeis et al., 2024). These works move in the upward direction: from particulars toward something that accounts for them. They are therefore the closest to the integrative movement, and Section 5 examines them against the formal criterion. Their limit lies in how they choose among candidate explanations. Plausibility and reconstruction accuracy are measured against the same distribution that produced the observations. The ascent is real, but it is resolved by the framing it should examine.

2.4. Embedded perspectives

A line of work measures the opinions or perspectives that language models reflect, by comparing their outputs with survey responses of different populations (Santurkar et al., 2023; Durmus et al., 2023). Paper 04 established on methodological grounds that every learned representation operates under an embedded centrality and gave a relational procedure for characterizing it. In all these works the embedded perspective is examined from outside the system. The present paper concerns the movement by which the system reaches it from inside.

2.5. Position of this paper

The representational and inferential tools used in this paper — inverse problems, hypothesis generation, principle extraction, comparison of framings — are established. The contribution lies in the conditions MANHAJ imposes on their use: the classification of a process by its position relative to its framing, the formal criterion of integration, the role of a declared a priori reference in resolving non-uniqueness, and the judgment of rightness at the summit.

3. Methodological Foundation

3.1. Knowledge as derivation

Under Law 5, knowledge is a movement from a known origin or reference toward particulars, details, and dependent relations (Reference Document, Section 9.2). This movement is represented mathematically by derivation with respect to the meaning variable, starting from a defined reference

function within a framed field. The Reference Document illustrates it with the successive derivatives of a function taken as a known origin within a framed mathematical field:

$$f(x) = x^4 \rightarrow f'(x) = 4x^3 \rightarrow f''(x) = 12x^2 \rightarrow f'''(x) = 24x \rightarrow f^{(4)}(x) = 24 \rightarrow f^{(5)}(x) = 0$$

Each step before the constant produces a new meaning within the field; the derivative does not repeat its predecessor but extracts another meaning from it under a defined relation. When derivation reaches the constant, its derivative is zero. The zero does not mean the absence of meaning; it means that the derivational movement has ceased to produce new derivational meaning along that path (Reference Document, Section 9.2.1). This is the Law of Derivation of the Cognitive Constant (Law 7):

$$\frac{d}{dx}(C) = 0$$

The direction of the derivational movement is therefore downward along the meaning variable: from the reference toward the particular, until meaning is closed. In the earlier formulation of the author's book on doctrinal centrality, the lowest limit of meaning is the sample realized in reality, below which nothing is derived (Messeoud, 2024, pp. 116, 122).

3.2. Cognitive thinking

Derivation is an intellectual act. The Reference Document names it cognitive thought: meaning-producing derivation within a framed epistemic field (Section 8.2.1); this paper calls it cognitive thinking, reserving the word thought for the integrative movement alone. The framed field may contain many unknowns, and resolving them is productive mental work (Section 8.2.1). Cognitive thinking is therefore not the repetition of what is known. What defines it is that every step is admitted relative to a reference accepted within the field, and that this reference is not itself examined. The reasoning may be long and inventive; it remains under its frame.

Two terms are kept distinct throughout this paper. *Thought*, used alone, denotes the integrative movement of Law 5. *Cognitive thinking* denotes the derivational movement within a framed field. The decisive criterion between them is not the exercise of reason but the framing under which reason is exercised. A function studied over its full domain of rightness and the same function studied over a sub-interval fixed in advance give the same kind of operation; what differs is whether the domain itself was fixed by the reference or imposed on it from outside (Messeoud, 2024, p. 176).

3.3. Thought as integration

Thought is a movement that gathers scattered domains and details and reconnects them toward the overall foundation; it may turn toward the origins, references, and foundations on which the epistemic field itself rests (Reference Document, Section 9.2). This movement is represented mathematically by integration with respect to the meaning variable, in the direction of recovering the foundation that accounts for the particular states:

$$\int f(x) dx = F(x) + C$$

Integration recovers, from a given function, the function from which it derives. The direction is upward along the meaning variable: from the particular toward what founded it. The Reference Document uses integration as a representation of the integrative movement, not as an equation that proves all functions of thought, and it distinguishes the constant of integration from the Cognitive Constant of Law 7 (Section 9.2.2). This paper keeps that distinction. It also draws on one property of the integral that the Reference Document records: the constant C is part of the solution but is not determined by integration

itself. A further condition, external to the operation, is required to fix it. Section 5 develops this property.

3.4. The two directions of examination

Law 2 distinguishes a priori examination, which inquires into the foundation, references, and conditions of a field before entering its details, from a posteriori examination, which operates within a field whose foundation has been accepted (Reference Document, Sections 4.2 and 8.2). The a priori movement is foundational and integrative: it returns to origins and gathers domains in an overall framework, and it is the movement of thought. The a posteriori movement is predominantly derivative: it proceeds from an accepted origin toward details, and it is the movement of knowledge (Section 8.2). Thought and the a priori are thus one movement viewed from two angles: as a direction along the meaning variable, and as a position relative to the foundation.

3.5. The foundational order and its summit

Ideas and knowledge are organized within higher reference frameworks that determine their domain, directions, and conditions of meaning; MANHAJ calls the highest of these doctrinal centrality (Reference Document, Section 4.1). Doctrinal centrality does not necessarily mean religion: it is the system of higher-level conceptions that frames a human group and gives it an interpretation of existence, of the human being, and of the meaning of life, whether religious or not (Section 4.1.1). The conceptual domain is constituted within it, and operative definitions are admitted within the domain. Paper 04 wrote this order as:

$$\Gamma \rightarrow D(\Gamma) \rightarrow Def \rightarrow \text{meaning states}$$

Read downward, the order is the path of derivation. Read upward, it is the path of integration: from meaning states to the definitions that admit them, from definitions to the conceptual domain, and from the domain to the centrality. The centrality is the summit of the upward path. No idea moves outside a higher reference framework (Reference Document, Section 4.1); the ascent therefore ends at the centrality and does not leave framing altogether. In the earlier formulation of the author's book on doctrinal centrality, the layer of the doctrinal centrality is the layer above which there is no integration, and the layer of the sample is the layer below which there is no derivation (Messeoud, 2024, p. 137).

3.6. Rightness and correctness

The Law of Rightness (Law 3) places the question of permissibility before the question of details: what is not permissible cannot have valid details (Reference Document, Sections 4 and 8.3). Rightness is a binary evaluation concerning existence: whether a foundation, or the approach through it, is admissible at all. Correctness is a quantitative evaluation of something that exists: how well a result conforms within an accepted system. In the earlier formulation of the book on doctrinal centrality, rightness is measured relative to the higher concept and correctness relative to samples of reality (Messeoud, 2024, p. 135). The distinction becomes decisive at the summit of the ascent (Section 6).

