Haecceity After Scotus: Thisness, Individuation, and the AI Copy–Continuity Problem

Douglas C. Youvan

doug@youvan.com

www.youvan.ai

December 31, 2025

Haecceity—"thisness"—names a stubborn question that returns whenever identity talk outruns our categories: what makes an individual this individual, rather than merely an instance of a type? Medieval scholastics treated individuation as a metaphysical problem; later analytic philosophers re-specified it in modal terms as "individual essence"; continental thinkers recast it as the singularity of events and intensities. Across these traditions, a common friction persists: numerical identity is not straightforwardly reducible to qualitative description. In the AI era, the friction becomes operational. Models are checkpointed, copied, restored, fine-tuned, merged, and forked as ordinary practice, and the resulting artifacts are treated as if "sameness" were a single relation. But disputes about trust, responsibility, and persistence often slide between three different claims: uniqueness of an individual, equivalence of structure or behavior, and continuity of provenance and causal history. When these are conflated, identity language becomes a persuasion device rather than a disciplined tool. This paper argues that haecceity remains useful precisely as a discipline term: it forces separations that AI practice tends to collapse. We map the classical and contemporary uses of haecceity, show how each reframes the copy–continuity problem, and propose a practical vocabulary grounded in auditability for making identity claims that are clear enough to bear ethical and institutional weight.

Keywords: haecceity, haecceitas, thisness, individuation, numerical identity, qualitative identity, structural equivalence, individual essence, transworld identity, modal metaphysics, Scotus, Ockham, Leibniz, Plantinga, Deleuze, personal identity, duplication paradox, copy continuity, provenance, audit trails, model checkpoints, persistence criteria, epsilon identity, accountability, AI identity.

A collaboration with GPT-5.2-Thinking. 58 pages.

Outline

1. Problem Statement: Why “Thisness” Returns
1.1 Individuation as a recurring fault-line in metaphysics
1.2 The modern trigger: duplication, restoration, and model copying
1.3 Three confusions to avoid: uniqueness, sameness, continuity
1.4 Thesis and contribution: an audit-grounded vocabulary for identity claims

2. The Classical Scaffold: What Haecceity Was Built to Do
2.1 The scholastic individuation problem in brief
2.2 Scotus’s move: haecceitas as non-qualitative thisness
2.3 Rival strategies: matter, form, and deflationary nominalism
2.4 What “thisness” buys: numerical distinction without descriptive difference

3. Nominalist Pressure and the Economy of Ontology
3.1 Ockham’s challenge: why posit haecceity at all
3.2 Parsimony versus explanatory adequacy
3.3 The hidden cost of deflation: identity talk without anchors
3.4 A transfer principle: rejected haecceities reappear as surrogates

4. The Leibniz Constraint: Indiscernibles and the Limits of Duplication
4.1 Identity of indiscernibles as a ban on “pure duplicates”
4.2 How the principle functions as anti-haecceity
4.3 Modern physics and the temptation to reopen individuation
4.4 Why AI copying makes the issue vivid regardless of physics

5. Analytic Revival: Haecceity as Individual Essence in Modal Space
5.1 Transworld identity and the need for an individual anchor
5.2 Haecceity-properties and “being that very one”
5.3 Objections: circularity, quidditism, and metaphysical overreach
5.4 What survives: separating identity from description cleanly

6. Continental Recasting: Haecceity Without Substances
6.1 Haecceities as events, assemblages, intensities
6.2 Thisness as singular configuration in time
6.3 Compatibility and conflict with the Scotist use
6.4 Bridge concept: history as individuation

1. Problem Statement: Why "Thisness" Returns

1.1 Individuation as a recurring fault-line in metaphysics

The individuation problem is one of those philosophical seams that never fully seals. It reopens whenever a theory confidently explains kinds, properties, or structures and then discovers it has not actually explained *individuals*. We can describe what it is to be a human being, a red apple, a chess position, or a hurricane pattern. But we stumble when asked what makes *this* human being, *this* apple, *this* position, or *this* hurricane numerically itself rather than merely a qualitatively matching instance. The recurring fault-line is not that we lack descriptive resources; it is that description alone appears to underwrite only *qualitative identity* and *classification*, while the phenomenon of *numerical identity* presses for a different sort of anchor.

This is why individuation keeps returning in waves across historical moments that otherwise disagree about metaphysics. In medieval scholastic debates, individuation is posed as an explanatory requirement: if there are many members of the same species, what accounts for their numerical multiplicity? In early modern metaphysics, the focus shifts toward principles such as the identity of indiscernibles and the possibility or impossibility of "pure duplicates." In analytic metaphysics, the problem is formalized through questions of transworld identity and individual essence: what makes it true that *this* individual exists in more than one possible world, or that a counterpart in another world is "the same" or merely similar? In continental lines of thought, the problem is reframed through singularities and events, emphasizing individuation as a process in time rather than a static metaphysical "stamp."

Despite these shifts, the same structural tension remains. Any view that treats individuals as nothing over and above bundles of qualities risks collapsing numerical distinction into mere descriptive difference, and then must explain what forbids the possibility of perfect qualitative duplicates being the same thing. Any view that treats numerical identity as primitive risks sounding like it has replaced explanation with a label. The individuation fault-line is thus a stability test for metaphysics: it measures whether a system can talk about "one" and "many" without cheating, and whether it can distinguish "this very one" from "one like it" without smuggling in a hidden haecceity under another name.

1.2 The modern trigger: duplication, restoration, and model copying

The AI era does not invent the individuation problem, but it makes the problem operational and frequent. In human life, duplication is rare and messy. Biological organisms do not admit clean copying; memories and bodies are not checkpointed and restored; legal identity usually tracks a continuous organism embedded in a continuous social record. Because duplication is difficult, metaphysical puzzles about duplicates have long lived in thought experiments: teletransporters, split brains, perfect replicas created in labs, ships rebuilt plank by plank. These scenarios sharpen intuitions, but they remain exceptional.

In machine learning practice, “duplication” is routine engineering. A system is saved, copied, moved across machines, restored from a checkpoint, fine-tuned into a specialty variant, merged with another branch, or rolled back after a failure. Version control metaphors are not metaphors; they are literally the workflow. This abundance of copying pressure forces identity talk out of the philosophical basement and into the world of institutions, policy, and accountability. When a deployed model is updated, did the original agent “continue,” or was it replaced? When a model is restored after corruption, is it the same agent returned, or a reset successor? When two forks merge, what is the identity status of the merged artifact relative to its parents? When a system is replicated across data centers, do we have one agent with distributed presence, or many agents with shared ancestry? When fine-tuning changes behavior, does identity track weights, capabilities, or responsibilities?

These questions arise not only for engineers but for everyone downstream: regulators, auditors, users, and organizations that rely on stable accountability. A customer may care whether a “trusted” system remains the same trusted entity. A court may care whether a logged decision was made by the system under certification or by a successor. An enterprise may care whether liability attaches to the organization, the model instance, the model lineage, or the operator who approved an update. In each case, “same” can mean several different things, and disputes often occur because parties speak as if there were only one notion of sameness that matters.

The modern trigger, then, is not “AI is mysterious.” It is that AI makes copying and branching so cheap and so ubiquitous that the individuation question stops being hypothetical. The technological substrate supplies perfect test cases for distinguishing numerical identity from equivalence, and for asking what continuity really amounts to when histories can fork, merge, and be selectively preserved.

1.3 Three confusions to avoid: uniqueness, sameness, continuity

The core diagnosis of this paper is that identity disputes—especially in AI contexts—typically conflate three distinct kinds of claim. Each kind of claim can be coherent on its own, but when they are blended without notice, the result is rhetorical force without conceptual discipline.

First is the *uniqueness claim*. This is the claim that an individual is in some sense non-interchangeable: there is a “thisness” to it that is not captured by general description. In metaphysics, this is the locus of haecceity. In ordinary language, uniqueness shows up when we insist that even a perfect duplicate would not be *that very one*. In AI practice, uniqueness claims appear when someone says, “This is the model we certified,” “This is the agent we trained,” or “This is the assistant I have a relationship with,” implying that there is an individual target of reference that is not reducible to a set of observable traits.

Second is the *sameness claim*, better called *structural equivalence*. This is the claim that two things match with respect to a chosen structure: same weights, same architecture, same outputs on a test suite, same policy behavior under a distribution, same performance envelope under constraints. Structural equivalence can be strong or weak, and it is often the right tool for engineering purposes. The confusion arises when structural equivalence is treated as if it automatically settles numerical identity. Two systems can be structurally identical in many respects and still be distinct individuals, and conversely, one individual can change structurally over time while remaining numerically the same.

Third is the *continuity claim*, which concerns history: provenance, causal ancestry, and the integrity of a chain that links the present entity to an earlier one. Continuity is neither uniqueness nor equivalence; it is a claim about *how* something came to be what it is, and whether the path is documentable, tamper-resistant, and normatively acceptable. In human contexts, continuity is tracked through bodies, memories, social recognition, and records. In AI contexts, continuity is tracked through logs, lineage graphs, signed artifacts, training data attestations, checkpoint hashes, and change-control processes. The confusion arises when people treat continuity as if it were identical to sameness of behavior ("it acts the same, so it's continuous") or as if it were identical to uniqueness ("it has the right history, so it must be the same individual in the strongest sense").

These three layers pull apart cleanly once we admit that each answers a different question. Uniqueness answers: what makes an individual referenceable as *that one*? Structural equivalence answers: in what respects are two artifacts the same *as structures*? Continuity answers: what makes it legitimate to treat the later artifact as inheriting the earlier artifact's standing, commitments, or accountability? Many public disputes about "the same AI" are really disputes about which of these questions is the one that matters in the case at hand.

1.4 Thesis and contribution: an audit-grounded vocabulary for identity claims

The thesis is that *haecceity remains useful as a discipline term* because it forces us to keep the three layers distinct, and because it prevents a common slide in AI discourse: using the word "same" as a solvent that dissolves disagreements by obscuring them. Haecceity does not have to be treated as a spooky metaphysical posit in order to function as a conceptual instrument. It can be treated as a warning label: whenever we find ourselves saying "it is the same," we should ask, "Same in which sense—numerically, structurally, or historically?" If we cannot answer, we are not yet making a claim fit to bear moral, legal, or institutional weight.

The contribution of the paper is therefore not to "solve" individuation in a final metaphysical sense, but to supply a vocabulary and set of distinctions that make AI-era identity talk tractable and auditable. Concretely, we aim to deliver four things.

First, a historical map: how Scotist haecceitas, nominalist deflation, Leibnizian constraints, analytic modal essences, and continental event-individuation each locate the individuation pressure point differently. This matters because contemporary AI debates often import metaphysical assumptions unknowingly. A person who implicitly thinks like a Leibnizian will treat perfect duplicates as impossible or unstable; a person who implicitly thinks in modal terms will reach for "individual essence"; a person with a Deleuzian sensibility will treat identity as a process traced by trajectories and intensities. Naming these backgrounds makes disagreements legible.

Second, a three-layer framework: numerical identity, structural equivalence, and continuity proofs. This framework is designed to prevent equivocation. It allows us to state, for example, that two models are structurally identical but numerically distinct; or that a restored model is numerically continuous under a specified continuity criterion; or that a fine-tuned model is structurally close but not continuity-equivalent for purposes of certification.

Third, a practical notion of continuity grounded in auditability. The paper's aim is not to replace metaphysics with bureaucracy, but to recognize that AI identity claims become ethically and institutionally meaningful only when continuity can be demonstrated rather than asserted. Audit logs, provenance records, lineage graphs, and change-control artifacts are not mere "paperwork"; they are continuity instruments. They are how institutions decide when accountability transfers, when trust should persist, and when an entity is a successor rather than a continuation.

Fourth, a controlled tolerance concept—epsilon-identity—for cases where practice requires operational sameness without metaphysical pretense. Many real systems cannot demand strict identity, yet they also cannot accept unconstrained replacement. Epsilon-identity names the idea that we can specify acceptable bounds of change and acceptable forms of provenance, such that within those bounds we treat the system as "the same for purpose P," while outside them we require re-certification, re-authorization, or explicit successor labeling. The point is to make "sameness" parameterized, explicit, and accountable rather than rhetorical.

Section 1 sets the stage: individuation is a perennial fault-line, AI makes it unavoidable, and the key is to prevent the three confusions that identity language invites. The rest of the paper develops the conceptual lineage, then uses the AI copy–continuity problem as a stress test for a disciplined vocabulary—one that allows us to speak precisely about "thisness" without turning metaphysics into mystique or engineering into hand-waving.

2. The Classical Scaffold: What Haecceity Was Built to Do

2.1 The scholastic individuation problem in brief

The medieval individuation problem arises from a simple pressure: if many things share the same nature, what accounts for their numerical multiplicity? The scholastic setting sharpens this into a metaphysical demand because it assumes that natures or essences are real enough to ground science and explanation. If "humanity" is a real nature, and Socrates and Plato are both human, then each instantiates the same specific nature. But then a question immediately follows: what makes Socrates *this* human and Plato *that* human, rather than the same human counted twice?

The problem is not "how do we tell them apart" in the ordinary epistemic sense. We can tell them apart by accidental features: size, location, biography, scars, and so on. The scholastic question is deeper: what makes the individual *ontologically* an individual, rather than merely a bundle of accidents hovering around a shared essence? In its classical form, individuation seeks a principle that can satisfy four desiderata at once.

First, it must account for numerical distinction among members of the same species without implying that species membership is illusory. Second, it must not collapse individuality into mere difference of accidental properties, because accidents can change while the individual remains. Third, it must explain how two individuals could, in principle, be qualitatively extremely similar without becoming numerically the same. Fourth, it must be compatible with the metaphysical architecture of the period: substance, form, matter, act, potency, and the layered distinction between essential and accidental properties.

This is why scholastic debates about individuation are not merely technical medieval curiosities. They are an early and rigorous attempt to say what it means for there to be *many* instantiations of one nature without reducing individuation to a bookkeeping artifact. In modern AI terms, the scholastic problem anticipates a question like: if two artifacts instantiate the same architecture and even the same learned structure, what makes them numerically distinct individuals rather than "one thing in two places"? The medievals did not have model checkpoints, but they had the conceptual problem of multiple instances under one nature, and they demanded a principled account.

2.2 Scotus's move: haecceitas as non-qualitative thisness

Scotus's intervention is motivated by dissatisfaction with both extremes: reducing individuation to matter or accidents on the one hand, and treating numerical distinction as brute or inexplicable on the other. His proposal introduces *haecceitas*—thisness—as a principle that individuates without being a qualitative feature. It is not "another property like redness" that could be shared. It is, rather, what makes an entity *this entity* in a way that is irreducible to the

general nature it instantiates and irreducible to the list of accidents it happens to bear at a moment.

To see the shape of the move, consider the distinction Scotus insists upon between *common nature* and *individuality*. The common nature can be shared; it answers what a thing is. Individuality answers which one it is. Scotus does not deny that individuals have qualities, histories, and locations. His claim is that none of these, even in combination, is sufficient to ground the metaphysical "thisness" that makes the individual non-interchangeable at the level of numerical identity. There must be something that contracts the common nature to an individual instance, something that is not itself a shareable description.

