

Microsoft Build 2024 Recap

Virtual Boston Azure
June 3, 2024

Bill Wilder, Veronika Kolesnikova and Jason Haley

Jason's Favorites

Five lessons Etsy learned building on Microsoft AI

<https://build.microsoft.com/en-US/sessions/ae85ca15-f1e5-4196-b397-48c1678800bb?source=/favorites>

Lessons:

1. Use the right platform for the job
2. Learn by building it
3. Always optimize for impact
4. Constraints Matter
5. It's a journey, not a destination

From the Book of News:

Browse Credentials | Microsoft Learn

New Applied Skills credentials that support developers who build AI and cloud solutions include:


- Accelerated app development using GitHub Copilot.
- Developing AI agents using Microsoft Azure OpenAI Service and Semantic Kernel.
- Automated Microsoft Azure Load Testing using GitHub.

The current portfolio of Microsoft credentials, which has more than 20 Microsoft Applied Skills and 50 Microsoft Certifications in solution areas like Data and AI, Infrastructure, Digital and App innovation, Business apps, Modern Work and Security, is now enhanced by these new credentials. Certifications offer the flexibility to grow the skills needed for critical roles and Applied Skills offer the ability to expand the skills needed for key business scenarios. Together they bring verifiable skill sets aligned to AI and cloud job roles and projects. Explore all [AI Microsoft Credentials](#).

Learning Path: Develop generative AI solutions with Azure OpenAI Service

- <https://learn.microsoft.com/en-us/credentials/applied-skills/develop-generative-ai-solutions-with-azure-openai-service/>








Prepare for the assessment



LEARNING PATH
Develop Generative AI solutions with Azure OpenAI Service
5 hr 34 min • 7 modules

[Continue learning path >](#)

Modules in this learning path

-  **Get started with Azure OpenAI Service**
43 min • Module • 10 units
-  **Build natural language solutions with Azure OpenAI Service**
57 min • Module • 7 units
-  **Apply prompt engineering with Azure OpenAI Service**
1 hr 3 min • Module • 7 units
-  **Generate code with Azure OpenAI Service**
50 min • Module • 7 units
-  **Generate images with Azure OpenAI Service**
38 min • Module • 7 units
-  **Implement Retrieval Augmented Generation (RAG) with Azure OpenAI Service**
33 min • Module • 7 units
-  **Fundamentals of Responsible Generative AI**
50 min • Module • 9 units

[Add](#)

RAG at scale: production-ready GenAI apps with Azure AI Search

<https://build.microsoft.com/en-US/sessions/ff81130c-db16-4120-bc08-2061f911e6e1?source=/favorites>

Mostly about Azure AI Search enhancements

- New Expanded Capacity
- Quantization
- Binary Quantization
- AI Studio Model Catalog for Embeddings

New Expanded Capacity

SKU	Previous vector limit	New vector limit	Previous storage limit	New storage limit	Previous max service size	New max Service size
Basic	1	5 (5x)	2	15 (7.5x)	2 GB	45 GB
S1	3	35 (11.5x)	25	160 (6.4x)	300 GB	1.88 TB
S2	12	150 (12.5x)	100	512 (5x)	1.2 TB	6 TB
S3	36	300 (8.3x)	200	1024 (5x)	2.4 TB	12 TB
L1	12	150 (12.5x)	1024	2048 (2x)	12 TB	24 TB
L2	36	300 (8.3x)	2048	4096 (2x)	24 TB	48 TB

→ Basic SKU now supports up to 3 partitions, up from a single partition
→ Limits are GB/partition

Take an Azure OpenAI chat application from PoC to enterprise-ready

<https://build.microsoft.com/en-US/sessions/9babe697-a4be-4a86-9995-f34c08e1d06e?source=sessions>

- Resources:

- <https://github.com/microsoft/rag-experiment-accelerator>
- <https://learn.microsoft.com/en-us/azure/architecture/ai-ml/guide/rag/rag-solution-design-and-evaluation-guide>

