Mapping Governance Strategy to Technical Architecture in Public-Sector AI

Sotiris Leventis[1]

Qognity. Hellenic OCR Team. Hypernetica (Greece)

Received: June 15, 2026
Accepted: July 1, 2026

https://doi.org/10.26422/REPA.2026.0100.lev

Abstract
Public organisations such as parliaments, ministries and courts seeking to adopt artificial intelligence should pursue a holistic yet incremental strategy. This requires a technical action plan that prioritises practical software applications and services designed to integrate consistently into existing organisational workflows. The underlying architecture should remain vendor-agnostic and deployable on-premises or within sovereign environments, allowing organisations to retain control over data, models, security and regulatory compliance, while preserving human oversight and institutional accountability. This paper argues that a multimodal, multi-agent integration platform provides a suitable foundation for such a strategy, enabling coordinated use of diverse AI capabilities across organisational systems. The proposed approach supports early, low-risk use cases, including evidence-grounded information access and accessible digital assistance delivered through familiar operational tools. The paper further outlines how an integration platform can be utilised to implement this approach, orchestrating multimodal components and enabling responsible, scalable AI adoption without institutional disruption.

Keywords: public-sector AI, sovereign AI, multi-agent systems.

1 The author is an enterprise architect and AI specialist with over 25 years designing intelligent systems and enterprise software. Founder and CEO of Qognity, where he leads multi-agent systems development, he has published on AI in governance and law and serves as Head of Technology for the Hellenic OCR Team.
sotiris.leventis@qognity.com / https://orcid.org/0000-0003-4556-8102

Correspondencia entre la estrategia de gobernanza y la arquitectura técnica en la inteligencia artificial del sector público

Resumen

Las organizaciones públicas, como los parlamentos, los ministerios y los tribunales, que buscan adoptar la inteligencia artificial deberían seguir una estrategia holística pero incremental. Esto requiere un plan de acción técnico que priorice aplicaciones y servicios prácticos diseñados para integrarse de forma coherente en los flujos de trabajo existentes de la organización. La arquitectura subyacente debería permanecer independiente del proveedor y poder desplegarse de forma local o en entornos controlados por la institución, lo que les permitiría a las organizaciones conservar el control sobre los datos, los modelos, la seguridad y el cumplimiento normativo, preservando, al mismo tiempo, la supervisión humana y la responsabilidad institucional. Este artículo sostiene que una plataforma de integración multimodal y multiagente constituye una base adecuada para dicha estrategia, al permitir el uso coordinado de diversas capacidades de inteligencia artificial en los sistemas de la organización. El enfoque propuesto admite casos de uso temprano y de bajo riesgo, incluido el acceso a información fundamentado en evidencias y en la asistencia digital accesible prestada a través de herramientas operativas conocidas. El artículo describe, además, cómo una plataforma de integración puede utilizarse para implementar este enfoque, orquestando componentes multimodales y posibilitando una adopción de la inteligencia artificial responsable y escalable sin disrupción institucional.

Palabras clave: inteligencia artificial del sector público, inteligencia artificial soberana, sistemas multiagente.

1. Introduction

Artificial intelligence has moved from the margins of public-sector planning to the centre of it (Inter-Parliamentary Union, 2024). Parliaments, ministries and courts now face a common question, not whether to adopt AI, but how to do so without compromising the obligations that distinguish a public institution from a commercial one (Fitsilis & De Almeida, 2024). These obligations include the protection of confidential information, accountability for decisions and the integrity of the public record. These matters should be addressed in the design of the system (Fitsilis, Von Lucke, et al., 2024). This paper offers an architectural framework rather than an evaluated system, and its claims are about how the structure of a deployment bears on governance. Although the framework is developed for parliaments, its foundations are domain-neutral by design, and the integration mechanisms and design choices are intended to transfer to the wider public sector and in many cases to private organisations. Throughout, *artificial intelligence* is used in its broad sense, covering established techniques beyond just generative models, and the architecture is designed to govern all techniques under one discipline.

In the public sector, just as in the private sector, architectural strategy is itself a governance decision. The structure of AI-enabled capability determines whether an institution can meet its obligations at all, since it is the overarching architecture that handles where data is processed, which models run where, how outputs are reviewed and how the system integrates with the tools staff already use. A capable model deployed through the wrong architecture is a governance failure waiting to surface, whereas a modest capability deployed through the right architecture delivers value immediately, while remaining defensible to auditors, courts, and citizens.