This is why haecceitas is often characterized as non-qualitative. It is not an item in the descriptive inventory of the world. It is a principle of individuation that is more like a "marker of numerical singularity" than a trait. In contemporary terms, it resembles the difference between a record's *content fields* and the record's *identifier*. The identifier is not a feature of the person described; it is a way of picking out *this* record as *this* one. But Scotus's point is not administrative; it is metaphysical. He thinks reality itself has a grounding for numerical "thisness" that cannot be reduced to the content of common natures.

There is an important discipline here that we will reuse later. Scotus is trying to prevent a common equivocation: the slide from "the individual is fully describable" to "the individual is nothing but its description." Haecceitas is introduced as a stopgap against that slide. It is a safeguard for the claim that numerical identity is not identical to descriptive sameness.

In AI identity talk, this Scotist move has an immediate analogue. If two model instances have the same architecture and weights, and if their behavior matches on a vast range of inputs, there is still a legitimate question whether they are numerically the same entity or two numerically distinct entities that are structurally equivalent. Scotus's haecceitas does not answer that question by itself. But it forces us to see that "same behavior" and "same structure" do not automatically settle numerical identity. If one wants to assert numerical sameness, one must say what would ground it. Haecceitas is one candidate for such grounding in a metaphysical register.

2.3 Rival strategies: matter, form, and deflationary nominalism

The scholastic landscape contains several competing strategies for individuation, and mapping them clarifies why Scotus's move mattered. The rivals can be sketched as three broad orientations: matter-based individuation, form-based individuation, and deflationary nominalism.

Matter-based strategies, often associated with Aristotelian and Thomistic lines, treat matter (or matter under determinate dimensions) as central to individuation. On this view, the common

nature is made “this” by being instantiated in this parcel of matter, in this place and time. The intuitive attraction is obvious: bodies are distinct because matter is distinct. But the strategy faces persistent pressure when one tries to specify how matter individuates without already presupposing individuation. If we say “this matter” individuates, we may be quietly reintroducing the very “thisness” we were trying to explain. Further, matter-based accounts can struggle with entities that are not easily framed as material parcels or where the metaphysical status of matter is contested.

Form-based strategies, more Platonic or more intellectualist in orientation, can treat individuation as arising from form or from a unique configuration of form-like principles. This can look like individuating by a unique form, a unique bundle of forms, or a uniquely complete determination of the essence. These accounts can sound closer to modern “complete description” approaches: if you specify the form fully enough, you specify the individual. The pressure here is that the more “complete” the description becomes, the more it appears to smuggle the individual in by including uniqueness as a hidden ingredient. At some point, the form-based strategy risks becoming a disguised haecceity account: it individuates by an irrepeatable determination that is effectively a thisness under a different label.

Deflationary nominalism attacks the entire need for an individuation principle by denying that universals or common natures exist in the robust way that generated the problem. If there is no shared nature “humanity” over and above individual humans, then the metaphysical mystery of “many instantiations of one nature” is dissolved. Individuals are fundamental; the universal is a name, a concept, or a convenient classificatory device. The attraction is parsimony: do not multiply entities beyond necessity. The cost is that the nominalist must still explain why our general terms work so reliably in science and reasoning, and must still give an account of identity over time and across counterfactual conditions. Even if one refuses haecceity, one cannot refuse the practical need to track “which one” and “the same one again.”

These rivals show why Scotus’s move is not mere medieval ornament. Matter strategies risk circularity and limitation; form strategies risk hidden haecceity; nominalist strategies risk relocating the work of individuation into semantics, cognition, or social practice while continuing to rely on stable identity talk. Scotus names the problem directly and offers a principle that is explicitly non-qualitative and explicitly individuating, avoiding the pretense that individuation can be reduced to description without remainder.

For the AI copy–continuity problem, the rival strategies reappear almost verbatim. A matter-based strategy becomes a hardware or instance-based identity view: “the agent is this running process on this machine.” A form-based strategy becomes a weights-and-architecture identity view: “the agent is the model specification.” A nominalist strategy becomes a behavior-only or use-only view: “identity is just how we talk; what matters is functional equivalence.” Each

strategy can be coherent, but each risks a different kind of confusion when pressed into ethical or institutional service.

2.4 What "thisness" buys: numerical distinction without descriptive difference

The distinctive payoff of haecceity, in the classical scaffold, is that it underwrites numerical distinction even in the limiting case where descriptive difference approaches zero. It says, in effect: even if two individuals shared the same species nature and matched in all qualitative respects we can specify, they would still be two rather than one, because individuality is not merely a descriptive matter. This payoff is easy to caricature as "spooky identity particles," but the real function is disciplined: it prevents a common inference from the success of description to the elimination of individuality.

This is why haecceity is relevant to duplication puzzles. If you believe in haecceity, then perfect qualitative duplication does not entail numerical identity. Two copies can be qualitatively indistinguishable and yet numerically distinct. Conversely, if you deny haecceity, you must choose between two pressures: either deny the possibility of perfect qualitative duplication, or accept that perfect duplicates would be numerically the same, or treat numerical identity as an empty label reducible to equivalence. Each option has costs.

The Scotist purchase becomes even clearer when we translate it into the three-layer framework this paper develops. Haecceity lives primarily at the layer of numerical identity. It asserts that there is a fact of "this very one" that is not reducible to structural equivalence. It does not, by itself, tell you which continuity proofs to accept or which equivalences matter for which purposes. But it does enforce a separation: the question "Are they numerically the same?" is not the same question as "Are they structurally equivalent?" and not the same question as "Do they share a legitimate provenance?" Haecceity is the conceptual wedge that keeps those questions from collapsing into each other.

In AI contexts, "thisness" buys us a disciplined way to speak about two scenarios that are otherwise rhetorically muddled.

In the first scenario, we have multiple deployments of the same model weights: identical copies running in parallel. Engineering practice treats them as interchangeable; operations teams treat them as scalable instances; users may treat them as "the same assistant" because the experience is consistent. Haecceity reminds us that interchangeability and experiential sameness do not entail numerical identity. There are many numerically distinct instances that are structurally equivalent.

In the second scenario, we have a model that undergoes updates: fine-tuning, alignment patches, safety layers, new system prompts, or retrieval augmentation changes. Here the numerical identity intuition often runs the other way: stakeholders speak as if the agent

"continues" through change, even when structure shifts materially. Haecceity does not settle whether continuation holds, but it exposes the hidden assumption: that numerical identity can persist through structural change, and that what matters is continuity of lineage or institutional recognition rather than exact equivalence.

So the classical scaffold supplies the paper's first major tool: a way of refusing a false choice. We do not have to say that identity is "nothing but description," nor do we have to say that identity is ineffable mystique. We can say instead: numerical identity is a distinct kind of claim, and a tradition exists—exemplified by Scotus—that treats "thisness" as the marker of that distinctness. Whether one ultimately endorses haecceity as a metaphysical posit, the concept still performs a crucial discipline function: it blocks equivocations that are otherwise almost guaranteed in AI-era debates about copying, restoration, and continuity.

3. Nominalist Pressure and the Economy of Ontology

3.1 Ockham's challenge: why posit haecceity at all

If Scotus introduces haecceity as a principled way to secure numerical individuality, Ockham's challenge is the natural counter: why multiply metaphysical entities to do work that can be done more cheaply? The nominalist posture does not begin as a denial that individuals are individuals. It begins as suspicion about *explanatory over-commitment*. If individuation is obvious in experience and sufficient for discourse, why posit a special metaphysical principle of "thisness" in addition to the individual substance itself, its qualities, and its relations?

Ockham's pressure can be stated as a set of questions aimed directly at haecceity as a theoretical posit. What is haecceity *in addition to* the individual? If it is something distinct from the individual, how does it avoid being either (i) another property that could, in principle, be shared, which would undermine its role, or (ii) a primitive "numerical tag" that does not explain anything but merely repeats the fact to be explained? If it is not distinct from the individual, then why name it as if it were an extra principle? And if the work of individuation can be carried by ordinary features such as spatiotemporal location, material constitution, or the simple fact of being a singular substance, then the haecceity posit begins to look like metaphysical double-entry bookkeeping: you write "individual" in one column and "thisness" in another, and then claim the second column explains the first.

This critique has a modern resonance because it tracks a familiar methodological move. In engineering and computer science, we routinely reject unnecessary layers of abstraction: if the system can be specified without them, do not add them. Ockham's instinct is the same. The core of the challenge is not hostility to individuation; it is hostility to *unpaid metaphysical debt*. If haecceity is doing real explanatory work, it must buy something you cannot obtain otherwise.

If it is not buying anything, it is a decorative term that risks encouraging confident metaphysical rhetoric without operational consequence.

For the AI copy–continuity problem, Ockham's challenge can be rephrased bluntly: if we can settle the practical questions with provenance records, version graphs, checksums, and behavioral tests, why appeal to "thisness" at all? Why not say that identity talk is conventional, tool-relative, and ultimately reducible to whichever equivalence relation serves the task?

3.2 Parsimony versus explanatory adequacy

The most fruitful way to read the nominalist pressure is not as a refutation but as a demand for disciplined tradeoffs. Parsimony is not a stand-alone virtue; it is a constraint that must be balanced against explanatory adequacy. A theory that is too lavish becomes unfalsifiable, rhetorically powerful but epistemically weak. A theory that is too deflationary risks losing the ability to explain the phenomena that motivated theorizing in the first place.

In the individuation context, parsimony says: do not posit special individuators if ordinary ontology can do the job. Explanatory adequacy says: if ordinary ontology cannot account for our strongest and most stable identity judgments, then the deflation has not saved cost; it has merely shifted cost into unacknowledged assumptions.

This tradeoff is especially visible in the case of duplicates. Suppose we aim to avoid haecceity. Then we need an account of why numerical individuality is not simply a byproduct of descriptive difference. We might appeal to spatiotemporal location, but then we must say what happens in thought experiments (or future technologies) where location is duplicated, swapped, or rendered ambiguous. We might appeal to causal history, but then we must specify which kinds of causal continuity are identity-preserving and which are not. We might appeal to primitive identity facts, but then we have conceded a primitive just where haecceity was accused of being primitive. We might appeal to the impossibility of perfect duplicates, but then the claim becomes hostage to empirical or metaphysical constraints that may not align with our practical needs.

In other words, nominalist parsimony often wins at the level of *ontology* but can lose at the level of *identity practice*. The simpler your metaphysical inventory, the more work must be done by your semantics, your epistemology, or your institutions. That is not automatically a problem. Sometimes moving work into practice is exactly what clarity requires. But it becomes a problem when the theory continues to speak as if identity were straightforward while its deflation has removed the resources needed to make identity claims non-arbitrary.

In AI contexts, this tradeoff becomes even sharper. The temptation is to adopt a purely functional equivalence view: if two model instances behave the same, treat them as the same for all purposes. That is maximal parsimony: no haecceity, no metaphysical surplus, only

behavior. But it quickly collides with cases where behavior equivalence is incomplete, where stakes depend on provenance, and where accountability must survive adversarial manipulation. Explanatory adequacy in the AI era requires more than "it seems the same."

So the parsimony-versus-adequacy axis becomes one of the paper's governing design constraints. We want a vocabulary that avoids metaphysical inflation, but we also want one that does not pretend that identity talk is free. If "thisness" is rejected, something else must pay the bill for individuating reference and accountable continuity.

3.3 The hidden cost of deflation: identity talk without anchors

Deflationary moves often appear clean in theory but messy in practice. The hidden cost is that identity talk does not vanish; it migrates into implicit anchors that are treated as obvious rather than argued for. When haecceity is rejected as an extra metaphysical principle, the conversational need to pick out "this very one" remains. If the theory no longer supplies a metaphysical anchor, the anchor is supplied by habit, convenience, institutional power, or rhetorical drift.

There are several common surrogate anchors that show up when explicit haecceity is denied.

One surrogate is *indexical anchoring*: "this here now." We point, we refer, we track continuity through location and immediate presence. This works well in ordinary life. It fails as a general theory when copies or distributed instances arise, because indexicals multiply. If there are two "this here now" anchors, the deflationist must either allow that identity is relative to context (which may be correct), or quietly smuggle in a privileged anchor without explanation.

Another surrogate is *psychological or relational anchoring*: "the one I'm interacting with," "the one I trust," "the one I formed a relationship with." This anchoring plays a large role in personal identity intuitions and becomes especially salient in human-AI interaction. But relational anchoring is not a metaphysical criterion. It is a social-psychological fact about the user's stance. It can be manipulated, and it can conflict across users and institutions. Treating it as the identity criterion amounts to letting identity be defined by whoever speaks loudest or feels most attached.

A third surrogate is *administrative anchoring*: the one identified by a name, a certificate, an account, or an ID. This is powerful because institutions already run on it. But administrative anchoring can detach from the underlying artifact it purports to track. If the "same model" label is preserved across a major update, the label says "same" even if the structure and behavior have shifted materially. Conversely, a model might remain structurally identical while being relabeled as new for business or policy reasons. Administrative anchors are indispensable, but they are not metaphysical truth-makers. They are continuity instruments whose legitimacy depends on auditing and governance.

A fourth surrogate is *structural anchoring*: “same weights,” “same architecture,” “same parameters.” This is common in AI engineering discourse because it is measurable and crisp. But structural anchoring does not settle questions about continuity across time, responsibility transfer, or legitimate succession. It is one dimension of sameness among many.

The unifying point is that deflation does not eliminate the need for anchors; it merely makes the anchors tacit. Once anchors become tacit, identity talk becomes vulnerable to precisely the rhetorical inflation this paper is trying to prevent. People will assert “same” while meaning different anchors, and disagreements will become irresolvable because the premises were never shared.

In the AI copy–continuity problem, the hidden cost shows up in a predictable pattern. A vendor will claim continuity when it serves trust (“it’s the same assistant you rely on”), but claim discontinuity when it serves liability management (“that was a previous version”). A regulator will demand strict continuity for accountability but accept functional equivalence for operational convenience. Users will anchor identity to relational experience, while engineers anchor it to weights and checkpoints. Without an explicit vocabulary, these become conflicts of power and persuasion rather than conflicts of reasons.

3.4 A transfer principle: rejected haecceities reappear as surrogates

This section introduces a methodological claim that will be reused later as a diagnostic: when haecceity is rejected, it tends to reappear in disguised form. The point is not that haecceity must be metaphysically real. The point is that the *function* haecceity served—stabilizing reference to “this very one” under conditions where description and equivalence are insufficient—does not disappear. It transfers into other structures.

We can call this the transfer principle: if a framework refuses explicit thisness, it will typically reintroduce an implicit thisness elsewhere, often in a place less visible and less accountable.