Veronika's Favorites

Phi-3 and Small Language Models

The screenshot displays the Azure AI Studio interface for the Phi-3-mini-128k-instruct model. The browser address bar shows the URL: `ai.azure.com/explore/models/Phi-3-mini-128k-instruct/version/5/registry/azureml?tid=17c5bbab-ad62-470e-aef3-890fb4b9b84c`. The page title is "Phi-3-mini-128k-instruct" with a "PREVIEW" badge. The navigation menu on the left includes "Home", "Model catalog" (selected), "Model benchmarks", "Prompt catalog", "Azure OpenAI", "AI Services", "All hubs", "Resources and keys", and "Quota". The main content area features tabs for "Overview", "Versions", "Artifacts", and "Security". Below the tabs, there are filters for "Task: Chat completion", "Fine-tuning task: chat-completion", "Languages: en", and "License: mit". Action buttons include "Refresh", "Fine-tune", "Deploy", and "View license". The "Description" section is titled "Model Details" and contains two paragraphs: "Phi-3 Mini-128K-Instruct is a 3.8B parameters, lightweight, state-of-the-art open model built upon datasets used for Phi-2 - synthetic data and filtered websites - with a focus on very high-quality, reasoning dense data. The model belongs to the Phi-3 model family, and the Mini version comes in two variants 4K and 128K which is the context length (in tokens) it can support." and "The model underwent a rigorous enhancement process, incorporating both supervised fine-tuning and direct preference optimization to ensure precise instruction adherence and robust safety measures. When assessed against benchmarks testing common sense, language understanding, math, code, long context and logical reasoning, Phi-3 Mini-128K-Instruct showcased a robust and state-of-the-art performance among models with less than 13 billion parameters." Below the description is a link for "Resources and Technical Documentation:". On the right, there is a "Try it out" section with a settings gear icon and a text input field with the placeholder "Start typing here".

Phi-3 Cookbook

The screenshot shows the GitHub repository page for 'Phi-3CookBook' by Microsoft. The browser's address bar shows the URL 'github.com/microsoft/Phi-3CookBook'. The repository name is 'Phi-3CookBook' and it is marked as 'Public'. It has 5 watchers, 55 forks, and 777 stars. The repository is on the 'main' branch with 1 branch and 0 tags. A search bar is present with the text 'Go to file'. There are buttons for 'Add file' and 'Code'. The commit history shows a recent commit by 'kinfey' updating the README.md file 18 hours ago, with 115 total commits. A list of files and folders is shown, including .vscode, code, imgs, md, .gitignore, CODE_OF_CONDUCT.md, LICENSE, README.md, and SECURITY.md. The 'About' section on the right provides a description of the book and lists various files like Readme, MIT license, Code of conduct, Security policy, Activity, Custom properties, and 777 stars.

github.com/microsoft/Phi-3CookBook

microsoft / Phi-3CookBook

Phi-3CookBook Public

Watch 5 Fork 55 Star 777

main 1 Branch 0 Tags

Go to file Add file Code

kinfey Update README.md 4bdebf4 · 18 hours ago 115 Commits

.vscode	Update Azure AI Studio Content	3 weeks ago
code	Update generate_dataset.py	3 days ago
imgs	update	4 days ago
md	Update Phi-3 Vision Inference content	18 hours ago
.gitignore	update settings	2 weeks ago
CODE_OF_CONDUCT.md	CODE_OF_CONDUCT.md committed	last month
LICENSE	LICENSE committed	last month
README.md	Update README.md	18 hours ago
SECURITY.md	SECURITY.md committed	last month