Because the architecture carries the governance load, institutions do not necessarily need to wait for a comprehensive digital transformation programme before beginning. They can adopt incrementally, starting with low-risk use cases that operate on internal data and keep humans in control, provided the architecture is built from the outset to scale without re-engineering. This is the meaning of holistic yet incremental. Holistic in that the architecture anticipates the full range of future capability; and incremental in that deployment proceeds one bounded use case at a time.

The paper proceeds as follows. Section 2 sets out the gap between governance requirements and prevailing technology. Section 3 articulates five governance principles that any deployment should satisfy. Section 4 presents the multimodal, multi-agent integration platform as the architecture that operationalises them. Section 5 develops two early use cases that should anchor a first wave of adoption, namely evidence-grounded information access and accessible digital assistance. Section 6 surveys a wider family of capabilities and the sovereignty boundary that governs them. Section 7 concludes with implications and open questions.

2. The Gap Between Governance and Technology in Public-Sector AI

The barriers to public-sector AI adoption are well documented and consistent across institution types and jurisdictions. Surveys of national parliaments and public administrations identify a recurring cluster of concerns, such as data protection and confidentiality, dependence on foreign-controlled infrastructure, opacity in model behaviour, the absence of adequate audit trails, and a shortage of in-house expertise (Fitsilis & De Almeida, 2024; Inter-Parliamentary Union, 2024). These concerns are not abstract, and each one becomes a binding constraint the moment an institution attempts to move from experimentation to production.

The *Guidelines for AI in Parliaments* devote substantial attention to privacy, governance, system design, and capacity building because each dimension hardens into an obstacle at the point of deployment (Fitsilis, Von Lucke, et al., 2024). A pilot that performs well in a sandbox can prove impossible to place into service because it cannot demonstrate where data is processed, cannot produce an audit trail a committee would accept, or cannot be operated by existing staff without specialist support.

The dominant commercial AI offerings sit at the opposite end of the design spectrum from these requirements. Frontier large language models (LLMs) are accessed primarily through cloud interfaces operated by a small number of providers. Any input submitted for processing transits, and may briefly be stored, outside the controlling institution's jurisdiction. Even where enterprise data protection terms are in place, the institution does not retain physical and perhaps not even legal custody of its data during processing. For a parliamentary committee handling a bill under consideration, for a regulator handling supervisory information under embargo, or for a ministry handling personal data about citizens, these are unacceptable conditions.

The regulatory environment seems to tighten this gap further. The EU Artificial Intelligence Act establishes obligations for high-risk systems covering human oversight, transparency, data quality and record keeping (European Parliament & Council of the European Union, 2024).[2] Some parliamentary uses fall within or are adjacent to the defined high-risk categories therein, for example AI systems used to interpret legislative text in a manner equivalent to judicial reasoning, tools deployed to influence citizens' voting behaviour in elections or referenda, and systems used for biometric identification or access control within parliamentary premises. These obligations cannot be satisfied solely by a model's accuracy but by the architecture surrounding it, the oversight mechanisms, the documentation and the traceability, among other things. Meeting them calls for an architecture designed around the full set of obligations rather than around the choice of model alone.

2 The AI Act entered into force on 1 August 2024 with staggered application. The AI Omnibus Regulation, in force since 27 July 2026, postpones the application of obligations for stand-alone high-risk AI systems from 2 August 2026 to 2 December 2027. The argument here does not depend on the precise date.

3. Governance Principles for Public-Sector AI

If the architecture carries the governance load, the first task is to state what governance demands. This paper focuses on five principles that constitute a governance strategy, drawn from the parliamentary AI literature and the EU AI Act, that define the strategy that the architecture must operationalise.

The first principle is vendor neutrality. Vendor neutrality means the architecture treats the model as a replaceable component, so that an institution can substitute one model for another, open weight or commercial, local or hosted, without rebuilding the workflow around it. No institution should be structurally dependent on a single model provider. Models improve and reprice rapidly, and a system bound to one vendor inherits that vendor's commercial decisions, availability and jurisdiction.