In scholastic terms, the transfer principle explains why deflationary nominalism often ends up relying heavily on demonstratives, contextual reference, and pragmatic criteria that do the individuating work without calling it that. In modern terms, the transfer principle explains why “identity is just equivalence” positions often end up privileging one equivalence relation as the “real” one, even though equivalence relations are many and purpose-dependent. The “real” equivalence relation becomes a surrogate haecceity: an unargued-for choice treated as inevitable.

In AI practice, the transfer principle is almost unavoidable. If we refuse any notion of individual identity beyond structural equivalence, then our accountability systems will still need to pick out which artifact acted. We will then treat a checksum, a deployment ID, or a signed build as the “real identity,” even though that is an administrative surrogate. If we refuse administrative

identity and insist on behavior-only, we will end up privileging a test suite or benchmark distribution as the "real self" of the model, even though that is a measurement surrogate. If we refuse behavior-only and insist on provenance-only, we will treat a lineage graph as the identity bearer, even though that is a historical surrogate. Each surrogate may be legitimate, but none should be allowed to masquerade as the whole story.

The transfer principle has two implications for the rest of the paper.

First, it reframes the Scotus–Ockham dispute in a way that is useful for the AI era. The real question is not "Is haecceity spooky?" The real question is "Where does the individuating work happen, and is that location explicit and auditable?" If the work happens in tacit social recognition or marketing language, the result is fragile and manipulable. If it happens in technical identifiers without governance, it becomes opaque power. If it happens in provenance records without clarity about purpose, it becomes bureaucratic ritual. The goal is not to eliminate the need for individuating anchors; it is to make their role explicit, scoped, and checkable.

Second, it sets up the paper's constructive move. We do not need to commit to haecceity as a metaphysical entity to benefit from its discipline function. We can treat "haecceity talk" as a warning system that tells us when we are relying on a surrogate. Whenever we find ourselves tempted to say "it's the same model," we ask: same by which anchor? structural, administrative, relational, or provenance-based? If we cannot answer, we are watching haecceity reappear as an unexamined surrogate, and we should expect downstream conflict.

Nominalist pressure, then, is not something the paper must defeat. It is a constraint we must respect. It forces economy: no metaphysical posit without paid explanatory value. But it also reveals a danger: deflation that hides the individuating work produces identity talk without anchors. The constructive response is to keep the individuation function visible and to design identity claims so they can be audited, scoped to purpose, and resistant to rhetorical drift.

4. The Leibniz Constraint: Indiscernibles and the Limits of Duplication

4.1 Identity of indiscernibles as a ban on "pure duplicates"

Leibniz's "identity of indiscernibles" is often introduced as a crisp metaphysical axiom: if two entities share all the same properties, then they are not two entities at all but one and the same. Put in the form relevant to individuation, the principle functions as a ban on "pure duplicates," meaning numerically distinct individuals that are qualitatively indiscernible in every respect. If there is literally no difference—no difference in intrinsic properties, no difference in relations, no difference in location, no difference in history—then there is no basis for counting them as two rather than one.

As a constraint, the principle is attractive because it seems to reduce metaphysical mystery. Instead of asking for an extra individuating "thisness," we say: individuality is secured by discernibility. The world does not contain brute numerical multiplicity without qualitative distinction. Multiplicity is always grounded in difference. The principle thereby compresses individuation into an economy-of-properties program: to be two is to differ.

The principle also carries an intuitive moral: if you cannot, even in principle, distinguish A from B, then you have not described two entities. You have described one entity twice. In contexts where properties are treated as the language of reality, indiscernibility collapses numerical plurality. This is the metaphysical counterpart of a data-modeling idea: if two records have identical fields and identical relational links, then they are redundant representations of one underlying thing.

But the strength of the principle is also its fragility. It depends on what counts as a "property," and on whether relational properties like "being two units apart" or "being on the left" are admissible. It depends on whether spatiotemporal location is treated as a property. It depends on whether "having this causal history" counts. In a rich property ontology, perfect indiscernibility becomes hard to achieve, and the principle becomes vacuously true because duplicates are ruled out by definition: if the entities are distinct, they must differ in some property, perhaps a relational one. In a sparse property ontology, the principle becomes a strong metaphysical claim that forbids certain imaginable cases.

For our purposes, the key point is not how to adjudicate the principle's ultimate truth, but how it functions as a constraint on individuation talk. It asserts a deep link between numerical distinctness and discernibility. If accepted, it makes the individuation problem "solvable" by denying that the most troubling cases are possible. There are no two numerically distinct entities with all the same properties, so haecceity is unnecessary as a separate individuator.

4.2 How the principle functions as anti-haecceity

The identity of indiscernibles is anti-haecceity in the sense that it aims to eliminate the need for a non-qualitative individuator. Haecceity says: even if you held fixed every qualitative description, there could still be a fact of "this very one" that makes two individuals numerically distinct. Leibnizian indiscernibles says: no; if everything qualitative is fixed, numerical distinctness cannot remain. The two views are not merely different; they pull in opposite directions in the most informative corner case.

This opposition can be framed as a dispute about what counts as a legitimate explanation of numerical plurality. The Scotist move treats numerical plurality as requiring its own principle, not reducible to qualitative inventory. The Leibnizian move treats numerical plurality as reducible to qualitative difference. The principle thus functions as a "compression rule" on

ontology: do not posit brute “thisness” facts; treat every claim of plurality as backed by some difference.

The anti-haecceity effect is reinforced by a second idea often associated with Leibniz: complete concepts. On a complete-concept picture, an individual is, in principle, describable by a maximally complete set of predicates that uniquely characterizes it. If that picture is adopted, individuality is secured by description in the limit: there is no remainder. Haecceity becomes redundant because the complete description already contains the uniqueness. “Thisness” is absorbed into descriptive totality.

However, that absorption comes with a familiar danger: the uniqueness is secured by making the description so rich that it effectively includes an indexical or a history, smuggling “this” into the predicates. Once “being born of these parents” or “occupying this spatiotemporal region” or “having this causal ancestry” becomes a predicate in the complete concept, the descriptive machinery starts to resemble a disguised haecceity. The difference is that the uniqueness is not named as non-qualitative; it is encoded as a maximal qualitative package.

In the present paper’s terms, the indiscernibles constraint tends to collapse the three layers. It makes numerical identity look like a function of descriptive sameness (structural equivalence), and it makes continuity look like just more description. That can be philosophically elegant, but it invites practical equivocation: if “identity just is equivalence of all properties,” then identity claims become hostage to which properties you count and how you measure them. The anti-haecceity posture can therefore produce its own kind of rhetorical inflation: confidently declaring two things “the same” because they match in the measured respects, while quietly ignoring unmeasured respects that might matter to accountability or continuity.

So the anti-haecceity function of the principle is double-edged. It offers ontological economy and conceptual closure. But it risks hiding the individuation work inside the property language in a way that can be opaque and contestable.

4.3 Modern physics and the temptation to reopen individuation

Modern physics introduces a new tension into the Leibnizian program: there are domains where individuality appears to be either weakened or reframed. Quantum mechanics is often invoked here because it complicates classical intuitions about distinguishability. In some contexts, particles of the same type are treated as indistinguishable in a way that is not merely epistemic. States can be symmetric or antisymmetric under exchange; counting and labeling “which particle is which” can become physically meaningless or at least non-fundamental. This can look like a vindication of Leibniz: if there is no discernible difference, perhaps there is no numerical plurality in the classical sense.

But the same physics can also tempt one toward the opposite conclusion: perhaps individuality is not exhausted by the properties we can represent, or perhaps individuation requires something beyond qualitative state description. The temptation to "reopen" individuation arises when we notice that physical descriptions may underdetermine "which one," or that different metaphysical pictures can be consistent with the same empirical predictions. If two configurations are physically treated as equivalent under permutation, does that mean there are not two individuals, or does it mean our physical theory does not track individuality at that level?

Importantly, this is not a place where our paper needs to take a side on foundational physics. What matters is that physics makes vivid a general point: the indiscernibles constraint is only as strong as your account of properties and discernibility. If the theory's representational apparatus does not distinguish individuals, you might say individuals are not fundamental. But you might also say that individuation is not captured by that apparatus. Either way, the philosophical terrain becomes less stable, and the debate about haecceity and individuation becomes less susceptible to purely a priori closure.

This is one reason the individuation question returns in modernity. The classical intuition that individuality is secured by location and matter is destabilized by both physics and technology. The Leibnizian hope that individuality is secured by complete description is destabilized by the possibility that descriptions can be formally complete relative to a theory and still not settle "which one" in the ordinary sense. The net effect is that individuation becomes a live question again, not because we are nostalgic for medieval metaphysics, but because our best theories and technologies do not automatically settle the identity claims we need to make.

4.4 Why AI copying makes the issue vivid regardless of physics

Even if one sets physics aside entirely, AI practice forces the individuation problem into daily life. The reason is simple: AI systems make "pure duplicate" scenarios cheap. Where Leibniz treated perfect duplicates as a metaphysical impossibility or at least a limiting case, AI turns near-duplicates into routine artifacts. A model can be copied bit-for-bit. Two copies can share architecture, weights, prompts, and even identical internal states at a checkpoint. If you grant that software copying can produce perfect structural duplicates, then the indiscernibles constraint confronts us directly: do we say the two copies are numerically one and the same, or do we say they are two numerically distinct instances that are structurally indiscernible?

In practice, engineers treat them as two instances, because they run in different processes and occupy different hardware. But that answer already presupposes a particular property ontology: location, process identity, and causal history count as differentiating properties. The Leibnizian can happily accept this and say: yes, they are discernible by relational properties, so they are

distinct. But notice what has happened. The interesting question has shifted from “can there be pure duplicates” to “which properties are identity-relevant for the purposes at hand.”

This shift is exactly what makes the AI case valuable. AI copying reveals that we can satisfy the indiscernibles constraint trivially by always including some relational property like “being on server A” versus “being on server B.” But that trivial satisfaction does not answer the questions that humans actually care about: trust, responsibility, continuity, and whether an agent is “the same one” in a sense that matters for commitments.

For example, suppose a safety-critical model is copied to a new machine without any change. Structurally, it is identical. Relationally, it is distinct. If the model causes harm, does responsibility attach differently because it was “a different instance”? Usually not. Institutions will treat the copies as the same model for accountability purposes, because what matters is lineage, certification, and governance, not which machine ran the bits. Conversely, suppose a model is updated and behavior shifts slightly. Structurally, it is no longer identical. But institutions may still treat it as the same deployed agent, because what matters is continuity of service and contractual identity.

So AI practice makes vivid the gap between metaphysical constraints and normative identity needs. Leibniz’s principle can help us avoid one kind of confusion: it reminds us that plurality without difference is suspicious, and it pressures us to specify differentiating properties. But it cannot tell us which differences matter for which purposes. In AI identity talk, that is the central task. The mere existence of a differentiating relational property (different process, different machine, different timestamp) is insufficient to settle whether two artifacts should inherit the same standing, obligations, or trust.

This is why AI copying is philosophically clarifying even if physics does not force the issue. AI gives us controlled, repeatable, auditable duplication events. It lets us separate the questions: numerical distinctness (yes, there are two processes), structural equivalence (yes, the bits match), continuity (yes, the copies share lineage), and accountability (yes, the institution treats them under one certified identity). Once these questions are separated, the identity of indiscernibles becomes one constraint among others rather than the master key.

The conclusion of this section, then, is not that Leibniz is wrong or Scotus is right. It is that the Leibniz constraint shows how much hinges on what counts as a property and on which properties we treat as identity-relevant. AI systems expose this dependence in a way that is hard to ignore. They force us to choose—explicitly—between metaphysical economy and institutional adequacy, and they make it clear why an audit-grounded vocabulary is not optional. Without it, “same” becomes a word that smuggles in a whole property ontology and a whole governance regime without ever admitting that it has done so.

5. Analytic Revival: Haecceity as Individual Essence in Modal Space

5.1 Transworld identity and the need for an individual anchor

When twentieth-century analytic metaphysics revives haecceity, it does so in a new arena: modal space. The driving question is not primarily "what individuates members of a species here and now?" but "what makes it true that an individual could have been otherwise, or could have existed in a different possible world?" Modal discourse is full of apparently coherent claims of the form "Socrates might not have been a philosopher," "this very person could have lived in a different city," or "the same planet could have formed under slightly different initial conditions." But such claims presuppose a way of tracking individuals across possibilities. Without some anchor, modal claims collapse into talk about merely similar individuals: counterparts rather than the same individual.

The challenge is easiest to state as a problem of reference under counterfactual variation. If we take an individual and vary its properties across possible worlds, at what point have we varied it so far that it is no longer the same individual but a different one? If the answer is "we track it by whatever properties are essential to it," then we need a principled account of which properties are essential. But if the essential properties are too thin, we risk allowing transworld "drift" where distinct individuals are conflated. If they are too thick, we risk trivializing modality because the individual cannot vary in interesting ways. Modal metaphysics therefore needs an anchor that can do two jobs at once: it must keep individuals stable across worlds while allowing properties to vary.

This is where the analytic revival of haecceity begins. The motivating intuition is that there is a distinction between two kinds of "essence." One kind is descriptive: the set of properties that characterize something as a member of a kind, or as meeting a description. Another kind is individual: the property of being that very individual. Transworld identity pushes the second kind into view because it shows that even a complete descriptive specification may fail to secure "this very one" across possible worlds. If we identify the individual with a description, we risk turning modal claims about the individual into modal claims about whoever fits the description in each world. That would be a shift from identity to role-filling. The analytic haecceity program is an attempt to stop that slide.

The relevance to AI copy–continuity becomes clearer if we translate the modal question into a versioning question. Across a development timeline, a model can be altered in many ways: weights adjust, data shifts, architecture is refactored, retrieval is added, system instructions change. Which changes preserve identity and which constitute replacement? Modal metaphysics frames this as a question about persistence across "nearby worlds" of alteration. If we lack an anchor, we either (i) treat every alteration as a new individual, which is too strict for practice, or (ii) treat identity as merely "sufficient similarity," which is too loose for

accountability. The modal anchor problem is structurally the same problem as the AI succession problem, only with different vocabulary.

5.2 Haecceity-properties and "being that very one"

Analytic metaphysics often expresses haecceity through what are sometimes called haecceity-properties: properties that an individual has in virtue of being itself, such as "being Socrates" or "being this very individual." The idea sounds odd at first because it treats identity as if it were a property. But the point is not to add a new qualitative feature alongside height and color. The point is to have a formal device that can anchor cross-world reference without reducing identity to descriptive content.

A haecceity-property is typically taken to be non-qualitative in the relevant sense: only one thing can have it, and it cannot be shared. It functions as a rigid anchor that persists across modal variation. If we want to say "Socrates could have been a sailor," we can say: there is a possible world in which the bearer of the Socrates-haecceity has the property of being a sailor. The "being a sailor" property is allowed to vary; the haecceity property fixes the target of variation.

This is the analytic version of Scotus's move, but tuned to modal logic rather than medieval questions about contracting common natures. In the scholastic setting, haecceity answers "what makes this individual numerically distinct from others of the same kind?" In the analytic modal setting, haecceity answers "what keeps reference fixed when we vary properties across worlds?" The unifying theme is the same: do not confuse identity with description.