About

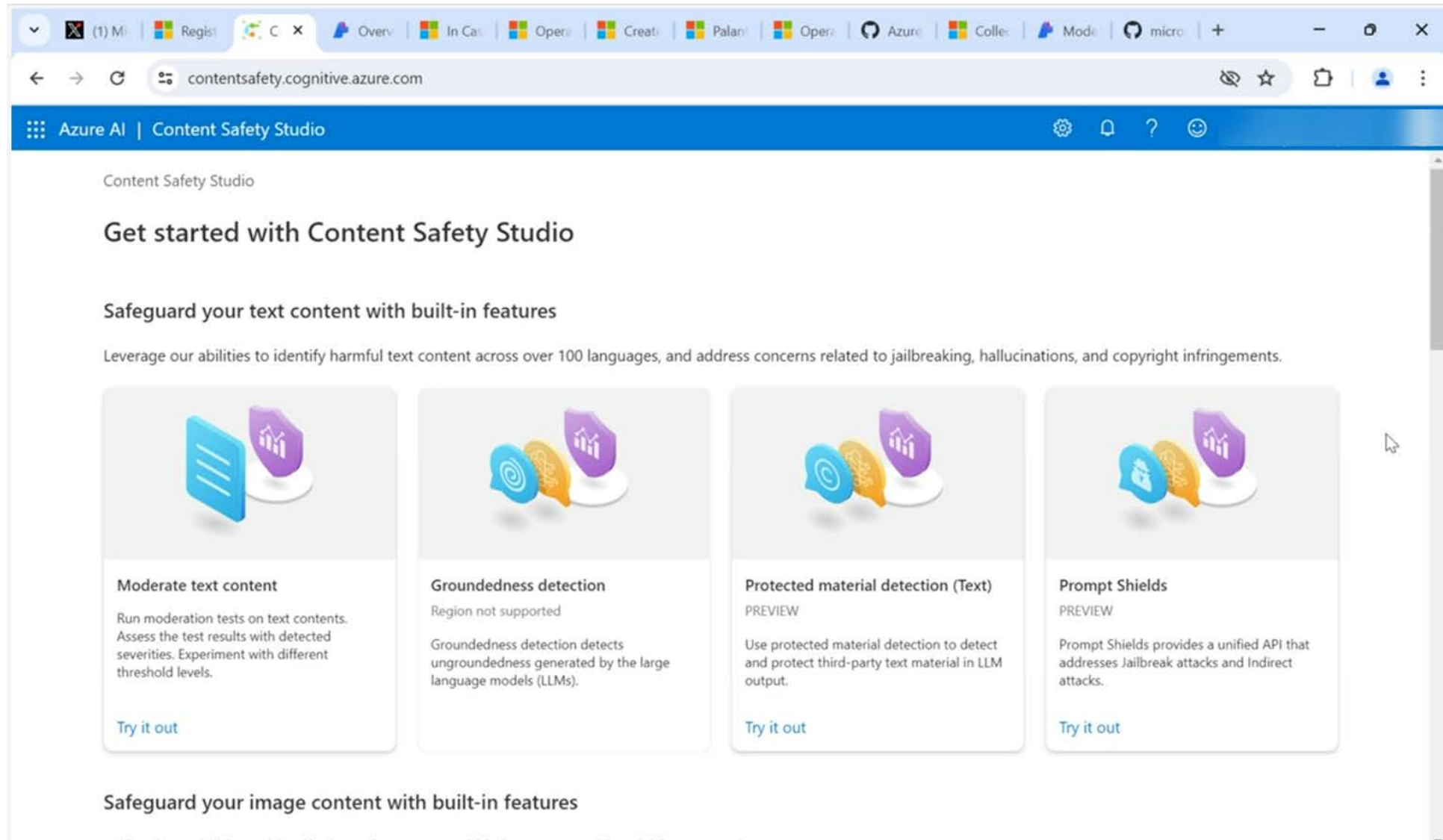
This is a Phi-3 book for getting started with Phi-3. Phi-3, a family of open AI models developed by Microsoft. Phi-3 models are the most capable and cost-effective small language models (SLMs) available, outperforming models of the same size and next size up across a variety of language, reasoning, coding, and math benchmarks.

