AI answers on a public real-estate website: reviewer pack

Prepared by Daniel Zivkovic, Magma Inc., 2026-11-09, from 26 source documents. Visibility: public.

> Proposed verdict: **CAVEAT**, a conditional yes: three items must close first. This page, prepared before the security, architecture and privacy reviews of one AI use case, summarizes the evidence and open questions for reviewers.

Before you read: a note from the author

This page reviews the AI answer box on jasminahomes.ca/search-ai, a real-estate site I built and run, where visitors ask about buying, selling and owning a home in the Greater Toronto Area.

I wrote this review method for other people's AI systems. Asking anyone else to sit in that chair before my own system had been through it would be backwards, so my own system is the first real one I put through it. It has not passed yet: three conditions are still open. I publish it as it stands, because a real review with its gaps on view teaches more than a sample that only shows a clean pass.

How it works: a visitor types a question in the Ask tab. An agent hosted by Algolia searches the guides already published on the site, and an OpenAI model writes a short answer that cites them. It takes no actions and never asks for a name, phone number or email. If it fails, the page falls back to plain keyword search. The contact form is a separate path and never reaches the AI.

The verdict, CAVEAT, is a conditional yes. The conditions are what must close first, and each names who closes it. The review also gives a score: the share of recommended safeguards in place, where a full score means all of them. This design has a little more than half. The score says nothing about how good the answers are.

Since the first run in October, three conditions have closed, each on my written confirmation as the owner. I named myself the risk owner and stated that Jasmina Homes is the data controller. I recorded a dated decision to leave Algolia's Model Context Protocol (MCP) setting on, because it only changes the shape of the agent's internal search tool; the configuration snapshot matches. I confirmed that the AI credit is a hard stop, and that an automated check (a canary) asks the agent a real question three times a day and emails me when it fails. That alert comes after the fact, and the hard stop is still to be checked on Algolia's plan page at the next review. I also confirmed that the agent runs on one named model snapshot rather than a moving alias, which raised the score, and I wrote the kill-switch runbook; its condition closes once I have run it and recorded the date.

If you run an AI system and want the same review for it, write to me at daniel@magmainc.ca.

Setting the stage

- **What was reviewed.** An AI answer box on jasminahomes.ca, a public real-estate website, that answers visitors' questions about buying, selling and owning a home in the Greater Toronto Area, using only guides already published on the site and citing them. It runs as an agent hosted by Algolia (Agent Studio) with an OpenAI model behind it, takes no actions, and never asks for a name, phone number or email.
- **How.** A structured intake of 26 questions, answered from 26 of the project's own documents. Each answer is marked as stated in a document, proposed (inferred; a named person must confirm it), or missing (an owner is named to supply it).
- **The risk catalogue.** FINOS (the Fintech Open Source Foundation) is an umbrella organization under the Linux Foundation. Its AI Governance Framework, developed by FINOS members, is an open catalogue of AI risks and the mitigations that reduce them. Each risk below has a plain name; its catalogue ID and link are in the appendix.
- **The verdict words.** PROCEED means yes, with no open conditions. CAVEAT means a conditional yes: the named items must close first. REFUSE means no, with the reason stated.

The verdict

CAVEAT: a conditional yes. The assistant only reads public pages and takes no actions, but three conditions on privacy, vendor terms and shutdown must be met first, and two of them could still change the verdict.

The proposed hosting category is AI built into a vendor product that is already approved, where the vendor's model reads data the company already gave it, inside the vendor's own tenant. This category and its review route are the intake method's example rules, not an authorization; an institution applies its own routing policy. It would be configured as a vendor-hosted agent. What could change it: the personal data question decides the route. The intake method's example rules do not permit any personal information to go out to a model in a vendor's tenant. If privacy rules that the personal details visitors type count as personal information under that rule, the verdict changes.

Approval conditions and what we ask of you

Approval is proposed on these three conditions. For each one, please reply with one of three answers: **accept** (the condition is enough for you), **evidence** (name the evidence you need to see first), or **blocker** (it stops approval for you; say why).

1. A formal privacy assessment rules on personal information that visitors volunteer, and the data-flow contract is signed off or a filter is added in front of the model. Owner: privacy.
2. Vendor terms on file for this route: Algolia's data region, OpenAI's retention and no-training terms, and any data processing agreement (DPA). Owner: third-party risk.