The second principle is sovereign deployability. The institution must be able to decide where data is processed and where models run. Where institutional rules permit, a sovereign cloud or a compliant hosted region may suffice. Where they do not, the same software must run unmodified on hardware the institution controls. Sovereignty means that confidential material remains on premises, lower-sensitivity work may draw on external resources, and the choice is made deliberately rather than imposed by the tooling.

The third principle is workflow integration. AI capabilities deliver value when they appear inside the tools staff already use, not as separate applications that compete for attention and require behavioural change. Decades of public-sector digital transformation show that tools demanding new habits face a far steeper adoption curve (Koryzis et al., 2021). For parliaments, the natural carriers include email, transcription, registry handling, document management systems and the standard office suite in general.

The fourth principle is human oversight and accountability. Systems must be designed so that humans remain in control of consequential decisions, so that the artefacts of that control are recorded and auditable, namely what was retrieved, what was generated, on what basis, and who approved it. This requirement has specific normative weight in public institutions, where the legitimacy of a decision depends on its attribution to an identified human actor. Architecturally, it means that every agent, a software component that perceives its environment, acts on it with a degree of autonomy and communicates through explicit messages (Wooldridge & Jennings, 1995), and every output must be designed for inspection.

The fifth principle is low-risk first use cases. The most successful early adoptions automate well-defined, repetitive, high-volume tasks where errors are easy to detect and the stakes of any single decision are bounded (see, for instance, Von Lucke et al., 2024). Characteristic examples include digitising historical records, indexing legislative materials, transcribing debates and answering common procedural questions. These "low-hanging fruits" are the proving ground on which institutional trust and institutional capability are built. Table 1 maps AI governance principles to their corresponding architectural properties and implementation mechanisms in the proposed reference architecture.

Table 1
Mapping governance principles to architectural properties

Governance principle	Architectural property	Principal mechanisms
Vendor neutrality	Model as a replaceable component	Domain-organised connectors with vendor selection at deployment. The model is used and integrated as a tool like any other
Sovereign deployability	Location-independent deployment	Same software on premises, in a sovereign cloud or against a hosted model. Local models behind the internal trust boundary. Placement decided per workflow
Workflow integration	Connection to existing systems	Connectors as adapters to existing systems. Views surfaced through embedded applications and a portal
Human oversight and accountability	Review gate and provenance	Read, Extract, Review, Process discipline. Routing and execution separately recorded per step. Sensitive steps reserved to human execution. Append-only audit stream
Low-risk first use cases	Bounded, composable capability	Specialised agents and versioned workflows with explicit contracts. Capability added by composition rather than re-architecting

Source: Authors own elaboration.

4. Architectural foundations

The architectural form that satisfies all five principles can be conceived as a *multimodal, multi-agent, integration* platform. Each term corresponds to a specific technical choice, and each choice maps back to a governance principle. The subsections elaborate on the structural overview that includes the platform's structure, its integration layer, its agents and workflows, the handling of multiple modalities and the design choices that make the whole system governable.

4.1 Platform architectural overview

The platform is organised around various kinds of elements: the *surfaces* through which staff encounter it, the *actors* that perform work, the *workflows* that define the work and the *tools* through which work touches systems. The systems of record, such as the document repositories, registries, case-handling platforms or databases the institution already operates, remain outside the platform and retain their authoritative role.

Surfaces present work. Work is defined by workflows, workflow steps invoke tools, and tools reach external systems only through connectors. The arrangement confines the new capability to a layer that can be extended, audited or withdrawn without disturbing

what lies beneath, and users interact with the platform within the applications they already use.

Surfaces take complementary forms. Embedded applications place the platform's views, its forms, queues and dashboards, within a tool staff already use, such as a word-processor add-in or a team-messaging client, and they behave deliberately non-intrusively. They suggest rather than act, presenting retrieved passages, candidate formulations or flagged inconsistencies beside the document, and nothing enters the text or the record until the user accepts it. The portal, by contrast, is the institution's operational view of the platform itself, managing corpus-scale and batch work, namely ingestion runs, review queues, registers, and the administration of workflows. The two surfaces draw on the same agents and workflows. Only the mode of presentation differs, which is why a capability built once can appear wherever the work happens.

