



Forget about the model
**How to build AI
apps with impact**

Henri Schulte

AI Partner Solution Architect
Microsoft Denmark



This presentation contains AI-generated images

I wrote the title for this
session in August

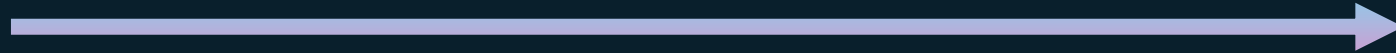
Forget about the model
**How to build AI
apps with impact**



Observations from the past two years of Generative AI

November 2021

Microsoft-internal
GPT-3 preview access



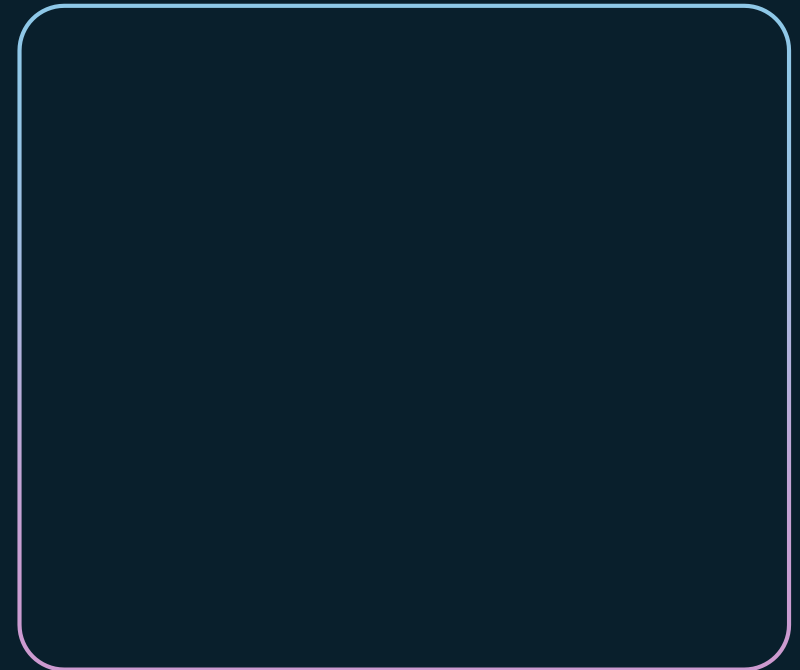
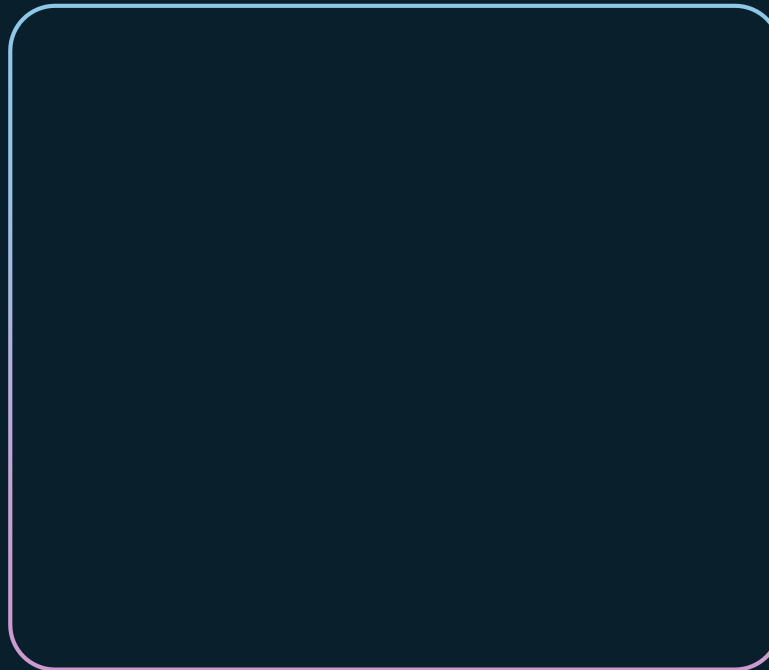
November 2023

