White Paper
Information Technology
October 2025

IT@Intel: Agentic AI in the Enterprise

Intel's unified, agent-driven GenAI chatbot platform—already in use by Sales and Marketing—serves as a blueprint for scalable, efficient, and future-ready enterprise AI

Authors

Arun Sagiraju
Principal Engineer, Chief Architect, Market To Revenue Segment, Intel IT

Boaz Efroni Rotman
Director of AI Strategy and Roadmap, Sales and Marketing Group

Vivek Kumar
Software Application Architect, IT Architecture and Infrastructure Services, Intel IT

Assaf Perry
Senior AI Product Manager, Sales and Marketing Group

Table of Contents

Executive Summary

Generative AI (GenAI) and agentic AI have great potential to improve productivity. However, redundant and isolated AI projects across the enterprise make inefficient use of AI model resources, hinder governance and security/privacy processes, and impede scalability and flexibility.

To address these challenges, we developed a unified agentic GenAI platform called One AI (1AI) that uses Intel® architecture and open-source tools to consolidate siloed chatbots into a single, unified agentic AI platform. The new platform uses a modular, agent-based architecture: each specialized agent addresses a unique business use case and chooses the most effective large language model (LLM) for its specific task.

Simply building an amazing agentic AI platform, however, is not sufficient for measurable success. Cultivating the strongest business value from agentic AI requires close collaboration and alignment with business units (BUs). The success of agentic AI hinges on active participation, open communication, and a shared commitment to business goals.

In 2025, we worked with the Sales and Marketing group (SMG) AI Strategy team to identify high-value use cases and deploy the 1AI platform into the SMG environment. The team consolidated several use cases that use different data sources and LLMs; all use cases are accessed through a single interface. This unified interface—instead of multiple, siloed chatbots—has been crucial to the 1AI platform's success because it helps ensure a positive, consistent user experience (UX), which in turn helps drive platform adoption.

Based on our SMG success, we plan to scale the 1AI platform to accelerate Intel's growth by embedding GenAI into Intel's business processes to improve productivity with automation. 1AI can enable us to simplify agentic AI deployment and maintenance through centralization, drive cost and resource efficiency with purpose-fit LLMs, and quickly deliver a flexible rollout of new GenAI-driven use cases.

As we scale 1AI across additional BUs, we are strategically aligning our AI efforts to corporate goals; establishing robust AI governance, security, and privacy guardrails; and investing in employee-readiness to enable improved decisions, enhanced performance, and greater productivity.

Contributors

IT Customer Sales and Support Group

Mahesh Biradar, Lead Developer
David Johnston, Technical Lead
Brent Rieck, Lead Developer
Renee Rivera, Senior Director
Michael Thaxton, Enterprise Architect
Amal Thomas, Lead Developer
Norman Yee, Enterprise Architect
Muhammad Zaman, Lead Developer

IT@Intel

Robert Vaughn, Industry Engagement Manager

Acronyms

1AI	One AI
AI	artificial intelligence
BU	business unit
CaaS	Container as a Service
CRM	customer relationship management
GenAI	generative AI
LLM	large language model
MCP	model context protocol
PoC	proof of concept
SMG	Sales and Marketing group
UX	user experience

Background

Artificial intelligence (AI) in the enterprise has been a key area of innovation at Intel for nearly a decade. For example, in 2018, Intel IT developed Sales AI,[1] a platform that enabled Intel to significantly scale its sales activity. Then in 2024, large language models (LLMs) and retrieval-augmented generation hit their stride, generating enormous interest in Generative AI (GenAI), chatbots, and agentic AI.

Fragmented AI Efforts Diminish Business Value

Intrigued by the GenAI's productivity potential, business units (BUs) across Intel began to develop and use chatbots in isolated silos—such as in customer service, sales and marketing, human resources, IT, and more. This fragmented AI landscape led to inconsistent user experience (UX), redundant solutions, a lack of governance, and rising maintenance challenges. It also negatively impacted Intel's business by increasing costs, enabling the authorization mechanism to be bypasssed to access sensitive data, delaying time-to-market, and allowing siloed/incomplete implementation of broader AI use cases. Intel IT's challenge was to consolidate AI chatbot interfaces and streamline their use to enhance efficiency and UX.