Actors are the platform's baseline representation for all the entity types that perform the work. More specifically, humans and agents are peers under a single abstracted entity. Any step of any workflow can be routed to either, and the platform separately records who routed each action and who performed it, so that every action is attributable to an identified actor. Agents themselves may be either model-driven, reasoning via a language model and invoking tools during that reasoning, or deterministic, executing fixed code; the two are governed identically. Representing humans and agents as peers enables their actions to be tracked and managed uniformly. Likewise, the approval or review of those actions can also be treated as routing decisions regardless of the actor type.

Figure 1 provides a graphic overview of the architecture with standalone and embedded applications at the top, the engine with its capability agents, the workflows they orchestrate, and the model options they draw on in the middle, and the external systems and data sources they integrate at the bottom. The agents shown are illustrative capability agents (cf. Table 4) and are distinct from the workflow-internal executors of Table 3. Text generation appears as a capability in Table 4 but is held outside the illustrated agent set, for the sovereignty reasons presented in Section 6.

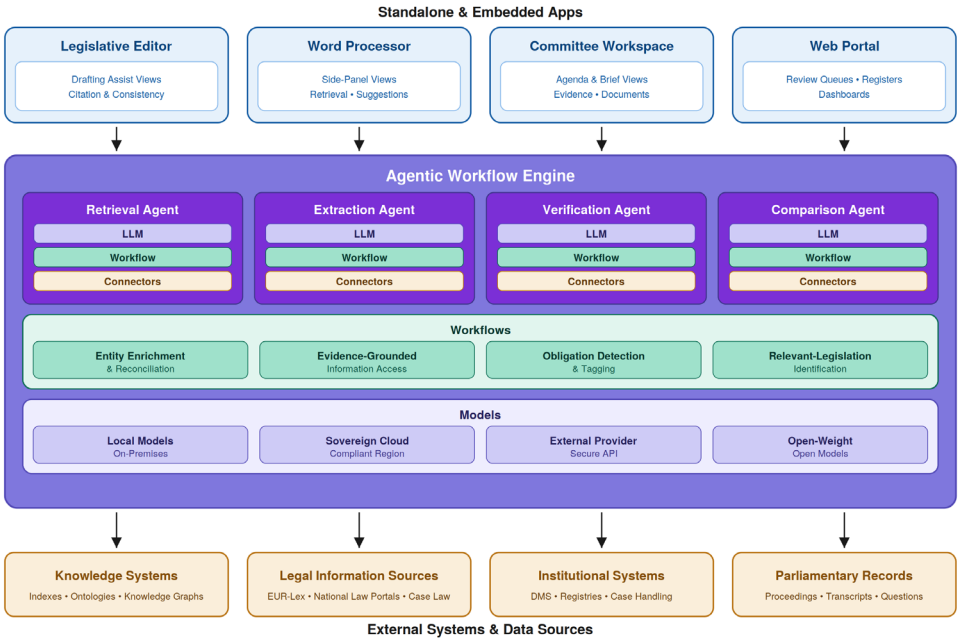

Figure 1. Architecture overview. Source: Author's own elaboration.

4.2 Integration layer

An agentic tool is a single named operation with declared input and output schemas. Every such tool reaches the world in one of two ways: either through a connector, which integrates an external system, or as an internal service that is used within the platform itself. The distinction is also a trust boundary, since connector traffic crosses to external systems while service execution never leaves the deployment. Connectors are organised by business domain rather than by vendor. A workflow references an email operation or a records operation, not a particular provider, and the vendor behind each domain is selected when the platform is deployed. This has the following consequences:

- Workflows are vendor-neutral by construction, which is the implementation of the first governance principle (see Section 3), and
- Access to external AI systems, including language models, is itself mediated like any other integration, reached either through a connector to an external provider or through an internal service running a locally hosted or open-weight model, which is a concrete form of the sovereignty decision (see Section 6).

The same operation is invocable by a human or an agent without redefinition. A person encounters it as a form, a model encounters it as a tool-use description, and a step can therefore be assigned to either a human or to an automated execution as routing requires.

The design descends from enterprise integration patterns, the catalogue of message-oriented endpoints and adapters developed for joining heterogeneous systems (Hohpe & Woolf, 2004), together with the adapter and facade patterns of object-oriented design (Gamma et al., 1994). Table 2 maps the patterns employed in this paper to these sources. Earlier parliamentary integration studies suggested these patterns (Leventis et al., 2020) and used the term software agent for the integration component itself, adopting the actor-oriented vocabulary of its time (Leventis et al., 2021). This paper reserves the term *agent* for the model-bearing workers described in the next subsection and the term *connector* for the integration component, since the two concepts have diverged as model-driven capability entered the platform.

