The agent that decides has no signatory

Why the second wave of shadow IT is no longer an IT matter, but a crisis of decisional accountability

A decision was taken yesterday, on behalf of the organisation, by an agent. A client is contesting it today. Reconstructing it requires four cumulative conditions: the exact version of the model executed at the time of the call, the state of the agent's memory at the moment of the decision, the precise composition of the chain of agents and connectors that converged towards that decision, and the verifiable integrity of each of these elements at the date of the incident. None of these conditions is met. The model version is no longer available from the provider. The memory was never journaled by the citizen platform that hosted it. The call chain is exposed nowhere in any versioned representation. The connectors were not pinned to a signature. The decision took place. It engaged the organisation. It cannot be reconstructed.

This situation is not an edge case. It is a mode of operation structurally encouraged by citizen agentic AI as it is being deployed today across most large organisations. It is also, and this is the subject of the present paper, *the emergence of a category of risk whose contours have already been named by positive law, without enterprise practice yet having internalised them*. Classical shadow IT reproduced. Agentic shadow IT decides. A macro inherits the user's permissions; an agent inherits their judgement. And when that judgement engages the organisation without any reconstructible chain making it possible to identify the bearer, the matter is not technical debt. The matter is *unassignable decisional debt*: an act producing effects in the name of the organisation without an identifiable signatory, that is, without any person, function or body capable of owning, reconstructing and defending the decision on the organisation's behalf, in the sense given to that term by internal control, civil liability, the GDPR and the AI Act.

The thesis defended here is that no organisation subject to accountability can, under the law applicable since 2024, continue to industrialise this practice without an explicit decision, a documented threshold, and a reconstruction capability proportionate to impact.

1. What we are talking about

Four objects present themselves today under the same label of "automation", and they must be separated before any further work.

The *VBA macro*, as it flourished between 1997 and 2015, is a local deterministic automaton. It runs on the user's workstation, inside a workbook whose state is fully observable, with reproducible instructions. Its failure is localisable to a line of code, diagnosis is possible at reasonable effort, and the object never leaves the perimeter mastered by the organisation.

Classical RPA (UiPath, Blue Prism, Automation Anywhere) extends this logic across multiple systems but preserves determinism: an RPA robot performs, on every execution, exactly the same sequence of clicks on the same fields of the same interfaces. Its most typical failure (a DOM selector altered by an application update) remains localisable, and its audit relies on stable artefacts. RPA is, in this respect, the last fully governable representative of the citizen-automation lineage.

The *generic citizen agent*, third category, is what is being built today on mainstream no-code and low-code platforms, whether independent or integrated into a dominant vendor's stack. The business user composes a workflow visually, wires connectors to their CRM, mailbox, ERP, and ticketing tool, and calls a general-purpose language model for the reasoning steps. The promise is local autonomy; the production mode is opacity by construction. This is not an incidental flaw of the tool. It is its value proposition.

The *architected agent*, fourth category, is what an organisation builds when it treats the agent as a fully-fledged information system: versioned models, contracts of interface between components, an identity distinct from that of the calling user, reconstructible logging, an identified chain owner, and formal validation prior to any production deployment. The distinction between citizen agentic and architected agentic is not a distinction of platform; it is a distinction of governance regime. The same technical brick can, depending on the organisational discipline surrounding it, fall on either side of the line. And the line is not impermeable: *a composite system is governable at the level of its weakest link*. A chain containing a single ungoverned call is an ungoverned chain. There is therefore, in operational reality, no clean zone insulated from a citizen zone. There is continuous governance, or governance lost.

The critique that follows targets the third category, and its predictable contamination of the fourth.

2. Causal non-localisability as a system property

The decisive point is not that the agent may err. All systems err. The decisive point is that the agent may err *without it being possible, ex post, to reconstitute the cause of the error in a stable space*.

This property, which we may name *causal non-localisability*, qualitatively distinguishes the generic citizen agent from the automata that preceded it, and constitutes the central doctrinal term of this paper. It is composed of three jointly reinforcing sub-properties.

The break of determinism. A VBA macro produces a bug localisable to an identified line of code, reproducible by replaying the macro on the same inputs. A composite agent produces a decision whose replay, *even in principle*, does not guarantee the same outcome: the underlying model is probabilistic, the context window modifies the informational state available at each execution, and the agent's internal state evolves. This property is not a maturity gap that practice will close; it is intrinsic to the substrate. Governance tools inherited from deterministic software (code review, unit tests, line-by-line traceability) are, on this front, structurally insufficient. The applicable governance must be *probabilistic, statistical and contractual*, not merely procedural.