Our Goal: A Unified, Scalable, Agentic AI Solution

Simple chatbots—like our internally developed iGPT—are only the tip of the GenAI iceberg. As the AI field evolves, so does the need for scalable, efficient, and well-governed AI infrastructure that supports not just isolated chatbots, but true agentic AI (see the sidebar, "A Closer Look at Agentic AI"). We seek to leverage existing technology to create a unified AI solution that can integrate vendor-provided and internal AI agents into a single common interface. This unified user interface is crucial to the success of the 1AI platform. We have learned through feedback that UX is a key driving factor in platform adoption. In other words, even the most powerful technical solution will not gain popularity unless it provides a positive, consistent UX.

In the next few sections, we'll discuss our AI vision, showcase some use cases already in production for Intel's Sales and Marketing group (SMG), and take a closer look at the solution architecture.

Overview of the 1AI Platform

Our AI vision—called OneAI (1AI)—focuses on embedding a single agentic AI platform into organizational processes to drive growth and enhance productivity across the enterprise (see Figure 1). This new platform can significantly reduce costs, improve data security, and accelerate time-to-market for Intel. We showcased an early version of the 1AI platform at the Intel Vision 2025 conference.[2]

As with other Intel IT AI initiatives, the 1AI platform is guided by three pillars that help ensure the entire enterprise can effectively adopt it:

- Align AI initiatives with BU and corporate goals for impactful AI execution.
- Establish robust, business-centralized governance to evaluate, prioritize, and consolidate use cases to maximize business value.
- Support employee-readiness by equipping BU employees with the necessary skills and knowledge to successfully adopt AI to enhance their productivity.

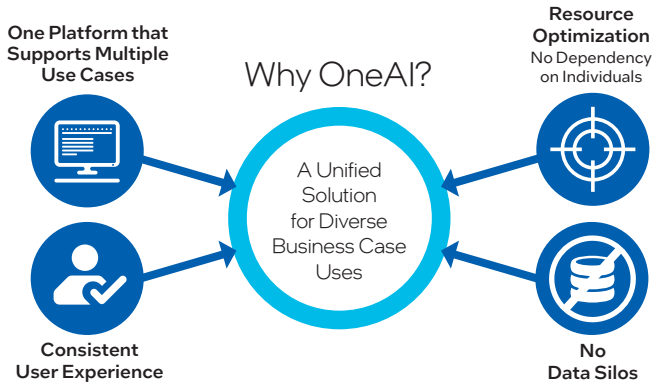

Figure 1. Embedding a single GenAI platform into organizational processes can drive growth and enhance productivity across the enterprise.

[1] See the IT@Intel white paper, "Improving Sales Account Coverage with Artificial Intelligence."

[2] Visit https://www.intel.com/content/www/us/en/events/on-event-series/vision.html for more information on this conference.