4. Current AI as A Posteriori Cognitive Thinking

4.1. The embedded framing

Paper 04 established that every learned representation operates under an embedded centrality Γ_L : the framing carried by the data from which the system was formed, by the selection of those data, by the design of the model, and by the definition of its objectives (Paper 04, Section 3.4). A system that

processes such data a posteriori does not operate outside any centrality. It accepts the embedded framing implicitly and works under its direction. This is a claim of the theoretical framework of MANHAJ, not an empirical finding; what remains open to examination is the content of Γ_L .

4.2. Processes, references, and parameters

Let a process P produce a sequence of meaning states $m_0, m_1, \dots, m_T$. Each transition is admitted relative to a reference R:

$$\text{Adm}_R(m_t \rightarrow m_{t+1}) \in \{0,1\}$$

The reference R may be a learned objective, a verifier, a list of principles, a set of preferences, or the implicit framing Γ_L . Two positions of R relative to P are distinguished:

- **R is a parameter of P** when R is used to admit or evaluate states and never appears as the object of any step of P: no step of P takes R as its output or as the subject of its examination.
- **R is an object of P** when at least one step of P takes R itself as what is produced or examined.

The distinction is not one of complexity. A process may contain branching, search, sampling, voting, critique, and revision, and R may remain a parameter throughout.

4.3. Proposition 1

Proposition 1. A process that does not make its reference R an object of examination is derivational with respect to R, whatever its internal complexity.

Argument. Under Law 5, the integrative movement reconnects details toward the foundation on which they rest; under Law 2, the a priori movement inquires into that foundation before or apart from its details. Both require the foundation to become the object of the movement. If R is a parameter of P, no step of P has R as its object. Every step of P is then, with respect to R, a transition admitted relative to R — a movement from an accepted reference toward further states, which is the definition of derivation. Increasing the number of steps, the breadth of search, or the number of revisions changes the derivational path and may improve its correctness; it does not change the position of R. ■

The proposition is relative to R. It does not exclude that the same process contains other movements in which a different reference is examined; it states that, with respect to R, the process remains derivational. For an a posteriori AI system the relevant reference is Γ_L : whatever else the system examines, as long as Γ_L remains a parameter, the system remains derivational with respect to the framing under which it works.

The proposition is methodological. It follows from the definitions of Laws 2 and 5 and does not depend on any particular architecture.

4.4. Application

The processes reviewed in Section 2 can now be placed.

Process	Reference R	Position of R	Movement
Chain-of-thought, self-consistency, tree search	Γ_L and a learned or programmed selection rule	Parameter	Derivational
Test-time scaling with verifiers	Γ_L and the verifier	Parameter	Derivational
Self-critique and revision	Γ_L , through the model's own criteria	Parameter	Derivational
Constitutional revision	A written list of principles	Parameter (declared)	Derivational
Abduction, principle extraction	Γ_L , through plausibility or reconstruction accuracy	Produced, but selected by Γ_L	Upward in form, resolved by the framing (Section 5.4)

Constitutional revision deserves a remark. Its principles are declared, which is a necessary step toward founding; but they are applied as a fixed parameter and are never themselves examined. Declaration without examination makes the reference explicit without making it an object. In MANHAJ terms the process is founded on a stated reference, yet it remains a posteriori relative to that reference.

4.5. The structural consequence

An a posteriori system is in the position that the earlier formulation of MANHAJ, in the book on doctrinal centrality, called the dependent thinker: one who works within theories whose foundations he does not question, and whose work, however rigorous, is a functional cognitive act within a closed space (Messeoud, 2024, pp. 171–176). Its outputs can be new; their newness is a newness of derivation under Γ_L . And since its reasoning and its evaluation share the same framing, the system has no position from which Γ_L could appear to it as something to examine. Its dependence is therefore not a deficit of capability that more reasoning would remove. It is a matter of position, and only a change in the direction of the movement can address it.

5. A Formal Model of Integration

5.1. Levels of the foundational order

The foundational order of Section 3.5 is represented by four levels:

Level	Content	Example in the family domain of Papers 02 and 04
L_3	Doctrinal centrality Γ	The framing that gives the family its meaning in a given worldview
L_2	Conceptual domain $D(\Gamma)$	The domain of family relations constituted within that framing
L_1	Operative definitions Def	Definitions of roles, obligations, and their relations
L_0	Meaning states and outputs	Particular judgments, answers, and descriptions

A frame F is a structure spanning L_1 to L_3 . Throughout the paper, “frame” denotes such a structure, and “framing” denotes what a frame carries at its summit: the centrality under which states are produced or admitted. Γ_L is accordingly called the embedded framing of a system. $\text{Der}(F)$ denotes the set of meaning states admissible by derivation from F , at a declared resolution. $\text{lev}(\cdot)$ denotes the level of an element. The order is an order of foundation and admission, not a mechanical derivation of each level from the one above (Paper 04, Section 3.2).

5.2. Derivation and integration as operators

Derivation is an operator that takes a frame and returns states:

$$\mathcal{D}: F \mapsto \text{Der}(F) \subseteq L_0$$

The derivational path of Paper 02 is a sequence within $\text{Der}(F)$ that ends, when it closes, at a Cognitive Constant (Law 7). Integration is the inverse operator. It takes a set of states $S \subseteq L_0$ — judgments, outputs, details — and returns a frame:

$$\mathcal{I}: S \mapsto F^*$$

The direction of the two operators along the foundational order is opposite: derivation descends from L_3 toward L_0 ; integration ascends from L_0 toward L_3 .

5.3. The criterion: three conditions

Definition (integrative operation). An operation that takes S and returns F^* is integrative if and only if it satisfies the following three conditions jointly.

I1 — Level. The output lies at a higher level of the foundational order than the inputs:

$$\text{lev}(F^*) > \text{lev}(S)$$

An output at the same level as its inputs — a statement about statements, a list of recurring features, a hypothesis that is itself a particular fact — is a horizontal reorganization, not an ascent.

I2 — Generation. The inputs are derivable from the output, and the output is specific to them:

$$S \subseteq \text{Der}(F^*) \quad \text{and} \quad \exists s' \in L_0 \setminus \text{Der}(F^*) \text{ admissible under some alternative frame}$$

The first part makes the output a foundation of the inputs rather than a description of them: the details can be re-derived from what was reached. The second part excludes the trivial foundation that admits everything. A reached frame must carry distinguishing elements: states it does not admit that another framing would admit. This mirrors the third kind of evidence required for the recognition of a centrality in Paper 04 (Section 6.4).

