DGF CONSULT

POSITIONING PAPER

Policy-Anchored AI

A pattern for compliance-grade public sector decision support.

AI assists the reasoning process.
Controlled logic governs the decision path.

Summary

Public sector institutions face a difficult choice in adopting artificial intelligence for compliance-sensitive domains. General-purpose generative AI is useful and increasingly mature, but produces outputs that are not defensible under review: a confident answer with no traceable path back to the policy authority that should govern it. The alternative — refusing to use AI at all — concedes the productivity gains and leaves officials worse-equipped than they need to be.

This paper describes a third path. **Policy-anchored AI** is a pattern in which compliance-sensitive conclusions are governed by a structured, versioned, traceable model of policy authority, while a language model is used for interpretation, explanation, and the parts of the work where natural language is the right tool. The boundary between the two layers is enforced by design.

The pattern is implemented today in a working application prototype: **GC Procurement Advisor**, a procurement decision-support tool tested against Government of Canada procurement scenarios. The prototype is available for structured walkthroughs with public-sector leaders, procurement officials, and potential pilot partners. It is intended to be applicable to other policy-heavy domains in which compliance, defensibility, and audit traceability are non-negotiable.

This paper describes the pattern, the operating disciplines that distinguish it from adjacent approaches, a worked procurement example, and the measures by which it should be evaluated. It does not describe the implementation. The implementation is the subject of separate technical material available under appropriate review conditions.

1. The problem

Public sector institutions are under increasing pressure to adopt artificial intelligence in domains where unconstrained generative models are not acceptable. Procurement, privacy assessment, security authorization, grants and contributions, regulatory triage, program eligibility, official languages, accessibility, and investment governance share a common pattern: officials need decision support, but the institution requires control, consistency, defensibility, and an audit trail that survives challenge.

A general-purpose language model is poorly suited to these domains, but not for the reason most often cited. The risk is not principally hallucination — the fabrication of plausible-sounding falsehoods. The deeper risk is *uncontrolled reasoning*: a model that produces a coherent, confident, well-written answer that quietly skips required steps, misses thresholds, invents obligations, or overstates certainty in a way that no reader can detect from the output alone.

The output of such a system may be useful most of the time, but it is not defensible any of the time. In a compliance-sensitive domain, that is the wrong trade.

There is a second, less visible problem. Most public sector decision pathways are not lookups against a single rule. They are intersections of overlapping authorities — a Treasury Board directive, an enabling Act, a delegated instrument, a procurement vehicle's terms, a trade agreement, a governance trigger. A retrieval-augmented chatbot that finds and summarizes the most relevant paragraph of policy will miss the intersections by construction. The relevant policy authority is rarely a single document; it is the relationship between several.

A system that cannot represent those relationships cannot reason about them, regardless of how fluent its prose.

2. What "policy-anchored" means

A policy-anchored AI system is one in which:

- **The authority for compliance-sensitive conclusions does not rest with a language model.** It rests with a structured representation of policy — a versioned, traceable model of rules, thresholds, authorities, and their relationships.
- **The language model is used where it is genuinely strongest:** translating informal natural-language input into structured facts, identifying what information is missing, explaining results in plain language, and drafting follow-on artifacts.
- **The boundary between the two layers is enforced, not advisory.** The language model cannot, by construction, override or invent the conclusions of the policy layer.

The phrase that captures the design is:

> *AI assists the reasoning process. Controlled logic governs the decision path.*

This is not a marketing distinction. It is an architectural one. A system either has this boundary or it does not, and the difference is visible to any reviewer who knows where to look.

The pattern is distinct from approaches that rely on prompting, fine-tuning, or retrieval to *encourage* compliant behaviour from a language model. Those approaches reduce the rate of failure; they do not bound it. A policy-anchored system bounds the failure mode by removing the language model from the path of authority entirely.

3. Three operating disciplines

A policy-anchored system is recognizable by three operating disciplines. These are framed here as outcomes, because that is what an official reviewing the system needs to verify.