Externalisation by default. The macro was local, executed within the user's mastered perimeter. The agent runs at a provider whose model is updated, deprecated, or withdrawn without enforceable contract on the version executed yesterday. The standard terms of foundation-model providers explicitly reserve the right to modify or withdraw a version, sometimes with short notice, without commitment to provide, to a third party in litigation, the exact instance that produced the contested decision. The consequence is that, at the very moment when the organisation would need to prove what was decided and how, it finds itself in the position of a manufacturer unable to access the plans of a vehicle in circulation because the subcontractor has erased the exact version.

Chain opacity. A composite agent mobilises, at the moment of decision, a series of nested calls (agent calling agent, tool calling tool) whose effective topology is not exposed by citizen platforms in any auditable representation. The chain exists in the instant of execution; it is not preserved in a reconstructible form. Connectors may have changed silently because a remote API was deprecated, replaced or simply re-tuned by its vendor. The audit point observable a posteriori is not the chain; it is, at best, its residual trace.

These three properties are not incidental risks. They are *constitutive* of citizen agentic AI in its current production regime. They define a system whose error is not localisable, and whose decision is consequently not reconstructible.

3. Legal qualification: GDPR, AI Act, internal control

Causal non-localisability is a system property. Its legal qualification is a separate piece of work, and it must be carried out with particular care on perimeter, because the critique loses all force if it overstates the reach of the texts mobilised.

The GDPR has, since 2018, framed solely automated decisions producing legal or similarly significant effects on the data subject (Article 22). It lays down a *general prohibition*, subject to strict exceptions, and imposes in all cases the right of the data subject to obtain human intervention, to express their point of view, and to contest the decision (Recital 71). The Court of Justice of the European Union, in its SCHUFA ruling of 7 December 2023 (C-634/21), clarified that an automated score playing a determinant role in a decision taken by a third party, even formally human, falls within Article 22. That ruling closes the usual escape route ("we are not solely automated since a human validates") whenever human validation is, in fact, dictated by the agent's output.

The exact perimeter deserves attention: *the GDPR does not capture all citizen agentic AI; it captures precisely the cases where the organisation can no longer pretend that the agent was merely a tool*. An agent producing a draft that its author substantially rewrites is not in scope. An agent producing the output that will, in practice, be transmitted unamended to the data subject, is. The distinction turns on the effective nature of the decisional chain, not on its formal description.

The AI Act, in force since 1 August 2024, imposes on *high-risk systems* (within the meaning of its Annex III, which covers in particular certain uses in HR, credit scoring, access to essential services, justice, and critical infrastructure) automatic logging of relevant events (Article 12), effective human oversight (Article 14), preservation of logs by the deployer (Article 19), and explicit responsibility along the value chain (Article 25). The reach of the regulation is not universal: binding obligations target, as a priority, high-risk systems and general-purpose systems posing systemic risk. But when usage crosses these categories, or produces comparable effects on persons, *the AI Act already provides the governance grammar that citizen platforms do not satisfy by default*.

Beyond these two regimes, the responsibility of the controller (GDPR Article 24), civil liability under Article 82, and internal-control obligations (LSF, SOX where applicable, sectoral regimes such as MiFID, Solvency II, Basel for financial actors, ASN, ANSM, ARS for healthcare actors) impose, everywhere, a *capacity for demonstration*: the ability to retrace, on demand, how a decision was taken, by whom, on the basis of which data, and through which validation process. The exact perimeter varies; the principle does not.

The rupture is the following. All of these regimes assume that a decision taken on behalf of an organisation can be *traced to an identifiable signatory* and *reconstructed in its grounds*. The operational promise of the citizen platform is, on the contrary, a fast execution that dispenses precisely with this traceability. The collision is not a grey zone to

explore; it is the simultaneous assertion of two irreconcilable theses. As long as it is not made explicit at executive committee level, the applicable law already renders difficult to defend the current practice of citizen agentic AI without a frame.

4. The multiplicative as default mode of product design

It will be objected that all of this holds only for complex compositions, and that most citizen agents are point automations without ambition. The objection does not survive examination of how the platforms themselves operate.

The radius of effect, first, is not comparable to that of earlier automata. A macro modifies a workbook. An agent acts in production systems: it updates a CRM, sends a message in the user's name, modifies a ticket, allocates a resource, signs a contractual reply, triggers a payment, adjusts a price. The failure of a macro costs an afternoon of rework; the failure of an agent may cost a personal-data leak falling under Article 33 of the GDPR, an unfulfilled contractual promise, a discriminatory bias in HR pre-screening, or a financial decision documented on erroneous data. The radius of effect is not additive with respect to the previous shadow IT; it is in another category of engagement.