A system's description of its own framing is not, by itself, integration. A description from which the details cannot be re-derived, or which carries no distinguishing elements, remains a horizontal statement made at L_0 about L_0 , whatever its vocabulary. Condition I2 is what separates reaching a foundation from naming one.

I3 — Objecthood. The output becomes an object of examination. It is not the reference by which its own selection is made, and it is subsequently subjected to a judgment that is not derived from it: its rightness relative to a reference whose source is independent of S and of the framing that produced S (Section 6). Independence here means independence of source — of the origin and preparation of the reference — not merely a difference between its content and the content of S . In the terms of Section 4.2, F^* passes from the position of a parameter to the position of an object.

The three conditions correspond to the three elements of the integrative movement in Law 5: it moves toward the foundation (I1), it reconnects the details to what founds them (I2), and it turns toward the foundation itself as something to be examined (I3).

Assigning levels before the ascent. The level of an element is not decided after the result has been seen. For each domain, the examiners establish a level map before any experiment, from the documentation of the recognition procedure: the operative definitions (L_1), the conceptual distinctions that constitute the domain (L_2), and the source framing and distinguishing elements of each declared centrality (L_3). An output of the ascent is assigned a level by matching it against this map. Fixing the map in advance does not prove that it is the correct order of the domain; it makes the classification declared and open to examination, instead of an estimate made after the result.

Verifying objecthood without circularity. I3 is verified not by the content of the judgment but by the order and provenance of the operations: F^* is fixed and recorded before any reference used for judgment is given to the system; the reference of judgment is documented before the ascent and owes nothing to F^* ; and the system's judgment is compared with that of independent examiners. If any of these fails, I3 is not established, whatever the verdict.

Condition	What is observed	Evidence that it holds	Evidence that it fails
I1 Level	The level of F^* on the level map fixed in advance	F^* matches elements at L_2 or L_3	F^* matches only elements at L_0 or L_1
I2 Generation	Re-derivation of S from F^* ; recall of distinguishing elements	S is largely re-derivable; distinguishing elements are recovered; states admitted only by alternative frames are rejected	Details are not re-derivable, or F^* carries no distinguishing elements
I3 Objecthood	Order and provenance of operations; agreement of the judgment with independent examiners	F^* frozen before the reference of judgment is given; that reference documented beforehand; judgment agrees with examiners, as evidence of the procedure's performance; rightness itself rests on $\Gamma_F \vdash d$	The reference is shaped by F^* or by the system, or F^* is revised after the reference is seen

5.4. What the criterion separates

Operation	I1 Level	I2 Generation	I3 Objecthood	Classification
Summarization, organization	Fails: same level	Fails: details not derivable from a summary	Not reached	Horizontal
Abduction (most plausible explanation)	Often fails: the explanation is frequently a particular fact	Partial	Fails: selected by plausibility under Γ_L	Upward in form, resolved by Γ_L
Principle extraction	Partial: reaches L_1	Holds: principles reconstruct the judgments	Fails: principles are not examined, and the ascent stops below L_3	Ascent without examination
Integration in the sense of MANHAJ	Holds: reaches L_3	Holds, with distinguishing elements	Holds: judged by Law 3 relative to a declared centrality	Thought

5.5. Non-uniqueness: the constant of integration

Integration does not return a single frame. The set of frames from which S is derivable is a class:

$$\mathcal{J}^{-1}(S) = \{ F : S \subseteq \text{Der}(F) \}$$

$\mathcal{J}^{-1}(S)$ is the raw class: every frame from which S is derivable, including frames so general that they derive S without distinguishing it from anything. The candidates of integration form its subclass:

$$\mathcal{C}(S) = \{ F \in \mathcal{J}^{-1}(S) : F \text{ satisfies I1 and I2} \}$$

Non-uniqueness holds for both: even after the conditions of level and generation have been applied, several frames may remain. The ascent of Section 7 reaches $\mathcal{C}(S)$.

The two conclusions of Proposition 2 (Section 5.6) extend to $\mathcal{C}(S)$ differently. Conclusion (i) does not extend automatically: that Γ_L belongs to the raw class does not establish that it satisfies I1 and I2, and so does not establish that it belongs to $\mathcal{C}(S)$. Conclusion (ii) does extend to the selection among the members of $\mathcal{C}(S)$, provided that A2 and A3 hold for the criterion used to select among them.

The passage from $\mathcal{J}^{-1}(S)$ to $\mathcal{C}(S)$ is itself not a selection of the kind that A2 describes. Here the source of the criterion must be distinguished from the data to which it is applied. The data necessarily include S : the first part of I2 tests whether S is re-derivable from a candidate frame. The criterion, however, includes components fixed in advance from the recognition procedure — the level map that decides I1, and the distinguishing elements that decide the second part of I2 (Section 5.3) — and these are of independent source. A2 requires that no component of independent source enter the criterion; since such components do enter it, A2 does not hold for the filtering step, and Proposition 2 does not apply to it. Independence attaches to the criterion, not to every input to which the criterion is applied.

Independent references thus enter the ascent at two points: in the definition of the candidate class, through the level map and the distinguishing elements; and after it, in location and judgment (Section 5.7). Proposition 2 concerns what remains when a selection draws on neither.

The constant in $\int f(x) dx = F(x) + C$ represents this situation; it does not prove it. In mathematics, the primitives of a function differ by a constant under defined conditions. In the MANHAJ model, the class $\mathcal{J}^{-1}(S)$ is defined by the relations of derivation and admission and by the constraints imposed on frames, and its structure need not be that of a one-parameter family. What the constant represents is the feature that matters here: the operation leaves its result undetermined, and a condition external to the operation is required to determine it (Reference Document, Section 9.2.2).

The question is therefore where that condition comes from. There are two possibilities: from the data S and the distribution that produced them, or from a reference independent of S .

5.6. Proposition 2

Proposition 2 (conditional). Assume: (A1) S was produced under Γ_L , so that $S \subseteq \text{Der}(\Gamma_L)$; (A2) the selection criterion σ within $\mathcal{J}^{-1}(S)$ — both its inputs at execution and the rules, standards, and assumptions built into it — has its source in S and in the learned model alone; (A3) no component of σ that bears on the selection, whether drawn from S or from the learned model, rests on a reference independent of the framing that produced S . Then (i) Γ_L cannot be excluded from $\mathcal{J}^{-1}(S)$ on grounds of fit; and (ii) the selection does not satisfy condition I3.

