ManTech
Securing the Future

2023 Generative AI LLMs

Enhancing Content Understanding and User Experience Across the Enterprise With Generative AI LLMs

A ManTech White Paper

Enhancing Content Understanding and User Experience Across the Enterprise With Generative AI LLMs

Introduction

The use of large language model (LLM) artificial intelligence/machine learning (AI/ML), such as that underlying ChatGPT, is poised to significantly impact and alter the enterprise, changing the skillsets needed to deploy and manage large, complex environments. Generative Pre-trained Transformer (GPT)-based AI is bringing AI to the enterprise—whether it is ready or not—leading to critical questions about how to best leverage its capabilities, the overall reliability of self-learning models, and integration complexity. Model governance will become a major area of management, as opening uncontrolled and unfettered access to GPT products presents problems of cost management, proprietary data exfiltration, and inappropriate use.

This white paper presents ManTech's new GPT AI platform, ManTech Enterprise Smart Assistant (MESA).

Problem Statement

Most organizations have a large corpus of knowledge—including business processes, procedures, policies, technical documentation, and financial information—locked up in documentation. Extracting insights from this content presents a complex challenge, requiring the ability to identify, organize, and analyze large amounts of data to uncover meaningful patterns and trends. This is often a difficult task, as the data may be spread across multiple sources, formats, and systems and may be unstructured, which hinders interpretation and analysis. The sheer volume of content can be overwhelming, making it difficult to identify the most relevant information.

Safely Introducing Generative Ai To The Enterprise

Generative AI aids in generating insights from large amounts of enterprise knowledge content and documentation. It uses natural language processing (NLP) to analyze and interpret the content and then generate insights based on the data, identifying patterns and trends in the data and uncovering hidden relationships between different pieces of information. Generative AI can also be used to generate summaries of the content, making it easier to quickly identify the most relevant information. Generative AI is an emerging technology that must be implemented safely, ensuring data security as well as accounting for human factors such as cultural shifts in adoption. This is where MESA can help.

MESA brings generative AI to the enterprise in support of several key areas, including Agile/DevSecOps projects, enterprise IT service management (ITSM), and content management. MESA has been integrated with ServiceNow to bring its capabilities to an enterprise-wide audience. This integration allows ServiceNow to use MESA's NLP capabilities to automate customer service tasks, reducing the time and effort needed to respond to customer inquiries. It provides NLP on text content, such as knowledge bases, allowing sophisticated queries and responses. MESA delivers personalized end-user service desk experiences by understanding user intent and providing tailored responses, which improves end-user satisfaction and loyalty.

MESA is a custom ManTech-developed platform built on an extensible architecture that allows the use of pluggable tools and services and different GPT AI providers, and uses Azure OpenAI's GPT AI engine as well as custom AI models. In addition to the ServiceNow integration, it provides out-of-the-box integrations with Git and Jira, along with comprehensive application programming interfaces (APIs) for custom integrations with existing tools and services. The platform's AI-based capabilities include code generation and explanations, IT artifact generation, dynamic service catalogs, and intelligent end-user assistance. These capabilities allow teams to efficiently apply the latest changes in machine learning (ML), security, and infrastructure automation while ensuring compliance with development standards and policies.

Large Language Models

LLMs are evolving rapidly. The introduction of GPT-3.5, followed soon after by GPT-4, and Google's Pathways Language Model (PaLM) API are upending the AI space. The rapid rise and adoption of LLMs is why ManTech architected MESA to remain LLM-agnostic. It is also why we decouple training data from the LLM to allow for an easy upgrade path—as the LLMs improve, they can reuse all of the prior work done to generate text embedding values. MESA provides support across the enterprise, expanding AI capabilities to the service desk, development teams, and operations.

Knowledge Management

Knowledge management is the process of capturing, organizing, and sharing knowledge within an organization, typically in a tool such as ServiceNow. Having a robust knowledge management system harnessed with Generative AI vastly improves organizational performance by unlocking the collective knowledge of an organization through plain language inquiries.

One of the key features of the new LLM GPT AI models is the ability to statistically interpret content and inquiries. Content can include instructions, project documentation, policy documents, knowledge base articles, etc. MESA's approach for contextual and advanced document searches is to ingest documents, categorize them, and catalog them by their contents. These contents run through a text embedding technique that calculates the data as numerical vectors, and then those values are stored in MESA's Redis database. Multiple OpenAI models are able to use these values, providing the flexibility of upgrading or downgrading the models deployed based on use case demand. MESA uses techniques, such as cosine similarity, to identify what content is relevant to a user's inquiry and then sends that data and the inquiry to the GPT engine for resolution.