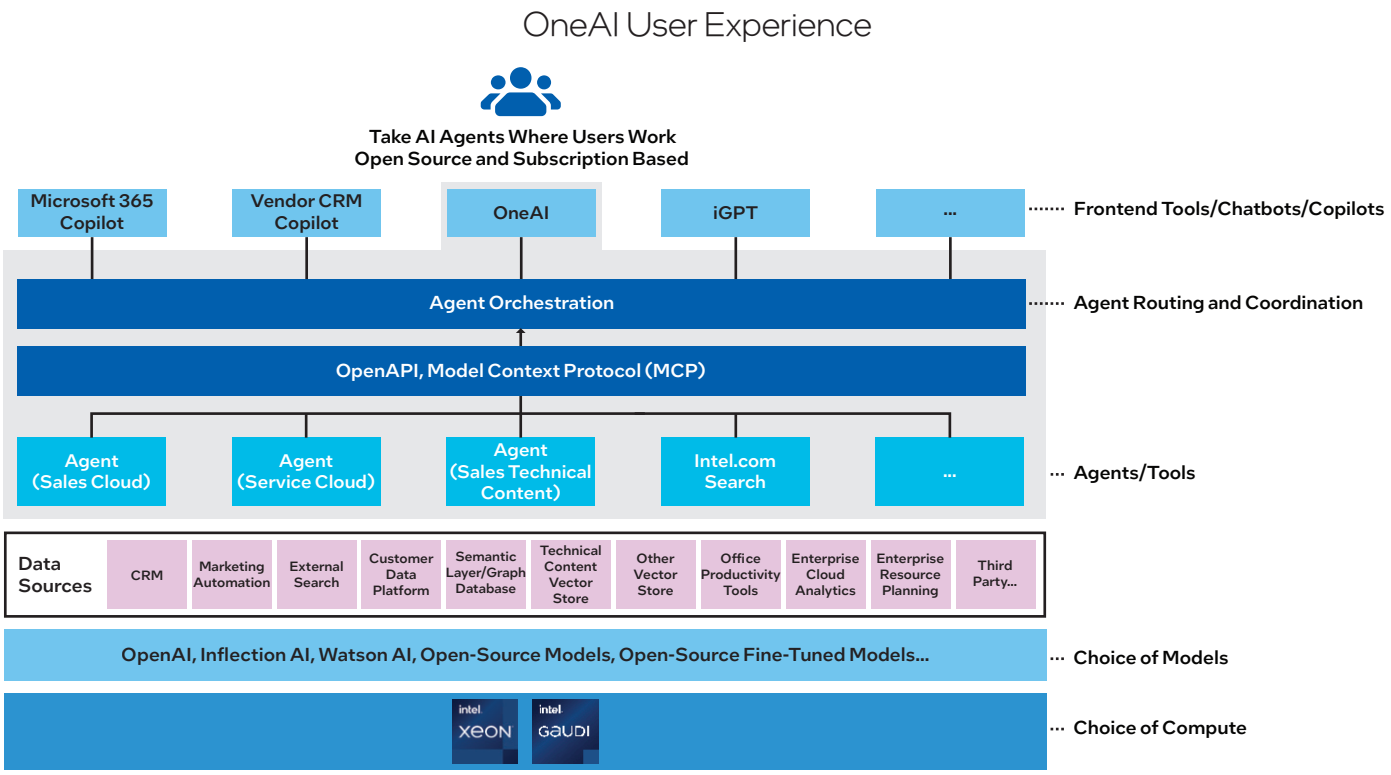

Figure 2. The 1AI platform enables agentic AI, where a prompt is routed to the most relevant AI agent and model.

Choosing a Deployment Option

When developing the 1AI platform, we considered several approaches, including creating a complete, custom solution; utilizing commercial offerings; and leveraging open-source software. After careful evaluation, we decided to use an open-source, out-of-the-box frontend, complemented by custom-built agentic AI bots hosted in Intel's private cloud. We selected this approach because it balanced flexibility, cost-effectiveness, and control over the solution, allowing us to tailor the AI bot to Intel's specific needs while leveraging the strengths of open-source software and Intel® architecture.

How It Works

As shown in Figure 2, the user interacts with the AI agents at a high level, simply entering a prompt. The platform passes the prompt to a semantic layer that dynamically routes the prompt to the most relevant AI agent, which in turn has access to multiple data sources and commercial frontier and open-source models. Which LLM is chosen depends on relevancy and cost. Multiple teams can independently develop AI agents without concern for the frontend, while adhering to common design principles. The platform is hosted in Intel's private cloud. The AI agents run on Intel® Xeon® processors and are configured to use private frontier LLMs or open-source LLMs running on Intel® Gaudi® AI accelerators.