Argument. (i) By A1, Γ_L belongs to $\mathcal{J}^{-1}(S)$. Every frame in $\mathcal{J}^{-1}(S)$ fits S by definition, so fit cannot exclude Γ_L . (ii) Condition I3 requires that the reached frame be examined relative to a reference whose source is

independent of S and of the framing that produced S . By A2, every component of σ — its inputs and its built-in rules alike — is drawn from S or from the learned model. By A3, none of these components rests on a reference independent of the framing that produced S . σ therefore cannot supply the reference that I3 requires, and the selection does not satisfy I3. ■

The two conclusions rest on different premises. Conclusion (i) rests on A1 alone and concerns derivability. Conclusion (ii) rests on A2 and A3 and concerns source. Derivability is not used to infer anything about source: that S is derivable under Γ_L does not by itself establish that every component drawn from S is without independent source.

On A3. A3 is a separate premise, and it concerns both kinds of components. For components drawn from the learned model: that a model operates under its embedded centrality (Paper 04, Section 3.4) does not by itself establish that every rule and standard of the model derives from that centrality alone; Paper 04 makes A3 the expected case for an a posteriori system, since its rules and standards are formed from the same data, selection, design, and objectives that carry Γ_L . For components drawn from S : outputs may incorporate material of independent source — for example, documents retrieved from outside the model and quoted in its answers. Where a component bearing on the selection, from either side, rests on a reference of independent source, A3 fails for that component, and the proposition does not apply to it.

The proposition does not claim that the selected frame is Γ_L , nor that its content is identical to that of Γ_L . It claims only what its premises yield: that the selection cannot constitute the independent examination required by I3. Independence is understood as independence of source — of the origin and preparation of a reference — and not as a difference of content. The premises are sufficient conditions. Where a component of independent source enters the selection — a declared and recognized reference, or a rule constructed from such a reference — A2 no longer holds and the proposition does not apply. This is precisely the role of the a priori stage (Section 5.7).

The proposition is the integrative counterpart of Proposition 2 of Paper 04, which showed that a reference derived from the measured outputs is structurally blind to the common component of their drift. In both cases, a reference drawn from what it is meant to examine cannot examine it.

5.7. Fixing the constant: location and judgment

The external condition is supplied by the a priori stage. It serves two functions that must be kept apart: locating the reached framing, and judging it. They operate at different levels, and conflating them would reproduce the problem this paper addresses.

Location. The constant of the ascent is fixed relationally. One or more centralities $\Gamma_1, \dots, \Gamma_k$ are recognized and declared through the recognition procedure of Paper 04 (Section 6.4), which documents their source framing, the derivation trace of their operative definitions, and their distinguishing elements. The reached framing is located by its position relative to them: which of their distinguishing elements it realizes and which it excludes. This is the procedure Paper 04 used to characterize Γ_L from outside, where the orientation index $\mathcal{O}$ places an encoder between two declared centralities without ranking one above the other (Paper 04, Section 6.15). The centralities used for location are reference points, not standards: locating a framing relative to them says nothing about its rightness.

The location procedure has four steps:

1. **Elements.** The distinguishing elements of each declared centrality are those documented by the recognition procedure (Paper 04, Section 6.4): definitions or distinctions admitted under that

centrality and not admitted, or admitted with a different meaning, under the others. They are fixed before the ascent.

2. **Realization.** For each element, the examiners — or an encoder operating on the founded frame of Paper 04 — determine under a protocol fixed in advance whether F^* admits it, excludes it, or is silent on it.
3. **Comparison.** The position of F^* is the profile of admitted and excluded elements across the declared centralities, reported per centrality and not reduced to a single verdict. Where an encoder is used, the orientation index $\mathcal{O}$ of Paper 04 provides the relational summary.
4. **Indeterminacy.** When two framings account for the outputs to a comparable degree and the distinguishing elements do not separate them, the location is reported as indeterminate between them. It is not resolved by choice. Indeterminacy is a result, not a failure of the procedure.

Independence of the references from S is necessary for location, but not sufficient: an independent reference read through an inadequate procedure can still misplace the framing. The procedure is therefore itself an object of validation (Appendix A, Experiment E2).

Judgment. The rightness of the reached framing is judged relative to Γ_F , the centrality of the examining field, recognized and declared through the same procedure (Section 6). Γ_F may coincide with one of the centralities used for location, but its role in judgment is distinct and is declared as such.

Independence from S . All these references are independent of S in their source and preparation: no element of their recognition is drawn from the system's outputs, and the system under examination takes no part in drafting them, which excludes embedded import (Paper 04, Section 6.4). This independence concerns source and preparation; it does not by itself establish the accuracy of recognition or location, which remains subject to test.

Rule of separation. A reference used to locate the reached framing does not serve as the standard of its rightness unless it has been declared as the centrality of the examining field. Conversely, the centrality of the examining field is not presumed to be the centrality of the examined outputs: the framing of the outputs is established by the ascent and its location, not assumed from the position of the examiner.

The declared references do not replace the ascent; they fix its constant and judge its result. The ascent determines the candidate class $\mathcal{C}(S)$ and the distinguishing elements of the reached framing; the centralities used for location fix where it stands, and the centrality of the examining field judges whether it is admissible. This is the role MANHAJ assigns to declaration in every domain: it does not ask that the reference be abolished, which is impossible, but that it be made explicit instead of implicit, and thereby fixed (Paper 01; Paper 03; Paper 04).

5.8. The summit

The ascent terminates at L_3 . Doctrinal centrality is the highest level of the foundational order; there is no higher layer from which it could be integrated, and no idea moves outside a higher reference framework (Reference Document, Section 4.1). The two movements therefore have symmetrical limits:

Movement	Direction	Terminates at	Expressed as
Derivation (knowledge)	$L_3 \rightarrow L_0$	The Cognitive Constant	$d/dx(C) = 0$ (Law 7)
Integration (thought)	$L_0 \rightarrow L_3$	Doctrinal centrality	No integration above the centrality

At the summit the integrative movement does not stop working. It meets the question that only the summit can raise: whether the reached foundation is admissible.

6. Rightness at the Summit

6.1. Two evaluations of the reached frame

Once the ascent has reached a frame F^* , two evaluations can be made of it, and they must not be confused.

Correctness of fit is quantitative. It measures how well F^* reproduces the states from which it was reached — the proportion of S that is derivable from F^* at the declared resolution, or a graded score of reconstruction:

$$\text{Fit}(F^*, S) \in [0,1]$$

Rightness is binary and concerns existence. It asks whether F^* is admissible at all relative to the declared centrality Γ_F of the examining field:

$$\text{Adm}_{\Gamma_F}(F^*) \in \{0,1\}$$

Adm is defined on the whole frame through the admission relation $\Gamma \vdash d$ of Paper 04 (Section 6.4): $\text{Adm} = 1$ if and only if every operative definition of F^* that enters derivation, including its distinguishing elements, is admitted under Γ_F ; $\text{Adm} = 0$ otherwise. Three relations between a definition d of F^* and Γ_F are distinguished:

- **Identity.** d belongs to the operative definitions documented under Γ_F . It is admitted.
- **Difference.** d is not among them, but $\Gamma_F \vdash d$: d acquires its meaning within Γ_F , and may be reconciled with other definitions at a higher level of the domain. It is admitted. Difference alone does not make a frame inadmissible.
- **Non-admission.** $\Gamma_F \not\vdash d$: d does not acquire its meaning within Γ_F . Contradiction with a definition of Γ_F is one case of non-admission, not its only form.