3. The written kill-switch runbook exercised once, with the date recorded. Owner: platform owner.

What blocks a yes

Each item matches the approval condition with the same number.

1 **Personal information that visitors volunteer: no formal privacy assessment yet.**

Why it matters: Visitors can type personal details. These details reach OpenAI in the United States (US), Algolia's 30-day conversation log and the owner's analytics copy. The privacy notice has disclosed this since 2026-10-15, and the data-flow contract is a draft. If privacy rules that this counts as handling personal information, the intake method's example rules do not permit it in a vendor's tenant, and the verdict changes.

Who acts: privacy closes it; the business owner and risk owner (Daniel Zivkovic) signs off that it is closed.

2 **Vendor terms not on file: Algolia data region, OpenAI retention and no-training terms, any DPA.**

Why it matters: If the vendor terms allow training on visitor text or long retention, the vendor-embedded route fails and the verdict changes.

Who acts: third-party risk closes it; privacy and security sign off that it is closed.

3 **Kill-switch runbook written, not yet exercised.**

Why it matters: Reviewers need proof that the assistant can be switched off, and the vendor left, quickly during an incident. The runbook was written on 2026-11-09 (stop the agent, remove the Ask tab, roll back the prompt, leave Algolia); nobody has run it yet.

Who acts: the platform owner closes it; security signs off that it is closed.

The facts the verdict rests on

Fact	Question	Answer	Status
Data sensitivity	What is the most sensitive class of data it touches?	Public	**Proposed: data steward**

Fact	Question	Answer	Status
Personal data	Does it touch personal, payment card or health information?	Personal information (PI): none indexed or asked for, but a visitor can volunteer PI and it is kept with the conversation record; the contact form never reaches Algolia or OpenAI; user memory is disabled and must stay disabled. Payment card information (PCI): no. Personal health information (PHI): no (proposed)	**Proposed: privacy**
Direction of data	Does data go out to a model, or output into a system?	Data goes out to the model (visitor text plus retrieved public passages go to OpenAI through Algolia); the output is display text with links; no output into a system of record; no outbound calls beyond the search tool	**Proposed: third-party risk**
Hosting and retention: Algolia, agent host	Where is data processed, in which country, and kept how long?	Software-as-a-service tenant. Conversations (metadata, messages and tool calls) kept 30 days; answer cache lifetime undocumented; data region unknown.	**Missing: third-party risk**
Hosting and retention: OpenAI, model provider		Processing in the US. Retention and no-training terms for this route unknown.	**Missing: third-party risk**
Hosting and retention: BigQuery, the owner's analytics copy		A copy of each conversation in the owner's own Google Cloud project (US multi-region), to improve the assistant and the guides; deleted on request; disclosed in the privacy policy (2026-10-15); the owner accepted the risk (2026-10-16).	Stated in a document
Autonomy	How much can it do without a human approving?	Read-only: one answer per visitor message, up to 10 steps, no actions; runs only on a click of the Ask tab (architecture decision record (ADR) ADR-0007, live 2026-09-15)	Stated in a document

Risks from the catalogue

Only the risks rated APPLIES (present in this design) or LIKELY (probably present; the sources do not settle it) are listed, each with its main gap. The safeguards in place, the full mapping to catalogue

mitigations and the risks rated not applicable are in the full record. The rating says only whether a risk exists in this design; the Main gap column says what is still missing in how it is handled.

Risk	Rating	Main gap
Data sent to the hosted model leaks (regulatory)	APPLIES	No filter on chat text, and OpenAI's retention and no-training terms are not on file
Copies in the search index, logs or caches leak (security)	LIKELY	Vendor encryption at rest is not on file for the conversation log and answer cache
Wrong or made-up answers (operational)	APPLIES	No automated answer grading between owner reviews
The provider changes the model underneath (operational)	APPLIES	OpenAI does not state whether the model behind the snapshot name can still change
The same question gets different answers (operational)	APPLIES	No automated answer grading
The model is down, throttled or runs up the bill (operational)	APPLIES	No spend-threshold alert: an exhausted credit shows only as a failed canary, after the fact. The hard stop rests on the owner's statement, to be verified on the plan page at the next review
Hidden instructions take over the model (prompt injection, security)	APPLIES	The vendor content-safety filter is off, and search-tool input checks are not described
Answers drift from the business purpose or tone (operational)	APPLIES	No automated answer grading between owner reviews
Use grows beyond what was approved (operational)	LIKELY	The vendor content-safety filter is off
Stale or poor content degrades the answers (operational)	LIKELY	No freshness or drift check on the guides
A public failure damages the brand (operational)	APPLIES	The vendor content-safety filter is off, and vendor terms are not on file
It falls short of the rules that apply when a person does the same work (regulatory)	APPLIES	No formal privacy assessment, and vendor terms and encryption are not on file
Inputs steer which tools the agent calls (security)	LIKELY	Search-tool input checks are not described