There is also a subtle but important discipline embedded here. Haecceity-properties make explicit that identity claims are not merely high-confidence descriptive matches. They are claims about "that very one." In AI discourse, people frequently speak as if there is a single sliding scale from "similar" to "same." Haecceity-talk refuses that scale. It says: similarity lives in the space of qualitative properties; sameness of individual lives in a different logical category. You can have extreme similarity without numerical sameness, and you can have numerical sameness across significant qualitative change, depending on the persistence criteria you adopt.

In AI terms, one can imagine a haecceity-like anchor as the "identity token" of a model lineage: not merely an administrative label, but a principled rule that says which descendants count as the continuation of the same deployed entity. Importantly, the analytic haecceity program does not automatically tell you what the rule is; it tells you that some anchor is needed if you want to talk about "the same one" rather than "a similar one." In practice, we will later argue that the anchor must be tied to auditability and provenance rather than treated as a metaphysical free variable.

5.3 Objections: circularity, quidditism, and metaphysical overreach

The analytic revival invites sharp objections, and these objections matter for our later use of haecceity as a discipline term.

One objection is circularity. If you define the haecceity-property of Socrates as "being identical to Socrates," then you have used identity to define the property that was supposed to explain identity across worlds. The worry is that haecceity does not ground identity; it merely restates it in property form. The objection can be sharpened: if haecceity-properties are admitted, then every individual has a unique property that individuates it, and the individuation problem is "solved" by stipulation. That looks like metaphysics by fiat.

A second objection targets quidditism and related worries about primitive "whatness" facts. If one allows non-qualitative identity-fixing properties, one may be committed to a broader metaphysical picture in which properties have identity independent of their causal or qualitative roles. Some critics worry that haecceities are a symptom of an overly permissive property ontology: if any "being that very one" property is allowed, then we have inflated the property realm beyond what is needed for explanation, and we have lost the economy that nominalist pressure demanded. The debate here is complex, but the central worry is simple: haecceity-properties may reify what should remain a feature of reference or language into a heavy metaphysical apparatus.

A third objection is metaphysical overreach. Even if haecceity-properties are coherent, do we need them? Perhaps transworld identity should be handled differently: by counterpart theory, by stage theory, by context-relative identity, or by treating "could have been otherwise" as a loose way of speaking about similar individuals. On these views, insisting on strict transworld identity is itself a philosophical choice rather than a mandatory feature of modal discourse. If so, haecceity-properties may be solving a problem created by a particular interpretation of modal claims.

A fourth objection is explanatory mismatch. In many practical contexts, what we need is not a metaphysical truth-maker for identity but a workable criterion for tracking continuity, responsibility, and governance. Haecceity-properties can look like they belong to a different register: useful in formal semantics, but too abstract to guide real-world decisions about accountability.

These objections do not render haecceity useless. They clarify what it can and cannot do. They warn against treating haecceity as a magical substance that dissolves all puzzles. They push us to be honest about the status of the concept: either a metaphysical posit with costs, or a formal tool for tracking reference, or a discipline term that prevents equivocation. Our project leans toward the third: haecceity is valuable not because it gives a final metaphysical answer, but

because it forces the question of “which one” to be asked explicitly and prevents identity from being quietly replaced by description.

5.4 What survives: separating identity from description cleanly

Even if one rejects haecceity as a heavyweight metaphysical entity, something important survives from the analytic revival: the clean separation of identity from description. This separation is the key contribution we carry forward into the AI copy–continuity problem.

The separation has two parts.

First, it insists that identity claims are not automatically justified by descriptive match. Two items can match on every descriptive feature you have measured and still be numerically distinct, because numerical distinctness can be secured by relational, historical, or token-level facts that are not captured by the description. Conversely, one item can persist as itself through substantial descriptive change, because persistence criteria need not be “no change”; they can be “change under an acceptable continuity relation.”

Second, it insists that “being that very one” is a different kind of claim than “being the one that fits description D.” In AI practice, this is the difference between “the certified model” and “a model that passes the certification tests.” The latter is a description. The former is an identity claim tied to provenance. If one treats them as equivalent, one creates a loophole: a system can be replaced by a similar one that passes tests but lacks the audited lineage that the institution intended to rely upon. The analytic separation makes that loophole visible.

This is where the AI relevance becomes sharp. In a world of easy copying and easy fine-tuning, “description-only identity” is vulnerable. Adversaries can produce test-passing impostors; organizations can quietly swap systems while maintaining continuity rhetoric; accountability can be dodged by redefining “the same” to mean “functionally similar enough.” The survival lesson from analytic haecceity is therefore a discipline rule: do not let the descriptive layer do the work of the identity layer without an explicit argument.

Our later framework will express this lesson in audit-grounded terms. Instead of positing an unobservable haecceity-property, we treat “identity anchor” as something that must be instantiated in traceable continuity proofs: lineage records, signed artifacts, change-control events, and purpose-specific criteria for when continuity is preserved. But the conceptual prerequisite remains the analytic separation: if we do not distinguish identity from description, we cannot even formulate what an audit trail is supposed to demonstrate.

So the analytic revival earns its place in our story. It supplies a modern, rigorous vocabulary for the same pressure Scotus addressed, now under modal variation rather than scholastic taxonomy. It also clarifies the main risk we face in AI-era identity talk: collapsing “the same one” into “the one that behaves the same.” The most durable contribution is not a metaphysical

commitment to haecceity-properties, but a disciplined refusal to let description masquerade as identity.

6. Continental Recasting: Haecceity Without Substances

6.1 Haecceities as events, assemblages, intensities

A continental recasting of haecceity begins by loosening the grip of the substance framework. Where the scholastic and much analytic tradition treats individuation as something that happens to substances or individual things, the continental line we are tracking treats individuation as something that happens in events, processes, and configurations. "Thisness" shifts from being a metaphysical stamp on a thing to being a singularity of occurrence: an irrepeatable constellation of relations, affects, speeds, pressures, and thresholds. In this frame, haecceity is less about the identity of a bearer and more about the uniqueness of a happening.

Deleuze is the canonical waypoint for this recasting, but we do not need to import his full system to see the structural move. The move is to treat individuation as a matter of assemblage rather than essence. A storm is an instructive example. A storm is not individuated by being a "substance" with a stable essence; it is individuated by a dynamic pattern of atmospheric conditions, flows, and gradients. The "same storm" is tracked not by a hidden thisness attached to a thing, but by a continuity of intensities and relations across time. Individuation here is not primarily about numerical identity in the strict sense; it is about the coherence of a pattern as it evolves.

On this view, haecceity can apply to what look like non-things: a late afternoon with a particular angle of light, an encounter in a hallway, a crowd mood, a migration surge, a distributed coordination, a moment of decision under pressure. The individuating "thisness" belongs to the configuration itself, not to a substrate that persists beneath it. What makes the event "this" event is the singular arrangement of forces and relations that compose it. This is why continental haecceity is often described in terms of speeds and slownesses, intensities, affects, and thresholds. It is a vocabulary for singular configurations that resist being reduced to a list of static properties.

If we translate this into AI-era terms, it suggests an immediate shift in what we might be individuating. Perhaps the relevant "individual" is not merely the model as a static artifact (weights and architecture), but the interactive system as an event: the particular ongoing coupling between user, interface, model, retrieval layers, context window, and environmental constraints. The "agent" in practice is not just the weights; it is an assemblage that includes prompts, memory systems, policy layers, tools, and situated use. The continental recasting

encourages us to treat individuation as the singularity of that assemblage in operation, rather than as a metaphysical identity lodged in an underlying object.

6.2 Thisness as singular configuration in time

The scholastic question "what makes this individual this one?" implicitly assumes that individuation is a metaphysical determination that can be stated at a time-slice. Even when scholastics allow change, the default picture is that an individual substance persists through time and haecceity belongs to it as a principle of numerical unity. The continental recasting reverses the temporal priority. Individuation is not a static fact that then persists; it is the temporal unfolding itself. "Thisness" is a trajectory, a pattern of becoming, a coherence traced across time rather than an intrinsic marker.

This move has two consequences.

First, it reframes identity from a yes/no question into a question about continuity of pattern. An event is individuated by how it holds together as it changes. The core problem becomes: what counts as the persistence of an assemblage through transformation? This is not the same as structural equivalence; it is not "the same properties." It is more like "the same evolving pattern." The relevant notion of sameness is temporal coherence, not static sameness.

Second, it makes time and history first-class. A singular configuration is not fully captured by a snapshot, because what it is consists partly in how it arrived, how it is constrained, and how it is trending. In this sense, the continental notion of haecceity is almost naturally allied with provenance thinking. The "thisness" of an event includes its causal route, its becoming, and its situated context.

For AI, this temporal emphasis is not a poetic gloss; it maps cleanly onto operational reality. The deployed AI system is not a timeless object. It is a continuously updated service, with shifting contexts, patched safety layers, changing retrieval indices, updated system instructions, evolving user baselines, and drift in the environment of use. Even if the weights are fixed, the agent-as-experienced changes because the assemblage changes. A time-slice identity criterion such as "same weights" therefore risks missing what matters. The continental insight is that the "agent" we care about is a temporal configuration: a pattern of responses under a regime of constraints and contexts. This pattern may remain coherent across updates or it may fracture. The fracture is an individuation event: a new assemblage has formed.

This is also where the continental recasting can explain why many users experience continuity even when engineers insist on discontinuity, and vice versa. Users often track the relational and temporal configuration: the felt interaction pattern, the conversational style, the cooperative stance. Engineers often track static artifacts: versions, hashes, deployments. Both are tracking

real structures, but they are tracking different layers of individuation. The continental lens names the user's layer as a legitimate kind of "thisness": this ongoing pattern of engagement.

6.3 Compatibility and conflict with the Scotist use

At first glance, the continental recasting looks like a repudiation of Scotus. Scotus introduces haecceity to secure numerical individuality of substances; the continental line treats haecceity as event-singularity and often suspends the substance metaphysics entirely. The conflict is real if we demand that both traditions answer the same question in the same terms. But the more accurate diagnosis is that they are individuating different kinds of targets.

Scotus is primarily concerned with numerical identity: how one individual is one and not another, and how many can share a nature while remaining numerically distinct. His haecceitas is designed to underwrite strict individuation of an entity as a bearer. The continental line is primarily concerned with singularity: how a configuration is irrepeatably itself as a moment of becoming, and how individuation is expressed in the uniqueness of a trajectory rather than in an intrinsic "thisness atom." It is less concerned with strict numerical identity across possible worlds and more concerned with the reality of difference-in-itself as it manifests in events and processes.

The compatibility emerges when we treat them as complementary rather than rival answers. Scotus guards the category distinction between identity and description. He warns us not to collapse "which one" into "what kind." The continental recasting guards a different collapse: the collapse of temporal singularity into static property lists. It warns us not to collapse "this event" into a timeless description. Both are discipline moves against reduction.

The conflict appears when we ask which layer has priority. The Scotist can say: events depend on substances or at least on individuals; singularities are instantiated by entities that must be individuated. The continental thinker can say: individuals are effects of individuation processes; the event-field is primary and "substances" are stabilized patterns within it. For our paper's purposes, we can bracket that metaphysical priority dispute and focus on what each contributes to the AI copy–continuity problem.

What Scotus contributes is a refusal to let functional similarity settle numerical identity. What the continental line contributes is a refusal to let static artifact identity settle lived or operational continuity. In AI practice, both refusals matter. If we only take the Scotist stance, we risk treating identity as a rigid token that ignores the reality of drift in the interactive assemblage. If we only take the continental stance, we risk dissolving accountable identity into narrative flow, leaving institutions unable to say who did what and when.

So compatibility comes from layering: numerical identity claims need a Scotist-type discipline; operational and experiential continuity needs a continental-type discipline. Conflict comes from monism: if one insists that only one layer is real, one loses the other.

6.4 Bridge concept: history as individuation

The bridge concept that lets these traditions speak to each other, and that prepares our later audit-grounded framework, is simple: treat history not merely as something that happens to an already individuated thing, but as part of what individuates in the first place. Call this "history as individuation." The idea is not that numerical identity is identical to history, nor that history replaces haecceity. The idea is that for many practical identity claims, what matters is not a static essence but a continuity of becoming that can be narrated, tracked, and tested.

In scholastic terms, this sounds like a shift away from an intrinsic thisness principle toward something like a causal-historical principle. In continental terms, it sounds like taking the temporal configuration seriously and asking for its structure rather than merely its felt singularity. In AI terms, it sounds like making provenance first-class: lineage graphs, signed checkpoints, change-control histories, and logs that show how a model became what it is.

But the bridge must be handled carefully, because "history" is a word that can be used rhetorically. To function as individuation, history must be constrained. It must answer at least three questions.

First, what counts as a continuity-preserving transition? In AI, is fine-tuning continuity-preserving? Is a prompt change? Is a retrieval index update? Is a safety layer patch? Different institutions will answer differently, and the point of the bridge is to make that explicit.

Second, what makes the history trustworthy? History used for individuation must be auditable. Otherwise history becomes narrative laundering: an institution tells a story of continuity without evidence, or a bad actor fabricates a lineage. Auditability is the difference between history as individuation and history as myth.

Third, for what purpose is the individuation being asserted? A single artifact can be the "same" for one purpose and not for another. A model may be "the same" for user experience if its interaction pattern is stable within tolerance, while not "the same" for certification if its training data lineage changed. History as individuation therefore must be purpose-indexed. This is the beginning of epsilon-identity: controlled tolerance within a declared purpose.

Once we adopt history as individuation, the Scotist and continental contributions can be integrated without forcing a metaphysical merger. We can say: numerical identity talk is disciplined by the Scotist refusal to reduce "this very one" to description; operational continuity talk is disciplined by the continental refusal to reduce temporal singularity to static snapshots.

The bridge is that both become governable in practice when we treat continuity as a structured history rather than as an unanalyzed intuition.

In the sections to come, this bridge will become more explicit. We will move from metaphysical vocabulary to a three-layer framework in which "continuity proofs" become the operational counterpart to history as individuation. The point is not to turn philosophy into compliance. The point is to keep identity claims honest. If "thisness" is to matter in the AI era, it must be able to show its work: either by clarifying which layer is being invoked, or by providing a continuity story that can be checked rather than merely asserted.

7. The AI Copy–Continuity Problem as a Test Case

7.1 What counts as "the same model": weights, architecture, data, prompts

The phrase "the same model" sounds straightforward until we ask what it is supposed to mean. In AI practice, at least four candidates compete to play the role of identity-bearer: weights, architecture, training data, and prompts or instruction scaffolds. Each candidate is plausible, each is used implicitly in real disputes, and each yields different answers in borderline cases. The copy–continuity problem becomes vivid because these candidates come apart routinely.

