
CODE
PRESENTS
BUILD 2024 Recap
By Markus Egger
May 29, 2024 - Online

About Our Presenter

- **Markus Egger**
- President and Chief Software Architect
  CODE Group
- Publisher – CODE Magazine
- International Author and Speaker
- Microsoft RD (Regional Director)
- Microsoft MVP 1995-2019
- Email: markus@codemag.com
- Twitter: @markusegger

CODE STATE OF NET

CODE CONSULTING

CODE STAFFING

CODE MAGAZINE

CODE TRAINING

30 years of

"Helping People Build Better Software"

About CODE Consulting

CODE STATE OF NET

Custom Software Development
(Web, Mobile, Desktop and Cloud Platform Apps)

Copilot Development, AI, GPT & Azure OpenAI

Project Rescue, App Modernization
(VB, VFP, Access, etc.)

Application Security Testing

Support & Maintenance for existing applications

AI Consulting Services

CODE STATE OF NET

Check out our new Executive Briefing offer!

We can help with your AI needs

What does AI mean for you?

“Skunk Works” Projects

codemag.com/AI
codemag.com/ExecutiveBriefing

State of AI Roadshow

www.codemag.com/StateOfAI

Houston, TX

June, 25th, 2024
Microsoft – at the Ion

Dallas, TX

June, 27th, 2024
Microsoft Offices

Upcoming:

Austin, TX
Boston, MA
New York, NY
Los Angeles, CA
San Francisco, CA
Denver, CO
Chicago, IL
...

CODE Staffing

CODE
STATE OF NET

Disrupting the world of staffing!

Giving our customers the ability to have staff on par with Silicon Valley companies ...

... and our employees a work environment in a bleeding-edge tech company with the **industry leading benefits!**

codestaffing.com

Agenda

CODE STATE OF NET

- The most important happenings and announcements from BUILD 2024 in Seattle
- No surprise: AI and Copilot everywhere!
- A look at the big trends
- More information about new models
- More information about how these models are integrated into products
- More information about tools
- Hardware news

The new era of AI has begun!

The Big Trends

CODE STATE OF NET

1. AI models grow into all directions and gain all kinds of abilities and capabilities
2. Software is being reinvented on top of this new paradigm
3. Hardware changes to enable this new way of computing
4. We are still at the early stages of the current AI era
5. Microsoft is in a great position to capitalize (as are some others).

CODE
STATE OF NET

Models

New Models

CODE STATE OF NET

- New AI Model development continues at a rapid pace
- Models are growing both larger and smaller
  - Smaller models are more efficient, yet still surprisingly capable
  - Larger models are even more capable and impressive
- Models are gaining new abilities
  - Multi-modal models are now common
  - Includes text, images, videos, and speech
- Models are gaining more agency
  - In other words: They can do things, rather than just talk

OpenAI "*Spring Update*" Event

CODE STATE OF NET

- OpenAI announced GPT-4o
  - "o" stands for "omni-modal", indicating a focus on multi-modal input (text, images, videos, audio)
- The new model is much faster and cheaper to use
- The model is also "smoother"
  - Communicating with the models is more "natural" as it can be interrupted and it responds quicker
- There also is more "free stuff"

- https://openai.com/index/spring-update

GPT4 to GPT-4o (1 ½ years)

CODE STATE OF NET

12x

decrease in cost

6x

increase in speed

Microsoft/OpenAI Partnership

CODE STATE OF NET

- Microsoft is heavily invested in OpenAI
- All of OpenAI's tech runs on Azure
- Developers can use OpenAI models either through OpenAI directly or through Azure
- Same-day availability of OpenAI models on Azure through the Azure OpenAI Service

