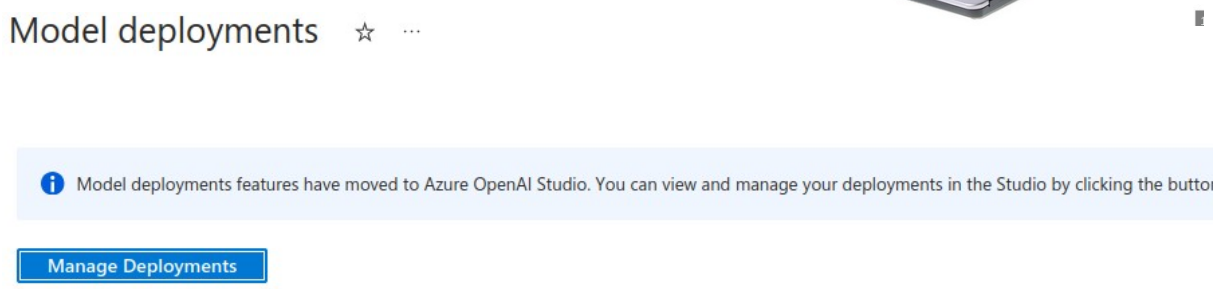
Authentication

Azure OpenAI provides two methods for authentication. You can use either API Keys or Microsoft Entra ID.

API Key authentication: For this type of authentication, all API requests must include the API Key in the api-key HTTP header. The Quickstart provides guidance for how to make calls with this type of authentication.

Microsoft Entra ID authentication: You can authenticate an API call using a Microsoft Entra token. Authentication tokens are included in a request as the Authorization header. The token provided must be preceded by Bearer, for example Bearer YOUR_AUTH_TOKEN. You can read our how-to guide on authenticating with Microsoft Entra ID.



Previously, building custom AI assistants needed heavy lifting even for experienced developers. While the chat completions API is lightweight and powerful, it's inherently stateless, which means that developers had to manage conversation state and chat threads, tool integrations, retrieval documents and indexes, and execute code manually.

The Assistants API, as the stateful evolution of the chat completion API, provides a solution for these challenges. Assistants API supports persistent automatically managed threads. This means that as a developer you no longer need to develop conversation state management systems and work around a model's context window constraints.

Currently, assistants, threads, messages, and files created for Assistants are scoped at the Azure OpenAI resource level. Therefore, anyone with access to the Azure OpenAI resource or API key access is able to read/write assistants, threads, messages, and files.

We strongly recommend the following data access controls:

Implement authorization. Before performing reads or writes on assistants, threads, messages, and files, ensure that the end-user is authorized to do so.

Restrict Azure OpenAI resource and API key access. Carefully consider who has access to Azure OpenAI resources where assistants are being used and associated API keys.

Routinely audit which accounts/individuals have access to the Azure OpenAI resource. API keys and resource level access enable a wide range of operations including reading and modifying messages and files.

Enable diagnostic settings to allow long-term tracking of certain aspects of the Azure OpenAI resource's activity log.