Table 2
Patterns employed by the platform and their sources

Pattern	Source
Message Endpoint	Hohpe & Woolf (2004)
Polling Consumer	
Aggregator	
Correlation Identifier	
Content Enricher	
Message Translator	
Channel Adapter	
Canonical Data Model	
Adapter	Gamma et al. (1994)
Facade	

Source: Hohpe & Woolf (2004), Gamma et al. (1994).

4.3 Agents and workflows

The agent is the platform's unit of work. The term is used in its established sense, namely a software component that perceives its environment, acts on it with a degree of autonomy and communicates through explicit messages (Wooldridge & Jennings, 1995), which is consistent with the definition standardised in ISO/IEC 22989 (International Organization for Standardization, 2022). Multi-agent systems treat the division of labour among agents as a first-class design concern, building on agent communication languages standardised by

FIPA, the Contract Net Protocol for task allocation and belief-desire-intention architectures (Foundation for Intelligent Physical Agents, 2002; Rao & Georgeff, 1995; Smith, 1980).

A workflow is a named, versioned directed graph of steps that captures one end-to-end outcome, with declared inputs and outputs so that it can be invoked as a unit. Each step performs one operation under explicit constraints. The workflow definition fixes which tool a step may invoke, what data it receives and produces, and who may execute it, so that a sensitive approval step can be reserved to human execution, which is how the review gate of this paper is enforced structurally rather than by convention. Authorisation is embedded by design in the flow of execution via the decision points gating the actions they govern. More specifically, governance rules such as separation of duties and role-based routing are evaluated where the routing happens. The steps of such workflows recur in multiple roles: monitoring an input channel for events worth acting on, extracting the relevant artefact into structured form, acting through models, or code to produce the output and validating the result against policy and evidence.

Composition between agents is itself explicit and recorded. Where a step exceeds an agent's remit, the work is delegated to a specialist agent that performs it under its own contract, in direct continuity with the Contract Net tradition of negotiated task allocation (Smith, 1980), and every such handoff appears in the record. A multi-stage process can therefore be expressed as one workflow or as several specialists delegating between each other, the same outcome under different decomposition strategies. Workflow definitions are themselves versioned institutional artefacts, changed through review like any process document, so that an auditor can establish not only what the system did but which version of which process did it and who approved that process.

Recent analysis finds that a large fraction of multi-agent LLM systems fail in deployment, with improper task decomposition the most common failure mode, which is the problem the classical engineering tradition was built to solve (Cemri et al., 2025). In the architecture proposed here, a language model reasons inside a step whose inputs, outputs and permitted tools are fixed by the workflow definition. The design therefore inherits the discipline of the older tradition and the flexibility of the newer one.

4.4 Multimodality as a pipeline property

Agents must, for example, extract text from scans, fields from forms, structure from layout and metadata from headers, and carry all of these through the workflow without collapsing each document to its weakest dimension. Each extraction converts into a common internal representation, so that downstream steps operate on one format regardless of whether the content began as a scan, a recording or a structured export. Where a domain open standard exists, it serves as the standard form. In the parliamentary and legislative domain such a standard is Akoma Ntoso (OASIS, 2018), which keeps the institution's structured textual

data, from legislative texts to the proceedings, transcripts and parliamentary questions, independent of any vendor's schema.

The digitisation of written questions from the first parliamentary term of the Hellenic Parliament (1974-1977) required custom optical character recognition over a fragmented corpus of variable quality, written in the polytonic Katharevousa variety of Greek (Fitsilis, Gatos, et al., 2024). Extraction performance in these cases varies, which is why every extraction output carries a confidence score and why low-confidence outputs are routed to the review gate (see Section 4.5). Lower quality of outputs should be mitigated by stricter demands for human oversight.

4.5 Knowledge, provenance and the review gate

The architecture becomes governable through a set of design choices, i.e., architecture-held knowledge and fixed workflow discipline. Rather than relying on a model to recall institutional facts from its training weights, the platform applies retrieval-augmented generation. Institutional sources are indexed in advance, the material relevant to a request is retrieved at the point of use and supplied to the model as context, and the model's role is confined to synthesising over what was retrieved (Lewis et al., 2020). The institution's knowledge stays within the institution, is updated by the institution and is independent of any model vendor, which is what makes vendor neutrality and sovereignty operationally real. The model becomes interchangeable while the knowledge index remains intact.