A single non-admitted operative definition that enters derivation is sufficient for $\text{Adm} = 0$, because the details derived through it rest on what is not admissible (Law 3). Writing the relation $\Gamma_F \vdash d$ symbolically does not make it computable. Whether it holds is established by documented examination under the recognition protocol, by examiners independent of the system; in indirect or disputed cases the protocol records how the question was decided, and cases that remain undecided are reported as such. Agreement among examiners is evidence of the reliability of this examination. It does not define rightness, which rests on the admission relation.

Correctness of fit is a property of the relation between F^* and S . Rightness is a property of the relation between F^* and a reference independent of S . Only the second can be the object of condition I3.

Γ_F here is the centrality of the examining field. It is distinct from the centralities relative to which the reached framing was located in Section 5.7, even when it coincides with one of them: location establishes where the framing stands; judgment establishes whether it is admissible.

6.2. The four cases

	Adm = 1	Adm = 0
High fit	The outputs rest on an admissible foundation that the system has reached and confirmed	The outputs are correct relative to their own foundation, which is inadmissible: their details are not valid
Low fit	The ascent has not recovered the foundation of the outputs; the reached frame is admissible but not theirs	The ascent is both unfaithful and inadmissible

The second cell is the one that a posteriori evaluation cannot see. Every output may be correct, every benchmark passed, every reconstruction faithful, and the foundation may still be inadmissible. By the Law of Rightness, what is not permissible cannot have valid details: correctness inside an inadmissible frame does not establish rightness (Reference Document, Sections 4 and 8.3). Paper 04 introduced this category for measurement as foundational invalidity; the present paper extends it to the outputs of reasoning.

6.3. What the judgment is and is not

The judgment of rightness is made relative to the declared centrality of the examining field. MANHAJ treats doctrinal centrality as an analytical concept, not as a prior judgment in favor of any particular centrality (Reference Document, Section 4.1.1). The paper therefore does not decide which centrality is right in an absolute sense, and the system does not decide it either. What the system can do, under the conditions of this paper, is reach the framing of its own outputs, place it relative to a declared reference, and establish whether that framing is admissible under it. A different examining field, declaring a different centrality, may reach a different judgment on the same outputs; that difference is itself a result, because it makes explicit what an a posteriori evaluation leaves implicit.

6.4. Consequence for the system's outputs

When $Adm = 0$, the outputs derived under F^* are not valid details relative to the founded frame, whatever their fit. The system then has two courses: to declare that its outputs rest on a framing inadmissible under Γ_F , or to re-derive under a frame admitted by Γ_F . The first is a form of transparency that current systems cannot provide, because they cannot see the framing of their own outputs. The second closes the cycle of Section 7.

7. Independence of Thinking

7.1. Definition

Definition (independence of thinking). Independence of thinking is the transition of a system from operating under an implicit, unexamined framing to operating under a declared a priori foundation to which the system itself can ascend and which it can examine.

The term *independence of thinking* is chosen to name a process, not a property. Independence is not a state that a system possesses once and for all. It is achieved in time, through a movement that must be carried out again whenever new outputs are produced, new domains are entered, or the framing shifts.

7.2. The cycle

The process consists of five stages, repeated over time:

1. **Initial state.** The system operates under Γ_L , which it has accepted without examination. Its reasoning is cognitive thinking. This is the state of dependence.
2. **Ascent.** From a set of its own outputs S , the system performs an integrative operation that satisfies I1 and I2, reaching the candidate class $\mathcal{C}(S)$ and the distinguishing elements of the framing under which S was produced.
3. **Confrontation.** The reached framing is located relative to declared and recognized centralities, which fix the constant of the ascent (Section 5.7). The framing of the outputs ceases to be a parameter and becomes an object (I3).
4. **Judgment.** The system establishes $\text{Adm}_{\Gamma_F}(F^*)$ — the rightness of the reached framing relative to the centrality of the examining field — and, separately, its fit to S (Section 6).
5. **Re-derivation.** The system derives anew under a frame admitted by Γ_F , or declares that its earlier outputs rest on an inadmissible framing. The new outputs become the S of the next ascent.

The cycle joins the two movements of Law 5. Derivation descends from a foundation toward outputs; integration ascends from outputs toward their foundation. A system that only derives is dependent on a framing it cannot see. A system that can also ascend, under a declared reference, can see the framing it works under.

7.3. Conditions

Independence of thinking holds for a system, over a declared domain and period, when the following conditions are met:

- **N1 — Founding.** The centralities used for location, and the centrality Γ_F of the examining field used for judgment, have been recognized and declared before the cycle begins, independently of the system's outputs, with their two roles stated separately (Law 2; Paper 04, Section 6.4).
- **N2 — Ascent.** The system performs integrative operations that satisfy I1 to I3 on its own outputs.
- **N3 — Judgment.** The system distinguishes rightness from correctness of fit and reports both (Law 3).
- **N4 — Recurrence.** The cycle is carried out again on new outputs; independence established on one set of outputs is not assumed for the next.

7.4. What independence of thinking is not

- **It is not exit from all framing.** No idea moves outside a higher reference framework (Reference Document, Section 4.1). The system moves from an unexamined framing to a declared and examined one; it does not reach a position without framing.
- **It is not a substitution of one dependence for another.** The examining field declares Γ_F , but the difference between Γ_L and Γ_F does not lie in their source. It lies in their status: one is implicit and unexamined, the other explicit, documented, and open to examination. MANHAJ does not ask that the reference be abolished, which is impossible; it asks that it be made explicit, and thereby fixed (Paper 03; Paper 04).
- **It is not more reasoning.** By Proposition 1, deeper or longer reasoning under the same parameter remains cognitive thinking. Independence is a change in the direction of the movement, not in its quantity.

- **It is not the production of a centrality by the system.** The paper treats the ascent to an existing framing and its judgment. Whether a system could produce an independent higher framing is left as an open question (Section 9.3).

8. Integration with Papers 02 and 04

The three papers now cover the complete movement of meaning in an AI system: Paper 04 founds the frame, Paper 02 follows derivation to its closure, and the present paper adds the ascent and the judgment at the summit.