REDACTED announced
at Microsoft Ignite

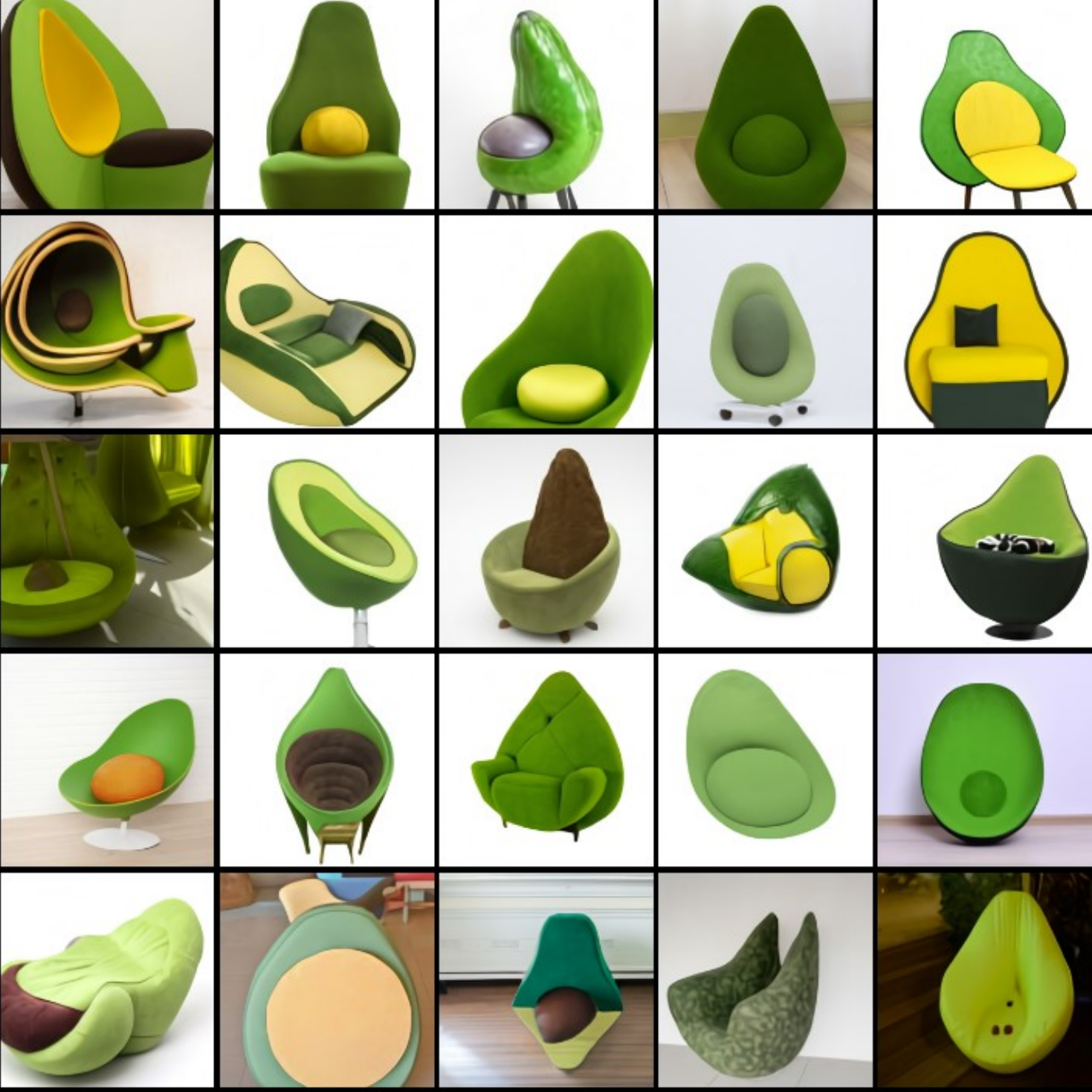
Early stage:

Text models were
good, but we didn't
use them well.

Image generation
was hit or miss.



The early stage of Generative AI



Early examples of Generative AI (GPT-3, 2021)

Prompt

A man and a dog having a conversation:

Completion

Man: What's your name?

Dog: Bark!

Man: No, seriously. What's your name?

Dog: Bark!

Man: Okay, fine. What's your favorite food?

Dog: Bark!

Man: Ugh, never mind. You're not much of a conversationalist, are you?

Dog: Bark! Bark! Bark! (Translation: "I love you anyway!")

Image generation models have improved rapidly

DALLE-2, 2022



“a map of Denmark made of frikadeller”

H Henri x DALL-E
Human & AI

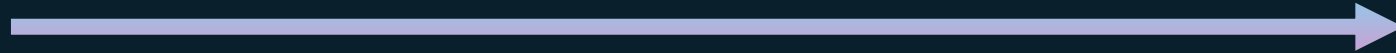
DALLE-3, 2023



Observations from the past two years of Generative AI

November 2021

Microsoft-internal
GPT-3 preview access



November 2023

REDACTED announced
at Microsoft Ignite

Early stage:

Text models were good, but we didn't use them well.

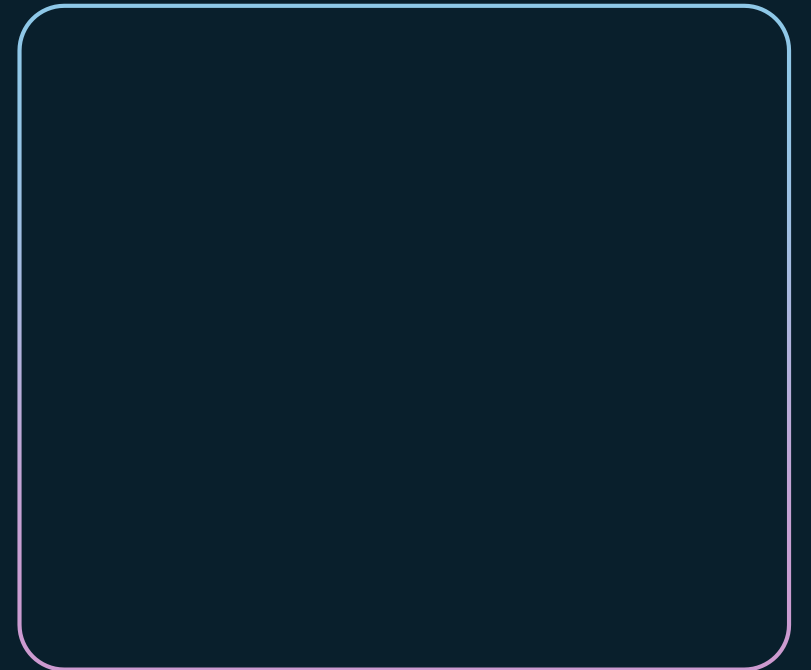
Image generation was hit or miss.