Azure Open AI Service

CODE STATE OF NET

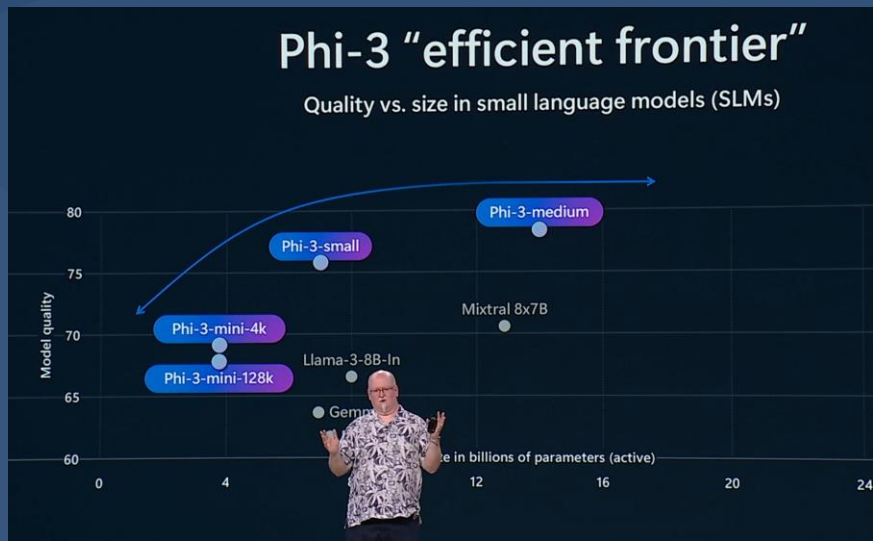
- GPT-4o is generally available through Azure OpenAI services
- New “auxiliary features”, such as Prompt Shields, fine-tuning, and many more
- Also: All other Azure features available in combination
- Used in production by 50,000+ organizations

- https://aka.ms/AOAIUpdates

Phi-3 – Models

- These are small-language-models (SLMs) developed by Microsoft
- “Large model quality at a fraction of the cost”
- Designed to run locally on devices and PCs,…
- …but can be combined with the Cloud in hybrid setups

- * There also is a Phi-3 Vision model



Phi-Silica

CODE STATE OF NET

- Small Language Model (SLM) specifically designed to run on specific hardware configurations
- Designed specifically to run locally
- Comes with every Copilot + PC

Hugging Face

- Microsoft is partnering with Hugging Face to bring those models to the overall Microsoft family

Azure AI model breadth

Offering a wide collection of frontier and open models

Azure OpenAI Service

- GPT-4o
- GPT-4-Turbo with Vision, GPT-4, GPT-3.5
- Embeddings
- DALL·E
- Whisper, Text to speech

Phi models

- Phi-3-mini
- Phi-3-small
- Phi-3-medium
- Phi-3-vision

Meta

- Llama-2-70b/70b-chat
- Llama-2-13b/13b-chat
- Llama-2-7b/7b-chat
- Llama-3
- CodeLlama

Mistral AI

- Mistral Large
- Mistral 7B
- Mixtral 8x7B – Mixture of Experts

Cohere

- Cohere R+
- Cohere R
- Embed v3-Multilingual
- Embed v3-English

Hugging Face

- Falcon/TII
- Stable Diffusion/Stability AI
- Dolly/Databricks
- CLIP/OpenAI

Databricks

- Databricks/dbrx-base
- Databricks/dbrx-instruct

NVIDIA

- Nemotron-3-8B-4k
- Nemotron-3-8B-Chat-SFT/RLHF/ SteerLM
- Nemotron-3-8B-QA

Snowflake

- Snowflake/arctic-base
- Snowflake/arctic-instruct

Azure AI Custom Models

CODE STATE OF NET

- “Co-innovation with Microsoft”
- Assisted fine tuning
- Training custom language models
- Ingest new knowledge domains
- Improve proficiency for specific languages

Hardware

The Scale Involved

Copilot+PC

Copilot + PC

- New hardware platform specifically designed for the era of Copilots and AI
- Very powerful local processing on top of ARM hardware
- Hardware from Microsoft as well as partners: Acer, ASUS, Dell, HP, Lenovo, and Samsung

Copilot + PC: Surface Laptop

CODE STATE OF NET

My Configuration:

- 15” display
- Snapdragon X Elite 12 Core
- 64GB Ram, 1TB SSD
- 22-hour battery life
- NPU (45 trillion operations/sec)

NPUs

- Neural Processing Unit
- Think “GPU but for AI” ;-)
  - Often smaller, more parallel, and less precise than GPU tasks, but LOTS of them!
- Specialized hardware to execute AI tasks
- Often integrated into the CPU or specialty chips right on the motherboard

Microsoft/NVIDIA Partnership

CODE STATE OF NET

- Blackwell Platform (available through Azure)
- NIM-optimized models
- Omnivers, DGX, Cloud + Fabric
- Windows 365 and Copilot GPU acceleration

- https://aka.ms/NVIDIAPartnership

Microsoft/AMD Partnership

CODE STATE OF NET

- ND MI300X V5 AI Accelerator Chip Support on Azure
- Leading price/performance on GPT-4
- Optimized for Microsoft Azure workloads

- https://aka.ms/AMDPartnership

Azure Maia

- Integrated system of silicon, racks, and software
- Designed for Azure OpenAI workloads

- https://aka.ms/AzureMaia

Azure Cobalt

CODE STATE OF NET

- Public preview of Cobalt-based VMs
- Used for things like video processing
- Most power-efficient compute offering in MS Azure

- https://aka.ms/AzureCobalt

Qualcomm Snapdragon Dev Kit for Windows

CODE STATE OF NET

- Ships June 18th for $899

Software

What is AI-driven Software?