- Readme
- MIT license
- Code of conduct
- Security policy
- Activity
- Custom properties
- 777 stars

Phi-3 Session

- Create Generative AI experiences using Phi
 - <https://build.microsoft.com/en-US/sessions/e6d21a49-2efb-4a39-8c26-f6eef1410c7a?source=sessions>

Responsible AI and Content Safety



The screenshot shows the Azure AI Content Safety Studio interface. The browser address bar displays 'contentsafety.cognitive.azure.com'. The page title is 'Content Safety Studio'. The main heading is 'Get started with Content Safety Studio'. Below this, a sub-heading reads 'Safeguard your text content with built-in features'. A descriptive paragraph states: 'Leverage our abilities to identify harmful text content across over 100 languages, and address concerns related to jailbreaking, hallucinations, and copyright infringements.' There are four feature cards, each with an icon and a 'Try it out' link:

- Moderate text content**: Run moderation tests on text contents. Assess the test results with detected severities. Experiment with different threshold levels.
- Groundedness detection**: Region not supported. Groundedness detection detects ungroundedness generated by the large language models (LLMs).
- Protected material detection (Text)**: PREVIEW. Use protected material detection to detect and protect third-party text material in LLM output.
- Prompt Shields**: PREVIEW. Prompt Shields provides a unified API that addresses Jailbreak attacks and Indirect attacks.

Below the feature cards, another sub-heading reads 'Safeguard your image content with built-in features'.

Responsible AI and Content Safety (contd)

- Sessions by Sara Bird
 - Operationalize AI responsibly with Azure AI Studio
 - <https://build.microsoft.com/en-US/sessions/de3dfc8d-3aaf-4dc2-97c3-5b2937961d54>
 - Next generation AI for developers with the Microsoft Cloud
 - <https://build.microsoft.com/en-US/sessions/226c764e-ab30-4ec3-b8ce-53eb230dcfe0>

Evaluations

The screenshot displays the Azure AI Studio Evaluation page. The browser address bar shows the URL: `ai.azure.com/build/evaluation?wsid=/subscriptions/353e25e6-f2ba-4b7d-9257-5ad187a56568/resourceGroups/rg-vkolesnikovaai/providers/Micro...`. The page title is "Assess and compare AI application performance".

The left sidebar contains navigation options: "Get started" (Model catalog, Model benchmarks, Prompt catalog), "Project playground" (Chat, Assistants, Images, Completion), "Tools" (Code, Prompt flow, Tracing, Evaluation, Fine-tuning), and "Components" (Data, Indexes).

The main content area is titled "Assess and compare AI application performance" and includes tabs for "Metric evaluations", "Manual evaluations", and "Evaluator library". A descriptive text states: "Evaluate the quality and safety of your generative AI applications with industry standard metrics to compare and choose the best version based on your need. [Learn more about metrics.](#)"