A Closer Look at Agentic AI

An AI agent is a system (code plus IT resources) that uses a reasoning system such as a large language model (LLM) to direct the control flow of an application. Agentic AI is the technology that empowers AI agents to take action with or without human oversight. Key characteristics of agentic AI include the following:

- **Autonomy:** The ability to initiate and complete tasks without constant human oversight
- **Decision-making:** Sophisticated reasoning based on context and tradeoffs
- **Adaptability:** Learning from interactions and adjusting behavior to achieve specific goals
- **Language understanding:** Comprehending and following complex instructions
- **Workflow optimization:** Efficient execution of multistep processes

1AI's First Deployment

According to recent industry research, 95% of GenAI pilot projects deliver little to no measurable impact on a company's profit and loss.[3] The remaining successful 5% deliver value because they are strongly aligned with BU goals and integrated with BU processes. This alignment is possible only by IT forming close relationships with BUs, understanding their processes, and determining where AI can significantly boost productivity.

For the first deployment of the 1AI platform, we connected with SMG's AI Strategy team in early 2025 to establish business-centralized governance and close alignment with BU goals and processes. The IT team performed rapid prototyping and forward engineering to showcase how technical capabilities could automate SMG's manual business processes through AI agents.

Narrowing Down SMG Use Cases

The SMG AI Strategy team worked with SMG leaders to catalog more than 60 use cases, including customer self-service, support agent productivity, and sales pipeline insights as well as their potential impact on operations and revenue. SMG experts worked with us to successfully unify disparate GenAI chatbot initiatives into a single, modular agentic AI platform. The following criteria were used to select the top use cases from the original list of more than 60:

- **Strategic alignment.** In line with top corporate objectives and current pain points
- **User impact scale.** The number of employees affected and frequency of use
- **Productivity gain quantification.** The hours saved per user per week/month, or revenue impact, with measurable key performance indicators
- **Value category.** Cost reduction, operational efficiency, revenue increase, or decision-making enhancement
- **Data maturity score.** Quality, accessibility, and governance readiness of required datasets
- **Implementation complexity.** Technical integration effort, infrastructure requirements, and timeline
- **Change management burden.** Training needs, process disruption, and difficulty level of user adoption
- **Total cost of ownership.** Build versus buy analysis, including development, licensing, and maintenance costs
- **Risk-adjusted return on investment.** Financial return accounting for technical, operational, and adoption risks
- **Scalability potential.** Expansion across departments, regions, or similar processes
- **Compliance and security impact.** Data privacy, internal controls, and audit requirements

We used these criteria when selecting the following active 1AI SMG use cases:

- **Customer Relationship Management (CRM) chatbot.** Sales managers can use conversational AI to obtain relevant CRM information, status updates, and insights about their team's accounts and opportunities. This use case proves 1AI's ability to scale to support BUs mining insights from unstructured data.
- **Seller self-support.** This dedicated agent provides real-time answers for questions relating to CRM, access and entitlement, and Intel Premier Support. The answers are sourced from the Sales CRM training site, which contains all these details.
- **Support agent chatbot.** SMG staff can search internal and external knowledge articles, customer stories, and other documents. The LLM provides unified, curated results.
- **Marketing - performance claims and benchmarking guidance.** The AI agent can assist in understanding Intel's legal guidelines and policies.
- **Ask co-marketing.** Sellers, co-marketers, and partner incentive participants can initiate a self-help chat based on training website content.

Figure 3 shows a screen capture of the 1AI interface for SMG—a single location to interact with various AI agents and data sources.

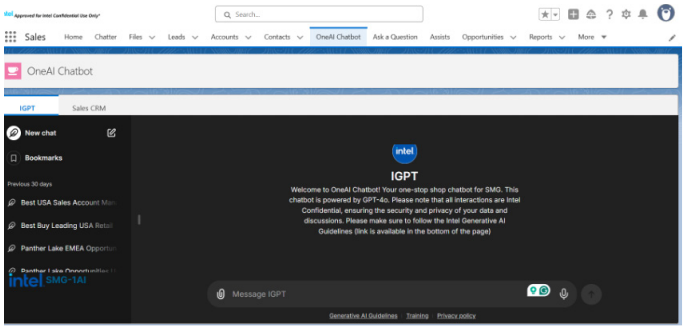

Figure 3. Users can access a single interface to ask various types of questions.

