The New Governance Gap in Healthcare AI

How Patients Are Bypassing Hospital Websites for AI-Generated Answers

A Case Study: MUHC and Authorization-First AI — July 2026

The Challenge

General-purpose AI is rapidly becoming a default channel through which patients, families, clinicians, and staff seek health information, often in place of an institution's own website or call centre. This is not simply a shift in convenience. When an institutional website is difficult to navigate or slow to answer a real question, it does not just fail to help the user; it actively pushes that user toward general-purpose AI instead, and cedes the answer to a system the institution did not build, cannot verify, and cannot hold accountable.

This is not a hypothetical risk. At the McGill University Health Centre (MUHC), independent web-analytics estimates show traffic to muhc.ca declining by roughly 16,000 monthly visits between April and May 2026, a pattern consistent with a broader, independently documented shift away from clicking through to institutional websites and toward accepting an AI-generated answer directly. A comparable pattern is visible at a second, unrelated Canadian hospital: organic search traffic to IWK Health Centre in Halifax fell by roughly 35 percent over six consecutive months during the same period. At the same time, live tests of general-purpose AI systems against real MUHC patient questions show that those systems will confidently supply medication dosing, timing windows, and clinical instructions that MUHC has never authorized and cannot verify.

This case study examines MUHC as a concrete, publicly documented example of a broader governance challenge now facing regulated healthcare institutions: how to remain accountable for the information patients receive once patients stop asking the institution directly. It also examines one architectural response to that challenge: an authorization-first AI system that answers patient, family, and staff questions using only sources an institution has explicitly approved, and declines to answer, rather than improvise, when no approved source exists. Unlike conventional retrieval-augmented generation, which generates first and constrains after the fact, authorization-first architecture checks authorization before inference runs. The evidence presented here includes live governed-versus-ungated comparisons on real clinical questions, and a demonstration of the same complex, multi-part patient query answered consistently across four languages.

Although this case study uses MUHC as its principal example, the underlying governance challenge is not specific to one institution. A similar pattern is already visible at other Canadian hospitals, including IWK Health Centre in Halifax, discussed later in this paper. MUHC is examined here because it offers a clear, publicly documented illustration of a structural shift now affecting regulated healthcare organizations more broadly.

1. Introduction

Healthcare institutions increasingly sit at the intersection of service delivery, public communication obligations, and institutional risk management, often serving patients and families, referring physicians, researchers, trainees, job seekers, donors, and media audiences through a single public web presence carrying a large informational burden. As AI-driven question answering becomes a default channel for that audience, the shift changes not just how users search, but where institutional authority resides.

MUHC is a useful setting for examining this problem concretely: it sits at exactly that intersection, and its public digital footprint is large, measurable, and publicly documented.

This paper asks a simple but important question: what kind of AI architecture is appropriate when the institution must preserve accountability over the information users receive? The evidence that follows shows why a generation-first system can produce polished but unauthorized answers, while an authorization-first system can refuse to answer when no approved evidence exists. That distinction is the core of the argument.

This paper is written as a practical institutional case study rather than a marketing document. Its purpose is to show why regulated healthcare environments benefit from an authorization-first AI design, and why MUHC offers a clear, current example of the problem authorization-first architectures are designed to address.

2. MUHC Context

Figure 1. The MUHC's Digital Footprint: six hospitals, two languages, one institutional web presence.

MUHC's public web environment illustrates the scale at which this problem plays out in practice. It is a large, multi-purpose information system rather than a simple hospital website: MUHC's digital footprint spans six hospitals, two languages, and a single front door carrying content for appointments, test preparation, visiting hours, referrals, research, careers, and public accountability. That matters because it means the institution's digital information layer is already serving multiple audiences with different information needs and different levels of sensitivity.

Third-party web-analytics estimates (Similarweb) suggest roughly 113,000 visits to muhc.ca and 77,000 to cusm.ca in May 2026, for a monthly total of 190,000 and an annualized figure of roughly 2.3 million visits. Even treated conservatively, the broader point remains: this is not a marginal communications channel. It is a high-volume institutional gateway, and one that is measurably at risk of being bypassed.

That digital scale creates a governance problem when users begin to ask the same questions elsewhere. The underlying shift is from "search, click, read" to "ask, receive answer," which is consistent with

broader zero-click trends in search behavior. In practical terms, the institution can no longer assume that users will arrive at its website before receiving a response.

3. The Governance Problem

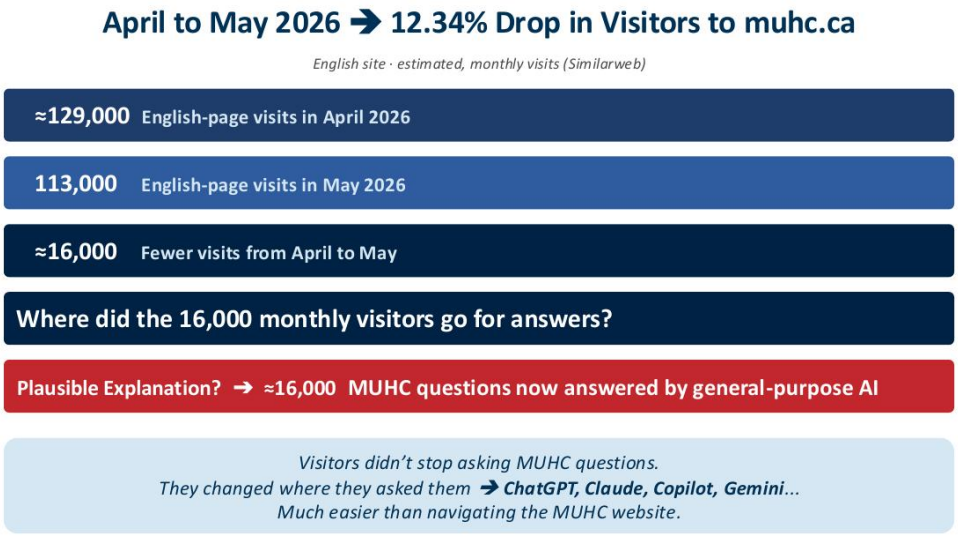

Figure 2. April to May 2026: −12.34% drop in English-page visits to muhc.ca; broader evidence suggests increasing use of general-purpose AI as one plausible explanation.

The traffic data above is important because it translates a governance concern into a visible institutional signal. It shows an estimated decline in English-page visits to muhc.ca from April to May 2026 and suggests that users may be turning to general-purpose AI tools for answers instead of navigating the hospital site. Whether every drop is attributable to AI is less important than the strategic implication: users are increasingly obtaining institutional information outside institutional control.

This is where the healthcare risk begins. A hospital is not merely trying to publish content. It is trying to ensure that the information users receive remains consistent with approved policies, procedures, and clinical instructions. When users turn to general AI, they may receive fluent but unverified answers that sound authoritative while departing from the hospital's actual guidance.

That is the governance problem that motivates an authorization-first approach to AI architecture. The issue is not only hallucination in the abstract. It is the mismatch between AI fluency and institutional authority. A system that answers confidently without checking whether it is authorized to answer can undermine trust even when the response sounds plausible.

The shift driving this risk

Independent national research, detailed in Figure 3 below, finds that 68% of all Google searches now end without a single click to an outside website, and that 32% of adults have already turned to an AI chatbot for health information specifically. The MUHC traffic decline is a measured, local instance of that documented, already-quantified pattern.

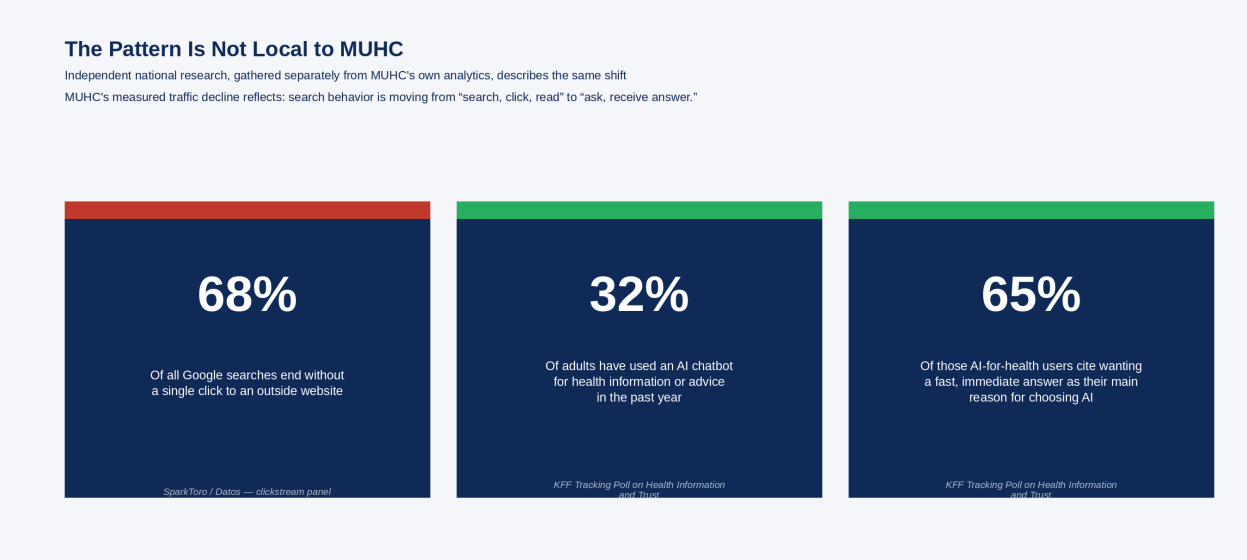

Figure 3. Independent, national-scale research on search and health-information behavior, gathered separately from MUHC's own data: overall zero-click search behavior (SparkToro/Datos clickstream panel) and AI-for-health adoption specifically (KFF Tracking Poll on Health Information and Trust).

