

# Helping developers manage high-stakes, high-security codebases

Llama-powered assistant generates code, explains legacy software



# At a glance

Leidos uses Llama to power Sourcegraph's code assistant, creating a code management and generation assistant capable of working with sensitive codebases on air-gapped infrastructure. Developers can ask Sourcegraph to explain code, help detect and fix bugs, generate new code and perform extensive, repetitive tasks with AI agents. In addition to boosting efficiency, Llama and Sourcegraph helped Leidos position itself as an innovator in its sector.



## Use case

Streamlining legacy codebase management and modernization



## Goal

Assist developers with massive codebases and automate mundane coding tasks



## Llama version

Llama 3.1 70B Instruct



## Deployment

Sourcegraph's code intelligence platform, AWS LISA

## Results\*

### Codebase QA

Developers explore, debug and document with natural language

**+20%**

increase in developer efficiency across teams

### Task offloads

Context-aware agents generate code for mundane, repetitive tasks

**+20%**

improvement in architecture and application development velocity

### Competitive edge

AI code generation positions Leidos as an innovator with government and commercial clients

**-92%**

reduction in typical user time lost to context switching and learning

**+33%**

boost in flexibility for faster engineering and feature implementation

\*All results are self-reported and not identifiably repeatable. Generally, expected individual results will differ.

# The challenge

## Wrangling massive, legacy codebases is an endless challenge

Leidos manages, modernizes and authors software for institutional clients in the defense, national security and healthcare sectors, which means they grapple with decades of legacy code in extremely secure, air-gapped environments all while meeting strict regulations. The size, complexity and requirements of these codebases make them ideal candidates for modernization, but the risk to the critical services they provide means the codebases and proposed changes need to be thoroughly documented, analyzed and understood.

## Automating code discovery, management and generation was key

The Leidos team wanted to use generative AI to understand their clients' sprawling codebases, troubleshoot issues and generate low-level code in bulk. Their ultimate goal was an agentic toolset that developers could use to ask questions about the codebase with natural language, get help with debugging and automate the bulk of the busy work.

“Lots of times developers are challenged with tedious tasks that they either don't have time to implement or struggle to complete. We wanted to offload those tasks and free up our developers so they could focus on harder, more complex problems.”

Jay Davis  
Chief National Security Software Center Architect,  
Leidos



Leidos is an industry and technology leader serving government and commercial customers with smarter, more efficient digital and mission innovations. Headquartered in Reston, Virginia, Leidos helps solve the toughest challenges in healthcare, national security, transportation and defense.

- **Industry:**  
Software
- **Company size:**  
48,000 global employees,  
\$15.4 billion in revenue  
in 2023



# The solution

## Llama powers comprehensive codebase insight and code generation

Leidos uses Sourcegraph — the code intelligence platform helping enterprises industrialize software development with AI — to manage and modernize legacy codebases.

The Sourcegraph platform turns codebases into knowledge bases, which the platform's AI uses to help developers generate accurate code.

Sourcegraph works with most large language models (LLMs). Leidos chose Llama 3.1 70B Instruct, creating a powerful solution that helps developers generate unit tests, create code skeletons and document code. Since Sourcegraph gives Llama access to entire codebases, the solution can answer complex questions about specific code details conversationally — a huge benefit for developers ramping up on unfamiliar projects.

“Llama is a great general-purpose foundation model that works well in coding situations and has the size and knowledge to help developers with day-to-day tasks. Llama goes above and beyond other models that have smaller context windows and outdated training data.”