Risk	Rating	Main gap
A compromised Model Context Protocol (MCP) server feeds in tainted data (security)	LIKELY	The MCP setting is on by the owner's decision, and how Algolia governs its MCP server is unknown

Proposed reviewers

- **Privacy** (proposed: full review): PI that visitors volunteer goes to OpenAI in the US, stays 30 days in Algolia's conversation log and is copied to the owner's Google Cloud analytics storage; the privacy notice discloses this; the data-flow contract is a draft; no formal privacy impact assessment (PIA).
- **Security** (proposed: full review): a public search key behind a domain allow-list that a forged Origin header gets past; the vendor's content-safety filter is off; the MCP setting is on, and the owner decided on 2026-11-09 to leave it on (it changes only the agent's internal search tool; the snapshot matches); no independent security testing.
- **Platform or cloud owner** (proposed: full review): the platform owner (Algolia app owner): the AI credit is a hard stop by the owner's statement, to be verified on the plan page at the next review; the alert is the canary's failure email, after the fact, not a spend threshold; the kill-switch runbook is written but not exercised. Cloud owner (proposed: light touch): the owner's own project holds the BigQuery analytics copy.
- **Data steward** (proposed: light touch): confirm that only published public content enters the search index.
- **Model risk** (proposed: light touch): a large language model (LLM) agent pinned to the snapshot gpt-5.6-luna; it makes no decision about a person.
- **Third-party risk** (proposed: full review): an Algolia to OpenAI chain inside a vendor product; the Algolia data region and OpenAI terms are unknown.
- **Architecture board** (proposed: light touch): check whether ADR-0007 (the paid AI runs only when a visitor clicks Ask) covers every application that calls the agent.

Legend

Abbreviations used above and not spelled out where they first appear.

- **ADR:** architecture decision record: a short written note of a design decision.
- **BigQuery:** Google's cloud data warehouse.

Appendix: catalogue references

The risks above, in the same order, with their FINOS AI Governance Framework risk IDs (ri-N) and

titles.

Risk on this page	Catalogue ID and title
Data sent to the hosted model leaks	ri-1 Information Leaked To Hosted Model
Copies in the search index, logs or caches leak	ri-2 Information Leaked to Vector Store
Wrong or made-up answers	ri-4 Hallucination and Inaccurate Outputs
The provider changes the model underneath	ri-5 Foundation Model Versioning
The same question gets different answers	ri-6 Non-Deterministic Behaviour
The model is down, throttled or runs up the bill	ri-7 Availability of Foundational Model
Hidden instructions take over the model (prompt injection)	ri-10 Prompt Injection
Answers drift from the business purpose or tone	ri-14 Inadequate System Alignment
Use grows beyond what was approved	ri-18 Model Overreach / Expanded Use
Stale or poor content degrades the answers	ri-19 Data Quality and Drift
A public failure damages the brand	ri-20 Reputational Risk
It falls short of the rules that apply when a person does the same work	ri-22 Regulatory Compliance and Oversight
Inputs steer which tools the agent calls	ri-25 Tool Chain Manipulation and Injection
A compromised Model Context Protocol (MCP) server feeds in tainted data	ri-26 MCP Server Supply Chain Compromise

Prepared by Daniel Zivkovic, Magma Inc. with the AI use-case intake method; approval remains with the institution's own functions. Risk IDs from the FINOS AI Governance Framework (CC BY 4.0). Related guidance: the NIST AI Risk Management Framework and OSFI's July 2026 bulletin on generative and agentic AI, which sets out sound practices that complement existing OSFI guidelines.