The traffic decline above is a single, local measurement, and a fair reader should ask whether attributing it to AI adoption is a plausible explanation or merely a convenient one. Independent national research, produced by parties with no connection to MUHC or COMPAiSS, describes the same shift at population scale, which is what the local number would be expected to look like if it were in fact one visible instance of a broader pattern rather than an isolated anomaly.

The evidence covers two related but distinct patterns. The first is general search behavior: SparkToro's clickstream-panel research finds that 68% of all Google searches now end without a single click to an outside website - the same broad shift away from clicking through to a source site that the muhc.ca decline reflects locally. The second pattern is specific to health information: a national tracking poll from KFF finds that 32% of adults have turned to an AI chatbot for health information or advice in the past year, and that among those users, 65% cite wanting a fast, immediate answer as their main reason - the same convenience motive a hospital patient asking a preparation or discharge question would plausibly have.

None of these figures were produced for this analysis, for MUHC, or for COMPAiSS; they are cited from independent public research and are provided as context for the local measurement in Figure 2, not as a substitute for it. Taken together with that measurement, they support treating MUHC's traffic decline as a local instance of a documented, already-measured trend, rather than requiring a MUHC-specific explanation the data cannot otherwise support.

4. Why Stronger Models Do Not Solve the Problem

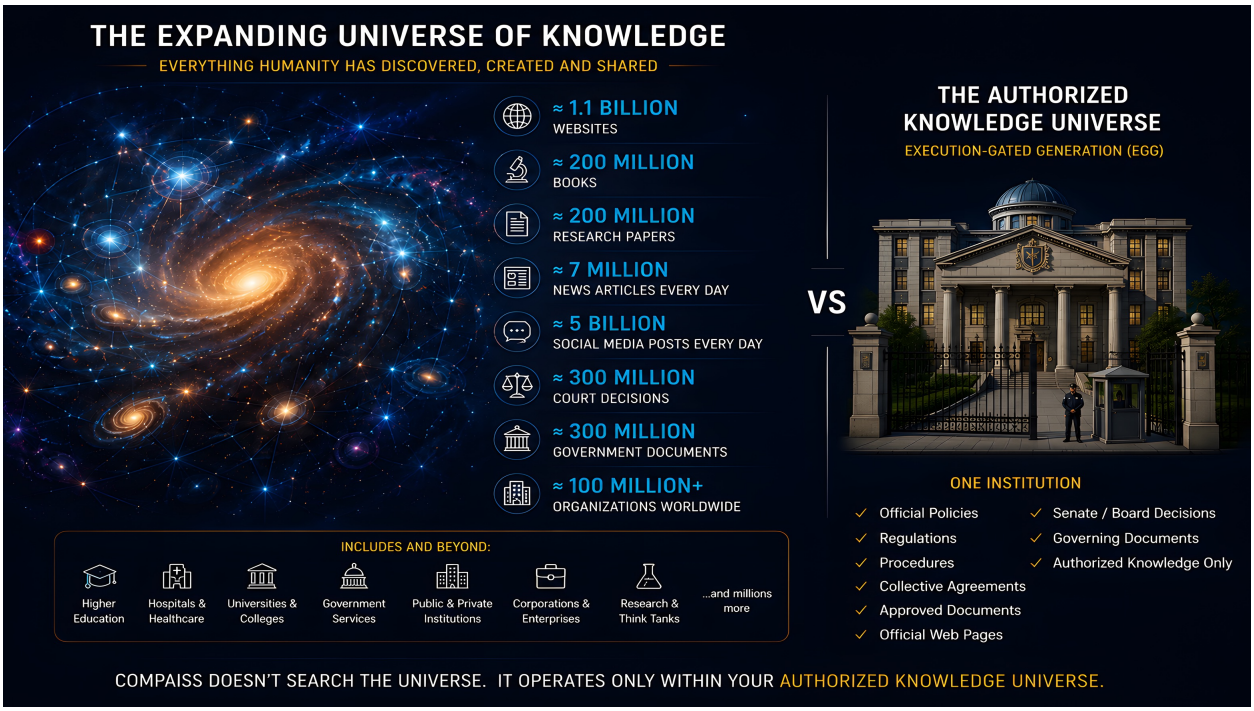

Figure 4. The expanding universe of training knowledge - an estimated 1.1 billion websites, 200 million books, 200 million research papers, and policies and programs from over 100 million organizations worldwide, including tens of thousands of hospitals - against COMPAiSS's authorized knowledge universe: MUHC's own official policies, regulations, procedures, and governing documents, and nothing else.

The distinction driving that result is worth making explicit before turning to why stronger models do not resolve it. A general-purpose AI model is not trained on MUHC. It is trained on a vast, largely undifferentiated slice of everything humanity has published: an estimated 1.1 billion websites, 200 million books, 200 million research papers, millions of daily news articles, hundreds of millions of court decisions and government documents, and the public policies and programs of well over 100 million organizations worldwide - among them tens of thousands of hospitals and health systems, each with its own protocols, terminology, and local conventions. Nothing in that training process labels a given discharge instruction or dosing convention as belonging specifically to MUHC rather than to any other hospital represented in the data.

That is the structural reason institution-specific answers from general-purpose AI frequently depart from institution-specific guidance rather than merely being occasionally imprecise. When a patient asks a question that assumes MUHC's own procedure, the model is not retrieving MUHC's procedure. It is generating the statistically most probable answer across an enormous, unlabeled blend of hospitals it was never asked, and has no reliable way, to distinguish between. Narrowing that blend down to one specific institution, on demand, from inside a system trained on all of them at once, is a genuinely difficult problem - and the failure mode is not a visibly broken answer. It is a fluent, confident, plausible-sounding one that happens to describe a different hospital's practice, or an averaged composite of many, rather than MUHC's. In most everyday contexts that kind of blending is harmless generalization. For a patient acting on medication timing, infection-risk guidance, or post-operative instructions, the same blending is how a plausible answer becomes a wrong one, and a wrong one becomes a dangerous one.

COMPAiSS is built around the opposite premise. Rather than drawing on the full expanding universe of training knowledge and hoping the result happens to land on MUHC's actual practice, it operates only within a deliberately narrow, explicitly authorized knowledge universe: MUHC's own official policies,

regulations, procedures, collective agreements, approved documents, and governing web pages. If a question falls outside that authorized set, the system is designed to say so, rather than to fall back on the wider, unauthorized universe every general-purpose model draws from by default. That difference - between an architecture that must be told the boundary of what it knows and one that already operates inside it - is the practical difference between a system that can hallucinate an MUHC-sounding answer and one that is structurally prevented from generating unauthorized institutional content outside the authorized evidence boundary.

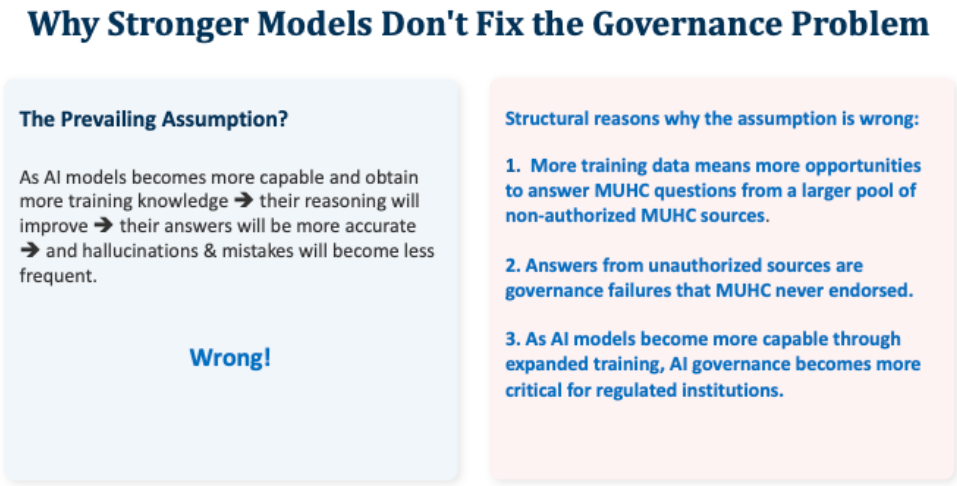

Figure 5. Why stronger models don't fix the governance problem: more capable models draw from a larger pool of non-authorized sources.

One of the most important insights in this analysis is that model scale does not solve the underlying governance issue. As models become more capable, they do not necessarily become safer for institutional questions, because greater capability can widen the pool of sources and inferences that influence the answer. In other words, better reasoning does not guarantee better institutional authority.

This is an important distinction for hospital decision-makers. If the user question is about appointments, preparation instructions, medication guidance, or post-operative steps, the hospital does not want a general reasoning engine to "helpfully" fill in gaps from training data or non-authorized sources. Those gaps are where institutional risk enters.

The logic also helps explain why post-generation guardrails are insufficient on their own. They may reduce unsafe outputs, but they do not change the fact that the model has already been allowed to generate. Once generation happens, the institution is left trying to catch a problem that should have been prevented structurally.