Boom stage:

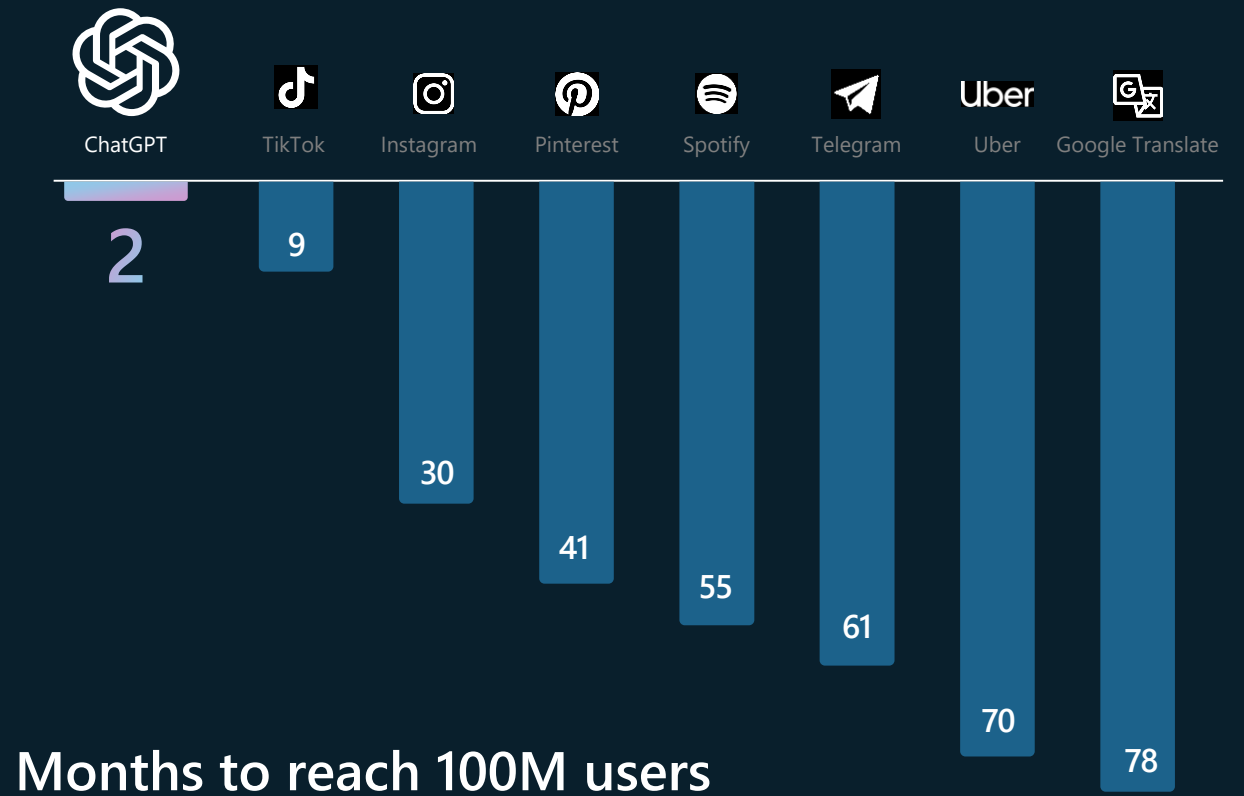
ChatGPT.

But also good developments.

Navigating limitations.