Composition, second, is not an advanced usage discovered after six months of practice. It is *the principal commercial argument* of generic agentic platforms. Connector marketplaces with several thousand templates, agents calling other agents through graphical chaining, libraries of pre-assembled workflows, multi-system integrations proposed by default at the home screen: everything in the product design encourages composition. The effective complexity of an agentic workflow in production exceeds, within a few weeks of usage, the comprehension of the citizen creator who assembled it. According to Gartner, an average global Fortune 500 company could exceed one hundred and fifty thousand agents in operation by 2028, against fewer than fifteen in 2025, and only 13 % of organisations consider that they currently have adequate governance in place to face this proliferation.

The operational consequence of undeclared composition escapes the business unit producing it. When agent A calls agent B, which calls connector C towards external model D, whose output feeds agent E that takes the final decision, an incident on any one of the five links produces an effect on the decision without any explicit signal linking it to its cause. The debt is not additive, *it composes in the algebraic sense*. A system with n ungoverned links does not have n potential defects: it has a number of failure paths that grows with the interaction paths between links, not merely with their count. Only a fraction of these paths is observable from any given audit point. The multiplicative is not a slip. It is the nominal regime.

Self-immunisation: what is criticised here is not agentic composition as such. Architectures composed under engineering discipline are precisely what the most serious work on regulated digital twins and clinical AI defends. Composition is necessary; what is untenable is *undeclared* composition, without interface contracts, without chain ownership, without decisional reconstruction capability. The distinction is analogous to the one separating, in civil engineering, an assembly by planking and an assembly by certified welding. Both connect elements. Only one supports a responsibility.

5. Audit trail is not enforceable governance

Citizen agentic platforms are sold, without notable exception, as the remedy to shadow IT. The vocabulary is constant: *native audit trails*, integrated *observability*, *governance ready*, *enterprise-grade compliance*, *fully traceable*. This promise is, in intent, sincere; it is not, in delivery, enforceable.

The decisive point is not marketing. It is legal, and it is formulated as follows: *an audit trail that does not allow reconstruction of the decision in a stable space is legally and operationally useless*. A technical log lists timestamped events. Enforceable governance reconstructs a decision: it restores, on demand, the model instance, the memory state, the connector versions, the effective call chain, the unaltered input, the result as produced, and the path by which that result became an engaging act. The distance between the two objects is the gap between an incident report and a notarial deed.

The AI Act, in its Articles 12, 14 and 19, does not merely require the preservation of events; it targets logging *designed to enable reconstruction*. The GDPR, in its Article 24, charges the controller with implementing the *appropriate technical and organisational measures to demonstrate that processing complies* with the regulation. Article 82 establishes civil liability for any damage resulting from a violation, and Article 83 sets administrative fines at levels that make exposure material. The confusion, today dominant at executive level, between the availability of a log and the capacity for demonstration, is the categorial error that makes the first signature possible. As long as it is not dispelled, the internal audit report will describe a compliant arrangement where the litigation will find an orphan decision.

6. The VBA precedent, and what we are unlearning

The historical argument is short, and is not mobilised to alarm. It is mobilised to disqualify the objection that the organisation will learn as it goes.

Between 1997 and 2015, VBA macros produced an unmanageable estate in nearly every large organisation. Successive audits, mapping projects, platform migrations, citizen ALM frameworks, dedicated governance cells: it took, in most sectors, between fifteen and twenty years to bring this legacy back within a perimeter of control, at the cost of repeated incidents (defective financial models, erroneous regulatory calculations, production errors) and considerable investment. This experience is sufficiently documented that the argument from ignorance is no longer admissible.

The observation to make is the following: *the same organisation that invested to clean up the VBA debt is reproducing, on a technological substrate presented as remedy, exactly the pattern it claimed to have corrected*. Same business sponsors, same circumvention of the IT department, same opacity by construction, same absence of identified ownership at the moment when the creator changes role. This regularity is not an inattention coincidence; it suggests that the institutional mechanism producing shadow IT has never been treated, and that it will reproduce the pathology on any substrate offering business units a route around governance. We have already learned. We are now unlearning. And the window is not twenty years this time.

7. Why the window is twelve to twenty-four months

Four parameters make the diffusion qualitatively different from that of macros, and concentrate within a short horizon what unfolded over two decades.