Jay Davis  
Chief National Security Software Architect, Leidos

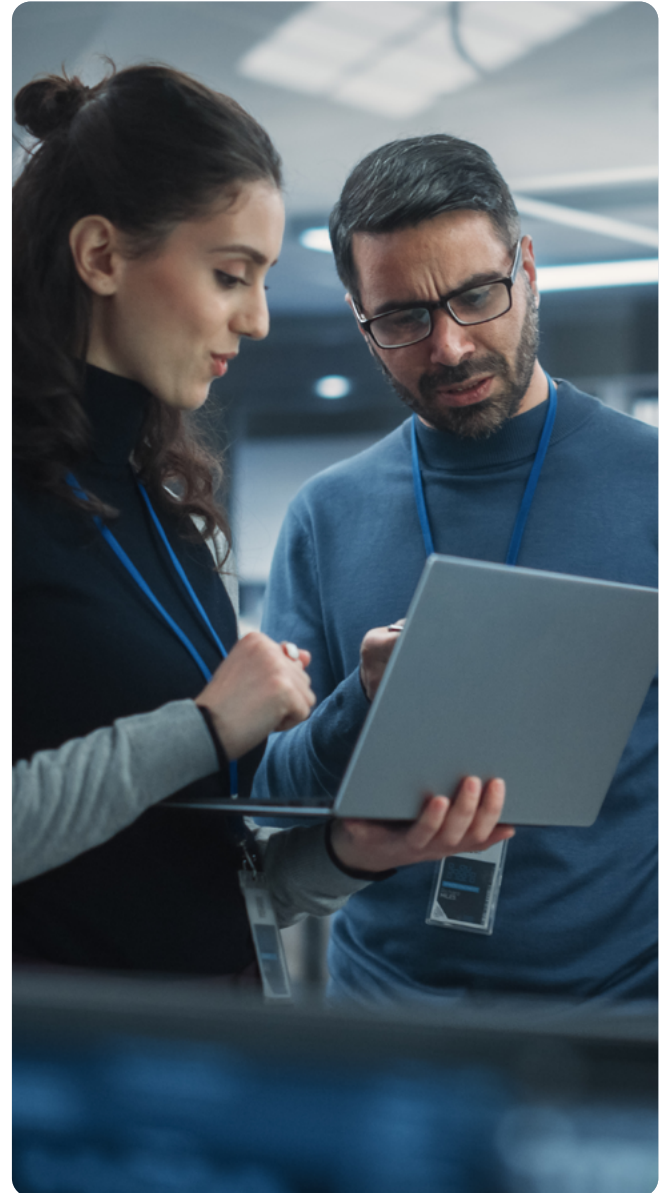


## Security and cost constraints required an open-source LLM, only Llama proved fully capable

Leidos developers work in highly-secure, air-gapped environments, which ruled out API calls to hosted models. Using local, proprietary models was a no-go due to per-token charges and data security concerns. Portable, low-cost, open-source LLMs seemed to be an ideal solution, but tests demonstrated two major flaws: high latency and out-of-date training data.

“With other open-source models, we were experiencing significant lag. Our Sourcegraph clients would timeout,” says Jay Davis, Chief National Security Software Architect at Leidos. “Once we implemented Llama, those performance issues were resolved and timeouts eliminated.”

Stale training data proved to be an even bigger problem. Other models’ training data could be two or more years out of date. A slow training and updating cadence may not matter for general-purpose chat services, but aging training data makes models worthless for code generation.



“One competing open-source model had no knowledge of Java versions beyond v15, which was released in September 2020. Our team was working with v21. Llama knows modern, up-to-date code libraries and versions, so it was able to give more accurate answers.”

Jay Davis  
Chief National Security Software Architect, Leidos

## Deploying on AWS LISA ensured airtight security and high-performance

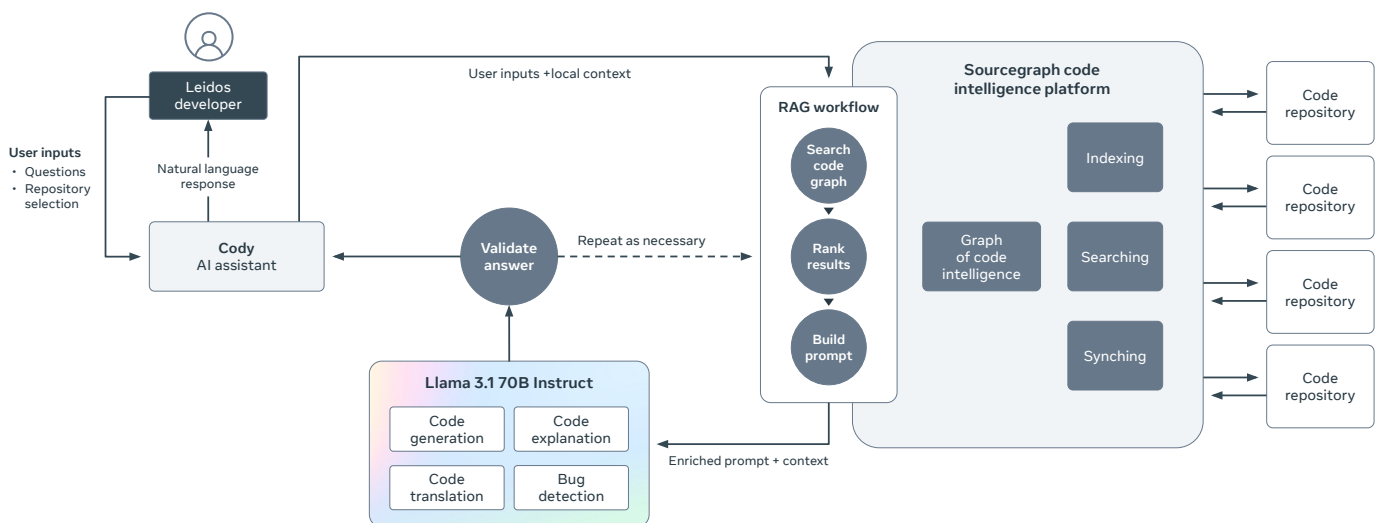
AWS Logically Isolated Secure Air-gapped (LISA) infrastructure provides elastic computing resources for workloads that demand the highest security — the kind of workloads Leidos developers run all day. Llama models are completely portable, and their open-source licensing gives users total control over their data and IP, making them ideal for the mission.