To support the above process, MESA has a series of tools, including an ETL processor that parses and organizes content using the aforementioned text embedding. Ingestible content includes knowledge base articles exported from ServiceNow, PDFs, Microsoft Word documents, text files, etc. This content is used in inquiries, searches, and questions. Rather than store information in the GPT server, MESA keeps the content, embeddings, and other metadata in a local store for increased security and portability. This content store runs on-premises or in the cloud, giving an organization complete control of its data.

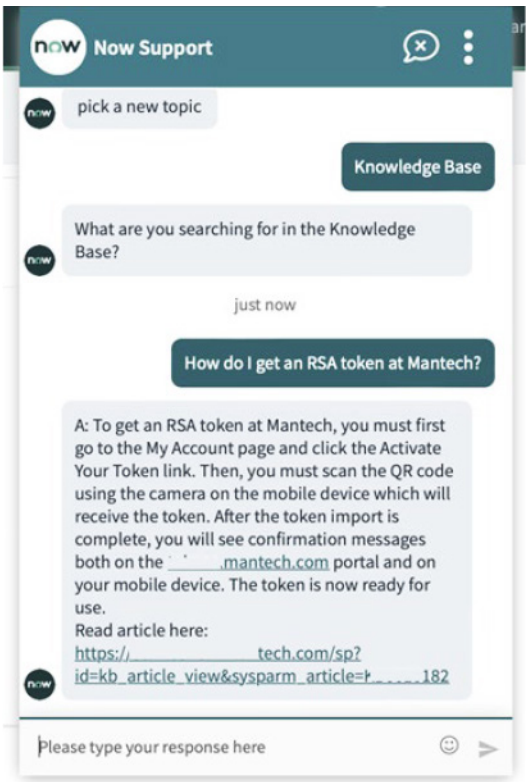


Figure 1. MESA-Powered Virtual Agent

By controlling the content sent to the GPT models, an organization is able to answer inquiries using their specific data rather than the general data found on the Internet. For example, an organization has a policy on traveling overseas with an organization-issued laptop. Searching for this topic on the open Internet—for example, through ChatGPT—will result in a generic answer instead of one compliant with the organization's actual policy. However, using MESA will deliver an organization-specific answer because MESA only uses policy content.

To make MESA's capabilities available to all end users of the service desk, our ServiceNow solution incorporates MESA into its Virtual Agent chatbot, shown in Figure 1. By integrating MESA as a topic block, end users are able to seamlessly leverage its advanced capabilities within the Service Portal to query the knowledge base, create tickets, and seek general assistance for work-related queries. The MESA-powered Virtual Agent messaging interface exposes users to world-class AI while allowing the enterprise to control and audit AI interactions. Additional MESA capabilities for advanced users are available through interfaces in ServiceNow, as is a Representational State Transfer (REST) API for custom integrations.

Visibility

Providing AI/ML model use visibility on a dashboard is important because it allows users to quickly and easily understand the performance of the model, see who is using the model and why, and view graphic depictions of associated costs. Because models are constantly being retrained and fine-tuned, the dashboard helps users identify any potential issues or areas of improvement that need to be addressed. By providing visibility into the model, users gain insights into how the model is performing and can make informed, data-driven decisions for improvement as well as identify any potential biases or errors in the model, which aids them in evaluating model usage.

TO PROVIDE THIS LEVEL OF INSIGHT, MESA HAS A DASHBOARD IN SERVICENOW AS PART OF OUR TOTAL VISIBILITY CENTER, SHOWN IN FIGURE 2.