CODE STATE OF NET

- We must fundamentally ask ourselves what software in the age of AI is like
- We do not have good answers for that

- Most software is currently being retro-fitted with AI (which works surprisingly well)…
- …but in the future, we envision completely different approaches to software solutions

CODE
STATE OF NET

All apps without Copilot/AI features are now legacy!

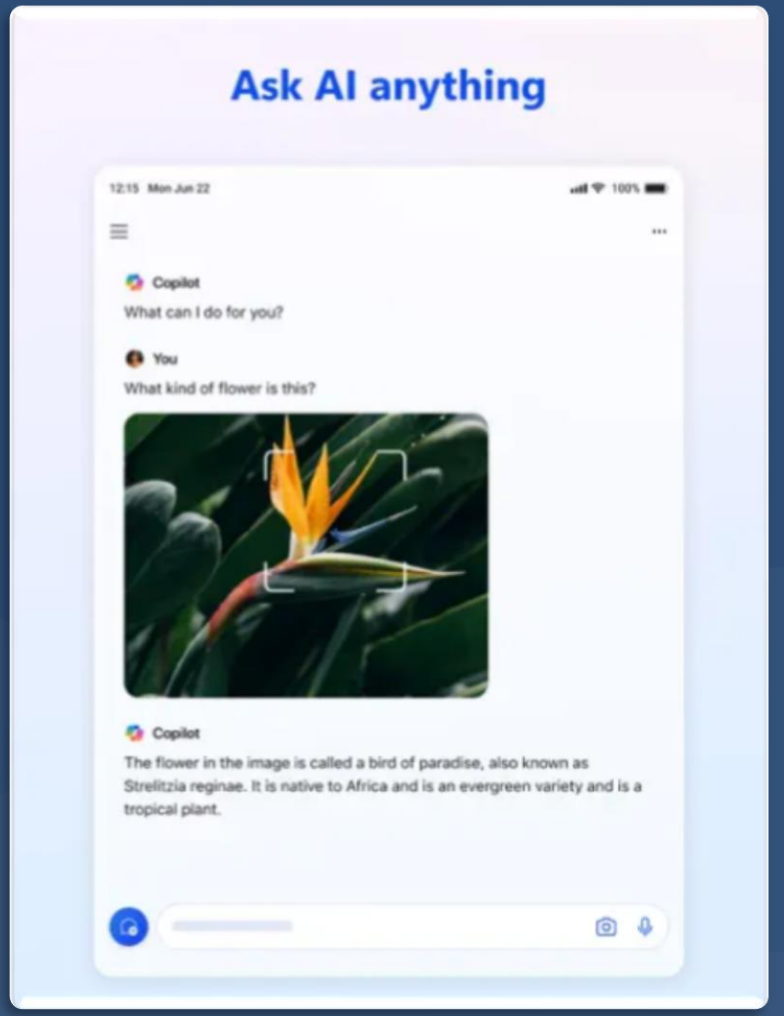
- Copilot is Microsoft’s branding of AI features
- Copilots appear in all Microsoft apps
- The term is also used as an interaction paradigm
  - Microsoft seems to encourage use of the term in our own apps

Microsoft Copilot for...

- Service
- Sales
- Azure
- Finance
- Security
- Teams
- Microsoft 365
- Bing
- Edge
- Windows
- Team
- ...

Copilot App

- Turns out Copilot is very appealing when used on its own, rather than inside anther application as an add-on



Copilots in our own Apps

CODE STATE OF NET

- IMO, the more specific the domain of an app is, the more powerful AI becomes
- I find that Copilots built into our own apps are better than the more generic apps in systems such as M365

