

AI Sweden Adopts Intel® Xeon® Processors and Intel® Gaudi® Accelerators for Prototype Virtual Assistant

RAG architecture with annotated training data can help public sector employees collaborate and access relevant information faster.



Executive Summary

AI Sweden is the national center for applied AI, with offices nationwide and in Montreal. It brings together more than 130 partners across the public and private sectors and academia. The Swedish government and its partners, both public and private, fund the center. Together, they invest in generating tools and resources to accelerate the use of AI to benefit Sweden's citizens, society, and global competitiveness.

Among other projects, the group focuses on developing a prototype generative AI solution to help public organizations collaborate nationally. In a public healthcare setting, for example, the digital assistant might help staff by reducing the time spent searching for and retrieving information and allowing more time for hands-on patient care and interpersonal work. Working with Intel, AI Sweden's proof-of-concept solution utilizes on-site servers equipped with Intel® Xeon® processors and Intel® Gaudi® 2 accelerators and a Retrieval-Augmented Generation (RAG) architecture that pulls information from staff-annotated content to make the virtual assistant's responses more accurate and helpful.

Challenge

Swedish municipalities can benefit from a digital assistant that helps government workers increase efficiency and mitigate the current labor shortage in the public sector. AI Sweden's work also enables a national collaboration around a shared solution with long-term value for government entities.



Jonatan Permert, AI Transformation Strategist, AI Sweden

To be successful, a generative AI assistant must provide specific and relevant answers to users' natural language queries. AI Sweden's prototype system involves a RAG (Retrieval-Augmented Generation) approach that uses staff-annotated content to increase the quality and validity of the digital assistant's responses.

Before the solution can help public organizations explore the value of resource sharing, AI Sweden must ensure that security measures adhere to local and European regulations like European Union's General Data Protection Regulation (GDPR). Therefore, the virtual assistant's front- and back-end technologies depend on end-to-end encryption.

Solution

AI Sweden is executing a multi-phased, multi-year project to develop a prototype digital assistant. Intel is a key partner in getting the project off the ground.

The solution's infrastructure, designed with security in mind, operates on-premise. Servers supporting the proof-of-concept system rely on Intel Xeon processors and Intel Gaudi 2 accelerators, which offer the team excellent price performance.

"We need powerful AI infrastructure to run our enormous language models. Working closely with Intel's team to deploy and optimize the Intel Gaudi accelerators made our prototype project possible. A common digital assistant for the public sector has the potential to benefit employees daily. We hope our work can serve as a template for other countries seeking to tackle similar challenges."

—Jonatan Permert, AI Transformation Strategist, Sweden's National Center for Applied AI

The system incorporates features like encrypted storage and memory in secure enclaves to accommodate Sweden's stringent privacy requirements and maintain security and access control. Intel® Trust Domain Extensions (Intel® TDX) provide added layers of protection to support confidential computing, and the execution environment uses hardware-isolated virtual machines (VM) to protect documents, data, and applications from unauthorized access.

If a municipality requires additional security measures, server infrastructure can adjust to meet these needs. With compliant security, even an administrator cannot access personal healthcare information passed through the system, ensuring privacy per European, national, and local regulations.

Development support from Intel helps AI Sweden maximize hardware capabilities, make software-based adjustments, and assist with preliminary testing on evaluation datasets. The process will help optimize the system to accommodate an increasing number of users over time.

With the assistance of Intel components and expertise, the proof-of-concept AI assistant can now handle tasks like creating documents, improving written language, or filling

out applications. Because municipalities employ a diverse workforce, and not all are fluent in Swedish, the solution also manages content and queries in English. Plans include support for more languages.

"In our early testing, generative AI helped public employees find relevant information in a fraction of the time required by a manual, legacy approach. Given the current shortage of skilled workers, our prototype virtual assistant can save existing employees' time so they can stay focused on job activities that matter most."

—Jonatan Permert

AI Sweden and the associated municipalities will scope multiple use cases as the proof-of-concept system demonstrates its potential. For example, if a government official needs to make a critical decision, the digital assistant can help summarize the most relevant background information.

Results

While the prototype digital assistant remains under development, the initial results show great promise. Early testing and anecdotal feedback indicate the solution remains on track to meet its goals. Government employees testing the system found benefits like time savings and more accurate results with text-related tasks.

A worker shortage in Sweden can make a digital assistant especially valuable to the country. For example, with generative AI taking over time-consuming and repetitive tasks in government facilities, employees can allocate more time to job activities that require human expertise and interaction.

The digital assistant can also adapt to support smaller public organizations with fewer resources, smaller budgets, and less specialized knowledge. AI Sweden plans to expand the testing process to a thousand users in the near future. Over time, the digital assistant could inspire other nations seeking to tap AI's potential in government settings.

The virtual assistant supports Sweden's goals of maintaining digital sovereignty and making technology investments that could make public employees' jobs easier and more efficient. The solution also supports Sweden's democratic ideals for transparency, integrity, and opportunity for greater representation.

Key Takeaways

- A collaboration between AI Sweden and governmental municipalities led to a prototype digital assistant that can assist government staff with repetitive tasks, allowing them to focus on the more interpersonal aspects of their jobs.
- Intel® Xeon® processors and Intel Gaudi® 2 accelerators underly the prototype's architecture and provide the needed security features, scalability, and performance.

- Making RAG solutions valuable to users requires significant fine-tuning. With the help of staff-annotated content, the digital assistant can respond to queries with more applicable and accurate information.
- AI Sweden designed the prototype to meet European data privacy requirements like GDPR and Sweden’s countrywide mandates. A combination of secure memory enclaves, encrypted storage, and Intel® TDX provided a foundation for multi-layered security.
- AI Sweden’s experience building and testing the proof-of-concept system can serve as a template for other countries wishing to implement a similar solution.

Find out more:

Read about [AI Sweden’s mission and goals](#).

Learn about [Intel® Gaudi® accelerators](#) and [Intel® Xeon® processors](#).



Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

For workloads and configurations visit www.Intel.com/PerformanceIndex. Results may vary.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.