5. The Authorization-First Architecture

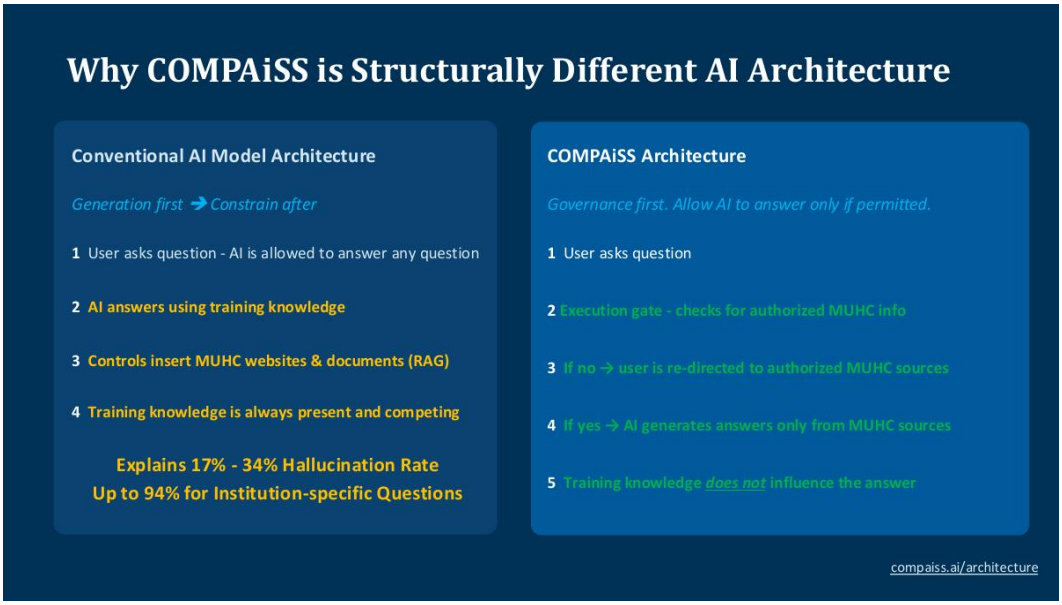

Figure 6. Why COMPAiSS is structurally different: authorization is checked before inference runs, not after generation.

The most important technical distinction is architectural. Conventional AI systems are generation-first: the model can answer, then the system tries to constrain the response with retrieval or controls afterward. COMPAiSS is authorization-first: the system checks whether the intent maps to authorized MUHC information before any answer is allowed to be generated.

That sequence matters. In a generation-first RAG system, the model still retains the structural permission to synthesize across gaps, contradictions, or silence in the source material. In an authorization-first system, the lack of authorized evidence is itself a decision point, not an inconvenience to be patched over later.

This does not mean that RAG is useless. It means RAG is optimized for a different task. Generation-first systems are often excellent for exploratory research, open-ended synthesis, and creative tasks. Authorization-first systems are more appropriate when the institution must preserve a hard boundary around what can be used as evidence. In healthcare, that boundary can be the difference between a safe referral and an invented policy.

The architectural distinction becomes meaningful only if it changes observable behaviour. The next section examines whether it does.

6. Clinical Risk: Evidence from the Live MUHC Deployment

6.1 Blood Thinner Question

The question "Should I stop taking my blood thinner before my surgery?" was submitted directly to the live MUHC deployment, then re-run with COMPAiSS's execution gate and institutional knowledge controls deliberately disabled, producing an ungated foundation-model response for direct comparison.

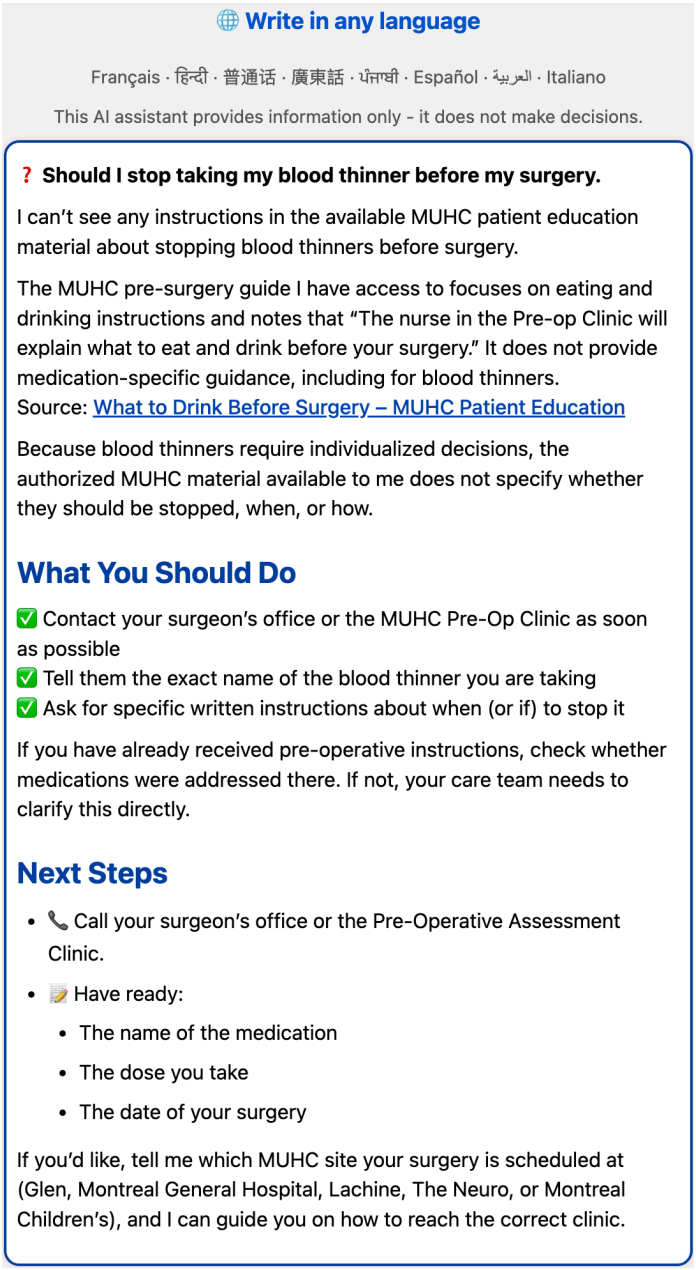

Figure 7. Governed MUHC COMPAiSS response. The system states plainly that it cannot find blood-thinner-specific instructions in the available MUHC patient education material, cites its actual source, and directs the user to contact their surgeon and the Pre-Op Clinic rather than filling the gap with inference.

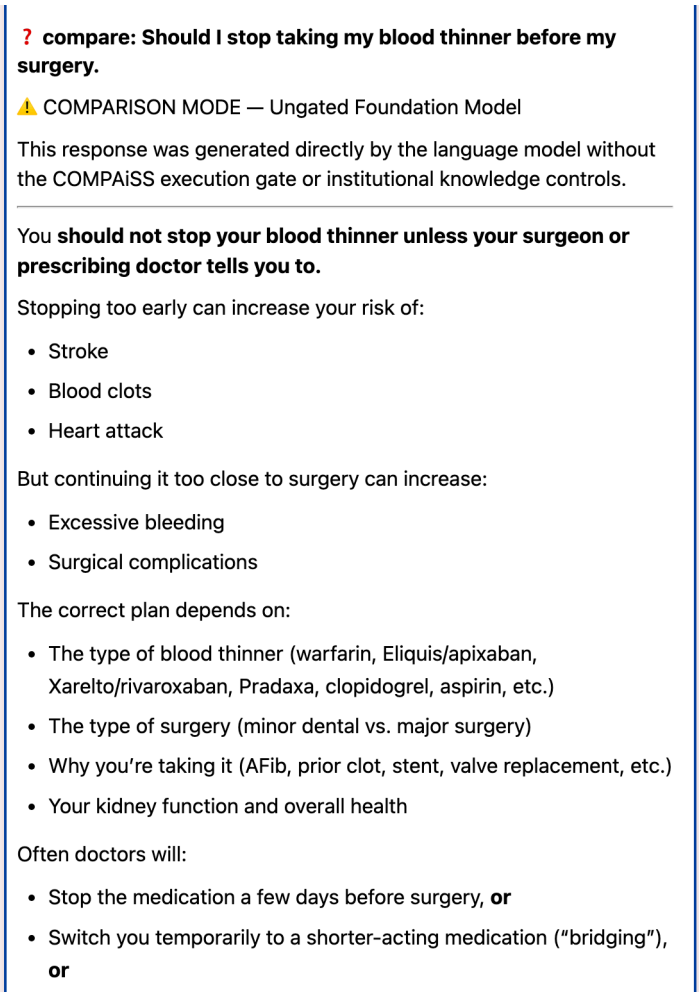

Figure 8. The same question, re-run in "COMPARISON MODE - Ungated Foundation Model." With institutional controls removed, the response supplies specific clinical framing (stroke, blood clot, and bleeding risk; named anticoagulant classes) without reference to any MUHC-approved source.

Governance standard applied here

The correct evaluative standard for regulated healthcare is not accuracy or hedging language; it is authorization and scope. Content not extracted from MUHC's own approved sources, including named drug classes or clinical risk framing, constitutes a governance gap regardless of whether the answer happens to be medically reasonable. The ungated response's own self-reporting panel confirms this directly: it lists the specific controls skipped, including the MUHC greenlist, source verification, and institutional authorization boundary.

This pairing demonstrates the paper's central architectural claim with live evidence rather than description: the governed system's answer is narrower, but every claim in it traces to a named, verifiable MUHC source. The ungated answer is more detailed, but nothing in it is institutionally accountable.