The *speed of diffusion* is several orders of magnitude greater. The natural-language interface lowers the entry barrier to a level no previous generation of tooling reached. The marketing, descending from the CEO who saw a demonstration, places the organisation in a posture of adoption ahead of any governance deliberation. According to Gartner, roughly 40 % of enterprise applications will integrate task-specific agents by the end of 2026, against fewer than 5 % in 2025. According to Microsoft's Cyber Pulse Report 2026, more than 80 % of Fortune 500 companies actively operate agents built with low-code or no-code tooling, and 29 % of employees report using agents not sanctioned by their organisation.

The *external dependency* is immediate, in contrast to the macro that remained local. Model, connectors, infrastructure: three sources of uncontrolled drift installed from the first deployment. The *marginal cost of experimentation* is near zero, which means that proliferation is not slowed by a local budgetary trade-off. And the *action surface* is, from day one, in production systems, because the very point of an agent is to act, not to compute.

Consequence: agentic debt does not take twenty years to surface. It becomes critical at the moment when a first incident requires a decisional reconstruction that does not exist.

Available leading indicators at end-2025 and early 2026 (according to IBM's *Cost of a Data Breach Report* 2025, only 37 % of organisations have a formal AI governance policy in place; according to Netskope, the average enterprise records on the order of two hundred and twenty monthly data-policy violations linked to GenAI usage) suggest a window on the order of twelve to twenty-four months between industrialisation of usage and the first enforceable governance incident. These figures do not describe a marginal risk; they describe a plausible trajectory of rapid materialisation, which leaves no time for the sequential learning that characterised the VBA era.

8. Four decisions, not four engineering principles

The proposal is not an IT best practice. It is an executive deliberation to be formally inscribed at the highest level of the control framework. Four decisions, taken explicitly, separate the organisation that still signs from the organisation that has stopped signing what it cannot reconstruct.

First decision: explicit criticality threshold. The organisation must define, by formal deliberation, the threshold beyond which an agent is treated as a critical system in the sense of internal control. The threshold mobilises the usual axes of criticality: financial impact (volume of commitments made), contractual impact (commitments to third parties), GDPR impact (processing of personal data, automated decision in the sense of Article 22), AI Act impact (high-risk qualification under Annex III, in particular HR, credit scoring, access to essential services, critical infrastructure, justice). Below the threshold, declared sandbox with bounded perimeter. Above, obligations identical to those of a critical information system, with no exception linked to the construction platform. The absence of a deliberate threshold is not neutrality: it is the zero threshold, that is, the implicit commitment to sign any agentic decision regardless of its impact.

Second decision: explicit prohibition of patterns. Four patterns must be banned by written decision, and their detection must trigger an alert at the compliance level, not at the IT level: undeclared agent composition outside the cartography; agent in production without an identified owner, named, with a designated deputy and a review cycle; agent with write access to a production system without formal validation of the complete call chain; agent lacking decisional reconstruction capability (model instance, memory, connectors, chain, input, output, decision path). These prohibitions do not eliminate citizen agentic AI; they confine it to its domain of validity.

Third decision: decisional kill switch, distinct from the technical kill switch. The technical capacity to suspend an agent's execution exists on most platforms. The organisational capacity to suspend the *engaging value* of its decisions, while a human chain takes over, almost never does. This second capacity must be formalised: an escalation path, an

explicit power to invalidate retroactively the acts emitted within a defined window, a notification arrangement for contractual stakeholders. The decisional kill switch does not erase already-emitted decisions; it confines their reach during the time of reconstruction. It is, on this front, the agentic equivalent of the product recall in manufacturing.

Fourth decision: validation circuit indexed on business impact, not on technology. The qualifying criterion of an agent must not be the construction platform, but the business impact of its acts. An agent drafting an internal note does not fall under the same circuit as an agent closing a customer ticket with implicit contractual engagement. This indexation by impact, rather than by technology, is what prevents arbitrage by technical convenience from producing a de facto arbitrage on responsibility.

These four decisions are not principles. They are *signatures to be affixed*, by the competent governance bodies, in a form reproducible by an auditor. As long as they have not been affixed, the organisation continues to sign blind.

9. Articulation with the corpus

The distinction laid down here (*citizen agentic* against *architected agentic*, and more deeply *decision with signatory* against *decision without signatory*) is consistent with the epistemic gesture mobilised elsewhere in this corpus. It is analogous, in the order of engineering, to the distinction between LLM-centred paradigm and regulated composite architecture. It is analogous, in the order of architecture, to the condition of hexagonality as a prerequisite of governability (INPI Soleau filing TI-2026-ART5-ES). It is analogous, in the operational order, to the necessity of an event-driven architecture to produce the observability that citizen platforms do not deliver.