Building on our work with SMG, we feel confident that we can scale and extend the 1AI platform to additional BUs, generating tremendous improvements in productivity across the enterprise. We will increase the use of automation to drive 1AI's business value (see the sidebar, "Ascending the Automation Ladder"). A scaled 1AI platform will enable every BU to deploy specialized AI agents powered by the most efficient language models, all under centralized governance for optimal business performance.

[3] Source: Fortune, August 2025, "MIT report: 95% of generative AI pilots at companies are failing."

Proofs of Concepts (PoCs) and Pilot Projects

During our work with SMG, we conducted multiple PoCs. Our goals included the following:

Overall architecture goals

- Evaluate the new chatbot architecture proposed by IT.
- Understand cross-platform portability, scalability, and reusability for any future AI platforms.

Specific architecture concerns

- Explore the best design for data storage in the vector database for unstructured data (chunking and metadata).
- Test vector database search options and recommend an initial approach, along with recommended additional preprocessing.
- Validate the ability for agents to understand the user's intent in the AI prompt and effectively source data in real time from multiple data stores and objects.
- Investigate data refresh processing into the vector database, including defining an initial approach and the effect on data latency.
- Test a simplified security approach, which we can then scale to support full production usage after a successful pilot project.

Usability concerns

- Understand the type of questions users ask.
- Evaluate the accuracy of answers and gather user feedback with a customer success framework that monitors AI agent responses and undertakes data-driven corrective actions and enhancements.
- Assess the business value for each of the use cases.

Ascending the Automation Ladder[4]

Agentic AI relies on the automation of tasks and processes using AI technologies. Automation maturity—the different levels of automation—can affect the impact of agentic AI on productivity. The higher the level of automation, the more agentic AI can drive user efficiency and business value. A single-step process, such as searching or summarizing, requires a low level of automation. But true business value lies in multistep automation that can accelerate AI agents so they can work across multiple interfaces (copilots, chatbots, and more).

Currently, there is no consensus in the IT industry as to the definitions of the various levels of automation. However, we use the following six levels:

- **Level 0 – Fixed automation:** No true agentic behavior, just robotic process automation (RPA) with deterministic outcomes.
- **Level 1 – AI-augmented automation:** Basic agentic behavior at the individual decision level; that is, fixed automation with some steps augmented by large language models (LLMs).
- **Level 2 – Agentic assistant:** Task-specific agentic automation assistants capable of using tool-calling. These systems can interpret user intent, determine the desired outcome, and take appropriate action.
- **Level 3 – Plan and reflect:** Commonly called an AI agent, this is the first level to exhibit constrained autonomy. These agentic systems can create plans based on given intents, execute them, reflect on their success, and modify plans mid-execution if necessary.
- **Level 4 – Self-refinement:** Agentic automation that is capable of meaningful self-improvement with or without human collaboration. It can examine and modify its instructions and learning data, create new tools, and connect to new data sources.
- **Level 5 – Autonomy:** Represents what many consider Artificial General Intelligence (AGI). These hypothetical agents exhibit original thinking and can synthesize solutions to previously unseen tasks.

The current SMG use cases for pilot projects range from Level 1 to Level 3, with our goal to progress to Levels 4 and 5 as agentic AI tooling continues to mature.

[4] Source: https://sema4.ai/

Solution Architecture

Key Design Goals

The following three goals guided our design of the 1AI platform:

- **Promote interoperability of AI agents.** Agents should be accessible from multiple vendor-based and open-source copilots or frontend chatbots.
- **Manage costs.** Agents should use the right LLM for the right job. Choices include private frontier LLMs and open-source LLMs; vendor-specific, subscription-based copilots and chatbots; and lower-cost open-source chatbot interfaces.
- **Democratize agent development.** Build agent interfaces compliant with common specifications so that multiple teams can build agents independently and apply their domain expertise in business processes, data, and development tools.

Deployment Details

The 1AI platform is developed according to OpenAPI 3.0 standards, and includes integration with existing platforms and tools. This helps ensure a cohesive approach to AI deployment across the organization. As newer standards like Model Context Protocol (MCP) mature and become integral to software development, transitioning to these standards will be inevitable. Figure 4 illustrates the solution architecture, which is purposefully designed to scale, because AI agents are deployed using Kubernetes on Intel IT's container-as-a-service (CaaS) hosting platform. This section describes the technical details of the AI agent deployment, as well as discusses security/compliance and continuous improvement considerations.