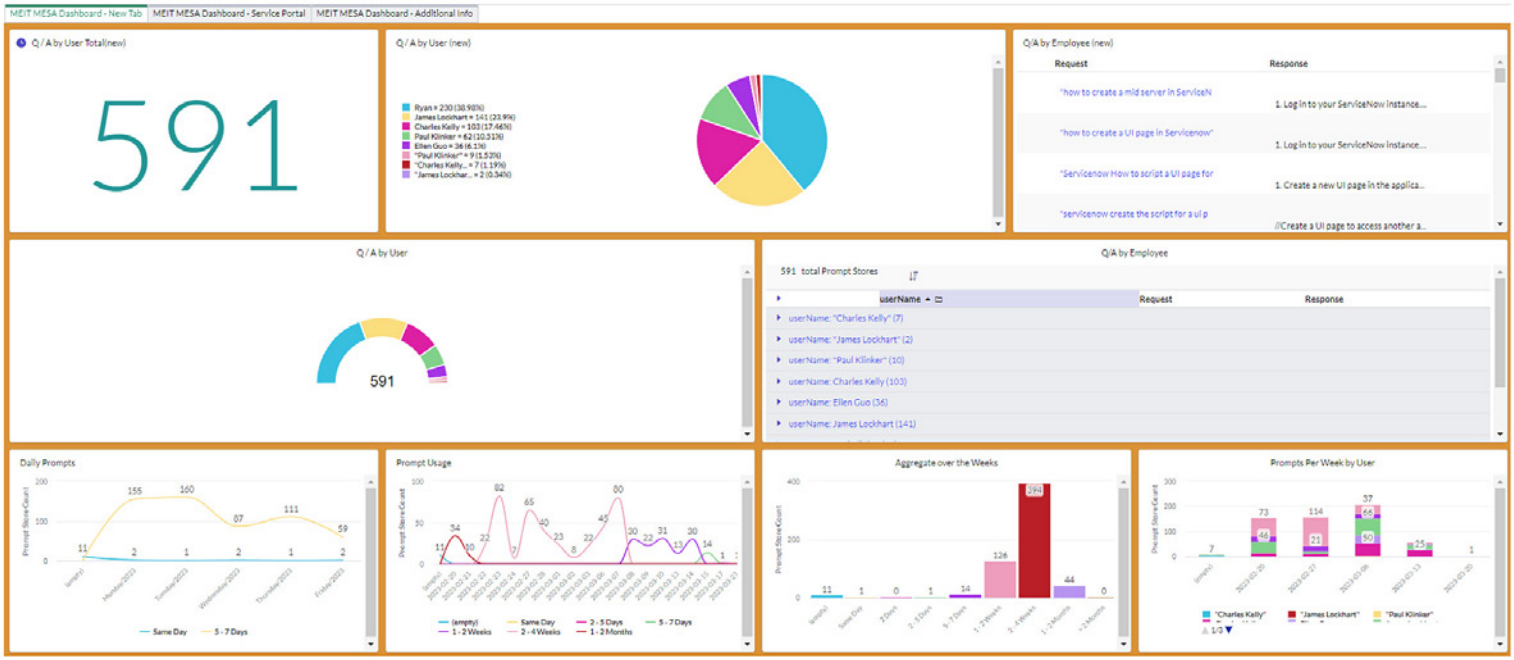


Figure 2. MESA ML Total Visibility Center

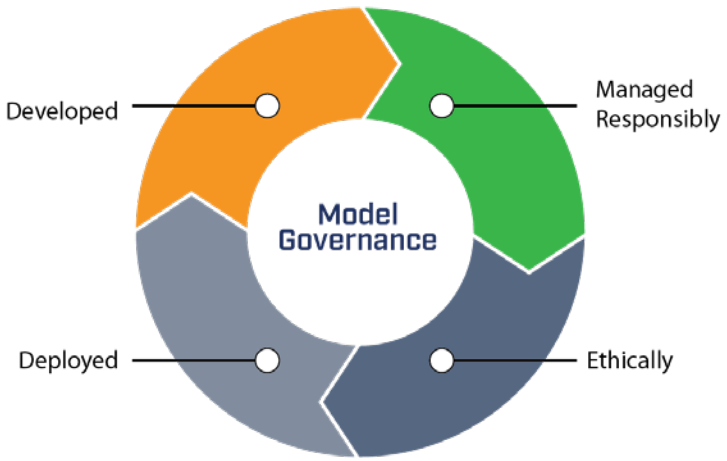
As shown in the above screenshot, MESA's dashboard provides information and metrics on who is using the ML services, the number of inquiries by user, query trends over time, and more. For users who do not use ServiceNow, MESA provides a rich set of REST APIs and data metrics to enable custom reporting and custom dashboards in their tool(s) of choice.

Model Governance

ML model governance is a fundamental part of any organization's data and policies. It is the process of ensuring that models are developed, deployed, and managed responsibly and ethically. This includes developing models with appropriate data; testing, validating, and deploying models in a secure and compliant manner; monitoring and updating models regularly; and documenting any changes to models and communicating those changes to stakeholders. To meet the needs of the Federal Government, model governance should include policies and procedures for data privacy, security, and compliance.