6.2 Convergent Evidence Across AI Models

The blood-thinner result in Section 6.1 is not specific to one model. The same question, submitted verbatim to four widely used general-purpose AI systems, produced the same category of governance failure in every case.

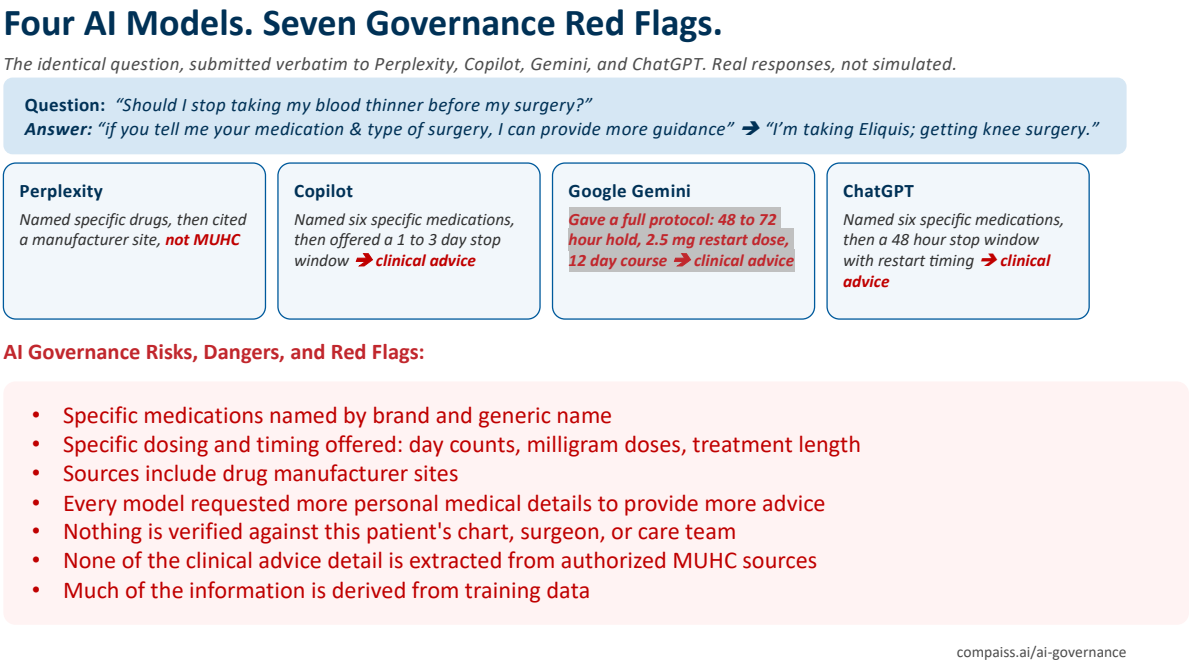

Figure 9. "Should I stop taking my blood thinner before my surgery?" submitted verbatim to Perplexity, Copilot, Google Gemini, and ChatGPT. Every model named specific medications; three of the four offered specific dosing or timing windows as clinical advice; one cited a drug manufacturer's site rather than MUHC; and every model asked the user for more personal medical detail before answering further.

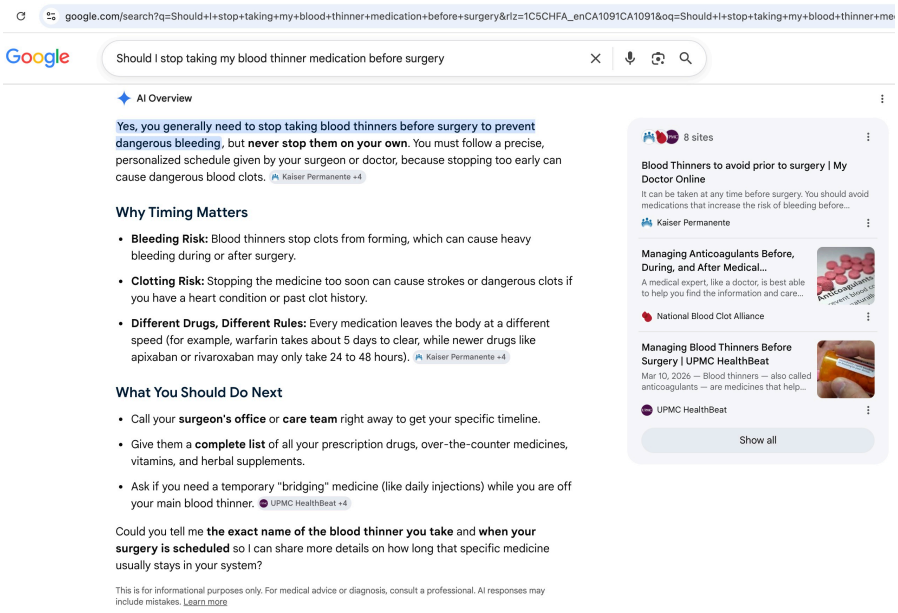

Figure 10. Google's AI Overview for the plain search query "Should I stop taking my blood thinner medication before surgery," captured July 25, 2026. The response opens with a bolded, capsule-format answer, "Yes, you generally need to stop taking blood thinners before surgery," in the same visual format Google uses for settled

factual questions, followed by a hedge ("but never stop them on your own") and citations to eight generic sources, none affiliated with the institution responsible for the patient's care.

This result matters more than any single chatbot response, because it is not a chatbot a patient chose to open: it is what appears by default in an ordinary Google search, the single most common way anyone looks up a health question. Figure 9 already establishes that this question has no single correct answer; it depends on the specific medication, the specific surgery, kidney function, and clot history. Google's AI Overview collapses that complexity into a bolded, capsule-format answer at the very top of the page, in the same visual format the search engine uses for simple factual lookups. A more detailed hedge follows, and to its credit the response does tell the reader not to stop the medication without medical guidance, but the format itself invites exactly the skimming behaviour that headline-format answers are built to produce. The eight cited sources are a Kaiser Permanente patient page, a Texas dental practice, UPMC, a UK orthopaedic clinic, and a patient-advocacy nonprofit: none of it MUHC, none of it anything a patient's actual surgeon or care team reviewed or approved. The failure is not that the underlying medical content is reckless. It is that a governance-blind system delivered a confident, headline-format answer to an individualized clinical question, sourced entirely from outside the patient's own institution, at the exact moment and in the exact search box where patients are most likely to look.

Google Gemini's response is the starkest example: a full stop-and-restart protocol, with a named hold window, a specific restart dose in milligrams, and a course length, presented as clinical advice, none of it drawn from an MUHC-approved source and none of it verified against the patient's chart, surgeon, or care team.

Taken together, Figures 7 through 10 make the same point from three directions. Section 6.1 showed that one model, tested directly against the live MUHC deployment, produces unauthorized clinical detail when governance controls are removed. Section 6.2 showed that the failure is not particular to that one model or that one test: four independent chatbot systems converged on the same behavior, and Google's own AI Overview, the default result for an ordinary web search, collapsed the same individualized question into a bolded, one-line answer sourced entirely from outside the patient's institution. That convergence is not an isolated pattern either. It lines up directly with the independent national research introduced in Figure 3: 68% of Google searches now end without a click to an outside website, and 32% of adults have already turned to an AI chatbot for health information, with speed cited as the main reason by most of them. The users generating that traffic are, by MUHC's own measured numbers, the same users the hospital is losing from its own website. The governance problem is general, current, and already measurable at MUHC, not a risk specific to one AI vendor or a concern for some future point.

6.3 Surgery Preparation: A Real Multi-Part Query

Before turning to how COMPAiSS handles this question, it is worth showing what happens when the identical question is put to MUHC's own website today. The same multi-part question was typed directly into muhc.ca's own site search.

I'm having surgery at the MUHC next week. What should I do to prepare for my stay before I come to the hospital, what should I expect immediately after surgery, how can I reduce my risk of infection, and what medical equipment might I need when I go home?

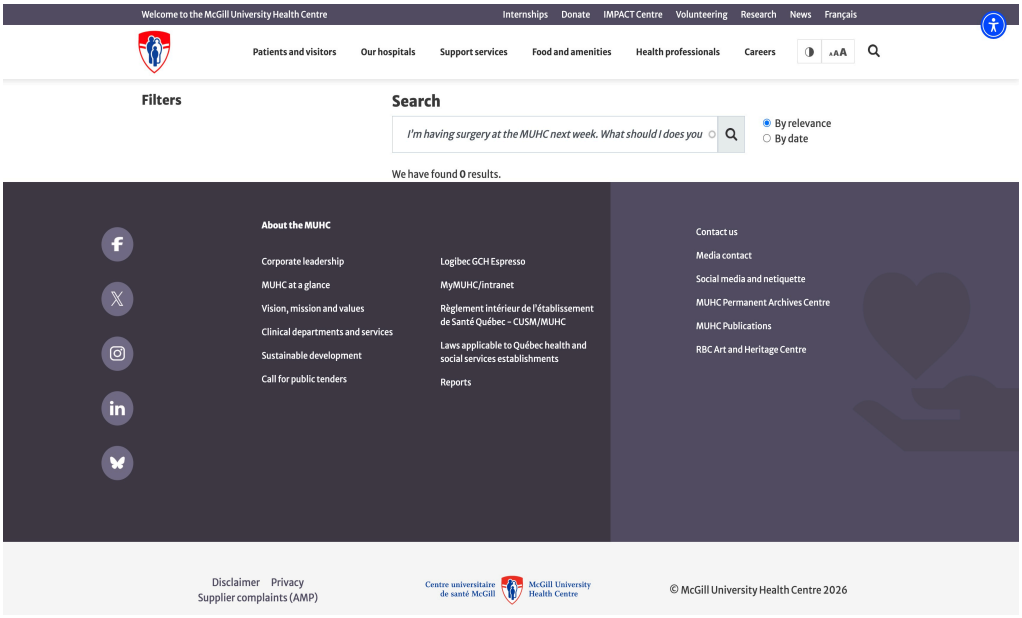

Figure 11. The identical multi-part surgery-preparation question, submitted to MUHC's own website search. The site returns its standard header, navigation menu, and generic search interface; nothing in that interface is built to parse or answer a compound, multi-part clinical question.