3.1 Fact discipline

The system treats facts as a first-class concern. Before producing a conclusion, it distinguishes:

- facts the user has provided;
- facts the system has confidently inferred from structured signals;
- facts that are missing;
- facts that are not applicable to this situation;
- conclusions that are *provisional* because the underlying facts are incomplete.

The corollary, and the one most often absent from AI tools in this category, is that the system must be willing to refuse. When the facts do not support a conclusion, the system says so, in plain language, and identifies what additional information would resolve the question.

In a retrieval-augmented chatbot, “I don’t know” is a failure mode the model is trained to avoid. In a policy-anchored system, it is a correct answer and a designed behaviour.

3.2 Traceable conclusion

Every conclusion the system produces is linked to the specific policy elements that support it, at the version they had on the date of the decision. A reviewer reading a recommendation six months later can:

- see the facts the conclusion was based on;
- see the rules that were applied;
- see the version of those rules at the time of decision;
- see what the system considered and rejected;
- reproduce the same conclusion from the same inputs.

This is traceability as a property of the architecture, not a feature retrofitted onto a log file. It is what makes the system defensible under review, under audit, and under access-to-information request.

3.3 Identity-stable output

Identical situations produce identical conclusions, regardless of who is asking, when they are asking, or how they phrased the question. This is a design invariant, not an aspiration.

The practical implication is that the system cannot be persuaded by phrasing, role, urgency, or social pressure to produce a different compliance conclusion than the policy supports. A user who rephrases the same request five times receives the same policy answer five times. A senior official receives the same answer a junior analyst receives.

This property is the one most often quietly missing from generative AI tools, and the one most consequential when officials begin to rely on a system. Where it is absent, the system can be gamed. Where it is present, it cannot.

4. A worked example: government procurement

A common public sector intake reads roughly:

> *"We need a cloud workflow tool for 200 users to manage internal case review tasks. The system may handle Protected B information including personal information. Estimated cost is $180,000 per year for three years."*

A general-purpose chatbot responding to this request will produce something plausible — a comparison of procurement vehicles, perhaps an answer about whether the requirement could be sole-sourced, perhaps a list of policy documents to read. Whether the answer is correct is, in practice, untestable: the user has no way to verify it, and the output carries no link back to the policy elements that govern it. The interaction feels helpful and produces nothing the institution can rely on downstream.

A policy-anchored advisor responding to the same request produces something different in kind.

It begins by translating the informal description into a structured account of the requirement: what is being acquired, the lifecycle value of the commitment, the data sensitivity, the presence of personal information, the user population, the apparent delivery model. It identifies what is known and what is missing — and, critically, it surfaces the missing items as questions the user can answer before the recommendation is finalized, not as silent assumptions baked into a confident-sounding output.

It then surfaces the governance implications that the structured facts trigger. The presence of personal information indicates a privacy assessment obligation. The Protected B sensitivity raises security assessment and authorization considerations and data residency questions. The lifecycle value crosses thresholds at which certain trade agreement obligations apply. The cloud delivery model directs the conversation toward specific procurement vehicles for which the requirement is eligible. Each of these conclusions is linked to the policy element that produced it, at the version current on the day of the recommendation.

Where the facts are incomplete, the system says so explicitly. If the user has not specified whether the personal information includes that of vulnerable populations, the privacy posture is marked as *provisional* and the question is surfaced. If the user has not confirmed whether the procurement is intended to be competitive, the path recommendation is marked as *conditional* on that confirmation.

What the user receives is not an answer dressed as advice. It is a structured account of the requirement, the policy obligations it triggers, the questions that remain open, and the path forward — every element of which a procurement officer can verify, challenge, or override, with a clear view of what would change downstream if any input is revised.

The procurement officer remains the decision-maker. The system makes the procurement officer’s job tractable on a request that, in current practice, often arrives without the information the officer needs to advise on it.

5. What this pattern is not

The pattern described here is sometimes confused with adjacent things it is not. The distinctions matter to anyone evaluating it.