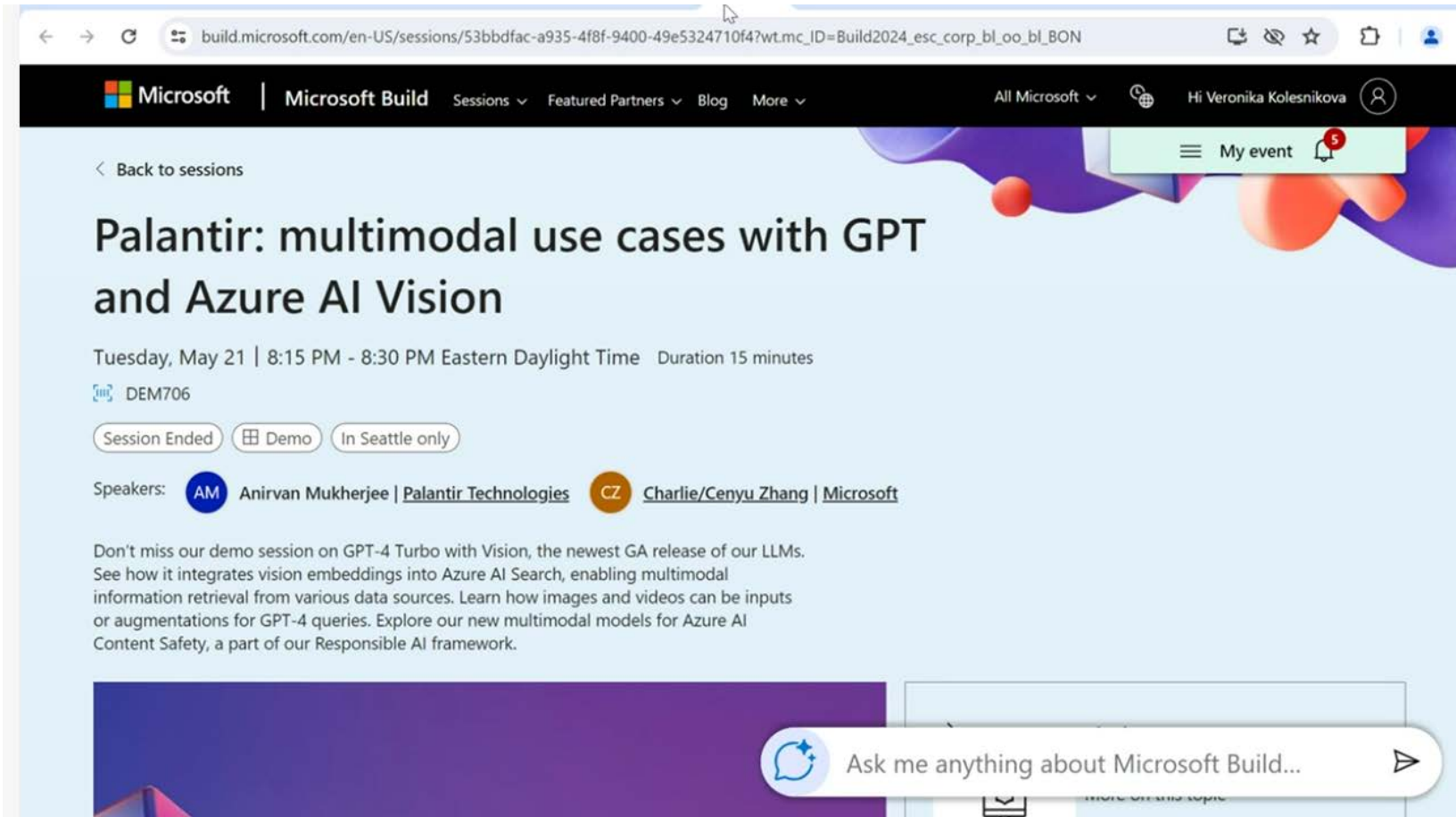
At the top of the evaluation area, there are buttons for "New evaluation", "Refresh", "Cancel", "Delete", and "Compare", along with a "Show batch runs" toggle.

The "Performance and quality check" section features a text box: "Use these industry standard metrics to gauge your AI application's effectiveness and efficiency, ensure optimal functionality, and identify areas for improvement." Below this text is a "View documentation" button.

Four bar charts are displayed, each representing a different metric with an average score and a 1-5 scoring scale:

- Groundedness**: Average score 3.78. The bar chart shows counts for scores 1 through 5.
- Relevance**: Average score 3.70. The bar chart shows counts for scores 1 through 5.
- Coherence**: Average score 3.21. The bar chart shows counts for scores 1 through 5.
- Fluency**: Average score 3.63. The bar chart shows counts for scores 1 through 5.