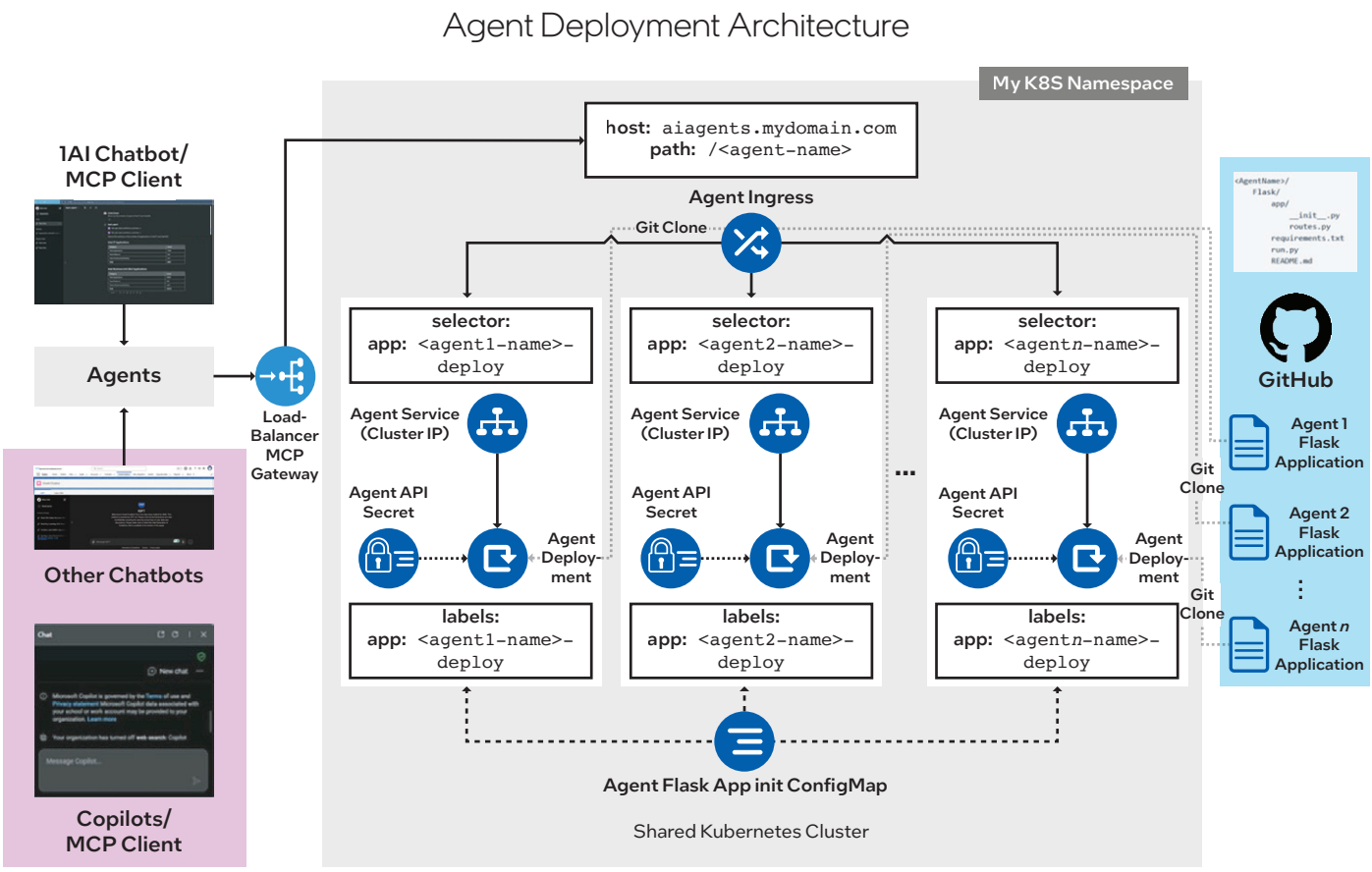

Figure 4. AI agent deployment solution architecture.