To simplify governance, MESA allows organizations to centrally control access to GPT-based technologies. By centralizing access to services such as Azure OpenAI, MESA ensures that:

- Only authorized personnel access the platform and all user activity on the platform is tracked and logged, which helps maintain regulatory compliance as well as provides visibility into how OpenAI is being used.
- Only approved and tested models are used, reducing the risk of introducing malicious or faulty models into production.
- New models and features are deployed expeditiously, without compromising innovation.



Architectural Overview

MESA is a containerized platform providing rich REST API and Kafka topics for custom integrations. MESA's internal components include an authentication and authorization module built on Keycloak, MongoDB for persistence, and a user interface for end user and administrative interaction. MESA can be integrated into a complex enterprise or as a brand-new implementation, based on the infrastructure and existing tools. A high-level depiction of the MESA platform architecture and current set of integrated tools is shown in Figure 3.

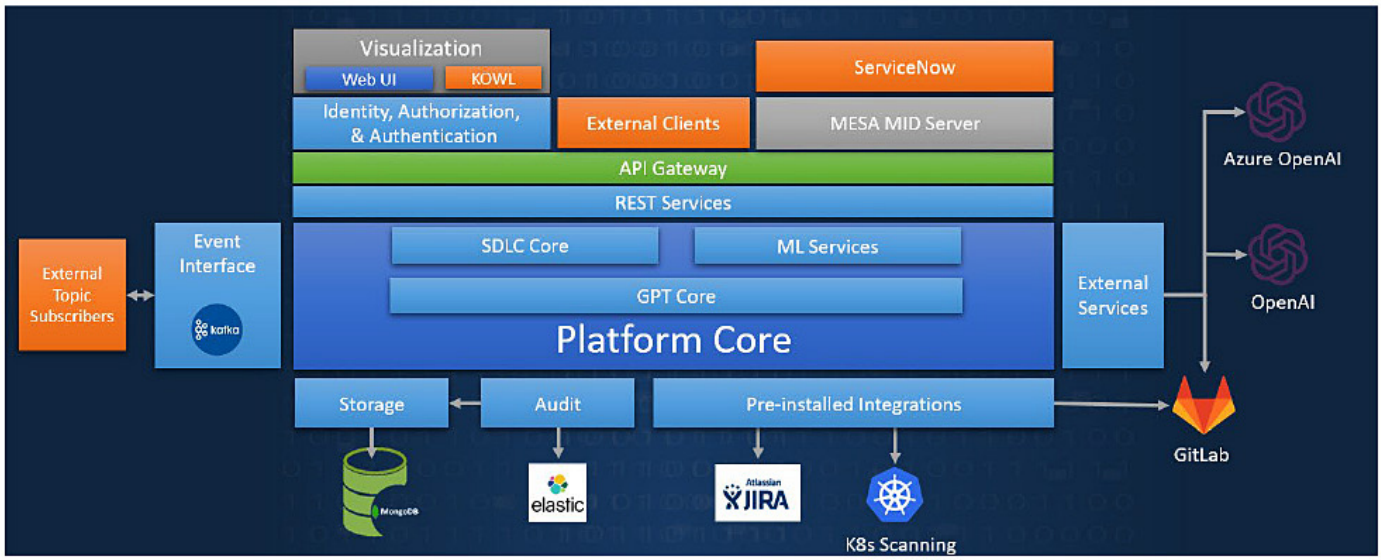


Figure 3. MESA High-level Architecture

real power of MESA is its ability to tap into external products. The number of enterprise tools used by organizations seems boundless—many products are common or de facto standards, and users have their products of choice. MESA is architected to work with a wide variety of tools and provides a suite of integrations out of the box. MESA's integrated tools provide the following capabilities.

- AI-enhanced content and knowledge management interactions
- Code explainer
- Container and Kubernetes YAML scanning
- Git repository interface
- Search and analytics on software development life cycle (SDLC) events
- Virtual agent support
- Visual data routing, transformation, and system mediation

Conclusion

MESA enables the enterprise to surface insights from knowledge captured within its systems by leveraging natural language interpretation with large-scale GPT AI models. These models have reasoning and comprehension capabilities that allow them to capture semantic similarity in text during text and code searches and can be customized to answer specific questions, summarize large amounts of information, or generate original content such policy and process documentation.

To learn how MESA can bring the benefits of generative AI to your organization, please reach us at:

ContactUs@ManTech.com