Recall

Find the red barn

Yesterday

Today

Now

Microsoft Excel

Microsoft PowerPoint

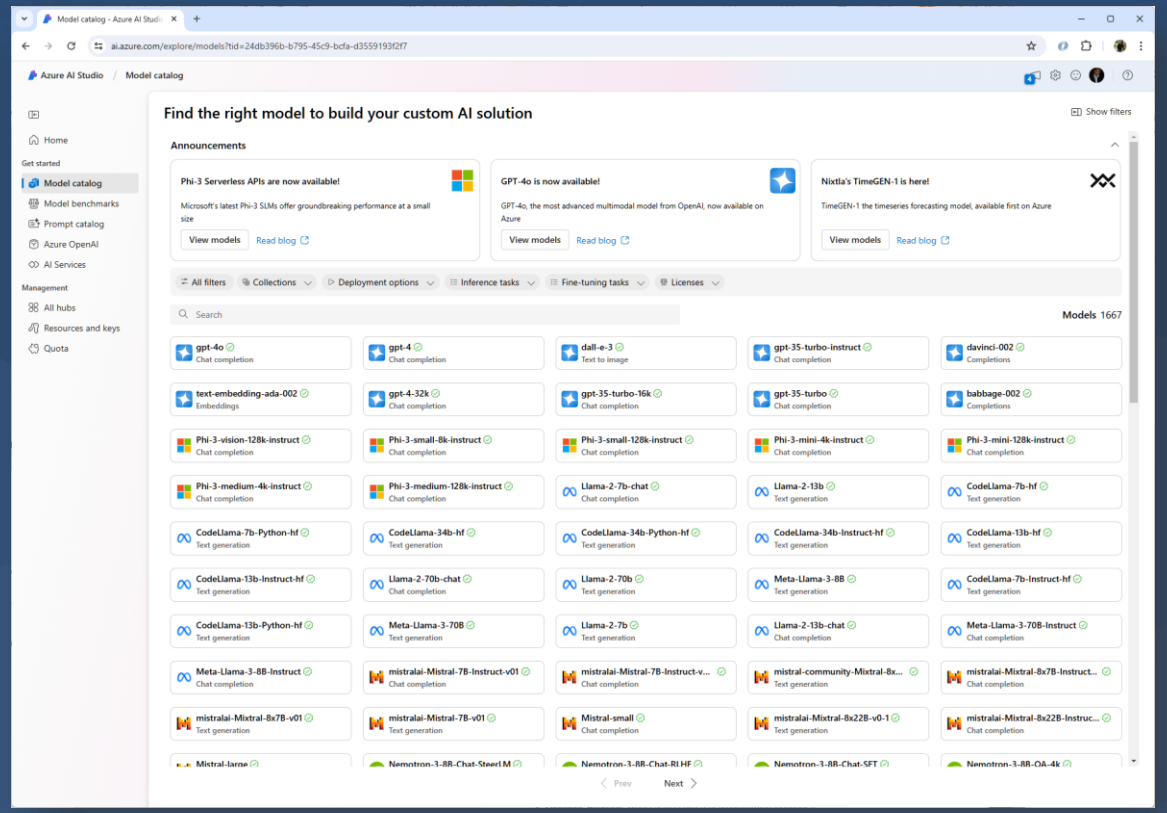
Microsoft Teams

Developer Tools

Azure AI Studio

- Comprehensive Platform for building and deploying custom AI and Copilots
- Now generally available

- http://ai.azure.com



Prompty:
Prompt Engineering Tool

CODE
STATE OF NET

- Specification
  - A standard format that unifies the prompt and how it is executed into a single asset
- Tooling
  - Developer friendly ways to interact with assets, create new assets, and manage supporting code
- Runtime
  - A simple way to transition this new asset into code

Prompty as a language-agnostic prompt asset

Specification

Prompty is intended to be a language agnostic asset class for creating prompts and managing the responses.

The goal is to simplify your workflow by creating a standard that can be used by any language, any framework, and any tool to create a prompt and manage the response.

Prompty can be imported to/exported from Azure AI Studio for smooth transition between local and cloud. Prompty assets can be also shared across organizations.

```
---
name: Basic Prompt
description: A basic prompt that uses the GPT-3 chat API to answer questions
authors:
  - sethjuarez
  - jietong
model:
  api: chat
  configuration:
    azure_deployment: gpt-35-turbo
sample:
  firstName: Jane
  lastName: Doe
  question: What is the meaning of life?
---
system:
You are an AI assistant who helps people find information.
As the assistant, you answer questions briefly, succinctly,
and in a personable manner using markdown and even add some
personal flair with appropriate emojis.

# Customer
You are helping {{firstName}} {{lastName}} to find answers to their questions.
Use their name to address them in your responses.

user:
{{question}}
```