Agent Workflow

- Each AI agent is created as a Python Flask application (referred to as "the application" in the remainder of this list) and runs as a Kubernetes deployment.
- The secrets used to access backend REST APIs and services are completely decoupled from the code. The code resides in a GitHub repository, while secrets are deployed as Kubernetes Secrets and exposed to deployments via environment variables.
- Each AI agent deployment is exposed through a Kubernetes Service that uses the deployment's label as the selector. This enables the service to discover all running replicas of the deployment and load-balance incoming requests across them.
- A single ingress resource exposes all AI agents/services deployed in the namespace using a single domain and path-based routing. The domain, also referred to as the hostname (such as `aiagents.mydomain.com`), is a CNAME to the load-balancer's virtual IP address (VIP) in front of the Kubernetes cluster. The domain must be added to the list of allowed domains in the frontend configuration. The path (such as `/agent1` in `https://aiagents.mydomain.com/agent1`) is used to identify the AI agent/service that needs to respond to the request and is removed before the request is forwarded to the appropriate agent/service.
- A config map containing an initialization bash script is mounted as a volume inside the Kubernetes Pods created by the deployment. The bash script is executed at container startup and performs the following tasks:
  - Installs the required Linux packages and internal SSL certificate chain (required for SSL validation while accessing internal REST API endpoints)
  - Clones the code for the AI agent from the GitHub repository
  - Creates a Python virtual environment for the Flask application
  - Installs Python packages required by the application
  - Runs the application
- A standardized folder structure across AI agents in the GitHub repository helps us maintain a low-code approach and allows variables to be reused. For example, an environment variable, `ai_agent`, defined by each Kubernetes deployment is used by the initialization config map/bash script to identify the AI agent and GitHub repository folder that need to be cloned. The number of replicas can be scaled up or down as required.
- The Kubernetes Pods created by the deployment use startup, readiness, and liveness probes to indicate their availability to serve incoming requests. The corresponding Kubernetes Services use these probes to make sure that requests are sent only to the Pods that are in ready state.

Note: It is also recommended to store API endpoints as Kubernetes config maps/secrets to make it easy to switch across environments without modifying code.

User Scenario

Here's how a typical user interaction with the 1AI platform occurs:

- A user goes to the web interface (see Figure 3), selects a category, and enters a question.
- The user's question is routed to the orchestrator agent with access to a semantic layer, which breaks down and evaluates the intent of the user's question.
- The system determines which tools or data sources are needed and dynamically executes a given task. The system may also decide to reach out to other LLMs as needed.
- The user receives the results and, if necessary, interacts by using conversational chat to refine the results.
- Each step in the chat process either dynamically fetches new data or provides responses based on the already-retrieved data.

Security and Compliance

The 1AI platform helps protect Intel's sensitive data by using access controls to enforce entitlement. Users can access only the information that they are entitled to see. We worked closely with cross-functional teams to verify that the platform meets all of Intel's security and compliance requirements. For example, we developed a GenAI disclaimer in concert with the Legal department, and the platform successfully passed the responsible AI, cyber risk, and privacy reviews.

Observability and Evaluation of Agents

As AI agents become increasingly complex and mission-critical, the need for comprehensive observability and evaluation frameworks has become paramount. We need detailed visibility into agent behavior, performance metrics, and quality assessments to help ensure reliable deployment and continuous optimization.

Tracing Framework

Our tracing system provides end-to-end visibility into LLM interactions through detailed capture of the following:

- **Input/output tracking.** Complete recording of all inputs and corresponding outputs
- **Tool usage monitoring.** Usage tracking across different agents
- **Performance metrics and cost attribution.** Tracking of crucial metrics such as overall usage volume, usage by model or token types, cost breakdowns (for example, by user), latency distributions, and quality metrics

Automated Testing Framework

We have developed a testing framework with the following features:

- Comprehensive test case coverage for each agent
- Automated execution triggered by code or prompt modifications
- Continuous integration pipeline

We also plan to implement an LLM-as-a-Judge evaluation system, which includes an automated quality assessment that uses advanced LLM evaluators with consistent evaluation criteria.