Multimodal use cases



The screenshot shows a web browser displaying a Microsoft Build session page. The browser's address bar shows the URL: `build.microsoft.com/en-US/sessions/53bbdfac-a935-4f8f-9400-49e5324710f4?wt.mc_ID=Build2024_esc_corp_bl_oo_bl_BON`. The page header includes the Microsoft logo, "Microsoft Build", and navigation links for "Sessions", "Featured Partners", "Blog", and "More". A user profile for "Hi Veronika Kolesnikova" is visible in the top right, along with a "My event" button and a notification icon with the number 5.

The main content area features a "Back to sessions" link and a large title: "Palantir: multimodal use cases with GPT and Azure AI Vision". Below the title, the session details are listed: "Tuesday, May 21 | 8:15 PM - 8:30 PM Eastern Daylight Time | Duration 15 minutes". A session ID "DEM706" is also present. There are three status tags: "Session Ended", "Demo", and "In Seattle only".

The speakers are listed as "AM Anirvan Mukherjee | Palantir Technologies" and "CZ Charlie/Cenyu Zhang | Microsoft". A short description follows: "Don't miss our demo session on GPT-4 Turbo with Vision, the newest GA release of our LLMs. See how it integrates vision embeddings into Azure AI Search, enabling multimodal information retrieval from various data sources. Learn how images and videos can be inputs or augmentations for GPT-4 queries. Explore our new multimodal models for Azure AI Content Safety, a part of our Responsible AI framework."

At the bottom of the page, there is a chat input field with a speech bubble icon and the placeholder text "Ask me anything about Microsoft Build...".

Bill's Favorites

Available at: <https://blog.codingoutloud.com/>

Some More Resources

Book of News

<https://news.microsoft.com/build-2024-book-of-news/>

MICROSOFT BUILD

BOOK OF NEWS

May 21 - 23, 2024

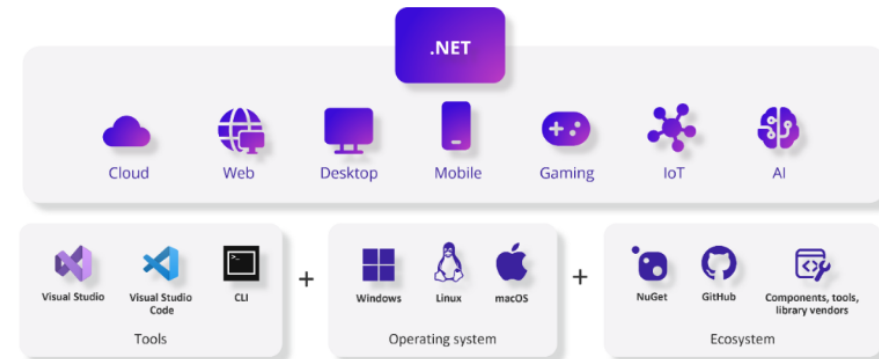
Welcome to Microsoft Build, our annual flagship event for developers, and to this year's edition of the Book of News. Here, you'll discover about 60 announcements, ranging from the latest AI features for Windows to the expansion of Microsoft Copilot and its new capabilities alongside novel tools for developers and cost-efficient and user-friendly cloud solutions for innovation.

Blog: .NET Announcements and Updates from Microsoft Build 2024

- <https://devblogs.microsoft.com/dotnet/dotnet-build-2024-announcements/>

At [Microsoft Build 2024](#), we're thrilled to unveil a new set of features and tools designed to make .NET development faster and easier.

Explore the [.NET sessions at Microsoft Build 2024](#) to see the new features in action, or [try them yourself](#) by downloading .NET 9 Preview 4 today.

