White Paper
May 2024

IT@Intel: Democratizing the Use and Development of Generative AI Across Intel

Our GenAI initiative empowers teams across the enterprise to rapidly and safely develop new tools powered by large language models

Intel IT Authors

John Hansen
Generative AI Platform Architect

Kyle Short
Director of Generative AI and Automation

Table of Contents

Executive Summary 1
Business Challenge 2
Solution 2
Solution Architecture 3
Governance and Control 3
Training and Communication 4
Results 4
Productivity Savings 4
Next Steps 4
Conclusion 4

Executive Summary

With the surge in the growth and adoption of Generative AI (GenAI) tools, Intel has seen hundreds of employees and teams begin using and developing tools based on commercial large language models (LLMs).

Recognizing the immense potential of these tools to enhance efficiency and innovation—as well as concerns around AI data security, ethical use, and accuracy of responses—Intel IT formed a working group to understand common usage patterns, coordinate knowledge sharing across teams, and establish guidelines for safe usage.

This proactive approach led to the development of a new platform, aimed at democratizing Intel's use of GenAI by empowering employees at all levels of technical skill to use these tools to improve their day-to-day productivity and unlock new efficiencies in Intel's business processes.

Our initiative is designed to facilitate the use of GenAI without the need to repeatedly rebuild common components, thereby reducing security risks and speeding time to value.

The rapid deployment of this platform has already demonstrated tangible value. Although the platform is currently in pilot, early users have reported notable increases in productivity. We estimate that approximately 75% of Intel employees will eventually use the GenAI platform.

Intel Contributors

Xochitl Monteon, Chief Cyber Officer and VP Cybersecurity Governance Risk & Compliance

Marla A. Gomez, Director of User Experience, Intel IT

Robert Vaughn, AI Industry Engagement Manager, Intel IT

Acronyms

BU	business unit
GenAI	Generative AI
LLM	large language model
NLP	natural language processing
RAG	retrieval augmented generation

Business Challenge

In 2023, Generative AI (GenAI) applications based on large language models (LLMs) from commercial vendors became widely available for the first time. Recognizing the immense potential of this technology to make various types of work more productive, numerous groups within Intel began using and developing tools aimed at harnessing its vast capabilities.

A key benefit of GenAI technology lies in its ability to enable employees to redirect their efforts toward higher-value tasks, thereby optimizing organizational resources. To meet Intel's ambitious productivity objectives, Intel IT recognized the need to enable employees across the enterprise to create and use their own tools based on LLMs.

In order to realize the full potential of GenAI, we needed to empower large top-down enterprise applications delivered by expert teams while also saturating the entire company with bottom-up productivity improvements. This involves capitalizing on long-tail opportunities by enabling end users and power users to address the numerous smaller projects that collectively contribute to substantial gains.

Solution

To better understand current GenAI use cases, we conducted a survey with teams across Intel to review over 200 proofs of concept to understand challenges and commonalities. This survey enabled us to coordinate feedback from various teams and inform our development of safe usage guidelines. Our analysis revealed three typical usage patterns:

- **Prompt engineering:** This pattern involves iteratively revising prompts to meet a specific need.
- **Code generation:** In this pattern, GenAI tools are used to create and revise programming code.
- **Retrieval augmented generation (RAG) and knowledge assistants:** RAG use cases involve using AI algorithms to retrieve relevant information from large datasets and generate content based on it. Knowledge assistants are AI-powered systems designed to help users find and infer new information.

In essence, we discovered that teams were repeatedly building the same things: setting up application servers, using AI orchestration tools, and building user interfaces. In addition, multiple groups were using the same datasets, increasing costs to prepare and store redundant data.

Based on this research, we embarked on a journey to develop a new GenAI platform that will enable teams to safely use GenAI without requiring them to rebuild common solution components, thereby reducing security risks and improving time to value.

The GenAI platform is designed to be used by anyone, regardless of technical skill level. With proper authorization, users can easily build GenAI solutions as well as consume them.

GenAI Applications at Intel

Below are three GenAI uses cases being implemented by Intel's AI Solutions Group:

- **Virtual Technical Assistant (VTA)** uses large language models (LLMs) to answer questions and provide reasoning capabilities. The VTA derives information from multiple documents and data sources, and scales the dissemination of technical knowledge by providing rapid, accurate answers to technical design inquiries. This helps to accelerate the design pipeline and increase efficiency and productivity across the organization.
- **Co-Design** brings the power of GenAI technologies to Design Engineering processes and tasks. With several capabilities already in production and many more in the works, Co-Design will help with tasks such as answering questions about Intel's technical design documents; generating, explaining, and debugging code; and creating validation and test artifacts like assertion rules and test programs.
- **Account Pulse** is a new AI capability in the Sales customer relationship management (CRM) platform that helps sellers save time when managing their accounts. Powered by LLMs, Account Pulse generates an auto-summary of the seller's accounts and opportunities based on Sales CRM data.