Prompty – Some Details

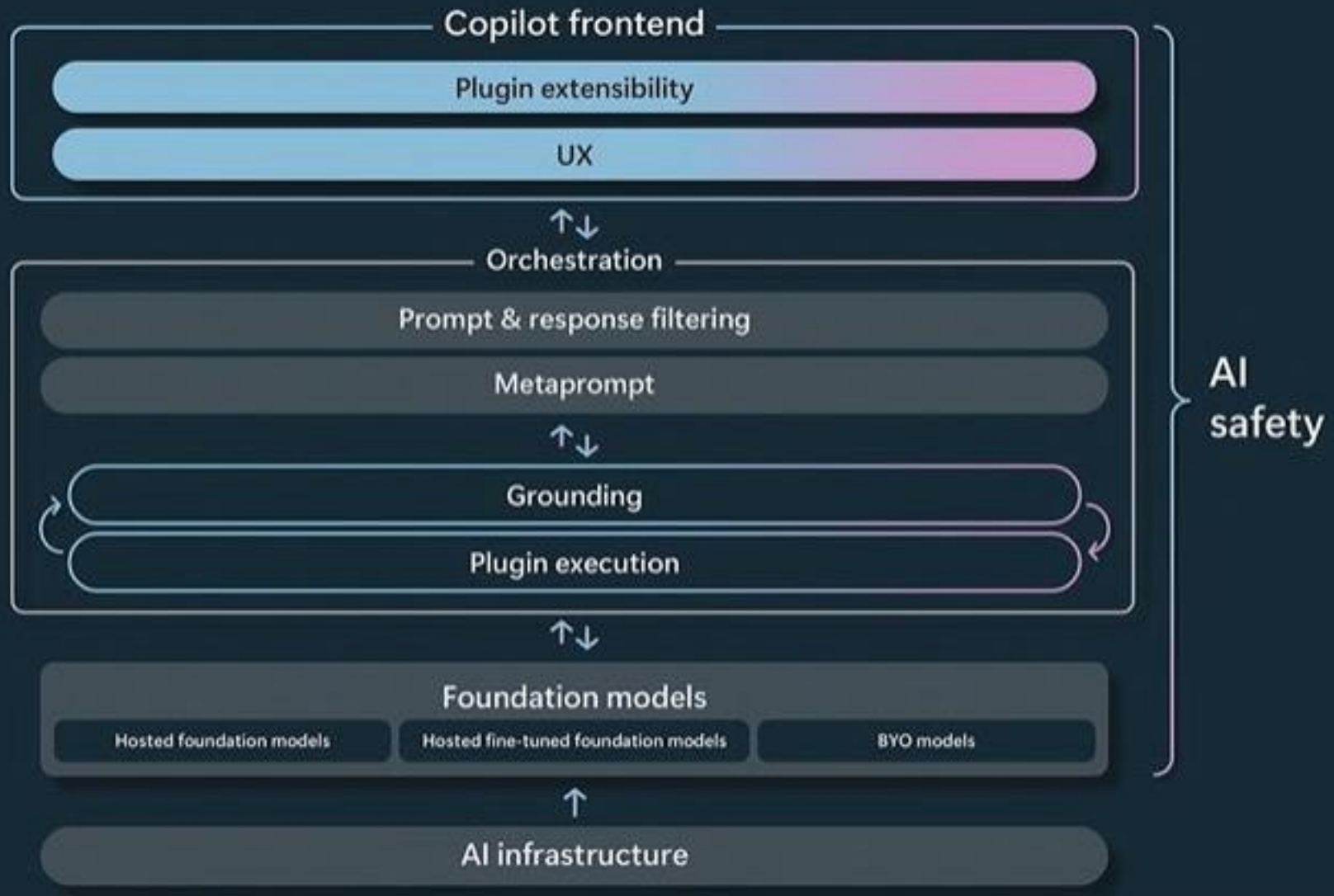
CODE STATE OF NET

- Langchain, PromptFlow, and Semantic Kernel are supported as runtime environments
- It’s essentially Markdown and Frontmatter
- VS Code has Prompty extensions

Copilot Studio

- Extend Microsoft Copilot and create custom copilots
- This is generally geared to M365 and Power Platform
- It’s a managed environment aimed at ease of use

Copilot Stack



Windows Copilot Runtime

DirectML

CODE STATE OF NET

- DirectML is a machine learning API
- Works seamlessly across a wide range of hardware platforms
- Supports DX12 GPUs and (soon) NPUs
- Powered by DirectML on Windows
  - PyTorch now native
  - Web Neural Network (WebNN) – Web-native machine learning framework – Available as developer preview today