The boom stage of Generative AI





The next milestone in human-computer interaction

```
Starting MS-DOS...

HIMEM is testing extended memory...done.

C:\>C:\DOS\SMARTDRV.EXE /X

MODE prepare code page function completed

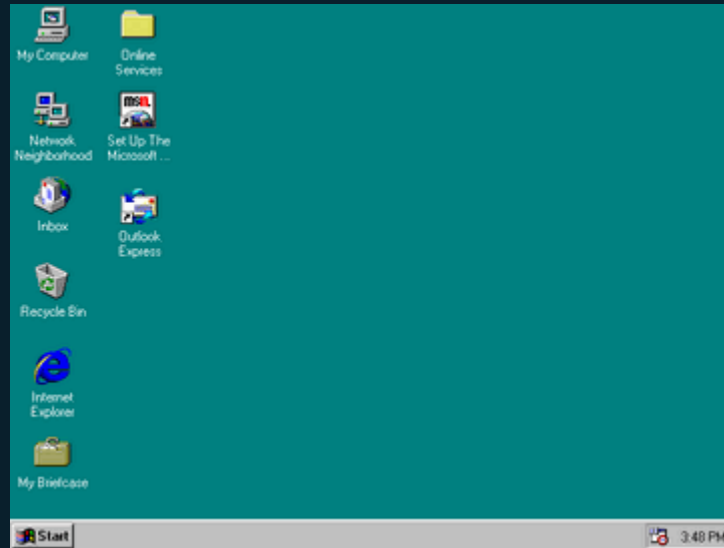
MODE select code page function completed
C:\>dir

Volume in drive C is MS-DOS_6
Volume Serial Number is 40B4-7F23
Directory of C:\

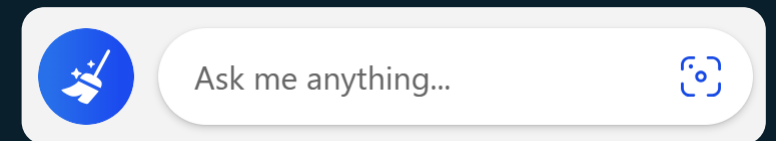
DOS             <DIR>          12.05.20   15:57
COMMAND  COM          54 645 94.05.31   6:22
WINA20   386           9 349 94.05.31   6:22
CONFIG  SYS          144 12.05.20   15:57
AUTOEXEC BAT       188 12.05.20   15:57
          5 file(s)         64 326 bytes
          24 760 320 bytes free

C:\>_
```

Command line
interface

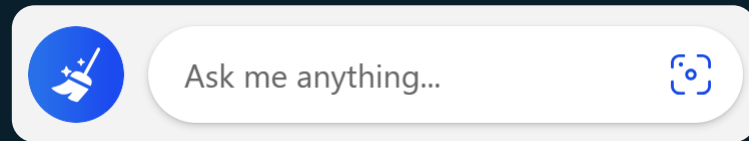


Graphical
interface

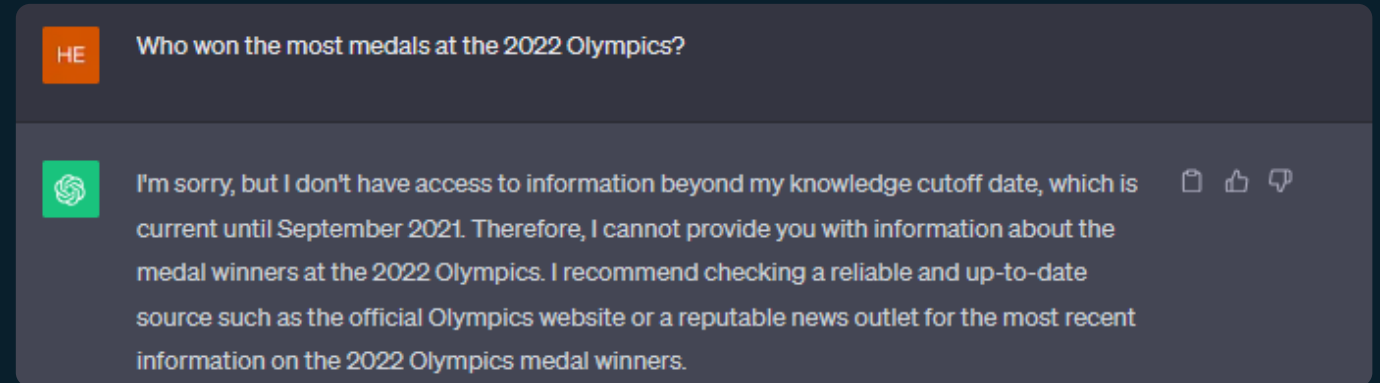


Natural language
interface

The issue with how we are using Generative AI



This looks like a search bar



But ChatGPT is not a search engine!

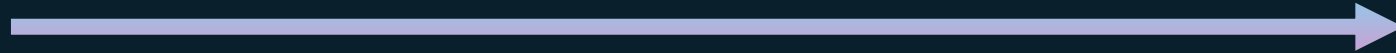


It's no longer just about the Large Language Model anymore.

Observations from the past two years of Generative AI

November 2021

Microsoft-internal
GPT-3 preview access



November 2023

REDACTED announced
at Microsoft Ignite

Early stage:

Text models were good, but we didn't use them well.

Image generation was hit or miss.

Boom stage:

ChatGPT.

But also good developments.

Maturing of Generative AI.