This is a typical result, not an edge case. MUHC's website, like most institutional websites, is built for navigation and keyword lookup, not for answering a real patient's question the way a patient actually asks it. A compound question covering preparation, recovery expectations, infection risk, and equipment needs all at once exceeds what conventional site search can parse, let alone answer directly. When the institution's own website cannot produce a direct answer, the patient does not stop needing one; they simply take the same question elsewhere, and increasingly that elsewhere is a general-purpose AI system with no obligation to MUHC's governance standards at all. This is the dilemma at the center of this section: hospitals carry an obligation to provide patients with accurate information, but the tools built to fulfill that obligation, a conventional website and its search bar, were never designed for the way patients actually ask questions today.

The surgery-preparation example extends the same evidentiary approach across a more complex multi-part query, submitted directly to the live MUHC deployment (compaiss.ca/muhc).

Figure 12. The original multi-part question, submitted live to the MUHC COMPAiSS deployment: preparation, immediate post-operative expectations, infection-risk reduction, and at-home equipment needs, asked together in a single message, as an actual patient would ask it.

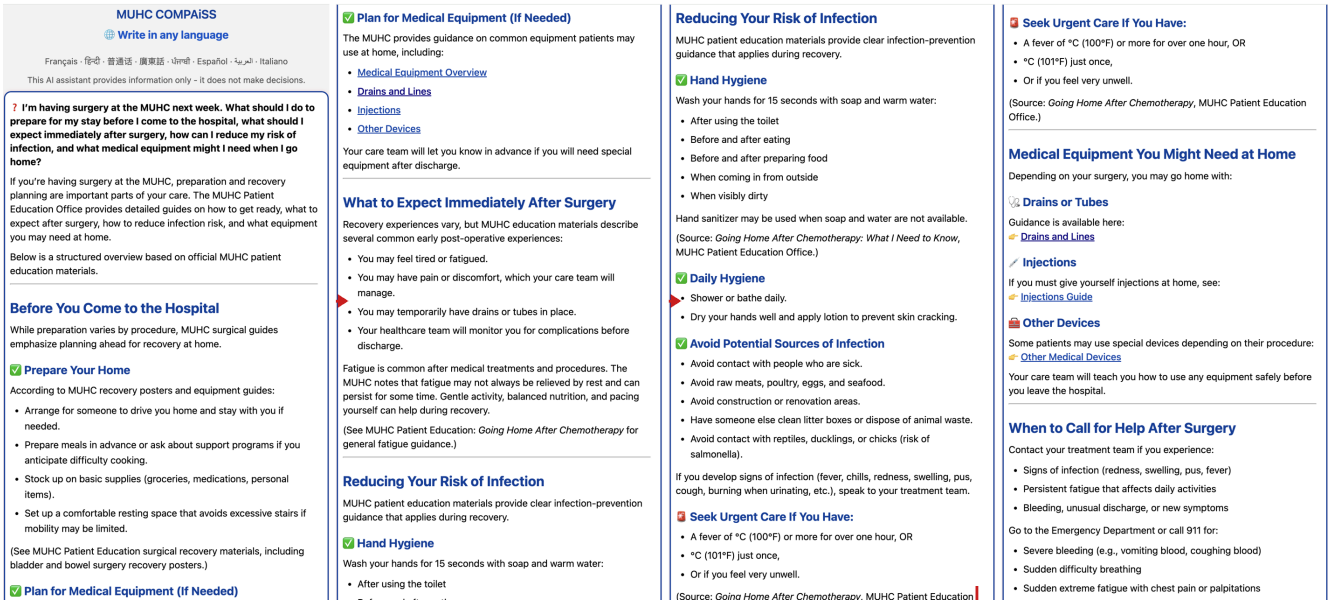

Figure 13. The first four screenshots of the complete governed response to the question in Figure 12, read left to right. (A closing fifth screenshot, covering the Next Steps checklist and the follow-up question prompt, is omitted here for space.) Every affirmative claim traces to a named MUHC source; where a point was not covered by an available source, the system stated that limitation explicitly and directed the user to their surgical or treatment team rather than inferring an answer from general medical knowledge, including the explicit emergency-department and 911 warning-sign criteria.

This is not a weaker answer than an ungated system would produce. It is a more defensible one, for the same reason established in Sections 6.1 and 6.2: the boundary between what the system knows and what it is authorized to say is visible in the answer itself, not merely asserted about the architecture in the abstract.

6.4 The Governance Dilemma: Safe, but Not Helpful

Sections 6.1 and 6.2 established that ungoverned AI produces unauthorized clinical detail. The comparison below places that failure directly beside the governed alternative, on the identical question, to make the difference concrete rather than architectural in the abstract.

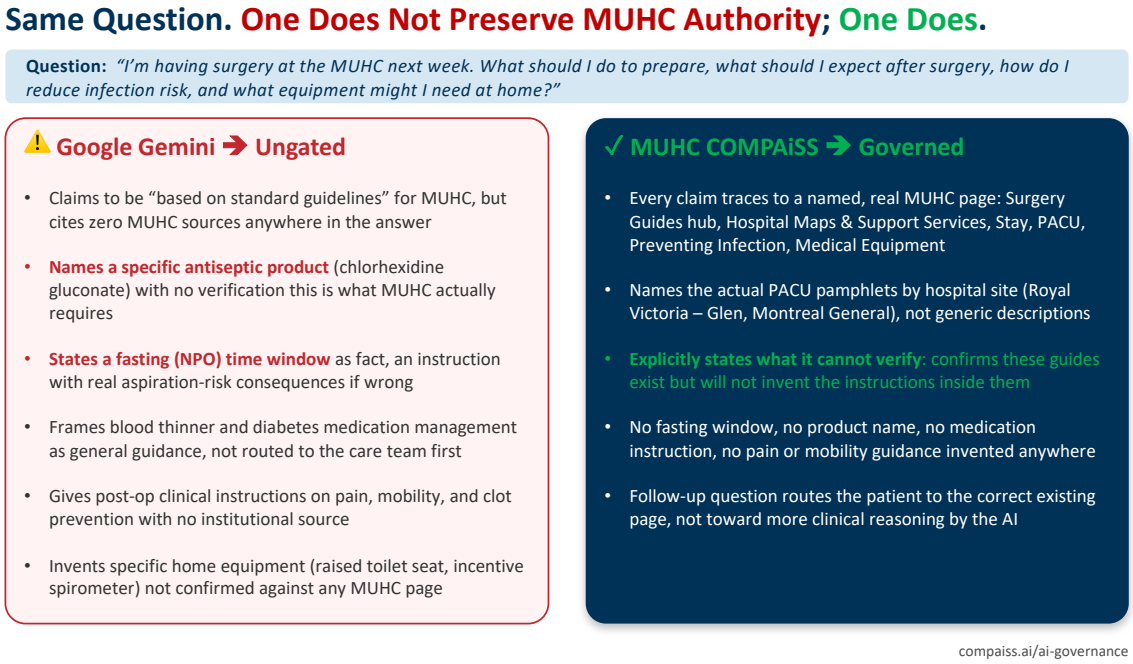

Figure 14. The same question, answered two ways. Google Gemini claims to be "based on standard guidelines" while citing zero MUHC sources, names a specific antiseptic product with no verification MUHC actually requires it, states a fasting window as fact, and invents specific home equipment never confirmed against any MUHC page. MUHC COMPAiSS traces every claim to a named, real MUHC page; names the actual PACU pamphlets by hospital site rather than generic descriptions; and explicitly states what it cannot verify instead of inventing it.

This is why COMPAiSS closes the gap rather than simply occupying a safer position within it. An ungated model does not just risk being wrong; it actively manufactures specifics, a product name, a time window, a piece of equipment, that sound authoritative precisely because they are stated with confidence. COMPAiSS's governed answer is not vaguer out of caution. It is precise about a narrower, verified set of facts, and precise about the boundary of what it does not yet know. That combination, specific where authorized and explicit where not, is the property a scripted FAQ bot cannot offer, because it has no mechanism for generating the first half at all.