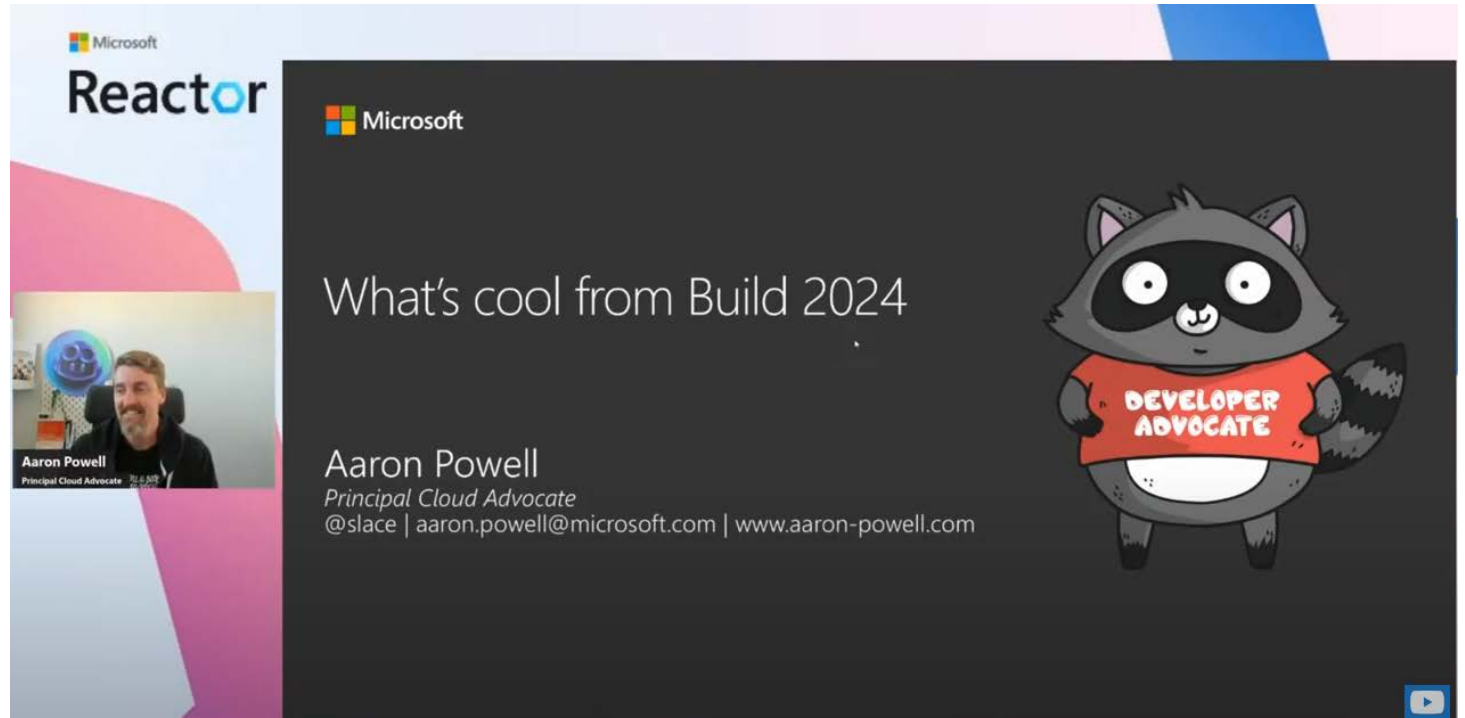


Youtube: Advancing AI - Microsoft Build 2024 AI Highlights

- <https://www.youtube.com/watch?v=UdkbjbC6ars>



Recap: Microsoft Build 2024 (by Aaron Powell)