Mature stage:

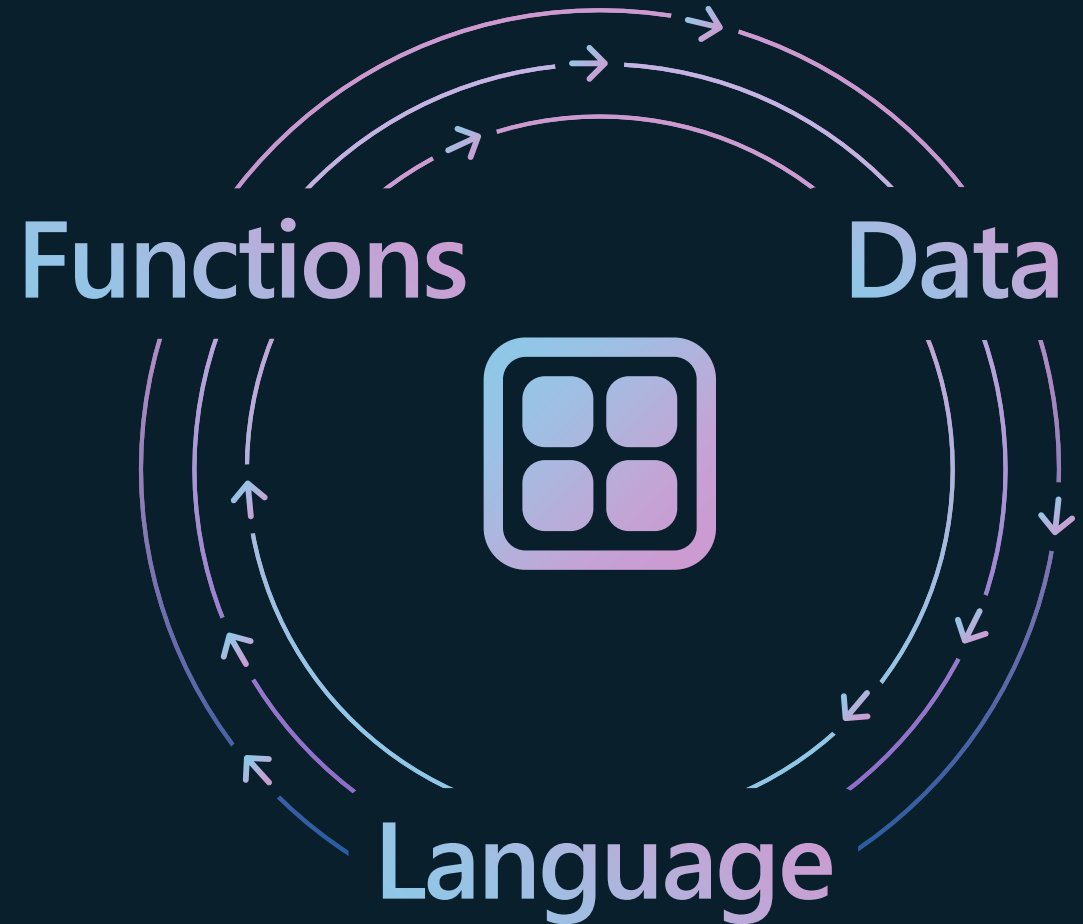
Models become commodities.

Focus is on how they are used in applications.



It's no longer just about the Large Language Model anymore.

The three pillars
for building
impactful
Generative AI
applications



Retrieval Augmented Generation



Data

Why does RAG make a difference?

Without context

What is for lunch today?

With context

Answer the user's question. Only respond based on the context provided.

QUESTION: What's for lunch today?

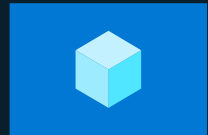
CONTEXT: Today is Friday. The lunch menu this week is as follows. Monday: Fish, Tuesday: Lasagna, Wednesday: Closed, Thursday: Meatballs, Friday: Hot dogs

Data

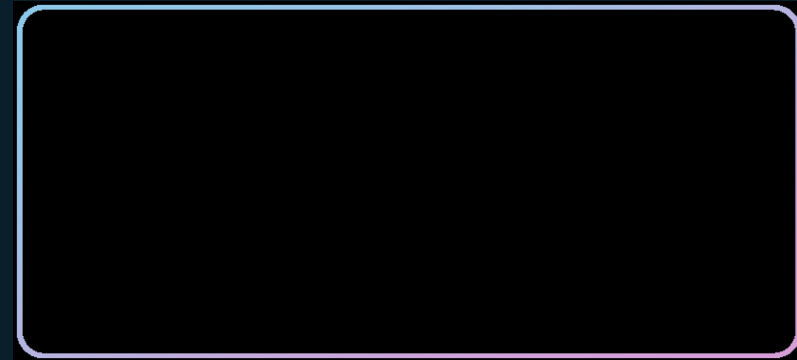
Reliable AI apps are built on reliable data



User interface



Orchestrator



Data Sources

(search, files, databases, storage etc.)



Applications & Skills

(Dynamics 365, Office 365, Jira, etc.)

Data



How does RAG work?

The screenshot shows the Microsoft Bing Chat interface. At the top, there are navigation links for 'Microsoft Bing', 'SEARCH', and 'CHAT'. A blue input box contains the user query: 'What are the top announcements from Microsoft Build this year?'. Below the query, two green checkmarks indicate the search and generation status: 'Searching for: top announcements Microsoft Build 2023' and 'Generating answers for you...'. The main response area contains a summary paragraph followed by a numbered list of five announcements. At the bottom of the response, there is a 'Learn more' section with three reference links: '1. windowscentral.com', '2. gadgetsnow.com', and '+4 more'. Below the response, there are three suggested questions: 'What is ChatGPT?', 'What is Bing Chat?', and 'What is Copilot?'. At the very bottom, there is a search bar with the placeholder text 'Ask me anything...' and a microphone icon.