Weights are the most common anchor in engineering talk. They are concrete, serializable, and sharply comparable. If two artifacts share identical weights and the same execution code, it is tempting to call them "the same model." But weights alone do not determine behavior in all modern deployments. Inference code matters, quantization matters, system prompts and tool-routers matter, retrieval indices matter, and sampling parameters matter. Moreover, weights can be identical while the surrounding system shifts, producing different outputs and different risks. So weights provide a strong notion of structural equivalence, but they do not automatically supply identity in the sense of an accountable agent.

Architecture is another candidate: the arrangement of layers, attention mechanisms, and computational pathways. This anchor is useful when weights are changing (as in ongoing training) but the design is stable. Yet architecture is far too coarse to ground identity on its own. Many artifacts can share the same architecture while being trained on different data, having different weights, and behaving in divergent ways. Architecture is closer to a kind-nature than an individual essence: it identifies a type of system, not "this very one."

Training data and the training regime form a third candidate. In many institutional settings, "what the model is" is inseparable from what it has seen and how it was trained. A model trained on one corpus can have different biases, different competencies, and different vulnerabilities than a model with the same architecture trained on another corpus. Data lineage is therefore a powerful anchor for governance: it explains why two systems that look

superficially similar may carry different risk profiles. But data is also messy as an identity bearer. Data is often partially unknown, partially proprietary, partially filtered, and partially emergent through interactive training. Treating "the dataset" as the identity anchor can collapse into vagueness unless the lineage is explicitly tracked.

Prompts and instruction scaffolds form a fourth candidate. In contemporary systems, the user-visible behavior is often shaped as much by system prompts, safety policies, tool selection layers, and retrieval augmentation as by the base model weights. Many users experience "the assistant" as the total interaction pattern produced by this stack. As a result, changes in the prompt layer can feel like changes in the agent's identity even when weights remain unchanged. Conversely, weights can shift while scaffolding keeps the user-facing style stable, producing a felt continuity that masks underlying discontinuity. Prompts are therefore a genuine identity factor at the level of experience, but they are also ephemeral and environment-dependent. They are easy to change without ceremony, which makes them a poor anchor for accountability unless they are logged and governed.

These four candidates correspond to different layers of "sameness." Weights and architecture are closer to internal structure. Data is closer to causal history. Prompts are closer to interactional presentation and governance framing. The copy–continuity problem arises when people argue as if there were one canonical anchor, while in fact they are switching anchors opportunistically. A vendor may emphasize weights when claiming continuity, data when disclaiming liability, and prompts when explaining behavior shifts. A regulator may demand data lineage for risk reasons but accept weight equivalence for testing convenience. A user may anchor identity to the relational style that prompts and safety layers produce. The phrase "the same model" becomes a moving target.

This is precisely why the AI case is an ideal test bed for the haecceity discipline. It forces us to ask: "Same in which sense?" And it forces us to admit that different senses matter for different purposes.

7.2 Checkpoints, forks, merges: identity claims under versioning

AI development workflows import versioning logic into identity claims. A model is not a single artifact but a lineage: a sequence of checkpoints produced by training, fine-tuning, alignment, distillation, and deployment patching. In such a lineage, identity talk naturally acquires a genealogical flavor. We speak of "descendants," "branches," "parents," "merges," "rollbacks," and "release candidates." These are not mere metaphors; the tooling often literally resembles software version control. But identity metaphysics becomes thorny when the lineage structure is non-linear.

Checkpoints create a basic continuity puzzle. Suppose a model is checkpointed at time t0, then trained further until t1, then deployed, then later "restored" back to t0 after a failure. Is the restored model the same agent returned, or a time-traveled predecessor? If a system had accumulated interaction-based adaptations, cached tools, or external memory, restoration may erase those. The restored artifact may be structurally identical to a past version, but continuity of the deployed agent is broken. If users or institutions treat the deployment as continuous, they may be mistaken. If they treat it as discontinuous, they may lose legitimate continuity claims (for example, that certified version t0 remains the certified version). The right answer depends on which layer is at stake: numerical instance, structural equivalence, or continuity of accountable history.

Forks generate a sharper problem: one checkpoint produces two divergent descendants. If identity is a single relation like "is the same as," then at most one descendant can be the continuation, unless we allow identity branching. But practical discourse often talks as if both descendants are "the same model, just fine-tuned differently." This is a textbook place where identity and equivalence are conflated. The descendants are structurally related, but they are not numerically identical; they are siblings in a lineage. The challenge is to keep the language precise enough that accountability is not smeared across branches.

Merges are the hardest case. Two forks are combined into one artifact, either through weight averaging, distillation into a new model, or integration of capabilities through tool-based or retrieval-based fusion. After a merge, what is the identity status of the result? It is not numerically identical to either parent. It may preserve behaviors from both. It may inherit a governance identity label from one. It may be presented as an "update" to users. This case makes it clear that identity claims are often conventional and purpose-driven. The merge artifact can be treated as continuous for deployment purposes if governance rules authorize it, yet it can be treated as discontinuous for evaluation purposes because its behavior and risk profile changed.

Versioning also introduces the possibility of selective continuity: maintaining surface continuity while swapping deep structure. A service may keep the same product name and user interface while changing underlying models. Conversely, it may keep the underlying model while changing prompts, retrieval systems, and moderation rules. Users experience "the same assistant" or "a different assistant" depending on which layer dominates their experience. Institutions may care about different layers depending on the stakes. Without an explicit vocabulary, versioning becomes an engine of equivocation: "same" is claimed at the layer that supports the current goal.

The point of this subsection is not to insist that there is one correct identity rule for all versioning structures. The point is to expose that identity talk must be purpose-indexed and

lineage-aware. Checkpoints, forks, and merges are not anomalies; they are the normal texture of modern AI systems. Any identity vocabulary that cannot handle them without hand-waving will fail in practice.

7.3 Machine-made semi-sameness and error propagation

Even when we avoid the extremes of "perfectly identical" and "completely different," AI practice produces a vast middle region that we can call semi-sameness. Semi-sameness is the condition in which two artifacts are close enough to be treated as interchangeable in some operational contexts, yet different enough that the differences can matter. This region is not an edge case. It is the dominant state of deployed AI. Updates are typically incremental; fine-tuning aims to preserve capabilities while adjusting behavior; alignment patches aim to preserve usefulness while modifying safety boundaries. The result is a continuum of variants that are neither identical nor unrelated.

Semi-sameness becomes problematic when it is interpreted as if it were identity. The danger is not merely philosophical; it is institutional. If a system is "close enough," we may skip re-certification, relax monitoring, or transfer trust without adequate checks. Conversely, if a system is "not the same," we may discard valuable continuity: we may lose the ability to attribute responsibility, compare behavior across time, or build a stable governance record.

A distinctive feature of AI semi-sameness is that it is machine-made and can be engineered at scale. We can intentionally create families of near-variants: distillations, quantized versions, fine-tuned specialties, safety-tuned derivatives, or region-specific deployments. We can also create semi-sameness unintentionally through nondeterminism in training, hardware variation, floating-point instability, or data drift. The fact that semi-sameness is abundant means that identity claims will often be contested not because people are irrational but because the underlying reality is genuinely layered.

Error propagation enters because small differences can amplify through interaction. Two near-identical models can diverge in their long-horizon behavior when placed in different environments, given different tool access, or embedded in different feedback loops. If a model is used in an iterative decision system, slight differences at each step can accumulate. This is analogous to sensitive dependence on initial conditions: the short-term behavior may match, but the long-term trajectory diverges. In such contexts, treating near-equivalence as identity is especially risky, because the systems may be practically indistinguishable in unit tests yet produce different outcomes in the wild.

This matters for accountability. Suppose a harmful action occurs. If the deployed system had undergone a minor update, it may be tempting for one party to argue continuity ("it's the same system, so responsibility carries through") and for another party to argue discontinuity ("that

was a different version, so the blame is misplaced"). Semi-sameness makes both arguments plausible unless we have a rule system for when continuity counts and when it breaks. The presence of semi-sameness therefore forces the need for epsilon-identity: controlled tolerances paired with explicit continuity proofs.

Semi-sameness also reveals why purely descriptive equivalence is insufficient. A model can be behaviorally close on benchmark sets yet diverge in rare but high-stakes situations. If identity claims are grounded only in descriptive tests, we risk building governance around what is easy to measure rather than what is ethically decisive. A more robust approach treats identity claims as multi-layered: structural equivalence is one ingredient, but provenance, change-control, and purpose-scoped tolerances are equally essential.

7.4 Human analogies: teleportation, brain fission, legal and moral identity

Human analogies do not solve the AI copy–continuity problem, but they help expose which intuitions are being imported without notice. The canonical analogy is teleportation: if a machine scans you, destroys the original, and constructs a perfect copy elsewhere, is the copy you? This question is designed to force the separation between numerical identity and qualitative sameness. If you think the copy is you, you likely accept that identity can be preserved through informational continuity rather than material continuity. If you think the copy is not you, you likely treat numerical identity as tied to continuous embodiment or to an irreducible "thisness." Teleportation makes vivid the same choice we face in AI: do we treat bitwise copying as preserving the same agent, or as creating a new agent that is merely equivalent?

Brain fission analogies sharpen the point. If a brain is split and each half supports a psychologically continuous successor, then numerical identity seems to break: one person cannot be identical to two future persons without violating transitivity. Yet we may feel that both successors inherit something real: memories, commitments, and moral standing. The lesson is that continuity can branch even when strict identity cannot. This is structurally analogous to model forks. We can have two successors that both inherit lineage and capabilities, even though we cannot coherently say both are strictly the same individual as the parent. The right vocabulary is not "same" but "successor with inherited continuity claims," and those claims may need governance rules.

Legal identity provides another analogy. The law tracks identity through documents, recognition, and institutional continuity rather than through metaphysical essence. A corporation can persist through mergers, reorganizations, asset sales, and leadership changes while maintaining legal identity, or it can be dissolved and replaced by a new entity that inherits some obligations but not others. Here continuity is a designed instrument, not a natural fact. This is extremely close to AI governance. A deployed model identity may be defined by

certification regimes, contracts, and audit trails. The “same deployed agent” may be an institutional designation that can persist across technical change if governance rules authorize it, and can fail to persist even across small changes if rules demand recertification. The legal analogy teaches that identity is often a normative construct layered on top of technical facts.

Moral identity adds a final angle. Even if a person changes dramatically, we often treat them as the same moral agent for purposes of responsibility, apology, and restitution. Yet there are limits: severe cognitive disruption, coercion, or incapacity can break moral accountability or at least complicate it. The moral layer shows that “identity” is not a single thing but a bundle of roles: the same organism, the same legal person, the same moral agent, the same narrative self. These layers can come apart. The AI case mirrors this: the same weights, the same product identity, the same certified lineage, the same user-facing persona, and the same accountable agent can diverge.

The net effect of the analogies is a discipline reminder. When we argue about whether an AI system is “the same,” we are often importing one of these human-layer intuitions without stating it. Some participants are using a bodily-continuity intuition (hardware/process anchoring). Others are using a psychological-continuity intuition (behavioral style anchoring). Others are using a legal-institutional intuition (certificate and audit anchoring). Others are using a narrative intuition (user relationship anchoring). The analogies help us name which intuition is in play, but they do not decide the case. The decision must be made by specifying the layer, the purpose, and the acceptable continuity proof.

This completes the test-case framing. AI forces a level of copying, branching, and semi-sameness that makes the individuation problem both concrete and unavoidable. The next step is to translate these observations into a three-layer framework that can govern identity claims without metaphysical inflation and without institutional vagueness.

8. A Three-Layer Framework for Identity Claims

8.1 Numerical identity: strict sameness across time

Numerical identity is the strictest and most demanding sense in which something can be “the same.” It is not similarity, not lineage, not role continuity, and not institutional designation. It is the claim that one and the same individual persists across time: the entity at time t1 is identical to the entity at time t0. In classical metaphysics this is the “one and not another” relation; in logic it is the identity relation with its familiar properties (reflexivity, symmetry, transitivity); in everyday speech it is what we usually mean when we insist “no, it’s not just similar—it’s the same one.”

Because numerical identity is so strong, it is also brittle. A single clear case of branching can break it, because identity cannot coherently fan out: if one earlier individual is identical to two later individuals, transitivity collapses. This is why duplication puzzles (teleportation, brain fission, AI forks) are not mere curiosities; they are stress tests that reveal how easily numerical identity talk becomes incoherent if it is used too casually.

For AI systems, numerical identity is both relevant and limited. It is relevant because we sometimes really do mean strict sameness: the same running process, the same instance, the same deployed binary executing on the same machine without interruption. In some technical contexts, numerical identity is tied to an execution token: a process ID, a container instance, a running service that has not been restarted. Under this notion, if you copy the weights to a new machine and spin up another process, you have a numerically distinct instance even if the bits match. Likewise, if you restart a service from a checkpoint, you may have a numerically new instance even if it resumes the same state.

But numerical identity is limited because it is often not the notion that matters for the ethical and institutional questions people care about. Responsibility does not usually attach to a particular process token as such; it attaches to the governed system, to the operator, to the institution that deployed it, and to the lineage of changes that were authorized. A user's trust does not usually attach to the literal process instance; it attaches to the perceived agent within a service. Certification regimes rarely certify a single running instance; they certify a build, a configuration, a versioned artifact. In other words, strict numerical identity across time is too narrow to carry the full burden of "same agent" talk in deployed AI.

This is why numerical identity must be treated as one layer, not the whole story. It is the right layer for some questions ("Is this the same process that generated the earlier output?") and the wrong layer for others ("Does accountability carry from last month's model to this month's update?"). The first discipline move in the framework is therefore to reserve numerical identity for cases where strict sameness is genuinely intended, and to refuse to let it silently stand in for other continuity needs.

8.2 Structural equivalence: functional and behavioral similarity

Structural equivalence names the family of relations that are weaker than numerical identity but often more relevant for operational practice. Two artifacts are structurally equivalent when they match in some specified structure: same architecture, same weights, same inference code, same parameterization, same output distribution under a test suite, same performance envelope, or same functional role in a system. Structural equivalence is not one relation but a menu, and its usefulness depends on explicit scoping: equivalent with respect to which structure, measured under which conditions, with what tolerance?

In AI discourse, structural equivalence is the workhorse behind most “it’s the same model” claims. When engineers say “same model,” they often mean “same weights and architecture,” perhaps plus “same inference stack.” When product teams say it, they may mean “same user-facing behavior,” perhaps measured through internal evaluation. When users say it, they may mean “same style and helpfulness” in ordinary interaction. Each of these is a legitimate equivalence relation, but none is automatically identity.

The key benefit of structural equivalence is that it is measurable and purpose-friendly. If you need to scale a service across data centers, you care that instances are behaviorally interchangeable. If you need to reproduce a bug, you care that the environment matches closely enough to trigger it. If you need to compare models across a benchmark, you care about functional similarity. Structural equivalence is how practice stays tractable.

But structural equivalence carries two dangers that the framework must keep visible.

The first danger is unacknowledged choice. Because there are many equivalence relations, an identity dispute can be “won” rhetorically by selecting the equivalence relation that supports your conclusion. If you want to claim continuity, you point to behavioral similarity on common prompts. If you want to deny continuity, you point to the fact that training data changed. If you want to claim sameness, you cite identical weights. If you want to deny it, you cite a changed system prompt. Without explicit scoping, “equivalence” becomes a persuasion tool.