End users, power users, developers, and data scientists can all assemble solutions that align to their skill levels.

- **End users** can find and use GenAI assistants developed by others through a common user interface. Any assistant developed on the platform is exposed as a simple chat client for easy consumption.
- **Power users,** or people with basic prompt engineering skills, can easily assemble and share both knowledge assistants and highly engineered prompts with other users.
- **Developers** can create more complex AI workflows that use the AI workflow engine to integrate with structured and unstructured data to take actions as directed by the user. These developers can then expose their solutions as APIs, which can be consumed by other applications.
- **Data scientists** can use their skills to build complex data ingestion solutions, as well as enhance the AI workflow engine with custom or Python-generated code to achieve specific high-value solutions. These solutions can then be automatically exposed to users in the form of chats or published as APIs for other applications to consume.

A future platform release will enable the sharing of knowledge assistants with vectorized data embedded into their workflows, along with security procedures to help ensure that data is not shared beyond the entitlements that protect the original data sources.

Solution Architecture

Figure 1 shows the high-level conceptual architecture of the GenAI platform, consisting of three main components which enable innovators of varying skill levels to assemble GenAI solutions.

- **Data:** These standardized components can be used to create vectorized storage of unstructured documents for RAG pattern workflows.
- **AI Workflow:** This component uses open-source software to provide a dynamically configurable workflow engine that combines data and LLM inference to achieve user goals.
- **LLMs:** This component abstracts a myriad of LLMs, hosted both on commercial cloud platforms and internal platforms, providing AI workflow developers with the option to use any approved model to meet their needs.

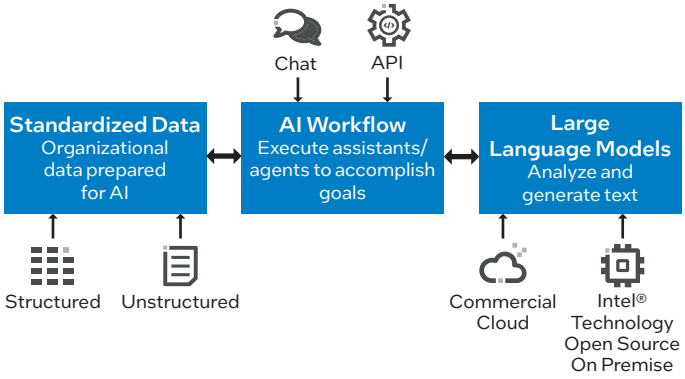

Figure 1. High-level architecture of Intel® GenAI platform.

Confidential computing is an important aspect of our GenAI platform. Certain use cases require highly secure hardware solutions, and these use cases must be delivered using on-premises hardware to protect data in accordance with Intel Information Security guidelines. The GenAI platform allows users to choose secure models as required by their use cases.

The GenAI platform is not tied to a single LLM vendor—rather, it is built to enable users to select the right model for desired business outcomes. The platform shows the cost difference between models to encourage the correct choice for a given set of cost and performance parameters. Business units (BUs) that can use lower-cost localized models can easily select them, while BUs that need high-cost, high-capability models can choose to use those instead.

Governance and Control

AI tools were once the exclusive realm of data scientists who were focusing on defined high-value projects. The popular availability of these tools introduces new challenges in governance, security, and data privacy.

Intel IT has developed a comprehensive GenAI governance process that encompasses widely available tools from internet vendors, solutions from enterprise cloud vendors, and solutions developed by BUs and enterprise IT.

The GenAI platform is built to enable Intel's policies for information security, ethical AI usage, and legal compliance.

- Each new standalone project that is implemented outside of the GenAI platform must include all security controls in their solution. Centralizing this policy through the platform helps decrease security risks.
- Intel's ethical AI team develops high standards for AI usage, which are applied on the GenAI platform.
- AI legal policies can also be applied at the GenAI platform level to help ensure that all of Intel's AI workloads adhere to regulatory requirements around the globe wherever we do business.

The governance process for new projects involves the following steps:

1. The process begins with Intel IT's standard application registration method. This process, which is systemized by an IT-supported tool, controls the flow of an application from idea to production.
2. Applications based on GenAI models are guided through a model and use case approval process, which is owned by Intel Information Security and Intel Privacy groups to ensure that use cases remain within approved policy.
3. Once the model and use cases are approved, segment architects review proposed solutions from the development teams before providing final approval.

The GenAI platform makes the approval process much more efficient. Many aspects of model approval, policy systemization, and information security control either are or will be implemented as part of the core GenAI platform. Projects implemented on the GenAI platform automatically inherit these controls.