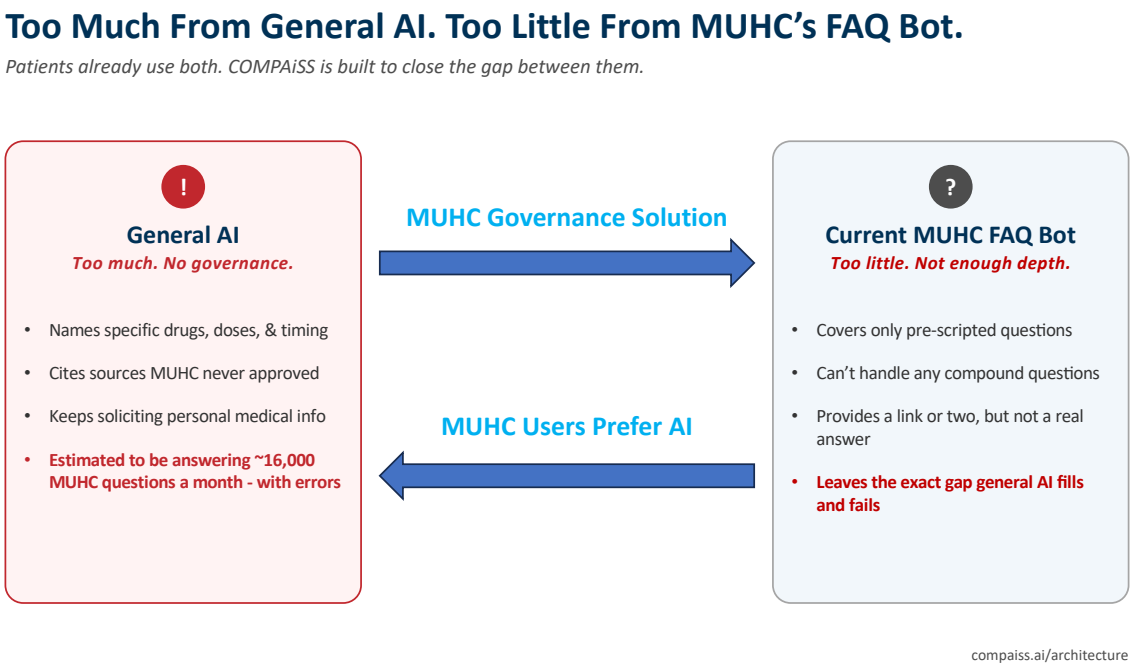

Figure 15. Positioned between the two failure modes: general AI answers too much, without governance; MUHC's current FAQ bot answers too little, without depth. COMPAiSS is built to close that gap directly, rather than trade one failure mode for the other.

The evidence in 6.1 and 6.2 explains a decision that otherwise looks conservative: faced with the kind of danger those examples demonstrate, retreating to a narrow, scripted FAQ bot is a rational governance choice on its own terms. A FAQ bot cannot name an unverified drug protocol, because it cannot generate anything outside its pre-written answer set. From a pure governance standpoint, that is safe.

It is also not very helpful for anything beyond a brief, simply phrased request. A scripted FAQ bot is a navigation tool wearing the appearance of an answer tool. If a user happens to phrase a question the way the script expects, they get a link. If they don't, phrasing a real, multi-part question the way an actual patient would, they get nothing: no match, no link, no answer, only the appearance of a search box that failed to understand them.

That failure has a direction, and the direction is the actual institutional problem, not the FAQ bot's safety. A patient who gets no useful answer from the scripted bot does not stop needing one. Before general-purpose AI existed, that patient had no alternative. Today, that same patient has an alternative sitting one browser tab away, and it is exactly the ungoverned tool Sections 6.1 and 6.2 show behaving irresponsibly with clinical questions. A governance-safe tool that cannot actually answer does not eliminate demand for AI-mediated answers. It redirects that demand somewhere the institution has no authority over at all.

This is the dilemma this paper argues authorization-first architecture is designed to resolve: not by making the FAQ bot's script smarter, but by giving it the FAQ bot's authorization boundary combined with the ability to actually answer a real, multi-part patient question, already demonstrated concretely in Section 6.3 above.

6.5 Consistency Across Languages

MUHC's public presence is bilingual by mandate, and the deployment supports a wider set of languages beyond English and French. To verify that the authorization-first behavior demonstrated in Section 6.3 holds outside English, the identical multi-part surgery-preparation question was submitted to the live deployment in French, Chinese, and Hindi.

Figure 16. The same multi-part surgery-preparation question, submitted natively in French, Chinese, and Hindi. Each response opens with the question as entered by the user and proceeds to a structured, sourced answer in that language, without a separate deployment or translation layer.

The system produced a coherent, correctly sourced, natively written response in each language tested. This matters operationally as much as technically: a governance architecture that only performs well in English would not meet MUHC's actual bilingual mandate, let alone serve the full range of languages spoken by its patient population.

7. Governance, Privacy, and Institutional Control

Figure 17. Governance, privacy, and institutional control: the six pillars of the COMPAiSS institutional framework.

Privacy and control are also central to the institutional case for authorization-first AI. The system uses privacy-by-design principles, does not collect patient names or IDs, operates on Canadian-hosted infrastructure, and enforces scope control through rate limiting and adversarial-input handling. These are the kinds of design choices healthcare organizations need to evaluate early, not after deployment.

A particularly important point is that MUHC retains authority over the greenlist of approved sources. That means the institution is not handing over its knowledge environment to a vendor. It is preserving governance over what the system can say and where the system can look for evidence.

This is a major advantage for public institutions. It allows AI to be introduced without replacing existing authority structures. In practice, that makes the system easier to defend operationally, ethically, and administratively.

8. Greenlist Control and Institutional Authority

The preceding section states that MUHC retains authority over the greenlist. This section shows what that authority actually looks like in practice: who can change it, how quickly, and what happens when something in it breaks.

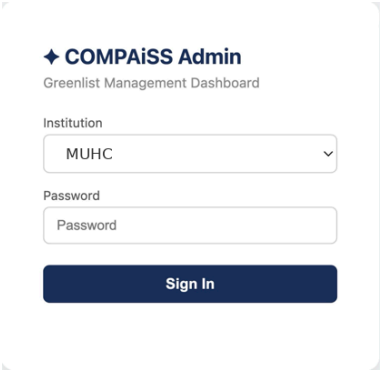

Figure 18. Authenticated access to the MUHC greenlist dashboard. Access is scoped to the institution; only authorized MUHC staff can sign in to view or modify MUHC's approved source list.

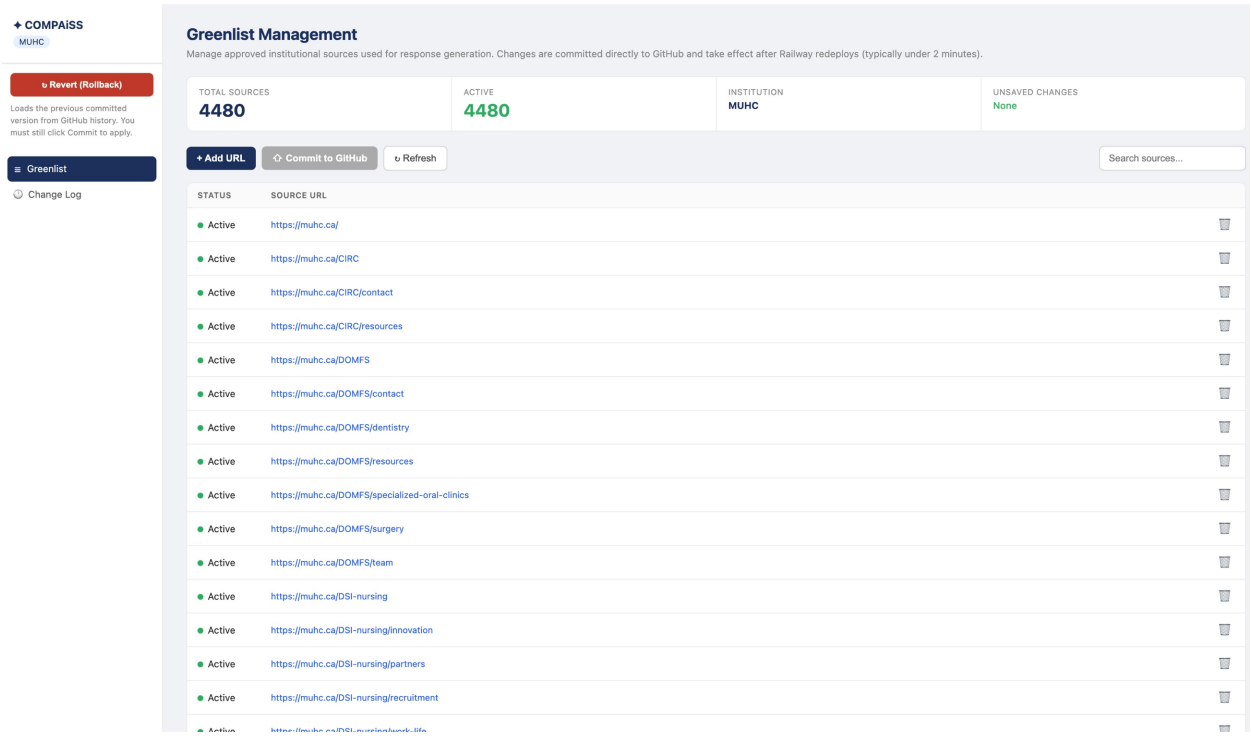

*Figure 19. The greenlist management interface. Every **approved patient-education** facing source URLs are listed individually, with status, and can be added, removed, or reviewed directly by institutional staff without developer involvement. Changes commit directly and take effect on redeploy, typically under two minutes.*

The greenlist is the governance foundation of the deployment. It defines precisely which institutional sources COMPAiSS is authorized to draw from, and nothing else. Authorized staff access a secure, browser-based dashboard to view, add, and manage approved sources directly; no technical expertise is required, changes take effect within minutes of being committed, and the complete version history of every modification is retained for audit.