It is not a chatbot. A chatbot's authority for its answers rests in the language model. A policy-anchored advisor's authority rests in the structured policy model behind it. The interface is conversational; the authority is not.

It is not retrieval-augmented generation. A retrieval-augmented system finds relevant text and summarizes it. A policy-anchored system resolves a structured situation against structured policy, and uses retrieval only to explain results. The difference is most visible when the right answer is "not enough information to conclude that." Retrieval-augmented systems do not naturally produce that answer; policy-anchored systems are designed to.

It is not an expert system with a chat interface. A traditional expert system requires the user to know what to ask and how to phrase it. A policy-anchored advisor uses a language model to translate informal user input into the structured form the policy layer requires, and to explain results in language the user can act on.

It is not a workflow tool. A workflow tool routes a request through a series of steps. A policy-anchored advisor reasons about what the request implies, what is missing from it, and what policy obligations it triggers — before any workflow step has been chosen.

It is not a replacement for human authority. It is decision support that produces traceable, defensible inputs for human decision-makers who remain accountable. Its highest aspiration is to ensure that when the official decides, they decide with the right facts on the table.

6. What success looks like

A policy-anchored system in a procurement setting should be measurable against outcomes that matter to the institution. Among the measures we consider load-bearing:

- the rate at which trade agreement obligations are correctly identified, particularly the rate of *missed* obligations, which is the dangerous direction of error;
- the rate at which governance triggers — privacy assessment, security authorization, accessibility, official languages, algorithmic impact, Indigenous procurement — are correctly identified at intake, before they are discovered late in the process;

- the proportion of requirements arriving at procurement teams that are complete enough to proceed without further clarification rounds;
- the agreement rate between the system and senior procurement officers on the appropriate procurement path, measured on a held-out evaluation set;
- the proportion of recommendations whose every claim can be traced back to a versioned policy element — a property the system either does or does not exhibit by construction.

These measures share a common property: they are *empirical*. They are claims the system can be evaluated against, rather than claims it can be marketed with. We intend to publish results against a constructed evaluation set, including failure cases. The objective is not to demonstrate that the system is always right. It is to demonstrate that it is honest about when it is not, and that its reasoning can be inspected when it is.

7. Why this matters now

Public sector institutions are under increasing pressure to adopt artificial intelligence in operationally meaningful ways. The path that has been most readily available, and most heavily marketed, is to deploy general-purpose generative models into productivity workflows. For routine work, this is reasonable. For compliance-sensitive decision support, it is not sufficient, and the cost of discovering the gap in production is asymmetric: a single confidently-wrong output in a procurement, a privacy assessment, or a security authorization is a substantially worse outcome than an honest “we cannot conclude that yet” delivered ten times.

Policy-anchored AI is the pattern that lets institutions adopt AI in these domains without lowering their standards for control, consistency, or defensibility. It is not the pattern for every public sector AI use case. It is the right pattern for the use cases where the current alternative — an unconstrained model loose in a compliance environment — is not yet safe enough to authorize.

The objective of the work described here is to make the pattern concrete, to demonstrate it in a domain that meaningfully tests it, and to make the resulting evaluation evidence available to the institutions that will need to make adoption decisions on the basis of more than vendor marketing.

ABOUT THIS DOCUMENT

This is a working draft of the public-facing positioning paper for the Policy-Anchored AI pattern, prepared for review prior to publication. It describes the pattern at a level appropriate for public dissemination. Implementation specifics — the structure of the policy authority model, the deterministic reasoning layer, the boundary-enforcement mechanisms, the evaluation set, and the prompt and orchestration design — are documented separately and made available only under appropriate review conditions.

The procurement advisor referenced throughout is implemented as **GC Procurement Advisor**, a working application prototype for Government of Canada procurement use cases. Demonstrations and structured walkthroughs are available on request. A 90-day pilot proposal is also available for organizations interested in testing the pattern against real intake scenarios.

CONTACT

Daniel Fallon

DGF Consult
danielfallon@dgfconsult.com
dgfconsult.com