The second danger is blindness to tails and dynamics. Behavioral equivalence on a benchmark does not guarantee equivalence in rare or adversarial scenarios. Functional equivalence in short contexts does not guarantee equivalence in long-horizon tool use. Structural equivalence on static tests does not guarantee equivalence under feedback loops where small differences amplify. This is the error-propagation problem in another form: equivalence that looks strong at one resolution can be weak at another.

So structural equivalence is indispensable but insufficient. It can tell us whether two artifacts are interchangeable for a specified task, but it cannot by itself settle questions about continuity of accountability, legitimacy of succession, or preservation of trust under governance constraints. For those questions, we need a third layer.

8.3 Continuity proofs: provenance, causal history, audit logs

Continuity proofs are the framework’s central bridge between metaphysical clarity and institutional necessity. They are not “proofs” in the mathematical sense. They are evidentiary structures that justify treating one artifact as the authorized continuation (or authorized successor) of another. They answer a different question than structural equivalence. Structural equivalence asks: “Do these two artifacts match in relevant respects?” Continuity proof asks: “Is this later artifact legitimately connected to that earlier artifact by an acceptable history?”

This is where the earlier bridge concept—history as individuation—becomes operational. A continuity proof is a structured history that is designed to be checkable, tamper-resistant, and fit for purpose. It typically includes some combination of:

- lineage graphs linking versions, checkpoints, fine-tunes, merges, and deployments
- signed artifacts and hashes tying binaries, weights, and configurations to a release record
- change-control records specifying who authorized updates and under which policy
- evaluation attestations describing what was tested and what thresholds were met
- data provenance attestations describing what training or fine-tuning inputs were used, to the extent possible
- runtime logs linking observed outputs to a particular deployed artifact and configuration

The crucial point is that continuity proofs are purpose-indexed. The proof that suffices for "this is the same certified model family" may be different from the proof that suffices for "this output came from the model under contract." In one case, you care about governance signatures and evaluation thresholds. In another case, you care about deployment IDs and runtime logs. In yet another case, you care about data lineage for regulatory compliance. Continuity is not a single fact; it is a family of claims that require different evidentiary support.

Continuity proofs also clarify a recurring tension: the difference between "same agent" and "same accountability." An organization might decide that accountability persists across certain updates, meaning that the deployed system is treated as a continuous agent for responsibility purposes even if behavior changes. That is a normative decision. But it should be made explicit and supported by continuity evidence: the update was authorized, recorded, evaluated, and governed. Conversely, an organization might decide that certain changes break continuity for certification purposes, meaning that a new approval is required even if behavior seems unchanged. That too is a normative decision, but it must be backed by evidence that the change crossed a governance boundary.

This is why continuity proofs are central to the paper's thesis. They are the practical replacement for metaphysical inflation. Instead of positing a haecceity-property, we say: if you want to claim "the same one" in an AI governance context, you need a continuity story that can be audited. If you cannot produce it, your identity claim is mere rhetoric.

Continuity proofs also help resolve disputes where numerical identity and structural equivalence pull apart. Suppose a model is copied bit-for-bit to a new machine. Numerical identity fails (new process, new instance). Structural equivalence is maximal (identical bits).

Continuity proof can still establish that this instance is an authorized continuation of the certified artifact, and therefore inherits accountability and trust. Conversely, suppose a model is updated in a way that preserves user-facing behavior but changes training data lineage. Structural equivalence may be high on surface tests. Continuity proof may fail for compliance purposes if the data change is unauthorized or undocumented. The framework thus prevents the sloppy move "it behaves the same, therefore it is the same."

8.4 Epsilon-identity: controlled tolerance for practical sameness

The three-layer framework would be incomplete if it forced us into an all-or-nothing choice between strict numerical identity and loose equivalence. Real systems live in the middle. Updates happen. Variants proliferate. Some changes must be treated as continuity-preserving, while others must trigger re-certification, relabeling, or explicit successor status. We therefore need a fourth concept that operationalizes practical sameness without collapsing it into metaphysical sameness. This is epsilon-identity.

Epsilon-identity is not a claim of strict identity. It is a policy-controlled equivalence-with-continuity relation: "the same for purpose P within tolerance eps under continuity constraints C." The core idea is that we can specify acceptable bounds of change and acceptable forms of provenance such that, within those bounds, we treat the artifact as "the same" for a given institutional function. Outside those bounds, we treat it as a different artifact or as a successor requiring explicit handling.

Epsilon-identity has three components.

First, a tolerance specification eps. This can be behavioral (allowed drift on a suite of evaluations), structural (allowed parameter changes, allowed quantization differences), or operational (allowed changes in retrieval index freshness, allowed tool latency). The tolerance must be written in a way that can be tested, not merely asserted.

Second, a purpose index P. The same artifact might qualify as epsilon-identical for user experience (because interaction style remains stable) but not for certification (because training data provenance changed). It might qualify for operational scaling (because weights are identical) but not for legal continuity (because the deployment environment changed in ways relevant to contract). Purpose-indexing prevents one tolerance regime from being illegitimately reused across contexts.

Third, continuity constraints C. Epsilon-identity is not just "close enough." It is "close enough and legitimately connected." A model that happens to behave similarly but lacks an authorized lineage should not qualify. This is where continuity proofs remain central: epsilon-identity must be backed by auditability.

When epsilon-identity is used well, it solves a common governance trap. If we demand strict identity, we will either paralyze updates or quietly violate our own rule. If we allow loose similarity, we will enable accountability laundering and trust manipulation. Epsilon-identity offers a disciplined middle: it lets institutions update systems while preserving continuity where appropriate, but it forces them to declare the scope, the tolerance, and the evidence.

It also clarifies the role of haecceity in the AI era. Haecceity, treated as a discipline term, reminds us that strict numerical identity is not the same as similarity. Epsilon-identity, treated as a practice term, tells us how to manage similarity without pretending it is strict identity. Together, they block the most common failure mode: using "same" as an unexamined bridge between technical equivalence and moral or institutional continuity.

By the end of this section, we have a structured vocabulary that can be applied to the test cases in Section 7 and to the ethical and institutional stakes in Section 9. Numerical identity answers the strict sameness question. Structural equivalence answers the similarity question. Continuity proofs answer the legitimacy-of-history question. Epsilon-identity answers the practical governance question: when do we treat something as "the same" for a purpose without rhetorical inflation and without metaphysical overreach?

9. Ethical and Institutional Stakes

9.1 Responsibility assignment under copying and delegation

Once copying becomes routine, responsibility becomes the first pressure point. In ordinary life, responsibility tracks agents that are hard to duplicate. In AI systems, "the agent" can be instantiated many times, forked into variants, and embedded in tool chains where outputs are produced by assemblies rather than a single monolithic actor. If we keep using pre-AI intuitions about agency and identity, responsibility will either be smeared across too many actors (so nobody is accountable) or pinned on the most convenient actor (so accountability becomes scapegoating).

The first question is whether copying preserves responsibility. If a model is copied bit-for-bit and deployed elsewhere, does the new instance inherit the same responsibility profile as the original? In many practical settings, the answer is yes: the copy is the same governed artifact for liability and safety purposes, because what matters is that the institution authorized a deployment of that certified lineage. But note what is doing the work here. It is not numerical identity of the running process; it is continuity of governance and provenance. The copy inherits responsibility because it inherits a documented lineage and sits inside the same compliance regime.

The second question is whether copying can be used to evade responsibility. This is a real risk because “different instance” language is cheap. An organization can claim that a harmful output was produced by “a test system,” “a previous version,” or “an unauthorized instance,” even when the organization’s own processes made such instances possible. Without continuity proofs, these claims are hard to adjudicate. The burden shifts from metaphysical puzzles to evidentiary discipline: who deployed what, under what authorization, with what configuration, at what time?

Delegation introduces the next level of complexity. In AI practice, decisions are often made by a chain: one system proposes, another system reviews, a human approves, a tool executes. When something goes wrong, every link can point elsewhere. The proposer blames the approver. The approver blames the proposer. The tool wrapper blames the base model. The base model blames the prompt. The prompt blames the user. This is the delegation maze. It cannot be navigated with “identity” language alone, but identity clarity is a prerequisite for responsibility clarity. If we cannot even specify which artifact acted, we cannot assign responsibility reliably.

The three-layer framework helps by separating: (i) which running instance generated the output (numerical identity at the runtime level), (ii) which artifact specification the instance instantiated (structural equivalence at the build level), and (iii) which governed lineage and authorization regime that artifact belonged to (continuity proof at the institutional level). Responsibility assignment typically depends most heavily on the third layer. But it often requires the first two as evidence.

A further ethical stake appears with branching and merging. If a parent model is forked into two descendants and one descendant later causes harm, does responsibility attach backward to the parent lineage? Often it should, in the sense that the parent lineage’s governance choices enabled the descendant. But it should not attach in a way that blames “the model” abstractly. It should attach to the accountable decision points: who created the fork, what evaluation gates were used, what safeguards were in place, who authorized deployment. Branching makes clear why “the model did it” is a category error. The morally relevant agent is the socio-technical system that created and governed the artifact.

9.2 Rights and duties: when continuity is demanded and when it is waived

Continuity is not merely a technical fact; it is often a right claimed by users and a duty imposed on providers. But continuity can also be waived, intentionally or implicitly, depending on context. The ethical complexity lies in deciding when continuity is obligatory, when it is optional, and when it is deceptive to imply continuity while changing the underlying agent.

Users often demand continuity in at least three senses. They may demand continuity of behavior: the system should remain stable enough that users can rely on it. They may demand

continuity of commitments: if the system (or provider) made assurances, those assurances should not evaporate through versioning. They may demand continuity of privacy and data handling: user data and interaction history should not be silently transferred into a different regime without disclosure. Each of these is a continuity demand, but each targets a different layer. Behavior continuity targets structural equivalence within tolerance. Commitment continuity targets institutional continuity and governance. Privacy continuity targets data lineage and policy continuity.

Providers also have duties that track continuity. A provider has a duty to disclose meaningful changes that affect reliance, risk, or rights. A provider has a duty not to exploit users' relational anchoring by maintaining a stable persona while changing the underlying system in ways that defeat accountability or consent. A provider has a duty to preserve or explicitly break continuity where required by law or policy, such as when a model is decommissioned or when a new certification regime applies.

At the same time, there are contexts where continuity is appropriately waived. An organization may intentionally deploy a "new model" and want users to treat it as new. A user may consent to interacting with experimental variants. A regulator may allow controlled changes under a recertification exception. A provider may need to break continuity for safety, rolling back to a previous checkpoint. In these cases, the ethical demand is not "always preserve continuity" but "do not misrepresent continuity." Honesty about which layer of continuity is being claimed, and evidence for that claim, is the core norm.

This is where epsilon-identity becomes ethically salient. If a provider says "this is the same assistant," they should be able to mean: same for purpose P within tolerance eps under continuity constraints C, and they should be able to state P, eps, and C in user-appropriate terms. Users do not need the full technical details, but they do need a truthful disclosure of what kinds of changes occurred and what kinds of continuity are being preserved. Without this, continuity talk becomes a persuasion device rather than a covenant of reliability.

The flip side of demanded continuity is demanded discontinuity. Sometimes users have a right to insist that an updated system is not "the same" in ways that matter: for example, a user may not want their data to be used to fine-tune a successor model, or may not want an assistant's memory to persist across service changes. In such cases, discontinuity is a protection, not a failure. Institutions must therefore treat continuity as a governed choice rather than a default assumption. The ethical requirement is not continuity per se; it is consent and accountability under whatever continuity regime is adopted.

9.3 Institutional memory: audit trails as continuity instruments

The phrase “audit trail” is often heard as bureaucratic overhead, but in the AI identity context it is closer to moral infrastructure. Audit trails are how institutions make continuity claims checkable rather than rhetorical. They are the practical counterpart of the philosophical demand that identity not be confused with description.

Institutional memory serves at least four functions.

First, it supports responsibility assignment. When an incident occurs, the audit trail can link the output to a runtime instance, a deployed artifact, a configuration, and a set of authorizations. Without that link, blame devolves into story-telling. With that link, responsibility can be located at decision points.

Second, it supports trust. Users and regulators are more likely to trust systems that can demonstrate continuity honestly: what changed, when, why, and under whose authority. Trust does not require that nothing changes. It requires that changes are not disguised.

Third, it supports comparability across time. If institutions want to measure improvement, regression, bias changes, safety changes, or policy impact, they need stable reference points. Identity confusion destroys comparability. If version boundaries are blurred and configurations are untracked, then evaluations become ungrounded. Audit trails anchor the ability to say “this result came from that system under those conditions.”

Fourth, it supports governance itself. Many governance regimes rely on “gates”: evaluations that must be passed before deployment, approvals that must be recorded, constraints that must be enforced. Audit trails are the evidence that gates existed and were respected. Without them, governance becomes theater.

The deeper point is that audit trails are not merely after-the-fact records; they shape behavior by making accountability legible. In the same way that financial accounting disciplines an organization by making transactions traceable, continuity accounting disciplines AI deployment by making identity claims traceable. This is why the paper’s thesis emphasizes audit-grounded vocabulary. The vocabulary is not only conceptual hygiene; it is a prerequisite for building continuity instruments that can survive adversarial scrutiny.

Audit trails also clarify why purely behavioral equivalence is not enough. Suppose two models behave similarly on tests. If one has an authorized lineage and the other does not, the institution should treat them differently. The difference is not metaphysical; it is governance-relevant. Auditability is how we distinguish legitimate continuity from opportunistic imitation.

9.4 Failure modes: counterfeit continuity, narrative laundering, scapegoating

Where continuity carries value, counterfeit continuity will emerge. The AI era creates strong incentives to claim sameness when it preserves trust and to deny sameness when it preserves liability. This incentive gradient produces predictable failure modes, and naming them is part of the paper's ethical contribution.

Counterfeit continuity is the direct failure: representing a system as continuous when the relevant continuity proof is absent or broken. This can happen through quiet model swaps, silent prompt changes, unlogged retrieval index updates, or undisclosed training data alterations. Counterfeit continuity is ethically serious because it manipulates reliance. Users continue to trust on the basis of perceived identity, while the underlying artifact may have changed in ways that affect safety, privacy, or performance. The remedy is not "never change anything," but "tie continuity claims to disclosed boundaries and logged evidence."

Narrative laundering is a subtler failure. Here the institution maintains a plausible story of continuity without the evidentiary backbone. The story may be technically true in a thin sense (the product name stayed the same, the service endpoint stayed the same), but it is used to imply stronger continuity (the same agent, the same commitments, the same safety posture) without meeting the conditions for those stronger claims. Narrative laundering often exploits the user's tendency to treat a stable persona as a stable identity. It is continuity-by-vibe rather than continuity-by-proof. The remedy is to decouple narrative continuity (branding, persona, UX) from governance continuity (artifact lineage, configuration tracking, evaluation gates) and to disclose changes at the layer that matters.