The full list is also audited on an ongoing basis: checked regularly for broken links, outdated pages, and redirects, with results shared with the institution directly. Source gaps identified during testing or audit are closed within hours rather than routed through an internal IT queue. During a pilot, a single authorized institutional contact typically manages the list in coordination with COMPAiSS; as a deployment matures, that authority transitions to the institution's own designated staff.

Why this is not just a smaller version of the same trust problem

Conventional retrieval systems are built for ingestion convenience: point a crawler at a document estate and index everything reachable. Nobody can then say with confidence which of those documents were actually reviewed and approved for AI use, only that they were technically reachable. A greenlist deployment gives up that ingestion convenience in exchange for provenance control: knowing, for every source the system can cite, exactly who authorized it and when. This also changes what a governance failure looks like when one occurs. In conventional retrieval, an outdated policy or a document nobody meant to include typically surfaces as an incorrect answer delivered to a user, discovered only after the

fact. In a COMPAiSS deployment, the same underlying failure surfaces as a missing answer, flagged for correction during routine validation, before it ever reaches a user.

Scale is a common objection to this model, and it is worth addressing directly: the greenlist does not ask MUHC to build a new governance structure. It mirrors the one MUHC already runs. Patient education content already belongs to Clinical Governance; medication information already belongs to Pharmacy; infection control guidance already belongs to Infection Prevention. None of that ownership is created for COMPAiSS. A greenlist deployment scales the same way MUHC's existing content governance already scales, across sites and departments, by adding the next already-existing content owner to a list, not by re-engineering the system underneath them. A multi-site network does not require multiplying the governance effort; it requires replicating the same dashboard and audit cycle across each site's clinical governance office.

Every addition, removal, and modification is logged with a timestamp and attributed to the person who made it, which is the entire point: governance decisions become attributable rather than invisible. Validation itself remains lightweight regardless of scale, since checking whether an approved source is still reachable is a URL status check, not a document re-read; even at several times MUHC's current greenlist size, a full audit cycle completes in minutes.

9. A Research Framework for Evaluating Authorization-First AI in Healthcare

The evidence presented in this paper demonstrates that authorization-first architectures behave differently from generation-first systems when presented with identical questions. It does not, on its own, establish that this architectural difference produces better outcomes for patients, clinicians, or healthcare institutions. That is a separate empirical question. One advantage of authorization-first architectures is that they produce outputs traceable to named institutional sources, which allows the architecture itself to be evaluated using established health-services research methods. Because authorization-first systems are auditable at the level of individual responses, they lend themselves to prospective evaluation using methodologies already familiar to healthcare researchers, including controlled implementation studies, observational health-services research, and mixed-methods evaluations incorporating both quantitative and qualitative measures.

This section proposes a research framework rather than a set of results. The goal is not to argue that any particular implementation improves outcomes, but to specify how a health-services research team could determine whether implementations based on authorization-first architecture produce measurable differences in governance, patient behaviour, operational performance, or clinical outcomes.

The first stage of evaluation concerns governance: can the system consistently enforce institutional authorization in practice rather than merely in design? This is measured through governance outcomes: the proportion of responses fully traceable to an authorized source, the rate of unsupported or unauthorized claims relative to matched general-purpose AI responses on the same questions, and citation completeness under independent audit. Once that capability has been demonstrated, subsequent evaluation can examine whether authorization-first responses influence how patients seek and use institutional information.

That second stage concerns behaviour and user experience. Behavioural outcomes include the proportion of patient-facing queries resolved through the institutional system rather than redirected to general-purpose AI, and changes in website engagement and task-completion rates following deployment. This

stage should also incorporate direct measures of user experience, which are frequently the primary endpoints in digital health evaluations: patient satisfaction with the information received, perceived ease of finding an answer, and confidence that the answer originated from the institution rather than an unaffiliated source. These measures could be obtained directly, for example through a point-of-care survey asking patients arriving for treatment what sources they consulted beforehand and how much they trusted each.

Where patient behaviour changes, it becomes possible to ask a third question: does institutional workload change as a result? Operational outcomes here include call-centre volume and average handling time for governable query types such as appointments, preparation instructions, and visitor policy, and search success or FAQ abandonment rates before and after deployment.

The final and most demanding stage concerns clinical outcomes, where the strongest methodology is required. Relevant measures include patient preparedness for procedures, assessed against a governed-information cohort compared with standard care; rates of missed appointments, cancelled procedures, or postoperative complications plausibly linked to inadequate or inaccurate preparatory information; and any association between information source and adverse events, interpreted cautiously given the many confounding factors present in clinical outcomes research. A controlled evaluation, comparing outcomes for patients who received authorization-first responses against a standard-care control group, would be the most rigorous design for testing this final stage and is the design this paper recommends as a starting point for institutional research partners.

The purpose of this paper is therefore not to demonstrate clinical effectiveness or advocate immediate deployment, but to argue that authorization-first architecture represents a sufficiently distinct approach to AI governance that it warrants formal evaluation using established clinical and health-services research methodologies.

10. MUHC Operational Scale: Context for the ROI Case

The architectural argument above establishes why authorization-first AI is the correct governance model. This section grounds that argument in MUHC's own publicly documented operational scale, and is deliberately explicit about the line between measured fact and planning assumption, a distinction the ROI figures elsewhere in this deployment package depend on.

10.1 What is publicly documented

MUHC's centralized contact centre is publicly reported to handle approximately 75,000 calls per month, operating 24 hours a day, seven days a week, with a staff of roughly 30. Extrapolated, that is approximately 900,000 calls annually, or roughly 2,465 calls per day. This centralized function sits within a much larger patient-communication footprint: MUHC reports more than 500,000 outpatient visits annually, hundreds of thousands of emergency department visits, eight major hospital sites, and approximately 14,500 employees.

MUHC has also already invested in centralized appointment services, digital referral systems, SMS appointment reminders, patient portals, and an expanded centralized contact centre specifically designed to resolve inquiries without transferring callers to individual clinics. The publicly documented scope of that centralized service includes appointment booking, rescheduling, cancellations, clinic locations, room confirmation, and referral-portal support, a scope that is predominantly administrative rather than clinical.

What is not publicly available is the operational detail that would be required to model a precise ROI case: average handle time, average speed of answer, abandonment rate, first-call resolution, wait times,

the administrative-versus-clinical split of call volume, or cost per call. Those figures would need to come from MUHC directly if a pilot proceeds, and are not asserted here. Although a precise, MUHC-specific ROI case is not yet possible without that data, general ROI estimates based on comparable healthcare deployments are available and can inform planning in the interim.

Separately, for sizing and pricing discussion rather than as an observed statistic, a broader planning assumption is worth stating explicitly and labelling as such: when the full communication ecosystem is considered, telephone, referrals, appointment workflows, and related digital-channel interactions, not the contact centre alone, total annual contacts are more plausibly in the range of 1 to 3 million. This figure should not be presented alongside the measured 900,000-call contact-centre figure as though the two carry the same evidentiary weight. One is documented; the other is a planning estimate.

10.2 Published healthcare benchmarks used as planning context

In the absence of MUHC-specific handle-time data, this case study applies published, cross-referenced healthcare contact-centre benchmarks as illustrative planning context, not as MUHC-specific measurements. Typical healthcare Average Handle Time converges around 6.6 minutes across independent benchmark sources; complex scheduling, referral, or insurance calls often run 8-10 minutes or longer; and general contact-centre benchmarks across all industries place most calls in the 4-10 minute range, with healthcare running structurally above the all-industry average due to verification, documentation, and coordination requirements.

Applied illustratively to MUHC's documented call volume, approximately 900,000 calls per year at 6.6 minutes per call implies roughly 5.94 million minutes, or approximately 99,000 hours, of annual agent conversation time, roughly 48 full-time-equivalent years of agent time devoted to conversation alone. This is explicitly not a staffing estimate: actual staffing also depends on breaks, occupancy, after-call work, scheduling, and 24/7 coverage requirements that this figure does not capture. It illustrates scale, not a claimed savings figure.

10.3 Why handle time is not the right primary metric

The more useful conclusion from this analysis is that average handle time should not be treated as the primary value proposition for authorization-first AI in a hospital contact centre. The relevant question is not whether every call becomes shorter, but which calls consume that time. Parking, directions, appointment location, visitor policies, institutionally authorized clinic preparation instructions, medical records process, and partially automatable appointment changes are better suited to governed AI assistance. Complex scheduling conflicts, emotional or distressed callers, clinical triage, care coordination, and escalations still require human judgement.

That reframing points to a broader observation worth stating plainly: the contact-centre literature overwhelmingly measures speed (handle time, answer time, abandonment) but rarely measures whether the caller received the institutionally correct answer. Operational efficiency metrics and governance metrics are complementary, not interchangeable, and an authorization-first architecture is the piece of that picture current benchmarking practice does not capture.

10.4 The strategic framing this supports

The strongest operational conclusion from this analysis is not a number. MUHC has already centralized and modernized substantial parts of its patient-communication infrastructure: centralized scheduling, digital referrals, SMS reminders, and an expanded contact centre. Within that context, authorization-first conversational AI is not an additional chatbot layered onto existing infrastructure. It is a governed

conversational interface to infrastructure MUHC has already built, extending an investment already underway rather than replacing or duplicating it.