Microsoft Intelligent Data Platform

Databases • Analytics • AI • Governance

Azure SQL DB

Azure Cosmos DB

Azure for PostgreSQL

MySQL

Microsoft Fabric

Azure Databricks

Azure AI

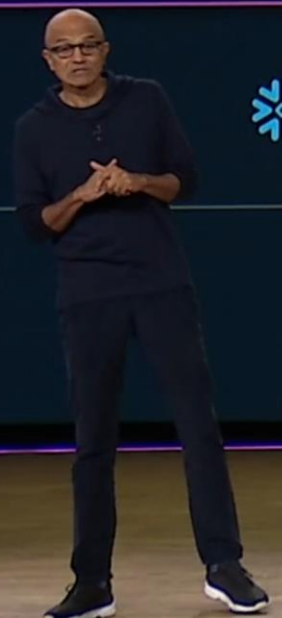
Microsoft Purview

snowflake

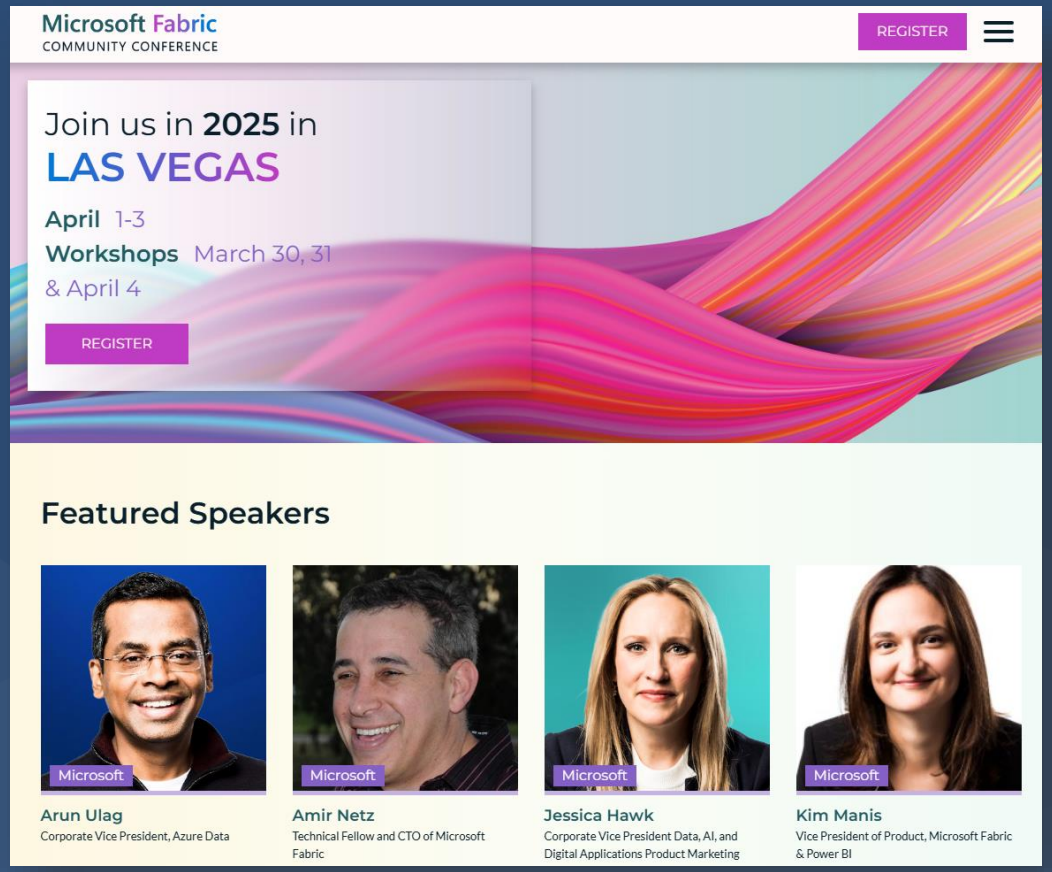
MongoDB.