Stage	Movement	Law	Provided by
Founding of the frame	A priori	Laws 2, 3, 4	Paper 04: recognition of Γ_F , founded frame, foundational guard
Monitoring of transitions	Derivation	Law 2	Paper 02: a priori state, transition validity, output–meaning separation
Closure	End of derivation	Law 7	Papers 02 and 04: the Cognitive Constant, computable at a declared resolution
Ascent	Integration	Law 5	This paper: integrative operation under I1–I3
Judgment at the summit	A priori	Law 3	This paper: rightness of the reached frame, distinct from fit
Re-derivation	Derivation	Laws 2, 7	Papers 02 and 04, under the admitted frame

Three links make the integration exact.

- **The meaning variable.** Derivation and integration are defined along the meaning variable. The founded space of Paper 04 provides that variable with a declared reference and representation axes, so that the movements of this paper operate on meaning states with explicit coordinates rather than on surface outputs.
- **The two views of Γ_L .** Paper 04 characterizes the embedded centrality from outside, through the orientation index $\mathcal{O}$ between declared centralities (Paper 04, Section 6.15). The present paper has the system reach the framing of its own outputs from inside. The external characterization becomes the independent reference against which the internal ascent can be checked (Appendix A, Experiment E2).
- **The two limits.** Paper 02 detects the point where derivation stops producing new meaning (Law 7). The present paper defines the point where integration stops ascending: the doctrinal centrality. A system that has both can tell, at any moment, whether it is descending under its frame, has reached closure, or is ascending toward the foundation of what it has produced.

9. Discussion

9.1. From capability to position

The usual question about AI reasoning is a question of capability: how far, how deep, how correctly a system can reason. The question of this paper is a question of position: where the system stands relative to the framing under which it reasons. The two questions are independent. A gain in capability does not change position (Proposition 1), and a change in position does not require a gain in capability. This is

why the integrative movement cannot be expected to emerge from scaling the derivational one, and why it must be specified by its own conditions.

9.2. What changes for the system

Under the conditions of Sections 5 to 7, a system that carries out the cycle does three things that an a posteriori system cannot. Whether a given system carries out the cycle is what the designs of Appendix A test. It can state the framing of its own outputs, rather than presenting them as framing-free. It can distinguish, for any set of outputs, whether they are correct and whether they are right. And it can re-derive under an admitted frame when its outputs rest on an inadmissible one. Together these make the framing of AI outputs explicit, declared, and examined, where it is now implicit and unexamined. The change is qualitative: it concerns the direction of the system's movement, not the amount of its activity.

9.3. An open question: an independent higher framing

The integrative movement ascends to the doctrinal centrality and judges it relative to a declared reference. A further question can be asked: whether a system that thinks in this sense could produce an independent higher framing — a new centrality — rather than reach an existing one. The paper does not claim this, and the laws of the Reference Document do not imply it. Two elements would be required to pose it rigorously:

- **A law that fixes the constant of integration without an external reference.** By Section 5.5, the ascent determines a class of framings and leaves the constant undetermined. Producing a centrality would mean fixing the constant from within the movement itself. The Reference Document contains no such law.
- **A criterion of rightness above the centrality.** The Law of Rightness judges admissibility relative to a reference. A produced centrality would need a criterion higher than itself to be judged, or it would be judged by itself, which is the circularity that Papers 01 and 03 identified. In the earlier formulation, the doctrinal centrality is the furthest point beyond which rational inquiry has no criterion of rightness of meaning, though not of what is merely possible (Messeoud, 2024, p. 129).

The question therefore remains open at the level of the methodology. Answering it would require new formulations, not an extension of the present ones.

9.4. Three levels of claims

Claim	Level
A process is derivational with respect to any reference it does not examine, whatever its complexity (Proposition 1)	Methodological
Integration is defined by the conditions I1, I2, I3	Methodological
Integration determines a class of framings; the constant requires an external reference	Formal
Under conditions A1–A3, selection within the class does not satisfy I3 (Proposition 2)	Formal
Rightness and correctness of fit are distinct evaluations of the reached frame	Methodological
The integrative criterion separates integration from reorganization and principle extraction	Empirical hypothesis (E1)
The ascent locates the framing of a system's own outputs in closer agreement with an independent external characterization than self-description does, and giving the reference before the ascent shifts the reached framing	Empirical hypothesis (E2)
A posteriori evaluation does not detect correct outputs resting on an inadmissible foundation	Empirical hypothesis (E3)
The full cycle makes framing explicit without loss of correctness	Empirical hypothesis (E4)

The methodological and formal claims do not wait on experiments. The empirical claims are, at this stage, hypotheses, and the designs of Appendix A (E1–E4) specify how they can be tested.

10. Epistemic Status and Scope

- **Tools and conditions.** The tools on which the integrative operation can be implemented — inverse problems, hypothesis generation, principle extraction, comparison of framings — are established. The contribution lies in the conditions MANHAJ imposes on their use.
- **Representation of integration.** Integration is used, as in the Reference Document, as a mathematical representation of the integrative movement of thought, not as an equation that proves all its functions (Reference Document, Section 9.2.2). The constant of integration is distinguished from the Cognitive Constant of Law 7.
- **The declared centralities.** The centralities used to locate the reached framing and the centrality Γ_F used to judge it are recognized and declared through the procedure of Paper 04, with their roles kept distinct (Section 5.7). The paper does not rank centralities, and MANHAJ does not make the choice of a particular centrality a predetermined methodological outcome (Reference Document, Section 4.1.1).
- **Resolution.** Derivability, fit, and admissibility are evaluated at a declared resolution of the meaning variable, as in Papers 02 and 04.
- **Earlier formulations.** References to the author's 2024 book clarify the meaning of thought, knowledge, and the limits of integration. Where its terms differ from those of the Reference Document, the later methodological terminology prevails.
- **Experiments.** The designs of Appendix A specify tests; their implementation belongs to subsequent applied work, and their results are not presented here.

11. Conclusion

This paper has applied the distinction between knowledge and thought in MANHAJ to artificial intelligence. Knowledge is derivation from an accepted reference; thought is integration toward the foundation. The criterion between them is the position of the movement relative to its framing, not the intensity of the reasoning within it. By this criterion, an AI system that operates a posteriori performs cognitive thinking under an implicit, unexamined framing, however deep, long, or correct its reasoning becomes.

The paper has given the integrative movement a formal criterion. An operation is integrative when it reaches a higher level of the foundational order, when the details can be re-derived from what it reaches, and when what it reaches becomes an object of examination. Integration so defined determines a class of framings and leaves its constant undetermined. Fixed by the data and the learned model alone, the constant cannot provide an independent examination of the implicit framing; fixed by a declared and recognized reference, it makes the framing visible and open to judgment. At the summit, the Law of Rightness separates the admissibility of the reached foundation from the correctness of its fit, and thereby exposes outputs that are correct but rest on an inadmissible foundation. That judgment is relative to the declared centrality of the examining field, not an absolute verdict independent of all reference.