A fixed workflow discipline, i.e., Read, Extract, Review, Process is applied. The platform reads from a monitored source, extracts the relevant content into structured form, presents the extracted data to a human reviewer, and only then processes it into a result. Review functions as a gate embedded within the process, controlling what may proceed before processing takes place. No further processing is performed until the institution confirms that the extracted data is correct. The audit trail is the engine's primary output. Every step execution logs who requested it and who performed it, and an immutable, i.e., append-only, event stream records the run from start to finish. These artefacts also serve as the operational records required by an AI management system conforming to ISO/IEC 42001, so conformance is evidenced through everyday operation, not retrospective documentation (International Organization for Standardization, 2023). This design choice operationalises the fourth governance principle and produces the audit trail that accountability requires.

5. Early use cases

Two use cases that could be suited to a first wave of adoption are evidence-grounded information access and accessible digital assistance. Neither is reported here as a completed deployment but as design walkthroughs that show how the governance principles could shape concrete workflows. Both are low risk, operate on internal data and keep a human at the decision

step. They also illustrate the surfacing modes, since the former is naturally portal-based and operates at corpus scale, while the latter is naturally embedded in the tools staff already use.

5.1 Evidence-grounded information access

Evidence-grounded information access means asking a question and receiving a synthesis in which every claim is linked to the source document it came from, with no fabricated citations. Delivering this reliably depends on the quality of the underlying knowledge, and that is where a generalised workflow earns its place. The recurring need across public institutions is to assemble a complete, trustworthy picture of an entity, whether a person, a case, a piece of legislation or a speaker, from fragments scattered across multiple systems and data sources. This is the entity enrichment and reconciliation workflow, and it generalises a pattern that appears throughout parliamentary data work.

The workflow has five steps (Table 3), and each step is the responsibility of an actor, whether agent or human. The sequence follows the fixed workflow discipline. The executors named here are workflow-internal roles, distinct from the capability agents of Section 6 and Figure 1.

Table 3
The entity enrichment, review and reconciliation workflow

Step	Function	Executor	EIP pattern	Phase
Poll/retrieve	Pull records about target entities from multiple source systems	Ingestion agent via source connectors	Polling Consumer. Message Endpoint	Read
Combine/ correlate	Match records referring to the same entity across sources, resolve identity	Reconciliation agent	Aggregator. Correlation Identifier	Extract
Enrich/ populate	Fill missing fields and metadata from corroborating sources, attach provenance to each value	Enrichment agent	Content Enricher	Extract
Review/adjust	Expert sees the unified entity with each value's source and confidence, corrects and resolves, low-confidence items flagged	Human reviewer via portal queue	Not applicable (human gate)	Review
Store	Write the validated, enriched record to a target system or data source, preserving the audit trail	Persistence agent via target connector	Message Translator. Channel Adapter	Process

Source: Author's own elaboration.

The target system or data source is chosen according to the institution's needs, for example, an institutional database suits structured operational records; a document management system suits cases where the enriched entity is a document with recovered metadata; and a knowledge graph suits cases where the value lies in the relationships between entities, since the enriched record becomes a node with typed links, each link carrying its provenance. The knowledge graph fits the goal of unified, more complete knowledge of entities most fully, because it preserves not only the completed record but also its connections to others.

5.2 Accessible digital assistance

The second use case brings the capability to the user, instead of the user to the capability. Accessible digital assistance means embedding AI features inside the tools staff already operate, most commonly a word processor such as Microsoft Word, though a registry interface or a legislative editor serves equally. Drafting suggestions, consistency checks, evidence retrieval and citation insertion appear as a familiar panel or annotation. The same retrieval-grounded agents that power evidence-grounded access in the portal are applied as in-document assistance, with suggestions the user accepts or rejects. Here, the retrieval and drafting-assist capabilities appear together as one assistive experience. Adoption cost is low, because there is no new application or parallel workflow. The human-oversight principle is satisfied as every suggestion is reviewed before it is committed, and the work stays within the institution's environment.