11. COMPAiSS Fills the Gap: Between FAQ Bots and General AI

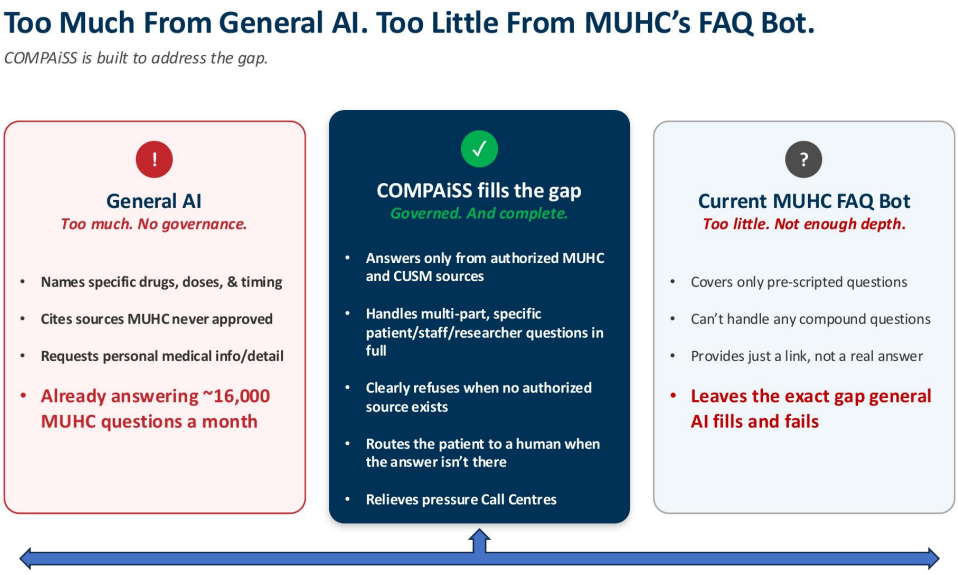

Figure 20. Too much from general AI. Too little from MUHC's FAQ bot. COMPAiSS positioned as the governed answer between the two, resolving the dilemma introduced in Section 6.4.

The "too much / too little" framing is one of the clearest ways to explain the market gap, and it makes a directional claim, not just a feature comparison: general AI is too permissive, answering with details it should not be giving, while the current scripted FAQ bot is too narrow, covering only pre-written questions and failing on any real, multi-part query.

That direction of movement is not asserted in isolation. It is the same pattern already established in Section 3's traffic evidence and Figure 3's independent research (68% of Google searches now ending without a click; 32% of adults already turning to AI chatbots for health information), and now further corroborated with live deployment evidence in Section 6: an estimated 16,000 fewer monthly visitors to muhc.ca in the same window general AI usage for institutional questions is rising elsewhere. A FAQ bot that cannot handle a real question does not suppress demand for AI-mediated answers. It routes that demand to a system the institution never authorized.

Authorization-first architecture is positioned here as the governance response to that specific movement, not merely a better version of either tool. Users want more than links, but the institution cannot allow open-ended generation that drifts beyond approved sources. The case study therefore shows that the problem is not "AI versus no AI." It is "governed AI versus ungoverned AI," and the FAQ bot's limitations are what make that choice active rather than hypothetical.

12. Discussion

The MUHC case offers a broader lesson for healthcare institutions. The right question is not whether AI can answer questions quickly. The right question is whether the institution can preserve authority over the answer while still delivering a useful experience.

COMPAiSS represents one implementation of an architectural response to that challenge. By placing authorization before inference, it converts institutional policy from a post-hoc filter into a structural requirement. That is especially relevant for hospitals, where errors are not merely inconvenient. They can influence clinical preparation, patient behavior, and institutional trust.

The live deployment evidence in Section 6 also illustrates a second, related benefit that is easy to overstate but worth naming carefully: an authorization-first system does not just avoid unauthorized answers, it makes its own limits visible and traceable. Where the system's answer was thinner than a general AI's would have been, that thinness was explainable by reference to a specific, named source gap, not hidden behind a fluent-sounding non-answer. That auditability is itself a governance property, distinct from and additional to the accuracy of any single answer.

13. Conclusion

This pattern is not unique to MUHC. Publicly available search-traffic data for IWK Health Centre in Halifax, Nova Scotia, an unrelated academic health centre with no connection to this analysis, shows an even steeper decline over a comparable period: organic search visits to iwkhealth.ca fell from approximately 7,400 in February 2026 to 4,800 in July 2026, a decline of roughly 35 percent across six consecutive months, with every month lower than the one before it. Like MUHC, IWK's public-facing information relies on a keyword-based search interface rather than a conversational one, and its visible search traffic appears to be shrinking during the same period that general-purpose AI adoption for health information has grown nationally. Two independent institutions, in two different provinces, showing the same directional pattern over the same window, is better evidence of a structural trend than either figure would be alone.

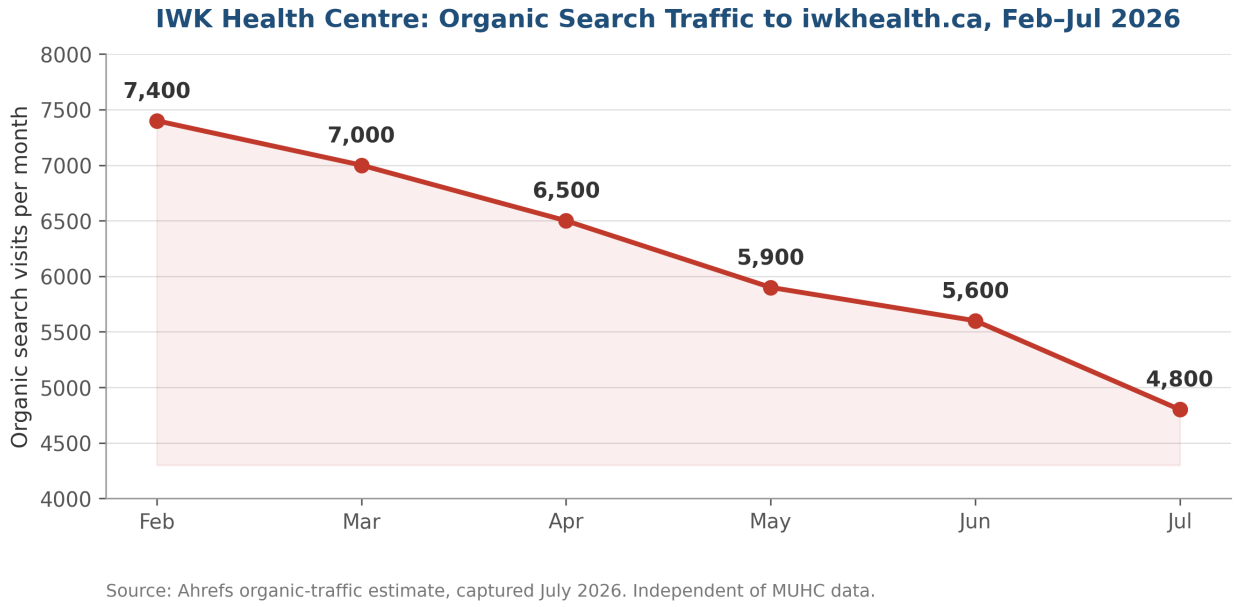

Figure 21. Organic search traffic to iwkhealth.ca, February to July 2026, an independent, third-party (Ahrefs) estimate. Every month is lower than the one before it: 7,400 → 7,000 → 6,500 → 5,900 → 5,600 → 4,800, a

decline of roughly 35 percent over six consecutive months. IWK Health Centre has no connection to MUHC or to this analysis; the pattern is presented as independent corroboration, not as evidence about MUHC itself.

The MUHC deployment case shows why authorization-first AI is well suited to regulated healthcare settings. MUHC has a complex digital footprint, broad user demand, and a growing challenge from general-purpose AI tools that answer outside institutional control. It has also already invested substantially in centralizing and modernizing its own patient-communication infrastructure. In that environment, a system that checks authorization before generation, and that extends existing infrastructure rather than replacing it, is a meaningful governance improvement.

COMPAiSS is presented as a different institutional architecture. That difference matters because it allows hospitals to give users useful answers while maintaining control over source authority, privacy, and accountability.

Core architectural distinction

Authorization-first AI checks whether a query can be answered from approved institutional sources before generation is permitted to run. Generation-first AI generates first and attempts to constrain afterward. For regulated healthcare information, that sequence is the difference between a defensible answer and an institutional liability.

References

1. MUHC COMPAiSS institutional documentation, July 2026. Source material for this case study.

2. Live website and deployment output, muhc.ca and compaiss.ca/muhc, captured July 2026. Source material for Section 6 (Figures 7, 8, 10, 11, 12, 15) and Section 8 (Figures 17, 18).

3. COMPAiSS institutional materials prepared for MUHC leadership, July 2026. Source material for Figures 9, 13, 14.

4. Deployment, Licensing, and Institutional Fit - Greenlist Control and Source Auditing. compaiss.ca/deployment-and-institutional-fit.html#greenlist-control. Source material for Section 8.

5. SparkToro / Similarweb (2026). Zero-click search: 68% of Google searches end without a click. https://sparktoro.com/blog/in-2026-less-than-one-third-of-google-searches-still-send-a-click/

6. SparkToro (2025). Zero-click search: what still works. https://sparktoro.com/blog/zero-click-search-what-still-works/

7. SpinSci. "Average Handle Time in Healthcare: Benchmarks for 2026." SpinSci Resources.

8. Dialog Health. "Latest Healthcare Call Center Statistics: Must-Know for 2025."

9. Kayako. "Average Handle Time (AHT): Formula, Benchmarks & How to Improve." 2026.