Case Study

Intel® Software Center of Excellence
Intel® Xeon® Processors

Integrated AI Solution Boosts Legal Firm's Productivity and Accuracy

At a Glance

Using a multi-layer generative AI solution in collaboration with Intel, Activeloop, and ZERO Systems, legal firm Ropers Majeski was able to:

- Boost knowledge worker productivity by 18.5%, saving 75 minutes per person per day.[1]
- Improve document ingestion rates by 2.3x.[1]
- Do this while meeting strict regulatory and security requirements.

Ropers Majeski partnered with Intel, Activeloop and ZERO Systems to create a powerful, dependable, and secure AI-driven solution, combining advanced hardware and data storage with a generative AI solution powered by a Large Language Model.

Executive Summary

Ropers Majeski, an established midsize law firm, sought to streamline their operations by automating various data processing tasks through an AI-based solution, while ensuring data security and compliance. They partnered with Intel, Activeloop, and ZERO Systems to create a multi-layer AI system powered by the high-performing 4th Gen Intel® Xeon® processors. The solution significantly improved worker productivity while maintaining data security.

Challenge: Processing Data Swiftly and Securely

The law firm of Ropers Majeski has provided high-quality legal advice to clients worldwide for over 70 years. They were searching for a way to automate a wide range of tasks, including documenting, filing, timekeeping, storing, and information retrieval. Their knowledge workers normally had to perform these tasks manually, which resulted in inaccuracies and consumed valuable non-billable hours for highly trained personnel.

Any automated solution would have to be fast, reliable, and able to handle a high volume of data to ensure regulatory compliance. It would also need to satisfy the security requirements involved in handling sensitive client information. For example, all data would have to stay in-house, and vendors could not have access to confidential documents.

Solution: Multi-Layer AI Approach

Ropers Majeski partnered with Intel, Activeloop and ZERO Systems to create a powerful, dependable, and more secure AI-driven solution, combining advanced hardware and data storage with a generative AI solution powered by a custom Large Language Model (LLM). ZERO Systems' proprietary AI engine, Hercules, powers the generative AI application for labeling, filing, and indexing confidential client material, as well as helping to generate content such as emails.

Secure, enterprise-grade storage came from Activeloop's Deep Lake, a database for AI. This made it possible to store both embeddings and multi-modal data for millions of sensitive documents and provided advanced role-based access controls to ZERO Systems' solution.

> "This kind of strategic partnership is just the beginning of integrating generative AI into the legal profession and other highly regulated industries."
>
> *—Gevorg Karapetyan, Co-founder and CTO of ZERO Systems.*

The 4th Gen Intel® Xeon® Scalable processors brought the compute power that enabled the customer to avoid investing in an expensive proprietary solution. Optimized for AI inference, these CPUs provided significant performance and efficiency gains through embedded accelerators such as Intel® Advanced Vector Extensions 512 (Intel® AVX-512) that helps with computation of vectors used in

contexts or applications outside of Deep Learning, and Intel® Advanced Matrix Extensions (Intel® AMX) with built-in hardware support for AI acceleration.

The connection to Activeloop and ZERO Systems came through the Intel Disruptor Initiative which supports its members by driving growth through technical enablement and multi-channel go to market activities.

Result: More Speed, Higher Productivity

Once the multi-layer system was up and running, Ropers Majeski saw the productivity of their knowledge workers increase by 18.5%, saving an average of 75 minutes per user per day. By automating tasks such as email processing, document filing, and report generation, the solution increased document ingestion rates by 2.3x.[1]

This increase in efficiency, including near real-time information and insights retrieval, freed staff to focus on billable work hours without sacrificing accuracy. In addition, the AI system helps create timecards based on worker activity as well as profile emails and documents, even when employees do not have access to internet.

> "The Intel disruptor team provides a very personal touch with all partners. We highly appreciate that as you don't often see this in partnering with other companies. In fact, you actually hear a lot of advice that doesn't make sense for a startup, to partner with large companies. And I can definitely tell you that Intel here is an exception."
>
> *—Davit Buniatyan, Activeloop CEO*

Thanks to Intel AVX-512, 4th Gen Intel® Xeon® Scalable processors could pack more operations into each clock cycle, improving the query retrieval speed of Activeloop's Serverless Tensor Query Engine by 100x.[2] The results met strict security requirements that are essential in highly regulated environments such as legal tech.

"We are excited about our collaboration with ZERO Systems, under the Intel Disruptor Program, in bringing strong value to the end customer," explains Arjit Bandyopadhyay, CTO of Enterprise Analytics & AI, Head of Strategy Cloud and Enterprise – CSV Group, Intel Corporation. "The performance and TCO optimization of the Gen-AI solution on Intel's latest generation CPUs building on in-silicon AI accelerators like Intel AMX, and other IA based software-hardware assets, go a long way to provide consistent benefits and a secure offering to the overall solution."

"This is where Intel CPUs shine," says Davit Buniatyan, CEO of Activeloop. When GPU availability is limited, Intel processors offer the compute resources to run and fine-tune LLMs. "With Deep Lake's performant dataloader and Tensor Query Engine, pioneers like ZERO can develop groundbreaking Gen AI products on Intel CPUs at a fraction of the cost," he adds.

"Moreover, enterprises can use the existing architecture they're already familiar with, without having to invest in expensive new hardware," says Gevorg Karapetyan, CTO of ZERO Systems. "Implementing your solution on CPUs not only makes it much more efficient and cost effective, but it also makes it very easy for customers to deploy this today, without having to wait."

"We're very happy with the fact that we're able to utilize our existing Intel systems, and we expect to continue to take advantage of the advanced AI technologies to streamline our operations," says Maks Agamir, Director of IT at Ropers Majeski.

By automating repetitive tasks, the AI solution provided by the collaboration of Intel, Activeloop, and ZERO Systems will only enhance the performance of Ropers Majeski's experienced knowledge workers, Agamir explained. "Technology is not our primary business—law is," he says. "Attorneys and legal professionals were happier because when their day runs out, instead of sitting there and profiling every email, every document, they just looked and said okay, it's done. AI did it all."

Where to Get More Information

Learn more about ZERO Systems.

Learn more about Activeloop.

Explore the capabilities of 4th Gen Intel Xeon processors.

Learn more about the Intel AI Disruptor Initiative.

[1] Performance claims provided by Ropers Majeski. Number of Embeddings: 100,000, Ndim: 384, Precision: float32, Cores: 64. Machines used for benchmark: Xeon 3 (64 cores) old set up, Xeon 4 (64 cores), Xeon 4 (64 cores) with AVX-512 enabled (the new setup delivered through the collaboration). Ingestion benchmark - the time to ingest 100,000 chunks of 512 characters was used as the benchmark. The average time elapsed to complete 1000 queries was used for query benchmarks.

[2] Query retrieval speed claim provided by Activeloop.

092023/RJMJ/JS/PDF ♻ PleaseRecycle 357368-001US