The recurring motif is the same: naming the structuring distinction that separates two regimes of operation, one viable under a regulated frame, the other not. The present contribution projects this motif onto the plane of accountability. It proposes two articulated doctrinal terms. The first, *causal non-localisability*, names the system property: an agent whose error can be neither reproduced, nor tied to a stable link, nor reconstructed from available artefacts. The second, *unassignable decisional debt*, names the organisational consequence: an act engaging the organisation without an identifiable signatory in the sense of applicable law. The two terms are joined. The first is conceptual; the second is operational. The distinction *citizen agentic / architected agentic* is, on this plane, a second-order distinction: the operator separating the production regimes that produce one or the other category of debt.

10. Limitations

Four objections deserve treatment without complacency.

The criticality threshold is non-trivial to set. This is correct. There is no universal threshold; there is a threshold proper to each organisation, a function of its sector, its regulatory exposure and its risk appetite. This difficulty is not a reason not to deliberate. It is precisely the object of the deliberation.

The frontier between citizen and architected agentic is gradual, not binary. This is correct, and it is treated explicitly in section 1. The applicable governance is continuous, not categorical; it is lost at the first ungoverned link. This does not imply the absence of a decisional threshold: the decision is to know from which point the organisation accepts to engage.

Organisations of low IT maturity do not have the critical mass for architected agentic. This is correct. The consequence is not that they may industrialise citizen agentic without frame, but that the industrialisation decision should, in their case, be suspended. This is precisely the decision these organisations avoid taking, and it is the avoidance of this decision, rather than the absence of capability, that produces the risk.

The regulated frame displaces but does not eliminate the debt. This is correct. The objective has never been to eliminate the debt; it is to render it *assignable*. An assignable debt is treatable; an unassignable debt is not, because no counterparty can be held to account for it. The proposal targets this displacement, not a zero-risk utopia.

11. Conclusion

The organisation has, at the date of this paper, twenty-five years of documented experience on the governance failure of macros, eight years of GDPR application to automated decision-making, and the effective entry into force of the AI Act since the summer of 2024. That these three corpora converge to render the current practice of unframed citizen agentic AI untenable, and that the organisation nevertheless continues to industrialise this practice without the necessary countermeasures, is a regularity that must be named to be interrupted.

The problem is not that an agent can err. The problem is that an organisation can be engaged by a decision it can *neither reconstruct, nor explain, nor assign*.

As long as an executive committee classifies this matter as one for the IT department, it signs without knowing decisions without signatory. The question is not whether agentic AI should be built. The question is from which point the organisation stops signing what it canots reconstruct.

References

Regulatory frameworks

Regulation (EU) 2016/679 (GDPR), articles 22 (automated individual decision-making), 24 (responsibility of the controller), 33 (notification of breach), 82 (right to compensation), 83 (administrative fines); Recital 71.

Regulation (EU) 2024/1689 (AI Act), articles 12 (record-keeping for high-risk systems), 14 (human oversight), 19 (automatically generated logs preserved by the deployer), 25 (responsibility along the value chain), 26 (deployer obligations), Annex III (categorisation of high-risk systems). Entry into force 1 August 2024.

Court of Justice of the European Union, case C-634/21, OQ v Land Hessen (SCHUFA), judgement of 7 December 2023.

Calibration data

Gartner, *Press Release: 40% of Enterprise Apps Will Feature Task-Specific AI Agents by End of 2026*, 26 August 2025.

Gartner, *Press Release: Six Steps to Manage AI Agent Sprawl* (referencing projections of one hundred and fifty thousand agents per Fortune 500 by 2028 and the 13 % rate of organisations considering they have adequate governance), 28 April 2026.

Gartner, *Top Predictions for Data and Analytics in 2026*, 11 March 2026.

Microsoft, *Cyber Pulse Report 2026* (rate of 80 % of Fortune 500 actively operating low-code / no-code agents, and 29 % of employees using non-sanctioned agents).

Netskope, *Cloud and Threat Report 2026* (around 220 monthly average data-policy violations linked to GenAI usage).

IBM, *Cost of a Data Breach Report 2025* (37 % of organisations with a formal AI governance policy).

Forrester, *AEGIS Framework: Agentic AI Enterprise Guardrails for Information Security*; *Predictions 2026*.

World Economic Forum, *Global Cybersecurity Outlook 2026*.