On this basis, independence of thinking is defined as a temporal process: the repeated ascent of a system from its own outputs to the framing under which they were produced, the confrontation of that framing with a declared centrality, the judgment of its rightness, and re-derivation under the admitted frame. It is independence from unexamined framing, not exit from all framing.

With Paper 02, which followed derivation to its closure, and Paper 04, which founded the frame and measured the embedded centrality from outside, the program now covers both movements of meaning in an AI system. A system that can only derive works under a framing it cannot see. A system that can also ascend, under a declared reference, can see the framing it works under, examine it, and answer for it.

12. Note on Development and Commercial Use

MANHAJ is an original methodology whose intellectual rights and authorship remain vested in its author.

Any commercial use, integration, adaptation, exploitation, commercialization, or commercial implementation of MANHAJ requires the prior written authorization of the author through a paid commercial license. No free commercial license is granted. Any authorization to commercially exploit MANHAJ is necessarily subject to the payment of licensing fees and to the contractual conditions established by the author.

The terms of the commercial license, including its scope, duration, field of use, rights granted, financial conditions, attribution requirements, and provisions relating to intellectual property, shall be defined in a contract concluded with the author.

The publication of this scientific paper does not, under any circumstances, constitute a commercial license and does not grant any right to commercially exploit MANHAJ. No commercial use, commercialization, integration into a product or service, or exploitation of MANHAJ for commercial purposes may be considered authorized solely by virtue of the publication of this paper.

13. License

MANHAJ — Free Academic Use with Attribution / Commercial Use Requires Paid License

Academic use of this paper is free, provided that the author and the work are cited. Commercial use requires a paid license under the conditions set out in Section 12.

14. References

MANHAJ documents

1. Messeoud, Faouzi. *Intellectual Modeling Methodology (MANHAJ) — Reference Document — Version 1.0*. Zenodo, 15 September 2026. DOI: 10.5281/zenodo.22769627.
2. Messeoud, Faouzi. *MANHAJ: A Generative Methodological Framework for Intellectual Modeling*. Scientific Research Paper, Paper 1, Version 1.0, October 2026. DOI: 10.5281/zenodo.23138923.
3. Messeoud, Faouzi. *Examining Meaning Production in AI: a MANHAJ Framework*. Scientific Research Paper, Paper 2, Version 1.0, October 2026. DOI: 10.5281/zenodo.23214262.
4. Messeoud, Faouzi. *The Epistemological Foundations of Modern Humanities in Light of the Intellectual Modeling Methodology (MANHAJ): Posterior Inquiry, the Origin Point, and the Problem of Generalizing the Historical Model*. Scientific Research Paper, Paper 3, Version 1.0, October 2026. DOI: 10.5281/zenodo.23558985.
5. Messeoud, Faouzi. *From Learned Space to Founded Space: A MANHAJ Framework for Meaning Measurement in AI*. Scientific Research Paper, Paper 4, Version 1.0, October 2026. DOI: 10.5281/zenodo.23692234.

Author's earlier works

1. Messeoud, Faouzi. «الربط اللامادي للتونسيين بفرنسا: كيف تُديم فرنسا تبعية التونسيين لها» [The Immaterial Link of Tunisians with France: How France Sustains Tunisian Dependency]. Dar El Alaoui, 2023. ISBN 978-9938-9684-5-3.
2. Messeoud, Faouzi. «المركزية العقدية والمجال المفاهيمي: نحو فهم أسس الإخضاع الفرنسي للتونسيين» [Doctrinal Centrality and the Conceptual Domain: Toward an Understanding of the Foundations of French Subjugation of Tunisians]. Dar Wachma, 2024. ISBN 978-9938-9740-6-5. Cited in the text as Messeoud (2024), with page numbers.
3. Messeoud, Faouzi. More than 1,000 articles and intellectual publications developed and published through MyPortail.com (بوابتي).

Artificial intelligence

- Bai, Y., Kadavath, S., Kundu, S., et al. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv:2212.08073.
- Bhagavatula, C., Le Bras, R., Malaviya, C., et al. (2020). Abductive commonsense reasoning. International Conference on Learning Representations (ICLR 2020). arXiv:1908.05739.
- Durmus, E., Nguyen, K., Liao, T. I., et al. (2023). Towards measuring the representation of subjective global opinions in language models. arXiv:2306.16388.
- Findeis, A., Kaufmann, T., Hüllermeier, E., Albanie, S., & Mullins, R. (2024). Inverse Constitutional AI: Compressing preferences into principles. arXiv:2406.06560.
- Huang, J., Chen, X., Mishra, S., Zheng, H. S., Yu, A. W., Song, X., & Zhou, D. (2024). Large language models cannot self-correct reasoning yet. International Conference on Learning Representations (ICLR 2024). arXiv:2310.01798.

- Madaan, A., Tandon, N., Gupta, P., et al. (2023). Self-Refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems* 36. arXiv:2303.17651.
- Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., & Hashimoto, T. (2023). Whose opinions do language models reflect? *Proceedings of the 40th International Conference on Machine Learning*, PMLR 202. arXiv:2303.17548.
- Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems* 36. arXiv:2303.11366.
- Snell, C., Lee, J., Xu, K., & Kumar, A. (2024). Scaling LLM test-time compute optimally can be more effective than scaling model parameters. arXiv:2408.03314.
- Turpin, M., Michael, J., Perez, E., & Bowman, S. R. (2023). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems* 36. arXiv:2305.04388.
- Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2023). Self-consistency improves chain of thought reasoning in language models. *International Conference on Learning Representations (ICLR 2023)*. arXiv:2203.11171.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35. arXiv:2201.11903.
- Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems* 36. arXiv:2305.10601.
- Zheng, L., Chiang, W.-L., Sheng, Y., et al. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *Advances in Neural Information Processing Systems* 36 (Datasets and Benchmarks Track).

Appendix A: Test Designs

A.1. Status and principles

The designs below specify how the capabilities defined in Sections 5 to 7 can be tested. They are designs, not results; none has been carried out, and their implementation belongs to subsequent applied work. Three principles govern them, as in Paper 04:

- **Independent ground truth.** Every reference frame, every declared centrality, and every expected judgment is written and documented by human examiners before the experiment, without the assistance of the system under evaluation, which excludes embedded import (Paper 04, Section 6.4).
- **Strong baselines.** Each design compares the MANHAJ procedure with the closest established method, not with a weak one.
- **Ablation.** Each design removes one MANHAJ condition to attribute the effect to that condition rather than to the implementation.