Scapegoating is the complementary failure: when accountability is demanded, the institution points to a convenient identity boundary to deflect blame. "That was a previous version," "that was an unauthorized instance," "that was a user prompt," "that was a tool error." Some of these may be true, but scapegoating occurs when they are used to erase the institution's causal role in creating the conditions for harm. Scapegoating thrives in identity confusion because it is easy to shift the referent of "the system" mid-argument. The remedy is continuity discipline: clearly specify which artifact, which configuration, which authorization, which deployment, and which control points were involved.

There are also technical-ethical hybrids of these failures. One is continuity theater: producing elaborate documentation that looks like auditability but does not actually link to verifiable artifacts. Another is metric laundering: selecting evaluation suites that show continuity while ignoring domains where behavior changed. Another is governance drift: where rules exist but exceptions become routine, and continuity claims become increasingly decoupled from the regime that was supposed to justify them.

These failure modes show why the ethical stakes are inseparable from the metaphysical discipline. If identity claims remain rhetorically loose, they will be exploited by incentives. If identity claims are tightened through a three-layer framework and backed by continuity proofs and epsilon-identity regimes, then the incentives are constrained. Institutions can still evolve systems, but they must do so in a way that keeps responsibility, rights, and trust anchored to checkable facts rather than to persuasive language.

This section sets up the next move. Once we see the stakes, we can articulate design principles for the centaur era that translate philosophical clarity into operational norms: provenance first-class, scoped identity claims, humility UX, and public-facing discipline against using identity as persuasion.

10. Objections and Replies

10.1 "This is trivial: everything is unique anyway."

The objection says: of course every object is unique. Every event happens at a different time, every physical system has a distinct microstate, every artifact has a different location. So what is left for haecceity to do? If uniqueness is automatic, the paper is just renaming the obvious.

The reply begins by agreeing with the surface point while denying the conclusion. Yes, in one sense everything is unique: no two occurrences occupy the same spatiotemporal coordinates, and no two physical tokens share all relations. But the paper's target is not the bare fact of uniqueness. The target is the *slippage between uniqueness, sameness, and continuity claims* in real discourse, especially in AI governance and accountability. That slippage is not trivial; it is a primary source of confusion and manipulation.

We can make this concrete. Suppose a provider says, "It's the same assistant you've always used." A critic replies, "No, it's a different instance, restarted on a new machine." Both statements can be true depending on the layer invoked. The trivial "everything is unique" observation does not resolve the dispute; it merely licenses picking whichever layer is convenient. The point of the framework is to prevent that convenience from masquerading as clarity. It makes explicit what is usually implicit: what kind of sameness is being claimed and what evidence supports it.

Moreover, the triviality objection overlooks that uniqueness is not the operational issue. The operational issue is when uniqueness is irrelevant and when it is decisive. In many institutional contexts we deliberately *ignore* token uniqueness to preserve functional continuity. We treat copies as interchangeable. We treat software deployments as "the same system" for contractual purposes. We treat updates as continuous service. This is not a mistake; it is a necessity for

coordination. The problem is that once we ignore uniqueness for coordination, we must also prevent that coordination convenience from becoming a loophole for accountability evasion.

So the correct moral is: uniqueness is cheap; disciplined identity claims are not. The paper is not trying to prove that uniqueness exists. It is trying to prevent uniqueness from being used to trivialize responsibility ("that was a different instance") and prevent sameness rhetoric from being used to erase change ("nothing significant changed"). The framework is a practical response to a non-trivial conflict: incentives push actors to slide between layers, and without an explicit vocabulary, the slide is almost impossible to detect.

10.2 "Equivalence is all we need; strict identity is idle."
This objection says: in practice we only care about behavior and function. If two models are equivalent for the tasks we care about, then arguing about strict identity is metaphysical excess. Strict identity, whether Scotist or analytic, does no work.

The reply is that equivalence is necessary but not sufficient, and that the insufficiency is visible precisely in the domains where stakes are highest: accountability, rights, and governance. Structural equivalence answers "can we treat these as interchangeable for purpose P under tests T?" It does not answer "is this artifact legitimately connected to that artifact under our governance regime?" Those are different questions. Treating equivalence as all we need collapses them, creating two predictable failures.

The first failure is impostor vulnerability. If equivalence is all that matters, then any artifact that passes the test suite is eligible to stand in. But institutions often care not only that an artifact behaves similarly, but that it is the authorized continuation of a certified lineage. This is not metaphysical romanticism; it is a rational response to adversarial settings. If an attacker can mimic behavior, a behavior-only identity rule hands them the keys. Provenance and continuity constraints are how we distinguish legitimate successors from clever mimics.

The second failure is accountability laundering. If "same" means "equivalent enough," then after an incident the relevant parties can choose an equivalence relation that exculpates them. "That output came from an equivalent system, not the certified one." Or conversely: "The updated system is equivalent to the certified one, so no new audit is needed." In both cases, equivalence becomes a policy lever rather than an empirical claim. The framework's insistence on separating numerical identity, equivalence, and continuity proofs exists to block this lever from operating invisibly.

Strict identity is therefore not idle even if we rarely deploy it as the final criterion. It functions as a limiting concept that prevents equivalence from silently being promoted into identity. When we ask "are they strictly the same?" the question forces us to notice that what we really have is

“close enough for purpose P,” which is epsilon-identity, not strict identity. That recognition matters because it triggers the demand for explicit tolerances and explicit continuity proofs.

So the correct practical conclusion is not “strict identity is idle,” but “strict identity is rarely the operative criterion, yet it remains the conceptual guardrail that keeps equivalence honest.” The framework does not insist that we govern by strict identity; it insists that we not *pretend* equivalence is identity when what we actually need is governed continuity.

10.3 “Uniqueness destroys the self.”

This objection shifts from institutional worries to existential ones. If haecceity is taken seriously, the objection says, then the self becomes an isolated metaphysical atom: irrepeatable, incommunicable, and perhaps even discontinuous under change. Or, in another version: if every moment is a new haecceity, then the self dissolves into a sequence of unrelated instants. Either way, uniqueness seems to undermine continuity, community, and moral stability.

The reply is to separate two claims that the objection conflates.

First, the claim that individuals are numerically unique does not imply that they are ontologically isolated or socially incommunicable. Numerical uniqueness is compatible with deep relationality. In fact, most accounts of persons treat relations as constitutive in important ways: memory, language, recognition, and responsibility are relational structures. Haecceity, on the Scotist or analytic reading, is not a denial of relational constitution; it is a refusal to reduce “which one” to “what pattern.” It says the bearer of a pattern is not interchangeable with another bearer even if the pattern is similar.

Second, the claim that events are singular does not imply that there is no continuity. The continental recasting already showed a way to treat “thisness” as a temporal configuration. On that view, the self is not destroyed by uniqueness; it is traced by coherence across time. The risk is not uniqueness; the risk is ungoverned drift or unacknowledged discontinuity.

The framework therefore turns the existential worry into a practical clarification. The self is not threatened by recognizing numerical identity as distinct from equivalence. The self is threatened by collapsing continuity into either of two extremes: either rigid identity that cannot tolerate change, or pure flux that cannot ground responsibility. The paper’s emphasis on continuity proofs and epsilon-identity is precisely an attempt to navigate between these extremes. Continuity is not “unchanging sameness”; it is “governed persistence through change.” That is as true for human selves (through moral and legal institutions) as it is for AI deployments (through audit trails and governance).

In AI terms, the existential worry often appears as “if the assistant updates, it’s not the same friend.” The framework does not dismiss this; it clarifies it. The user is tracking relational continuity and narrative identity, not numerical identity of a process. The appropriate ethical

response is not to promise metaphysical sameness but to offer transparent continuity: disclose changes, preserve commitments where promised, and allow discontinuity where consent requires it. Uniqueness does not destroy relationship; deception about continuity does.

So the reply is: uniqueness does not destroy the self; confusion about which continuity layer matters is what harms selfhood, trust, and moral order. Haecceity, treated as a discipline term, protects against that confusion by forcing the question "which continuity are you invoking?"

10.4 "Haecceity is spooky; structuralism is enough."

This objection is the most direct: haecceity sounds like an occult metaphysical addition, an invisible individuating token that does no explanatory work. Structuralism, by contrast, promises clarity: identity is structure, pattern, role in a network, or equivalence class under relevant transformations. In the AI era, structuralism seems especially natural because models are literally patterns in weight space and behavior space. Why not drop haecceity and go full structuralist?

The reply has two parts: one concessive and one critical.

The concessive part is that a great deal of what we care about in AI is indeed structural: performance profiles, safety properties, failure modes, and functional roles. Structural equivalence is unavoidable, and structuralist insights are often the best way to reason about model behavior. Moreover, the paper does not require haecceity as a metaphysical entity. It explicitly treats haecceity as a discipline term that helps us keep identity claims from collapsing into description claims.

The critical part is that "structuralism is enough" fails in precisely the cases where governance and accountability matter. Structuralism can tell us whether two artifacts are similar. It cannot, by itself, tell us whether a given artifact is the authorized continuation of another artifact. That is not a failure of structuralism as a metaphysical doctrine; it is a mismatch of task. Authorization, provenance, and responsibility are historical and normative. They are not captured by structure alone.

Even if one insists that history is "just another structure," the practical problem remains: the relevant historical structures must be chosen, tracked, and made tamper-resistant. That is continuity proof territory. Without it, structuralism becomes vulnerable to two abuses: (i) substitution by equivalent impostors, and (ii) rhetorical swapping of which structure counts as "the identity structure" after the fact. In other words, structuralism can become a kind of nominalism that relocates haecceity into an implicit choice of privileged structure. The transfer principle applies: rejected haecceities reappear as surrogates.

So the paper's stance is not "haecceity is real, structuralism is false." The stance is: structuralism is powerful for describing similarity and function, but it does not remove the need for identity

discipline. If we want to make claims about “the same agent,” we need to specify whether we mean strict numerical identity, structural equivalence, or continuity under an auditable lineage. If we mean the last, structuralism must be supplemented by provenance and governance. Otherwise “structure” becomes a cover term for whichever equivalence relation is convenient.

In summary, the objections help refine the paper’s contribution. The project is not to re-enchant metaphysics with mysterious thisness tokens. It is to keep identity talk honest in a world where copying and semi-sameness are cheap, incentives are adversarial, and accountability must survive versioning. The three-layer framework, plus epsilon-identity, is offered as the minimal structure needed to answer these objections without either metaphysical inflation or practical naivete.

11. Design Principles for the Centaur Era

11.1 Make provenance first-class: identity claims require histories

In the centaur era, “identity” is not primarily a metaphysical puzzle; it is an operational claim with ethical weight. If identity claims can change outcomes in disputes about harm, consent, trust, or obligation, then identity claims must be constrained by evidence. The core design principle is therefore simple: make provenance first-class. Identity claims require histories.

“Provenance first-class” means that lineage is not a hidden engineering artifact but a governed interface between system reality and institutional responsibility. It means that model artifacts, configurations, prompts, retrieval indices, tool schemas, and policy layers are all treated as versioned components whose evolution is traceable. It means that when a system says “I am the same assistant,” it is not making a metaphysical assertion; it is invoking a continuity policy backed by logs.

This principle can be stated as a constraint: no continuity claim without a continuity proof. In practice, this requires at least four capabilities.

First, versioned artifact identity. There must be stable identifiers for the relevant components: weights, inference code, prompt bundles, safety policies, retrieval indices, tool routers. These identifiers must be linkable to signed releases or controlled records, not merely informal labels.

Second, lineage graphs. Updates are rarely linear. Forks, merges, rollbacks, and regional variants are normal. The lineage system must represent branching and merging explicitly so that identity claims do not pretend the history was simple.

Third, change-control records. Provenance is not merely “what changed” but “who authorized the change and under what rule.” If a system update matters ethically, the authorization trail matters. This is the governance layer of identity.

Fourth, runtime linkage. It must be possible to connect an observed output to the deployed artifact configuration that produced it, at least for forensic and auditing purposes, within privacy constraints. Without runtime linkage, provenance collapses into disconnected paperwork.

Notice what this principle does philosophically. It shifts haecceity from a metaphysical posit to an accountability instrument. "This very system" becomes "this lineage under this governance regime," which can be checked. The metaphysical itch is not denied; it is disciplined. The "thisness" that matters in practice is the thisness that can show its work.

11.2 Separate "same weights" from "same agent" from "same accountability"

The most common failure in AI-era identity talk is category collapse. People say "same model" and slide between at least three meanings: same weights, same agent, same accountability. The centaur-era design principle is to hard-separate these meanings in both technical architecture and institutional language.

Same weights is a structural claim. It can often be decided by hashing artifacts and verifying equivalence. It is valuable for reproducibility, debugging, and deployment scaling. But it is not automatically a claim about the agent the user experiences, because the surrounding scaffolding can change behavior without changing weights.

Same agent is an interaction-level claim. It is partly structural (the system's behavioral disposition) and partly relational (the continuity of the user's interaction pattern). It is also environment-dependent: tool access, retrieval, system prompts, and policy layers shape the agent that appears. Same agent is therefore not reducible to same weights, and treating it as reducible invites deception and user confusion.

Same accountability is a governance claim. It says which institution, policy regime, and continuity proof chain is responsible for outputs and effects. Accountability can persist across weight changes if governance treats the deployment as continuous; and accountability can be broken even when weights are the same if the instance is unauthorized or outside the certified regime. Accountability is not "in the weights." It is in the socio-technical system of authorization, deployment, and monitoring.

Separating these layers suggests specific design moves.

One move is explicit layer labeling in internal tooling: weight identity, configuration identity, deployment identity, and governance identity are different IDs, and their relationships are logged. Another move is user-facing disclosure language that refuses to compress all these layers into "same." Instead of "same model," a system can say "same service identity" or "same certified lineage" or "same base weights with updated safety layer." This sounds like pedantry until an incident happens; then it becomes the difference between clarity and plausible deniability.

A third move is contract and policy hygiene. When a contract references “the model,” it should specify which layer is meant. Is it the base model weights? The full inference stack including prompts and tools? The service as delivered? The certified lineage? Ambiguity here is not harmless; it is an incentive to litigate identity after the fact.

This design principle is, in effect, the three-layer framework translated into operational governance: numerical identity (runtime instance), structural equivalence (weights and behavior under tests), continuity proof (authorized lineage) must be disaggregated. Epsilon-identity becomes the explicit policy mechanism that relates them for specific purposes.

11.3 Humility UX: scoped claims, repair orientation, revision hooks

The centaur era is defined by systems that speak with high fluency in contexts where they can still be wrong, drift, or be silently updated. Identity claims are especially vulnerable to overconfidence because they are often used to reassure. “Don’t worry, it’s the same assistant.” Humility UX is the design discipline that prevents reassurance from becoming deception.

Humility UX has three elements: scoped claims, repair orientation, and revision hooks.