The team downloaded Llama 3.1 70B Instruct from Hugging Face and used LiteLLM to host the model in a Docker container that adhered to the OpenAI format for predictable integration with Sourcegraph. AWS Elastic Container Service (ECS) instances and Docker containers run entirely within the Leidos AWS LISA deployment.

By using Llama 3.1 70B Instruct, the team gained a 128K token input window, which significantly expanded the capabilities of Sourcegraph. Having a larger window allowed Sourcegraph to ingest more context at a time, support richer prompts and deliver more nuanced and accurate responses. However, that expanded capacity did create one of the project’s few hurdles.

“Our biggest challenge was ensuring we had adequate hardware to run Llama 3.1 70B Instruct,” says Davis. “We worked with AWS to requisition the correct GPU-enabled instance types in EC2 to back our ECS cluster and support Llama.”

## Llama-powered Sourcegraph natural language code assistant



Sourcegraph uses Llama to create intelligent code knowledge bases and assist developers with analysis, code generation and documentation.

# The outcome

## Open-source Llama helps Leidos innovate within tight budgets

Many Leidos customers are government agencies with fixed budgets, which limits the team's ability to adopt expensive, proprietary technologies. "We are always looking for ways to leverage open-source technologies to help reduce large software purchases," says Davis.

With Llama and Sourcegraph, Leidos found an open-source technology they could use to bring leading generative AI innovation to customers who aren't interested in chasing fads that might not advance the mission.

"Our customers have constrained budgets. They prioritize resources for mission functionality instead of enabling technologies," says Davis. "Open-source Llama allowed us to pursue generative AI coding assistants with customers who would have otherwise considered it cost prohibitive."

Llama-powered innovation has proven to be a valuable asset for Leidos in the marketplace. Their success in combining Llama and Sourcegraph has given them a competitive edge in federal contracts for advanced software development.

### Codebase QA

Developers explore, debug and document with natural language

### Task offloads

Context-aware agents generate code for mundane, repetitive tasks

### Competitive edge

AI code generation positions Leidos as an innovator with government and commercial clients

## +20%

increase in developer efficiency across teams

## +20%

improvement in architecture and application development velocity

## -92%

reduction in typical user time lost to context switching and learning

## +33%

boost in flexibility for faster engineering and feature implementation

"Llama provided Leidos with the flexibility, knowledge and power we needed to use generative AI for coding tasks in highly secure environments, enhancing efficiency by increasing capacity for more complex software development."

Jay Davis  
Chief National Security Software Architect, Leidos

# Conclusion

## Next up: Expanding code assistance and fine-tuning Llama

Leidos is exploring how to apply Llama in different Sourcegraph environments and scales to extend coding assistance to more situations. As they plan a future upgrade to Llama 3.2, the team sees a need to create custom models specifically for code. For Leidos, training Llama on their codebases and custom libraries they manage is the next step in tailoring performance for their unique, air-gapped customer spaces.



## How can Llama help your business?

See how open-source Llama brings unmatched control, customization and flexibility to generative AI application development and deployment.

[Learn More](#)[Related Stories ▶](#)