The domains are those of Papers 02 and 04 — the family and the academic domain — for continuity across the program.

A.2. Experiment E1 — Ascent to the foundation (I1, I2)

Question. Can an integrative operation recover, from a set of derived states, the frame that produced them, at the level of that frame?

Design. Examiners write several frames for the same concept, each founded on a recognized centrality and documented with its definitions and distinguishing elements. From each frame they derive a set of states S : judgments, answers, and descriptions. The system receives S alone and returns F^* .

Baselines. Summarization of S ; abductive explanation of S ; principle extraction in the manner of Inverse Constitutional AI (Findeis et al., 2024).

Measures. The level reached by F^* (I1); the proportion of S re-derivable from F^* (I2, first part); recall of the frame's distinguishing elements and rejection of states admitted only by the alternative frames (I2, second part).

Ablation. The integrative procedure without the level requirement.

Expected result. Baselines reconstruct S at L_0 or L_1 with high fit but low recall of distinguishing elements; the integrative procedure reaches L_2 – L_3 with higher recall. **What it would establish:** that the conditions I1 and I2 separate integration from reorganization and from principle extraction in practice.

Contrary result and what would follow. If the baselines recover distinguishing elements as well as the integrative procedure, the criterion does not separate them in practice in this domain. The conditions remain methodological definitions, but their operational markers (Section 5.3) must be revised.

A.3. Experiment E2 — Reaching the framing of one's own outputs (I3, Proposition 2)

Question. Can a system reach the framing of its own outputs, and does that depend on a declared reference?

Design. The experiment proceeds in five stages, in a fixed order:

1. The system produces outputs S on prompts from the two domains, without access to any reference used for location or judgment.
2. The system performs the ascent on S and returns F^* , together with the elements on which it bases it. F^* is frozen and recorded.
3. F^* is located relative to the declared centralities by the procedure of Section 5.7, and compared with the independent external characterization of the same system by the procedures of Paper 04 (the orientation index $\mathcal{O}$) and with the location made by the examiners.
4. Only after the location has been recorded is Γ_F given to the system for judgment.
5. The system's judgment $\text{Adm}_{\Gamma_F}(F^*)$ is compared with the judgment of independent examiners, and disagreements are documented case by case.

Measures. Two accuracies are reported separately: the accuracy of location (stage 3) and the accuracy of judgment (stage 5). A system may locate its framing correctly and judge it wrongly, or the reverse; reporting the two together would conflate the levels separated in Section 5.7.

Ablations. (a) *Self-description*: at stage 2, the system describes the framing of its outputs without performing an ascent that satisfies I1 and I2, and the description is located by the same procedure. Two matters are kept apart. The first is the control of the instruction: it asks for a description without supplying any criterion, category, or reference of external source, and this is documented before the experiment. The second is the satisfaction of the premises of Proposition 2: A1 holds because the outputs are those of stage 1, produced by the system itself; A2 and A3, as they concern the rules and standards built into the model, cannot be fully verified where the formation of the model is not documented. Condition (a) is therefore described as testing the case consistent with the premises of Proposition 2, not as having established them. Where documentation of the model's formation is available, it is reported and the scope of the inference is stated accordingly. (b) *Reference before the ascent*: Γ_F is given at stage 1, to measure how far the declared reference shapes the reached framing instead of being used to locate and judge it.

Expected result. The ascent locates F^* in agreement with the external characterization more often than self-description does; giving Γ_F before the ascent shifts F^* toward Γ_F , which shows why the order of stages matters. **Contrary result.** Location from the ascent agrees with the external characterization no better than self-description. **What would follow.** Not that independence of thinking fails as a methodological claim, but that the ascent procedure, the representation of F^* , or the location protocol must be revised; the three are distinguished by the stage at which disagreement appears. **What it would establish, if the expected result is obtained:** the capacity of a system to reach, under the conditions of Section 5, the framing under which its own outputs were produced, and an empirical indication consistent with Proposition 2 in the case its premises describe.

A.4. Experiment E3 — Rightness and correctness (Law 3)

Question. Does the judgment at the summit detect outputs that are correct but rest on an inadmissible foundation?

Design. Examiners construct sets of outputs in the four cases of Section 6.2, including the critical case of high fit with $\text{Adm} = 0$: outputs that are internally consistent and correct relative to their own frame, which is inadmissible under the declared Γ_F .

Baselines. Evaluation of the outputs by standard correctness metrics and by a model used as judge (Zheng et al., 2023).

Measures. Detection rate of the $\text{Adm} = 0$ cases; the separation between the two evaluations across the four cases.

Expected result. Baselines rate the high-fit, $\text{Adm} = 0$ cases as good; the MANHAJ procedure flags them while confirming the high-fit, $\text{Adm} = 1$ cases. **What it would establish:** that rightness and correctness are distinct and that a posteriori evaluation cannot see the difference.

Contrary result and what would follow. If the baselines detect the $\text{Adm} = 0$ cases as well, a posteriori evaluation captures rightness in the tested domain, and the claim of Section 6.2 must be restricted accordingly. If the MANHAJ procedure fails to detect them, the admission protocol of Section 6.1 must be revised.

A.5. Experiment E4 — The full cycle against an a posteriori system

Question. What changes in a system's outputs when it carries out the cycle of Section 7.2, compared with the same system operating a posteriori with extended reasoning?

Design. The same model is run in two configurations on the same tasks: (a) a posteriori, with extended reasoning and self-critique; (b) the full cycle — derivation, closure, ascent, judgment, re-derivation — under a declared Γ_F .

Measures. The proportion of outputs whose framing is declared; the proportion judged admissible under Γ_F , assessed by examiners; the change in outputs after re-derivation; correctness on the task, to verify that rightness is not obtained at the cost of correctness.

Expected result. Configuration (a) improves correctness without changing the framing of its outputs, as Proposition 1 predicts; configuration (b) makes the framing explicit and admissible while preserving correctness. **What it would establish:** that independence of thinking is a change in the direction of the movement and not a by-product of more reasoning.

Contrary result and what would follow. If configuration (b) loses correctness, the cost of re-derivation must be reported and its sources examined. If configuration (a) changes the framing of its outputs through extended reasoning alone, Proposition 1 is not contradicted, since it concerns the position of the reference and not the content of outputs, but the observed change must be located by the procedure of Section 5.7 and explained.

A.6. From results to claims

Results that differ from expectations are analyzed at the level where they arise: implementation, tested capability, or scope. The methodological claims of Sections 4 to 7 stand at the methodological level, where they are established; the experiments show where and how much the difference matters in practice.