

Fictional demonstration

Researching an applied AI engineer for a support product

I'd look for an engineer who can evaluate a retrieval feature, investigate its failures and operate it in a product. A working demonstration is useful evidence, but this role also needs someone who can explain what happened after release. The research brief keeps those requirements separate.

The assumed role

This fictional B2B support product is preparing an assistant that answers questions from customer documentation. The engineer would own retrieval experiments, an evaluation set, failure analysis and the application integration. A platform engineer owns shared infrastructure. Every source excerpt and assessment below is invented; no real person, employer or candidate result is represented.

Work the role must own	Evidence to seek	What is insufficient on its own
Evaluate answer quality	A defined task set, failure categories and a reason for changing the system.	A model name or a broad claim that accuracy improved.
Build retrieval into the application	An explanation of indexing, permissions and behavior when evidence is missing.	A notebook that does not address product access rules.
Operate the released feature	A latency or cost tradeoff, monitoring and a response to failure.	A polished demo without operating evidence.
Work with product and support	A concrete decision about when the assistant should defer or ask for help.	General enthusiasm for AI or fluent terminology.

Useful adjacent backgrounds

Search applied ML product teams, search and relevance teams, and backend teams that have shipped retrieval features. I would consider a backend engineer with strong evaluation evidence alongside an applied ML engineer with product ownership. Foundation model training is outside this assumed role; requiring it would change the brief.

Completed fictional evidence packet

Record A01 combines three invented source excerpts. In a live map, each entry would include its accessible URL, publication date, access date and attribution. The relative timing here describes the fictional scenario only.

Source entry	What the excerpt actually says	Evidence limit
A01a / engineering article	The author built a 40-question task set. A retrieval revision changed answers judged supported from 24 to 31.	Seven more supported answers on the same small set. Task selection and independent review are unknown.
A01b / public project notes	The contributor added a document permission check before retrieved passages were returned.	Contribution is relevant. Production deployment and complete access coverage are not established.
A01c / product demonstration	The presenter shows an answer with citations and a fallback when no source is found.	The demo shows behavior for the examples presented. It does not prove reliability across other questions.

Interpret the result without enlarging the claim

In this fictional task set, 24 of 40 is 60% and 31 of 40 is 77.5%. The difference is 17.5 percentage points. Reusing the same questions may favor the revision, and the record does not define how a supported answer was judged. I would not describe this as a production accuracy result.

Research recommendation

Keep A01 as a lead for a focused evidence discussion. The record supports retrieval evaluation and a permission-related contribution. It leaves released-feature ownership, latency, cost and cross-functional decisions unknown. There is no basis here to call the person available, interested or ready for an offer.

A useful follow up

Ask how the task set was chosen, who reviewed the answers and what happened on questions excluded from it. Then ask for one release decision involving access, latency or cost and clarify the person's own responsibility. The hiring lead owns any progression decision.

Compare the next research lead

Evidence area	A01 / retrieval product work	A02 / model training work
Evaluation	Small task set and answer support counts described.	Training benchmark described; product task evaluation absent.
Product integration	Permission check contribution described.	Application integration not established.
Operation	Released ownership still unknown.	Training job operation described; customer-facing operation unknown.
Research action	Clarify the missing operating and product decisions.	Investigate product evidence before advancing the research priority.

A02 is another invented record, not a weaker engineer. Its evidence is farther from this particular brief. If the actual role needs model training, the search order and criteria would change.

Collection and uncertainty rules

Use public work available without a personal account, including engineering articles, project documentation and conference material. Attribute team claims carefully. Record contradictory dates or unclear contribution rather than selecting the most favorable version. Do not collect private contact details or infer personal characteristics from a profile.

What a buyer receives

A scoped research brief would include the agreed role criteria, employer groups to investigate, source-backed leads, evidence notes, uncertainty and recommended follow-up. A fit note explains the connection to the work; it does not establish motivation or availability. This sample contains no live shortlist or outreach.

Next step

Agree on the product task, the division of responsibility with platform engineering and the evidence needed for released-feature ownership. Those choices let the research focus on work the hire will actually do.