

AZURE DAY ROME 2025



Tieni sotto controllo i servizi AI su Azure con Azure Monitor











Thanks to





















Tieni sotto controllo i servizi Al su Azure con Azure Monitor

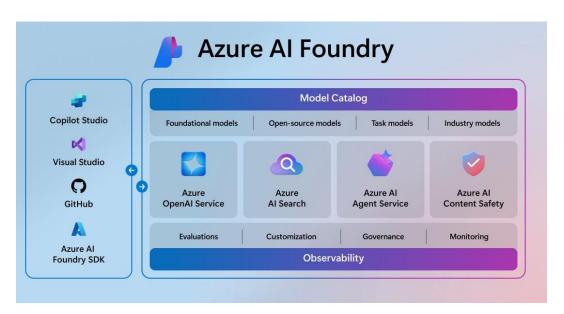


Azure Al Foundry (o Azure Foundry Open Al)

Entrambi rappresentano il giusto punto di ingresso nel mondo Al su Azure.

Foundry è il framework suggerito da Microsoft per mettere provare e sperimentare con i vari modelli (e con Agenti e pipeline ML)

La differenza principale è che Foundry è agnostico, mentre Foundry OpenAl è specializzato per progetti GPT su Azure PaaS.





Azure Monitor

- Entry point per il mondo LOG all'interno di Azure
- Permette un'analisi molto più approfondita grazie a KQL
- Lo strumento che andremo ad usare sarà Log Analytics Workspace
- ATTENZIONE: è un costo extra l'ingestion dei log





Primo sguardo: Dashboard

• La dashboard inclusa in entrambi permette di avere uno sguardo molto approfondito delle metriche essenziali quali

Monitor usage and performance for your application

- Tempi di risposta
- Token utilizzati
- Errori
- Utilizzo dei modelli

Resource usage Model deployment: gpt-4o ∨ () Refresh ⊞ 06/13/2025 - 06/20/2025 Last day 7D 1M Summary Total requests Total token count Prompt token count Completion token count 4.87K 98 4.77K 1622 avg per request 33 avg per request 1589 avg per request **Usage metrics** Input vs Output vs Total Number of requests 1800 1200 14. Jun 16. Jun 20. Jun Time (Local)



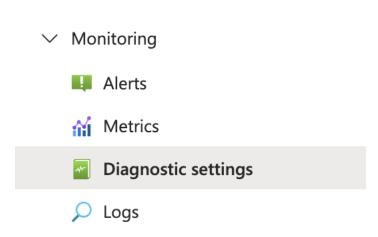
Aggiungiamo del logging esteso...

Oltre alle funzionalità base, Azure Foundry permette di aggiungere anche il logging su vari accentratori

- Log Analytics Workspace
- Storage accounts
- Event Hub
- Software di terze parti

Utilissimo per

- Analisi approfondite
- Alerting





E andiamo in profondità!

Query KQL (Kusto Query Language)

- AzureDiagnostics
- AzureMetrics

AzureDiagnostics |where Category == "RequestResponse"

ATTENZIONE: non tutti i modelli loggano i medesimi campi e i log RequestResponse appaiono solo con chiamate REST al momento

<u>ATTENZIONE2:</u> per ottenere log completi delle request fatte serve logging applicativo





Demo time



I costi su Azure Al Foundry

- Come possiamo controllare la nostra spesa su Azure AI Foundry (e la sua parte Open AI)?
- Cost Analysis
 - Cognitive Services
 - By Meters
- Usiamo i tag!





Attenzione alla spesa

- Il primo step è SEMPRE pianificare e testare, con
 - Playground
 - Azure Pricing Calculator
 - Tenere presente gli errori
- Utilizzare gli alert di spesa
- Utilizzare gli alert sui token
- Impostare budget
- Ricordarsi tutto il «contorno» (storage account, apim, etc etc...)







Demo time



Riferimenti

- https://learn.microsoft.com/en-us/azure/ai-services/openai/howto/manage-costs
- https://learn.microsoft.com/en-us/azure/ai-foundry/how-to/develop/trace-application
- https://learn.microsoft.com/en-us/azure/ai-foundry/how-to/costs-plan-manage
- https://techcommunity.microsoft.com/blog/finopsblog/managing-azure-openai-costs/4286830
- https://learn.microsoft.com/en-us/azure/ai-services/openai/how-to/monitor-openai



Questions & Discussion

Let's keep in touch







Please Vote for This Session





THANK YOU!!!