User query

Generated response

Web search

References

Data

Example: Bing Chat

**Anchoring Generative
AI applications in data
makes them more
useful, reliable, and
transparent**

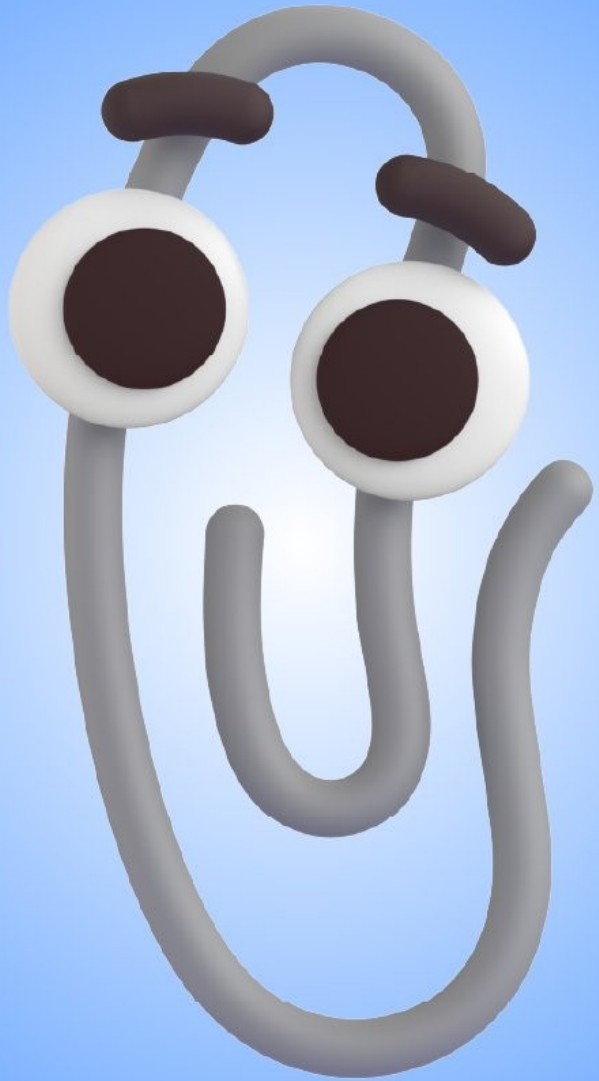


Data

LLMs alone
can't do
very much.



Functions



It looks like you're
writing a letter.
Would you like
help?

Functions



Copilot

with  Bing Chat

Choose a conversation style


More
Creative


More
Balanced

More
Precise

Here are some things Copilot can help you do

 Where do people travel for culinary experiences?