ORACLE

Redis

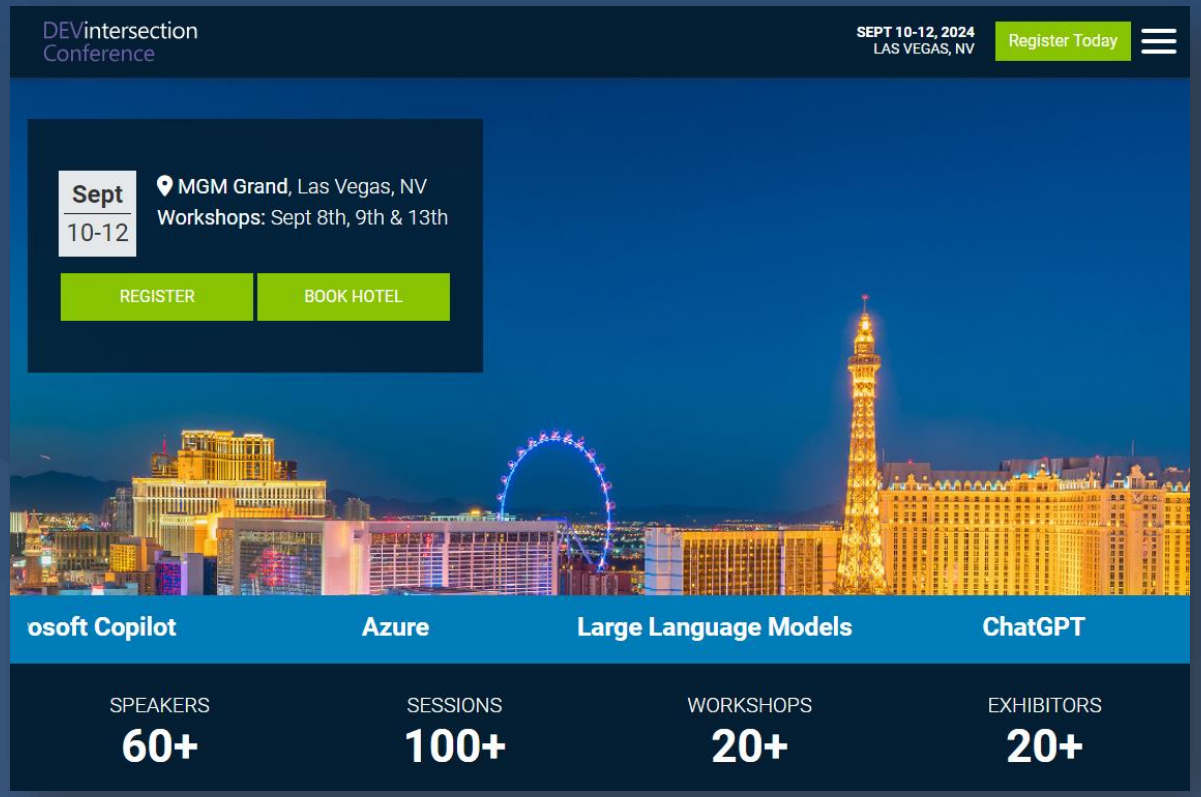


Microsoft Fabric Community Conference



www.fabricconf.com

DevIntersection – Las Vegas Featuring: CODE AI Track



www.devintersection.com

Copilot Capabilities in Azure Databases

CODE STATE OF NET

- AI-enhanced management and performance
- Public Preview:
  - Self-help for managing and operating Azure SQL databases
  - Convert natural language to Azure SQL Database T-SQL queries
  - Chat with Azure Database for MySQL technical documentation in natural language

Vector Search in Databases

CODE STATE OF NET

- It is probably safe to assume that any kind of database and storage technology will support vector search capabilities in the not too distant future

Azure AI Search

CODE STATE OF NET

- State-of-the-art search technology
- Feature-rich vector database
- Seamless data and platform integration
- Up to 12x more scale and performance

Secure Future Initiative

CODE STATE OF NET

- Secure by design
  - Protect tenants and isolate production systems
  - Protect identities and secrets
- Secure by default
  - Protect networks
  - Protect engineering systems
- Secure operations
  - Monitor and detect threats
  - Accelerate response and remediation

Azure AI Content Safety

CODE STATE OF NET

- Coming Soon: Custom Categories
- Preview: Prompt Shields
- Preview: Groundedness detection

- https://aka.ms/ContentSafetyUpdates

GitHub Copilot

CODE STATE OF NET

- 1.8 Million subscriptions across 50,000 organizations
- From personal experience, I ***highly*** recommend using GitHub Copilot for quality and productivity reasons

GitHub Copilot Extensions

CODE STATE OF NET

- GitHub Copilot now supports extensions
- https://aka.ms/GitHubCopilotExtensions

GitHub Copilot for Azure

CODE STATE OF NET

- GitHub Copilot Extension specific to Azure
  - Available in Visual Studio and VS Code

- https://aka.ms/GitHubCopilotAzure

Side-Note: Developer Tools

- Visual Studio + GitHub are the most widely used dev tools
- .NET 9 will release in November of 2024

State of AI Roadshow

www.codemag.com/StateOfAI

Houston, TX

June, 25th, 2024
Microsoft – at the Ion

Dallas, TX

June, 27th, 2024
Microsoft Offices

Upcoming:

Austin, TX
Boston, MA
New York, NY
Los Angeles, CA
San Francisco, CA
Denver, CO
Chicago, IL
...