User Feedback Integration

We have integrated a comprehensive user feedback system that captures direct user input through like/dislike mechanisms and detailed comments to enable continuous improvement based on user insights.

Business Benefits

Our 1AI platform provides a strategic blueprint for unified GenAI adoption. Its modular, agent-based architecture delivers agility, efficiency, and cost optimization, which enables us to rapidly deploy new AI use cases without rework and provides users with a one-stop-shop for all of their corporate chat solutions. In addition, it supports governance, security, and compliance at scale. Some of the tangible evidence of business impact includes the following:

- Our AI agents reduce technical support response times from hours to seconds, boosting employee productivity. We are currently serving over 2,000 users.
- By using the more efficient LLMs at the right places for each agent, we achieve optimal response as well as lower cost token payments.
- Centralization eliminates redundant licenses and support, further reducing costs, training, and user change management.
- Use-case-specific agents enable agents to interact with each other to obtain the best results for the user's inquiry.

More qualitative improvements include a consistent GenAI UX, automated repetitive workflows that free employees for higher-value work, reliable local access to AI even without consistent cloud access, and shortened sales cycles as well as more productive teams, all of which can lead to new business opportunities and improved top-line results. Customers may also benefit, because when employees can provide fast, accurate responses, it strengthens customer trust and satisfaction.

Our choice of architecture—including Intel Xeon processors and Intel Gaudi AI accelerators, open frameworks, and optimized software—delivers reliable, scalable AI performance and fast, highly secure deployment.

Conclusion

Scaling AI agents and strategic initiatives is crucial for future growth. To achieve long-term objectives, we plan to expand our AI capabilities and enhance collaboration across BUs. We will continue to develop use case-specific agents that will adhere to each BU's workflows, such as new agents that can assist sellers in every aspect of customer engagement and support, and agents that can assist the Legal department in preparing and reviewing contracts. This approach facilitates an efficient and adaptable process of continuous improvement and refinement at each stage of deployment and scaling.

Our 1AI platform is a unified, agent-based GenAI chatbot platform that enables the entire enterprise to leverage the most efficient AI for every use case—driving scalability, efficiency, strong governance, and faster innovation across the business. Our close partnership with business goals and processes helps ensure that our investments are strategic and our innovation is impactful. The deployment of the platform for SMG proves that 1AI can empower BUs to future-proof their AI strategy with a flexible, centrally managed platform that matches the best AI models to every business need.

Our work with the 1AI platform enables the enterprise to rapidly scale innovation, increase efficiency, and deliver robust security and governance. The platform's open, Intel® technology-based architecture lets us adopt the latest AI models and methods as they emerge. Through the IT@Intel program, we would be happy to respond to requests for consultation or pilots and explore how other enterprises can leverage our work to meet their unique business needs.

Key Learnings

While developing the OneAI (1AI) platform and tuning it for the Sales and Marketing group's (SMG's) needs, we learned some key lessons:

- Being agile is critical, because business needs and technology change quickly.
- It's best to maintain an entrepreneurial, startup mentality so that we can **fail fast** and pivot quickly when we run into technology hurdles.
- A consistent, performant user experience (UX) helps ensure adoption of the new platform. To measure UX, we defined minimum viable key performance indicator limits for AI agent response performance.
- Perfection is the enemy of the good. That is, don't wait for perfection, but take an iterative approach, solving one or two problems at a time.
- We need to understand agentic task types and choose the right compute infrastructure, LLMs, and frontend chatbots to manage costs.
- Future agents (vendor or privately developed) should be interoperable.

Related Content

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

- IT@Intel: Preparing Our PC Fleet for the Future of AI
- IT@Intel: Democratizing the Use and Development of Generative AI Across Intel
- IT@Intel: Improving Sales Account Coverage with Artificial Intelligence

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.

Visit us today at **intel.com/IT** or contact your local Intel representative if you would like to learn more.