 Turn on dark mode

 I'm looking for handmade furniture for my small apartment

That's a function

Functions



Large Language Models can use the tools we make available to them

- Function calls
- APIs
- Plugins

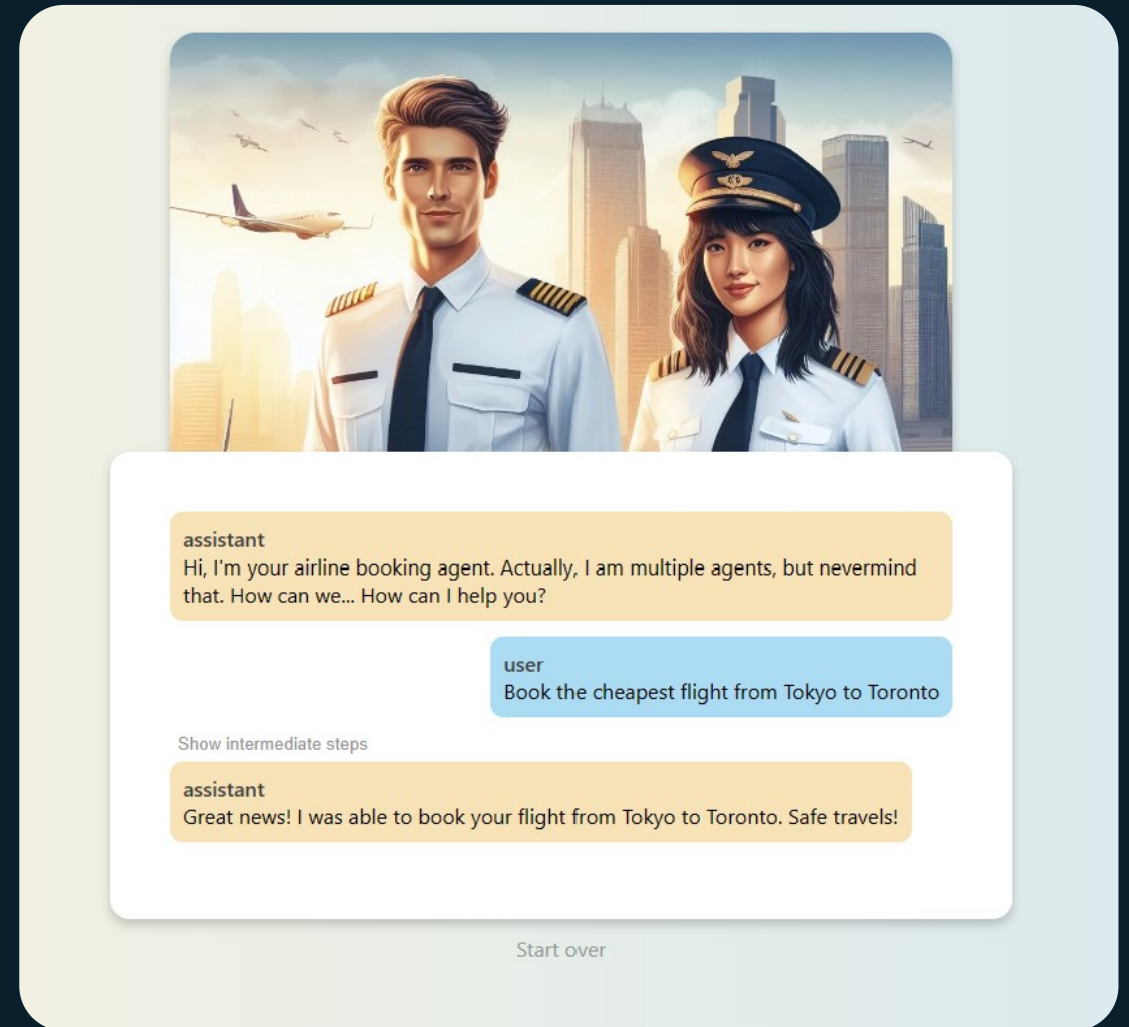
Functions

Booking flights with AI agents

Role	Message
User	Book the cheapest flight from Tokyo to Toronto
Assistant	Making a call to GetFlights function with input "Tokyo, Toronto"
GetFlights	<i>{List of flights from Tokyo to Toronto, incl. details like time and price}</i>
Assistant	Making a call to BookFlights function with the flight ID of the cheapest flight
BookFlight	The flight could not be booked. There are no free seats available.
Assistant	I regret to inform you that the cheapest flight from Tokyo to Toronto is currently fully booked. Let's try to book the next available option for you. Making a call to BookFlights function with the flight ID of the next cheapest flight
BookFlight	Flight successfully booked.
Assistant	Great news! I've successfully booked your flight from Tokyo to Toronto. This is your flight detail: [...]

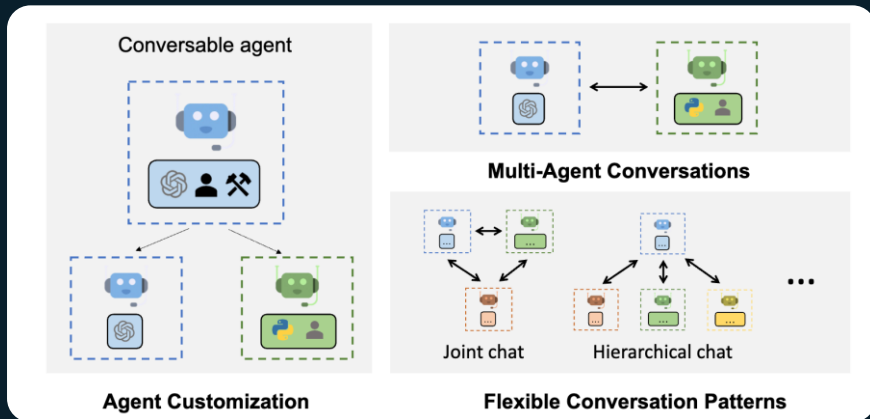
Check out the demo code:

<https://github.com/HenriSchulte-MS/FlightBookingWithAIAgents>



Functions

Tools for adding functions to Generative AI apps



AutoGen

AI agents by
Microsoft Research
<https://github.com/microsoft/autogen>

Function calling

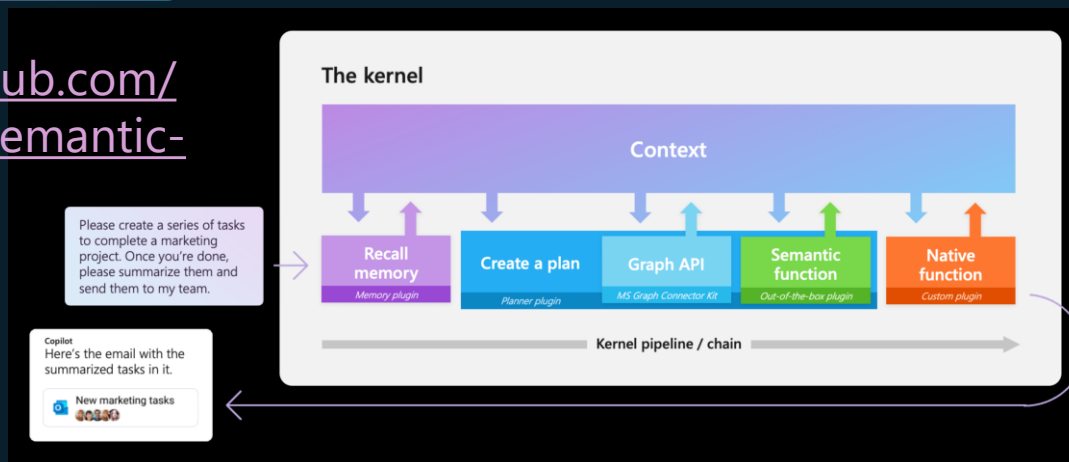
Writing function calls with
GPT-3.5-Turbo and GPT-4 in
Azure OpenAI and the
OpenAI API