CODE Training:
New AI Classes to be announced soon!
www.codemag.com/training
pment with OpenAI,
ChatGPT, Azure OpenAI, and more...

Other Announcements

AI Executive Briefings

CODE STATE OF NET

Check out our new Executive Briefing offer!

We can help with your AI needs

What does AI mean for you?

“Skunk Works” Projects

codemag.com/AI
codemag.com/ExecutiveBriefing

Event Survey – Win $100!

Complete this short 12 question survey for a chance at a **$100 Amazon Gift Card!**

Survey must be completed by
11:59pm ET on Friday 5/31/2024
to be eligible!

THIS SLIDE WILL BE REPEATED AT THE END
AND SURVEY LINK REPEATED IN THE CHAT WINDOW!

https://bit.ly/SODN52924Survey

Free Subscription

CODE STATE OF NET

The leading software development magazine, written by expert developers for developers.

All registered attendees will receive a **free digital subscription** to CODE Magazine!

Share this link to our free subscription:
https://bit.ly/SODN52924Subscribe

CODE Staffing

CODE
STATE OF NET

Disrupting the world of staffing!

Giving our customers the ability to have staff on par with Silicon Valley companies ...

... and our employees a work environment in a bleeding-edge tech company with the **industry leading benefits!**

codestaffing.com

CODE
STATE OF NET

WE NEED YOU!

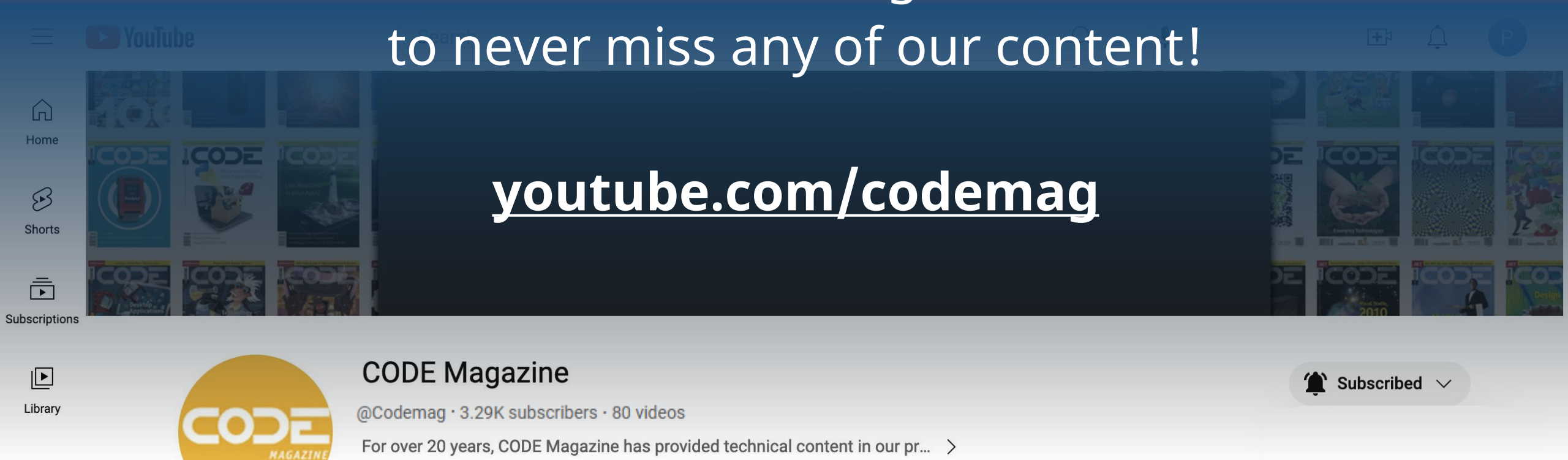
codestaffing.com/careers

CODE STATE OF NET

YouTube

Subscribe and “ring that bell”
to never miss any of our content!

youtube.com/codemag



CODE STATE OF NET

Be the first to know about our
webinars, workshops, and live events!

Follow us! www.linkedin.com/company/**code-magazine**

CODE
STATE OF NET

Q&A

Contact us with questions!

CODE/EPS Contact

codemag.com
info@codemag.com
facebook.com/codemag
twitter.com/codemagazine

Presenter Contact

markus@codemag.com
twitter.com/markusegger