6. Additional Capabilities and the Sovereignty Boundary

A study commissioned by the European Commission asked legislative drafting experts to prioritise candidate smart functionalities for a legal drafting environment (Fitsilis, Mikros, et al., 2024). The prioritised capabilities (see Table 4) map onto the platform proposed here. Re-expressed in editor-agnostic terms, they form a family that the same agents can deliver, either through an embedded editor application or through the portal, under a single sovereignty model.

Table 4
Prioritised capabilities, their agents, surfaces and sovereignty decisions

Capability	Agent	Surface	Sovereignty decision
Present relevant and related legislation while drafting	Retrieval agent	Embedded (e.g. Word panel)	Local index. Model interchangeable
Suggest formulations consistent with structure and institutional style	Drafting-assist agent	Embedded	Local. Small model sufficient
Verify entities, flag structures that risk misinterpretation	Verification agent	Embedded annotation or portal sweep	Local. No LLM required
Check internal alignment and cross-version divergence	Comparison agent	Portal (corpus), embedded (spot check)	Local. Semantic similarity, no LLM
Detect and tag obligations, rights, permissions, penalties	Extraction agent	Portal (register)	Local. Detection, not generation
Generate legal or policy text	Generation agent	Embedded, review-gated	On-premises model or secure external interface, decided per workflow

Source: Author's own elaboration.

Most of these capabilities do not require an LLM, since established techniques such as named entity recognition, semantic similarity and information extraction can deliver them locally, while language models can further strengthen them where the institution chooses to apply one. This is direct evidence for incremental adoption, since an institution can deploy the retrieval, verification, comparison, and extraction capabilities first, at low risk and within its own environment, before confronting any harder decision such as which cloud-based LLMs can or should be used.

Moreover, text generation is the single capability that forces the sovereignty decision. It is the one function that may require the institution to choose early on between an on-premises deployment and a secure external interface, weighing data sensitivity, regulatory exposure and infrastructure cost. For these reasons, the study recommended deferring the generation functionality from its near-term roadmap. Our architecture does not resolve the decision but it makes the decision explicit, isolates it to the workflows that require generation, and ensures that the surrounding workflow is unchanged whichever choice is made.

7. Conclusions

The case for treating architecture as a governance decision is, in the end, a case for sequencing. An institutional digital infrastructure that begins with a sovereign, vendor-neutral, integration-first architecture can adopt AI capability incrementally, deploying the local, lower-risk,

non-generative capabilities first, building staff confidence and institutional trust, and deferring the capabilities that force harder sovereignty and infrastructure decisions until it is ready to make them deliberately. An institution that begins instead with a capable model wired directly into a commercial cloud service may show faster early results, but it accumulates governance debt that comes due at the moment of audit, litigation or regulatory scrutiny.

The cited examples originate in a parliamentary background, but the entity enrichment workflow, the embedded assistance pattern and the sovereignty boundary apply equally to almost any public institution. As generative models grow more capable of imitating institutional style, signed and provenance-bearing outputs may become the institutional defence against forgery, such as of the kind specified by the C2PA content credentials standard (Coalition for Content Provenance and Authenticity, n.d.). A direction for future research is to systematically study how much operational time these capabilities save, measured in real public workspaces.

Such research would also help define what would count as evidence for the framework.

Bibliography

Cemri, M., Pan, M. Z., Yang, S., Agrawal, L. A., Chopra, B., Tiwari, R., Keutzer, K., Parameswaran, A., Klein, D., Ramchandran, K., Zaharia, M., Gonzalez, J. E., & Stoica, I. (2025). *Why do multi-agent LLM systems fail?* (arXiv:2503.13657). arXiv. https://doi.org/10.48550/arXiv.2503.13657

Coalition for Content Provenance and Authenticity. (n.d.). *C2PA technical specification*. https://c2pa.org/specifications/

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L, 2024/1689. http://data.europa.eu/eli/reg/2024/1689/oj

Fitsilis, F., & De Almeida, P. G. R. (2024). Artificial intelligence and its regulation in representative institutions. In Charalabidis, Y., Medaglia, R. & Van Noordt, C. (Eds.), *Research handbook on public management and artificial intelligence* (pp. 151-166). Edward Elgar Publishing. https://doi.org/10.4337/9781802207347