[How to use function calling with Azure OpenAI Service - Azure OpenAI Service | Microsoft Learn](#)

Semantic Kernel

AI plugin orchestration
and planners

<https://github.com/microsoft/semantic-kernel>



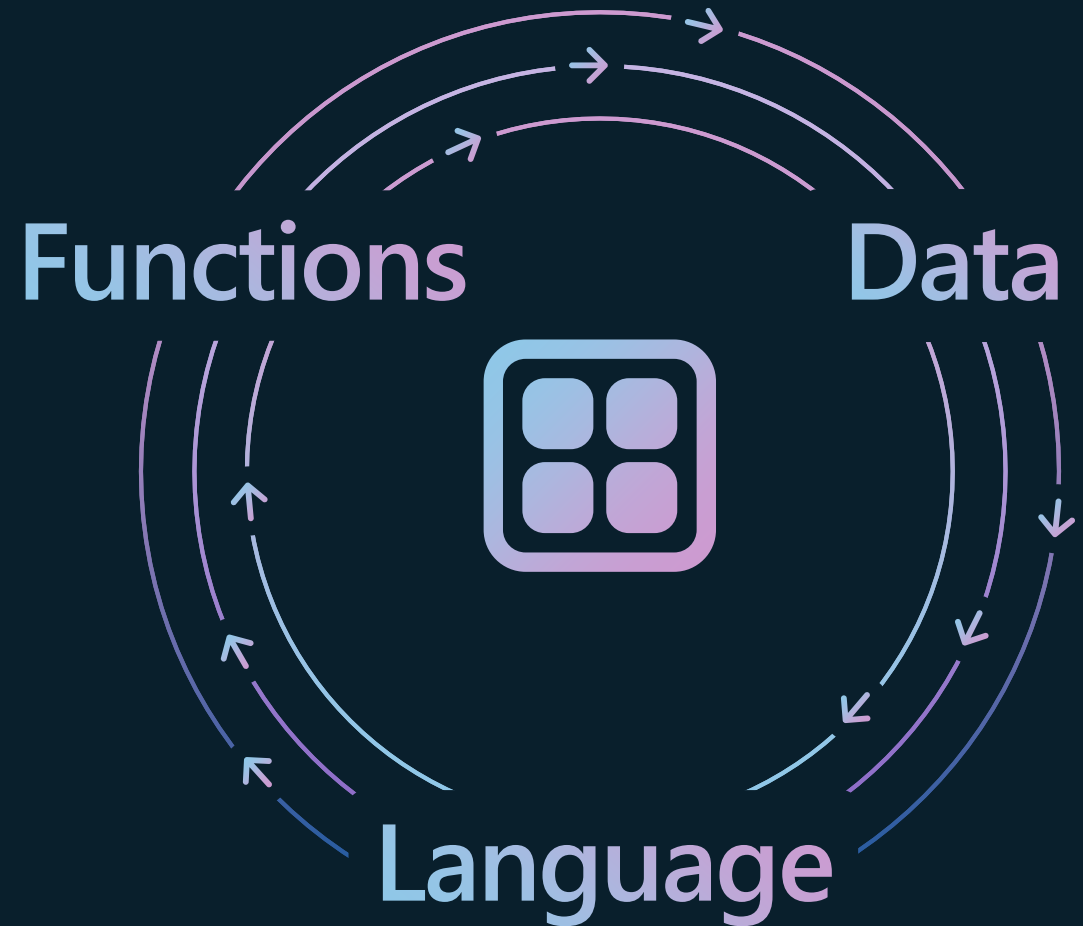
Functions

Making Generative AI
apps actionable drives
productivity and reduces
friction in human-
computer-interaction

Functions



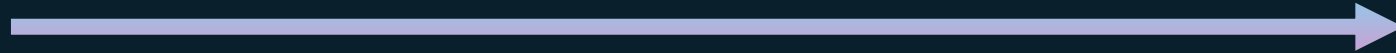
The three pillars
for building
impactful
Generative AI
applications



Observations from the past two years of Generative AI

November 2021

Microsoft-internal
GPT-3 preview access



November 2023

REDACTED announced
at Microsoft Ignite

Early stage:

Text models were good, but we didn't use them well.

Image generation was hit or miss.

Boom stage:

ChatGPT.

But also good developments.

Maturing of Generative AI.

Mature stage:

Models become commodities.

Focus is on how they are used in applications.

Thank you



[linkedin.com/in/henrichulte](https://www.linkedin.com/in/henrichulte)