Training and Communication

To promote the use of the GenAI platform, we've initiated a training and communication program called AI @ Work, which is headed by our Director of User Experience. This group holds focus groups and conducts surveys to find out what people want from GenAI.

Early research indicates that people who first use the platform without an adequate understanding of its capabilities are often unsatisfied, which underscores the importance of our training and communication program to help set expectations about what GenAI can and cannot do.

The program's operating premise is "We can give you the best platform in the world, but if we don't make you aware of it, inspire you to use it, and tell you how, we've failed."

Results

Although the platform is still in pilot, usage is growing quickly.

Figure 2 shows the growth in usage over a recent thirty-day period, with the number of engaged users almost doubling in that time. Engaged users are defined as those who have used the GenAI platform more than seven times in the past thirty days. This growth was entirely organic, taking place before the kickoff of outreach and training efforts.

Eventually, we anticipate that approximately 75% of Intel employees will use the GenAI platform. The basis for this internal estimate is that the majority of employee roles are technical or involve computer skills and the observation that employees across the company have expressed interest in GenAI activities.

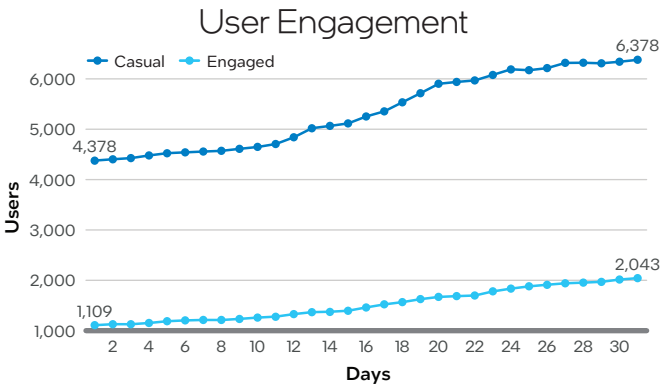

Figure 2. Intel® GenAI platform user engagement levels.

Productivity Savings

The overarching goal for Intel's AI strategy (encompassing all types of AI, not just GenAI) is to achieve productivity improvements of 30–40% across the board.

Quotes from Users

- "[GenAI] will drive a ton of innovation up and down the tech stack—which makes for a happy engineer flush with interesting problems and useful solutions."
- "I used it to analyze over 800 survey comments recently and saved myself at least 40 hours of work."
- "The ability to ask questions without tying down a human resource (i.e., another developer) and coming off as completely ignorant may help to reduce feelings of imposter syndrome."
- "AI is already saving time and allowing focus on more critical aspects of the job. Excited by all the possibilities as the capability continues to evolve!"

Complete and detailed data are not yet available, but teams are already reporting encouraging productivity enhancements with the use of GenAI-based tools.

Anecdotally, users relate that GenAI technology is changing how they feel about work. For example, an employee who is a non-native speaker of English reported less stress and better responses when using a GenAI tool for help with writing emails. Positive results have also been reported by users with visual and hearing impairments who find that using GenAI tools increases ease of communication.

Next Steps

We will be deploying model cloud capabilities in mid-Q2 2024 which allow GenAI developers to easily test and use models that run best on servers powered by Intel® Gaudi® processors and Intel® Xeon® processors. This will enable us to demonstrate how Intel® hardware can be used to build solutions that meet the needs of specific use cases while helping to control GenAI-related costs.

Conclusion

Intel IT is committed to empowering employees through innovative technologies, fostering a culture of collaboration, and ensuring responsible and safe GenAI use to drive transformative outcomes for Intel's business and beyond.

The Intel GenAI platform provides a standardized framework which unlocks value across BUs by enabling employees to innovate and build GenAI solutions in a standard way while adhering to responsible and safe AI practices.

This innovative and timely initiative helps democratize GenAI across Intel by enabling users of all levels of technical skill to contribute to Intel's business goals.

Related Content

If you liked this paper, you may also be interested in these related stories:

- Autonomous Quality in AI Model Productization: A Journey
- Transforming Intel with AI
- Smart Manufacturing Using Computer Vision and AI for Inline Inspection
- Transforming Manufacturing Yield Analysis with AI
- Minimizing Manufacturing Data Management Costs

For more information on Intel IT best practices, visit intel.com/IT.

IT@Intel

We connect IT professionals with their IT peers inside Intel. Our IT department solves some of today's most demanding and complex technology issues, and we want to share these lessons directly with our fellow IT professionals in an open peer-to-peer forum.

Our goal is simple: improve efficiency throughout the organization and enhance the business value of IT investments.

Follow us and join the conversation on **X** or **LinkedIn**. Visit us today at **intel.com/IT** if you would like to learn more.

Intel technologies may require enabled hardware, software or service activation.
No product or component can be absolutely secure.
Your costs and results may vary.
© 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.
0524/WWES/KC/PDF