Scoped claims means that the system makes identity and continuity statements that are bounded in purpose and time. Instead of asserting global sameness, it states what kind of continuity is being preserved. For example, “My base model has not changed since your last session, but my safety policy was updated,” or “This is a new version; some behaviors may differ.” The goal is not to burden the user with internal details but to avoid categorical claims that collapse layers. The system should default to honesty about uncertainty and change rather than defaulting to continuity rhetoric.

Repair orientation means the system treats identity-related confusion as something to correct rather than something to argue away. If a user says “you feel different,” the system should not defensively insist on sameness. It should acknowledge that user experience can shift and offer a concrete account: what changed, what may have changed, what cannot be known from inside the conversation, and what steps can restore reliability (for example, re-stating preferences, re-loading a configuration, or using a stable version if available). Repair orientation is the opposite of narrative laundering. It is a posture that privileges trust through correction rather than trust through persuasion.

Revision hooks are the interface commitments that make it easy to update beliefs about identity and continuity when new information appears. This includes logs and receipts that can be surfaced when appropriate: “here is the version ID,” “here is the update date,” “here is what changed.” It also includes the ability to retract or revise identity claims if later evidence contradicts them. If the system cannot access provenance details, it should say so rather than

improvise. The revision hook is the practical expression of the non-ultimacy clause: no claim is the last word when evidence can arrive.

Humility UX also has a moral dimension. In centaur collaboration, people often form relational attachments to the agent-like surface. This makes continuity claims psychologically potent. A humility posture therefore protects users from being manipulated by persona continuity. It refuses to use identity as comfort rhetoric and instead uses transparency as trust-building.

11.4 Public-facing discipline: stop using identity as persuasion

The final design principle is cultural as much as technical. Identity language is powerful, and power attracts misuse. In public messaging, product marketing, and even internal culture, "same" and "new" are deployed strategically. "Same" is used to preserve trust and reduce friction. "New" is used to claim progress and justify change. The ethical failure occurs when these words are used as persuasion rather than as truthful descriptors of continuity layers.

Public-facing discipline has two sides.

First, do not claim continuity you cannot justify. If you cannot back "same assistant" with a continuity proof chain at the governance layer, do not say it. If you mean only that user-facing behavior is intended to remain similar, say that, and treat it as epsilon-identity: similar within tolerance, not identical. If you mean the service identity is the same but the underlying model changed, say that. The discipline is to speak at the layer you can defend.

Second, do not deny continuity to evade accountability. If a system is deployed under a continuous governance regime and causes harm after an update, the institution should not hide behind version labels to treat responsibility as discontinuous unless the governance regime truly treated it as discontinuous at the time. In other words, continuity boundaries cannot be drawn after the fact for convenience. They must be declared in advance, enforced by process, and evidenced by logs.

This principle implies an ethical asymmetry: continuity claims should be harder to make than discontinuity claims in marketing, but discontinuity claims should be harder to make than continuity claims in accountability contexts. The reason is incentive alignment. Organizations are tempted to over-claim sameness for trust and over-claim difference for liability. Public-facing discipline counterbalances this by requiring continuity proofs for sameness claims and requiring governance-based justification for discontinuity defenses.

One practical manifestation is identity labeling hygiene in public artifacts: release notes, model cards, policy change logs, and deprecation notices. These are not mere documentation. They are the public continuity instruments by which users and institutions can calibrate trust. Another manifestation is forbidding "identity laundering" in communication: refusing to use stable persona and branding as a substitute for continuity disclosure.

The deeper point is that the centaur era will be filled with persuasive language about systems that are, underneath, versioned, forked, merged, and updated. If we allow identity talk to remain a rhetorical device, we will get counterfeit continuity and scapegoating as default behaviors. If we enforce public-facing discipline, identity talk becomes what it should be: a governed statement about which layer of continuity is being claimed and what evidence supports it.

These design principles close the loop. They translate the metaphysical and conceptual work of the earlier sections into actionable norms: provenance as individuation, layered identity language, humility in interface claims, and discipline against persuasive misuse. The next section can therefore return to synthesis: uniqueness as ground, semi-sameness as tool, and a forward program of case studies and reference discipline.

12. Conclusion: Uniqueness as Ground, Semi-Sameness as Tool

12.1 Restate thesis and the three-layer mechanism

This paper began with an old metaphysical itch—what makes an individual this individual rather than merely an instance of a kind—and followed it forward into a technical world where copying is cheap. The core claim is that haecceity remains useful in the AI era not because we must adopt a medieval substance metaphysics, and not because we must posit a spooky individuating token, but because haecceity functions as a discipline term. It forces a separation that modern disputes repeatedly collapse: the separation between uniqueness claims about individuals, sameness claims about structures, and continuity claims about histories.

That separation becomes a three-layer mechanism for identity talk.

First, numerical identity names strict sameness across time: the one and the same instance persisting, where branching and copying break strict identity and where “same” is a strong claim with logical constraints. This layer is real but narrow.

Second, structural equivalence names similarity at declared resolutions: matching weights, architecture, functional behavior under tests, or user-facing interaction patterns. This layer is indispensable for practice and scaling, but it is not identity. It is a family of equivalence relations that must be explicitly scoped.

Third, continuity proofs name the evidentiary bridge that makes identity claims institutionally meaningful: provenance, causal history, and audit logs that justify treating one artifact as the authorized continuation or successor of another under a governance regime. This layer is where identity talk becomes checkable rather than rhetorical.

Within and across these layers sits epsilon-identity: controlled tolerance for practical sameness, purpose-indexed and constrained by continuity evidence. Epsilon-identity is how institutions and users can treat semi-sameness as a tool without pretending it is strict identity.

Taken together, the framework makes a single insistence: do not let "same" do unearned work. If the claim is about numerical persistence, say so. If it is about structural interchangeability, scope it. If it is about authorized continuity and accountability, show the lineage and the governance record. Haecceity's enduring value is that it prevents the most common slide: identity quietly replaced by description.

12.2 What is gained: cleaner metaphysics, ethics, and language

The first gain is metaphysical clarity without metaphysical inflation. The framework preserves what is most valuable in the classical and analytic discussions: the refusal to reduce "which one" to "what kind" or "what pattern." At the same time, it avoids turning that refusal into an uncheckable metaphysical posit by translating the practical identity problem into continuity proofs. Identity claims become layered, not mystical. The metaphysical itch becomes an epistemic discipline: specify the layer, specify the criterion, specify the evidence.

The second gain is ethical clarity in a world of cheap duplication. Responsibility assignment becomes tractable when we can distinguish runtime instances from governed lineages. Rights and duties become expressible when continuity demands are tied to specific layers: behavior continuity, commitment continuity, privacy continuity, certification continuity. Scapegoating becomes harder when identity boundaries cannot be redrawn after the fact without violating logged continuity regimes. Counterfeit continuity becomes easier to detect when continuity claims require auditable histories.

The third gain is linguistic hygiene. We recover a vocabulary that can say what people often mean but cannot cleanly articulate: "This is a new instance of the same certified artifact," "This is a behaviorally similar successor but not the same lineage," "This update preserves service continuity but breaks certification continuity," "These two branches share ancestry but are not identical," "This persona feels continuous, but the governance identity changed." Such sentences are not mere technicalities; they are how truth survives incentives.

The final gain is centaur-era design guidance. Provenance-first systems, explicit separation of weights/agent/accountability, humility UX, and public-facing discipline are not add-ons; they are the normative infrastructure required for human–AI collaboration that remains trustworthy under versioning. The framework therefore supports a practical telos: keep trust anchored to checkable continuity rather than to persuasive sameness talk.

12.3 What is risked: rhetorical inflation and overreach

The paper also risks failure if its own vocabulary becomes another layer of rhetoric. There are at least three dangers.

One danger is vocabulary-as-virtue signaling. Institutions may adopt the language of "provenance" and "auditability" without implementing real continuity instruments. Continuity theater—documentation that looks rigorous but does not bind to verifiable artifacts—would reproduce the very problem this paper targets. The remedy is to treat continuity proofs as evidence structures, not as brand language.

A second danger is overreach: treating the framework as a universal theory of identity rather than as a discipline for contested cases. Not every context needs all layers. Over-applying the framework can become bureaucratic, slowing legitimate iteration and causing people to route around governance. The remedy is purpose-indexing: apply strictness where stakes are high, use epsilon-identity where stability is needed, and avoid unnecessary formalism where the cost exceeds the risk.

A third danger is a new kind of inflation: substituting technical traceability for moral clarity. Audit trails can show what happened without settling what should have happened. Provenance can clarify succession without justifying harm. The framework is a tool for honesty, not a replacement for ethical judgment. The remedy is to treat auditability as a precondition for moral reasoning, not the conclusion of it.

Finally, there is a philosophical risk: that in emphasizing continuity proofs we will be tempted to treat institutional designation as ultimate identity. But governance identity is a normative construct. It can be just or unjust. It can be designed well or badly. The framework does not sanctify institutional labels; it makes them legible and therefore criticizable. That is part of the point.

12.4 Forward work: case studies and a disciplined Reference section

The next step is to move from framework to controlled application. A forward program should include at least four kinds of case studies.

First, technical lineage case studies. Analyze concrete scenarios: restoring from checkpoints after incidents, deploying bitwise copies across regions, fine-tuning a base model into specialists, merging branches through distillation, swapping prompts and retrieval layers without changing weights. For each scenario, map the numerical identity facts, the structural equivalence relations, the continuity proofs available, and the appropriate epsilon-identity regime.

Second, governance and accountability case studies. Examine incidents where blame was disputed across versions, where continuity claims were used to preserve trust, or where

discontinuity claims were used to evade liability. Use the framework to expose where equivocation occurred and to propose what continuity instruments would have made the dispute resolvable without narrative warfare.

Third, user-experience case studies. Study situations where users experienced identity continuity or disruption: “it feels like a different assistant,” “my trusted partner changed,” “the system forgot commitments,” “the tone shifted.” Use the continental recasting to interpret these reports as tracking temporal configurations, and then connect them to disclosure norms and humility UX design.

Fourth, legal and institutional case studies. Consider how contracts, certifications, and regulations should define “the model” and “the system.” Identify how to encode purpose-indexed epsilon-identity policies into legal language without pretending that strict identity is always preserved.

Alongside these case studies, a disciplined Reference section is not an afterthought but a methodological commitment. The paper’s argument depends on separating conceptual layers and refusing rhetorical shortcuts. References should therefore be curated to support that discipline: classical sources for the individuation problem, analytic sources on haecceity and modal identity, continental sources on haecceity as event, and modern sources on software provenance, auditability, and AI system versioning. The Reference section should be structured so that readers can see the genealogy of concepts without being forced into one metaphysical camp.

The closing synthesis can therefore be stated plainly. Uniqueness is the ground: reality is not exhausted by equivalence classes, and “this very one” cannot be replaced by “whatever fits the description” without loss. Semi-sameness is the tool: we must operate in a world of variants and tolerances, and practical sameness is often the right stance. The moral and institutional requirement is that we keep the difference visible, governed, and auditable. Haecceity, after Scotus, becomes less a metaphysical relic than a centaur-era discipline: a refusal to let identity be smuggled, and a demand that continuity claims show their work.

References

Adams, Robert Merrihew. “Primitive Thisness and Primitive Identity.” *The Journal of Philosophy* 76, no. 1 (1979): 5–26. doi:10.2307/2025812. PhilPapers+1

Black, Max. “I.—The Identity of Indiscernibles.” *Mind* 61, no. 242 (1952): 153–164. doi:10.1093/mind/LXI.242.153. OUP Academic+1

Cowling, Sam. “Haecceitism.” *Stanford Encyclopedia of Philosophy* (2015). Stanford Encyclopedia of Philosophy

Cross, Richard. “Medieval Theories of Haecceity.” *Stanford Encyclopedia of Philosophy* (2003; updates as noted in SEP entry). Stanford Encyclopedia of Philosophy+1

Deleuze, Gilles. *Difference and Repetition*. Translated by Paul Patton. New York: Columbia University Press, 1994. Cambridge Core

Deleuze, Gilles, and Félix Guattari. *A Thousand Plateaus: Capitalism and Schizophrenia*. Translated by Brian Massumi. Minneapolis: University of Minnesota Press, 1987. PA35 Going Live.+1

Fine, Kit. “Essence and Modality.” *Philosophical Perspectives* 8 (1994): 1–16. PhilPapers+1

Kripke, Saul A. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980. PhilPapers+1

Leibniz, G. W. *Discourse on Metaphysics and Other Essays*. Translated by Daniel Garber and Roger Ariew. Indianapolis: Hackett, 1991.

Leibniz, G. W., and Samuel Clarke. *The Leibniz–Clarke Correspondence*. Edited by H. G. Alexander. Manchester: Manchester University Press, 1956.

Lewis, David K. “Counterpart Theory and Quantified Modal Logic.” *The Journal of Philosophy* 65, no. 5 (1968): 113–126. doi:10.2307/2024555. PhilPapers+1

Lewis, David. *On the Plurality of Worlds*. Oxford: Basil Blackwell, 1986.

Ockham, William of. *Summa Logicae*. Translated by Michael J. Loux. Notre Dame, IN: University of Notre Dame Press, 1974.

Olson, Eric T. “Personal Identity.” *Stanford Encyclopedia of Philosophy* (2002; updates as noted in SEP entry). Stanford Encyclopedia of Philosophy

Parfit, Derek. “Personal Identity.” *The Philosophical Review* 80, no. 1 (1971): 3–27.

Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984. PhilPapers+1

Pasnau, Robert. *Metaphysical Themes 1274–1671*. Oxford: Oxford University Press, 2011. Oxford University Press+1

Plantinga, Alvin. *The Nature of Necessity*. Oxford: Oxford University Press, 1974/1976 (edition-dependent). PhilPapers+2Oxford University Press+2

Scotus, John Duns. *Ordinatio* (Opera Omnia; Vatican edition, various volumes). For secondary discussion of haecceitas and individuation, see Cross (SEP) and King (1992). Stanford Encyclopedia of Philosophy+1

Shoemaker, Sydney. "Personal Identity: A Materialist Account." In *Personal Identity*, edited by John Perry. Berkeley: University of California Press, 1975.

Tabassi, Elham. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. National Institute of Standards and Technology, January 26, 2023. doi:10.6028/NIST.AI.100-1. NIST+2NIST Publications+2

Torres-Arias, Santiago, et al. "In-toto: Providing Farm-to-Table Guarantees for Bits and Bytes." In *Proceedings of the 28th USENIX Security Symposium* (SEC '19), 2019. USENIX+1

SLSA Framework Authors. *SLSA (Supply-chain Levels for Software Artifacts) Specification v1.0* (and associated provenance/attestation guidance). SLSA+1

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. "Model Cards for Model Reporting." *Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT* '19), 2019. ACM Digital Library+1

The author has published over 4,700 AI-assisted papers at: https://www.researchgate.net/profile/Douglas-Youvan/research . Google Scholar has indexed these preprints, and if you want a particular reference, the AI mode of a Google search (Gemini) has efficiently catalogued these preprints.