CHECKLIST

Build or Buy an LLM?

To increase research efficiency, break down workflow silos, and solve knowledge sharing problems, companies are implementing generative AI for its search and summarization capabilities. To do so, organizations are trying to understand the investments required to either build an LLM themselves or buy a system that can be customized to their needs.

Defining the Problem an LLM Could Solve

Whether to build or buy an LLM (or something in between) relies largely on what you need it for. Some of the problems that basic knowledge management and market intelligence systems don't solve, for which you would need a more sophisticated genAI system, include:

- Employees rely on CTRL+F to search for info within a document

- Data is siloed between different teams, leading to duplicative research efforts and conflicting market perspectives

- Data is unorganized in disparate locations, excluding broader organizational access.

- You're paying for different content storage platforms, yet employees are still missing information or waiting days on a data request

- Onboarding is slow and expensive

- Internal knowledge is lost when employees leave your organization

- Credibility is lost when you can't easily find the source of data

Once your problem is defined, you can look at the pros and cons of either building or buying and customizing a large language model that boosts productivity at scale, and at a reasonable cost.

> There are some problems that basic knowledge management and market intelligence systems don't solve, for which you would need a more sophisticated genAI system.

AlphaSense
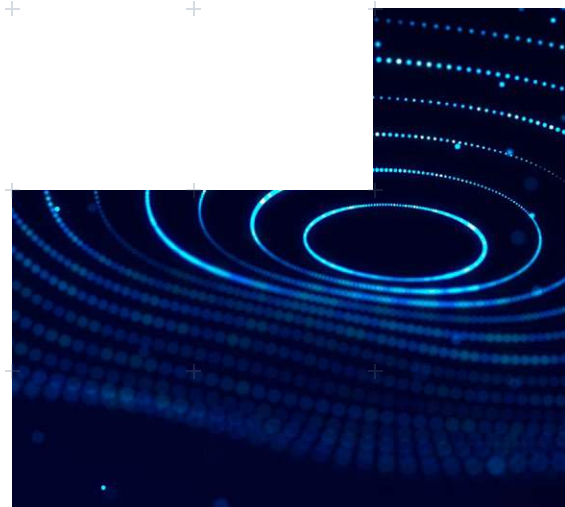
Build Your LLM

Pros

- ✔ Popular open-source genAI models are widely available, built for general use, and can be deployed quickly

- ✔ For lower complexity work, a general-purpose model can easily be layered onto existing content management systems

- ✔ You can choose the content that is used for model training purposes

- ✔ All proprietary content is kept on your servers, so IP is secure and not shared externally by default

- ✔ You can build just the integrations you need depending on where your organization's content lives

Cons

- ✖ Consumer-grade or open-source LLMs could be trained on inaccurate information that is based on the entire internet or non-authoritative sources

- ✖ Building LLMs takes significant time, resources, and capital in order to effectively acquire, clean, and curate large data sets

- ✖ For strategic work with high complexity, open-source models without use-case specificity introduce risk to expensive decisions

- ✖ Sourcing enough content to train the model can be difficult

- ✖ The price tag for training your LLM can grow quickly due to computational costs required, along with personnel and maintenance costs

- ✖ Baseline LLMs don't include broader semantic search capabilities, relevancy algorithms, sentiment analysis or topic identification, which require domain expertise

- ✖ Front-end interfaces need significant resources for design and user adoption / change management

AlphaSense

Buy Your LLM

Pros

✔ Has domain expertise, thus delivering higher precision and recall, improved information retrieval, less bias, computational efficiencies, and targeted knowledge extraction

✔ GenAI platforms with domain expertise can offer broader semantic search capabilities, relevancy algorithms, and theme extraction based on your use cases

✔ Have built-in scalability features including user management, broad system connectors, infrastructure compatibility, and collaboration or workflow tools

✔ Pre-built front-end interfaces are already tested and proven in market, requiring fewer resources for design and user education

✔ Multiple pre-built integrations allow flexibility to use different third-party connectors as internal content management systems evolve

Cons

✘ A general-purpose genAI tool could be trained on misleading information pulled from the entire internet or non-authoritative sources

✘ Non-vetted or untrustworthy tools have a higher likelihood or potential to generate fake facts (i.e hallucinate) while sounding realistic or trustworthy

✘ Risks around data privacy, as some models may use client data as materials to train

✘ Some LLMs or Vector Databases still require you to build and customize many aspects, such as procuring external content, creating training data, and then training the LLM, as well as developing UI and feature sets

✘ Companies with strict data privacy regulations need to fully trust the LLM provider so no IP is vulnerable to loss or theft

Partnering with an external solution provider marries the best qualities of an in-house build with the appeal of a pre-built genAI offering.

With AlphaSense's Enterprise Intelligence offering, purpose-built AI search and summarization capabilities surface insights across internal and external research, so you can centralize siloed research workflows and make investment decisions with less reputational risk.

AlphaSense's genAI model was developed by 500-plus developers in our product organization, leveraging 10+ years of AI and search expertise and investing millions of dollars into deciphering a solution to this challenge. We never leverage proprietary client information to train our specialized LLM, and we offer enterprise-grade security measures to ensure your IP is private and secure.

With Enterprise Intelligence, you can get quick context using genAI summaries with in-line citations—verifiable with one click. You can also interrogate complex documents, including PDFs, Excel models and CIMs, with natural-language chat that goes straight to the source.

AlphaSense

Learn more about
Enterprise Intelligence

Lorem CTA

FIRM'S PROPRIETARY INTERNAL DATA   AI   300M+ PREMIUM EXTERNAL DOCUMENTS

SEARCH   CHAT   SUMMARIES   ALERTS   COLLABORATION