- <https://www.youtube.com/watch?v=iVeE4oo4I6I>

Notes on next page ->

- Copilot + PCs (ARM based computers) - actually announced the day before MSBuild
- Developer deep dive on building plugins for Microsoft Copilot (Bob German & David Rousset)
 - Session: <https://build.microsoft.com/en-US/sessions/8822feb4-ebb8-4427-9e93-67011e4d0620?source=sessions>
 - VS Code Extensions for helping build copilots
- Teams copilot (mentioned in several sessions)
- AI Services - lots of announcements
 - GPT 4o now GA in Azure
 - Phi-3 SLMs
 - AI Search improvements
 - Azure AI Studio no GA
- Create Generative AI experiences using Phi (Sam Kemp, Yufeng Li)
 - Session: <https://build.microsoft.com/en-US/sessions/e6d21a49-2efb-4a39-8c26-f6eef1410c7a?source=sessions>
 - AI Toolkit VSCode extension (pre-release version), kind of like a local AI Studio
 - Onyx and Olive
- Announcing EAP for Vector Support in Azure SQL Database
 - Blog post: <https://devblogs.microsoft.com/azure-sql/announcing-eap-native-vector-support-in-azure-sql-database/>
- Azure OpenAI extension for Azure Functions
 - Blog post: <https://learn.microsoft.com/en-us/azure/azure-functions/functions-bindings-openai>
 - Triggers and Bindings for working with OpenAI
- Session: .NET Aspire development on any OS with the Visual Studio family
- Session: What's new in GitHub Copilot and Visual Studio

Azure AI @ MS Build 2024

- <https://github.com/azure/azure-ai-at-build>

Azure AI Session & Resources

Day 1

Time	Session	Resources
9AM	Microsoft Build opening keynote	
11:30AM	AI Everywhere – Accelerate your development from the edge to cloud	
11:30AM	Azure AI Studio - Creating and scaling your custom copilots	Learn Collection: Create custom copilots with Studio
11:30AM	Developer's Guide to Customizing Microsoft Copilot	Learn Collection: Enhance productivity with Microsoft Copilot
11:30AM	Building Plugins for Copilot for Microsoft 365	Learn Collection: Explore the possibilities with Microsoft Copilot for Microsoft 365
11:45AM	Employ AI on Snapdragon X Elite for code generation and image creation	
11:45AM	Build Generative AI voice bots with line of business data	Learn Collection: Create custom copilots with Studio
1PM	Learn how to interact with Large Language Models	Learn Collection: Learn how to interact with Large Language Models
1PM	Unlock potential on Azure with Microsoft Copilot	Learn Collection: The adaptive cloud approach on Azure
1PM	What's new with Microsoft Copilot Studio	Learn Collection: Build your own copilot with Copilot Studio

Blog: In Case You Missed It: Top Announcements and On-demand Sessions from Microsoft Build 2024

- <https://startups.microsoft.com/blog/in-case-you-missed-it-top-announcements-and-on-demand-sessions-from-microsoft-build-2024/>

Microsoft Build, our flagship developer conference, wrapped up last Thursday. From the opening keynote from Satya Nadella to over 55 product announcements, our team has gone through them all and curated the top announcements and on-demand sessions for startup founders.

GPT-4o generally available: Use your Azure credits to build and deploy multimodal apps on this groundbreaking model that integrates text, image, and audio processing all within Azure OpenAI Service. Visit the breakout session [here](#) to learn more.

Phi-3 model family generally available: Conserve runway and still build lifechanging AI experiences with Microsoft's small language models (SLMs). These more accessible models can be easily fine-tuned to fit your startup's needs while using limited compute. To find out more, watch the breakout session [here](#).

Azure AI Studio generally available: Get access to more than 1,600 models to explore, build, and deploy your AI solutions on a platform that empowers responsible generative AI development. Watch [this](#) breakout session to learn more.

Build with AI tab in Founders Hub: Don't know where to start? Save time and resources in the new Build with AI tab, where you'll find links to explore all these offerings and a curated set of flexible templates that make creating and using AI solutions faster and easier. View [this](#) breakout session to find out more.

New guidance for developing chatbots generally available: Design and deploy applications with the latest Azure reference architectures, service guides, and patterns to create a private chatbot. The latest guides make it easier for you to build a cost-efficient and compliant AI assistant. View the breakout session with Rob Bagby and Ritesh Modi [here](#) to learn more.