Fitsilis, F., Gatos, B., Palaiologos, K., Kaddas, P., Kyrkos, C., Georgoulea, M. E., Armenakis, Y., Tasouli, C., Mikros, G. K., Rozenberg, O., & Kiousi, E. (2024). Digitization of written parliamentary questions from the historical archive (1974-1977) of the Hellenic Parliament. In Mouchère, H. & Zhu, A. (Eds.), *Document analysis and recognition, ICDAR 2024 workshops* (Lecture Notes in Computer Science, Vol. 14935, pp. 103-117). Springer. https://doi.org/10.1007/978-3-031-70645-5_8

Fitsilis, F., Mikros, G., & Leventis, S. (2024). *Overview of smart functionalities in drafting legislation in LEOS, Augmented LEOS final report*. Directorate-General for Digital Services, European Commission. https://interoperable-europe.ec.europa.eu/sites/default/files/document/2024-06/Augmented%20LEOS%20with%20smart%20functionalities%20final%20report.pdf

Fitsilis, F., Von Lucke, J., & De Vrieze, F. (Eds.). (2024). *Guidelines for AI in parliaments*. Westminster Foundation for Democracy. https://www.wfd.org/sites/default/files/2024-07/wfd-ai-guidelines-for-parliaments-2024-english.pdf

Foundation for Intelligent Physical Agents. (2002). *FIPA ACL message structure specification* (SC00061G).

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns, elements of reusable object-oriented software*. Addison-Wesley Professional.

Hohpe, G., & Woolf, B. (2004). *Enterprise integration patterns, designing, building, and deploying messaging solutions*. Addison-Wesley Professional.

International Organization for Standardization. (2022). *Information technology, artificial intelligence, artificial intelligence concepts and terminology* (ISO/IEC Standard No. 22989:2022).

International Organization for Standardization. (2023). *Information technology, artificial intelligence, management system* (ISO/IEC Standard No. 42001:2023).

Inter-Parliamentary Union. (2024). *World e-Parliament report 2024*. https://www.ipu.org/resources/publications/reports/2024-10/world-e-parliament-report-2024

Koryzis, D., Dalas, A., Spiliotopoulos, D., & Fitsilis, F. (2021). ParlTech, transformation framework for the digital parliament. *Big Data and Cognitive Computing, 5*(1), 15. https://doi.org/10.3390/bdcc5010015

Leventis, S., Anastasiou, V., & Fitsilis, F. (2020). Application of enterprise integration patterns for the digital transformation of parliamentary control. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2020)* (pp. 738-741). Association for Computing Machinery. https://doi.org/10.1145/3428502.3428612

Leventis, S., Fitsilis, F., & Anastasiou, V. (2021). Diversification of Legislation Editing Open Software (LEOS) using software agents, transforming parliamentary control of the Hellenic Parliament into big open legal data. *Big Data and Cognitive Computing, 5*(3), 45. https://doi.org/10.3390/bdcc5030045

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459-9474.

OASIS. (2018). *Akoma Ntoso version 1.0* (OASIS Standard).

Rao, A. S., & Georgeff, M. P. (1995). BDI agents, from theory to practice. In *Proceedings of the First International Conference on Multi-Agent Systems* (pp. 312–319). AAAI Press.

Smith, R. G. (1980). The Contract Net Protocol, high-level communication and control in a distributed problem solver. *IEEE Transactions on Computers, C-29*(12), 1104-1113. https://doi.org/10.1109/TC.1980.1675516

Von Lucke, J., Fitsilis, F., & Gagnon, S. (2024). Comparative analysis of the relevance and priority for artificial intelligence tools, services and open questions in the Hellenic, Argentinian and Canadian parliaments. *International Journal of Parliamentary Studies, 4*(2), 182-212. https://doi.org/10.1163/26668912-bja10089

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents, theory and practice. *The Knowledge Engineering Review, 10*(2), 115-152. https://doi.org/10.1017/S0269888900008122

Transparency in the use of artificial intelligence

The author used Claude Fable for language correction and formatting. He checked and approved all outputs and retain full responsibility for the manuscript's accuracy, originality, integrity and conclusions.

Authorship roles notice

The author declares having performed all authorship roles for this article.

Conflict of interest notice

The author is the founder of Qognity, which develops the Qognity Suite, an agentic workflow automation platform designed according to the reference architecture presented in this paper. No other conflicts of interest exist.