

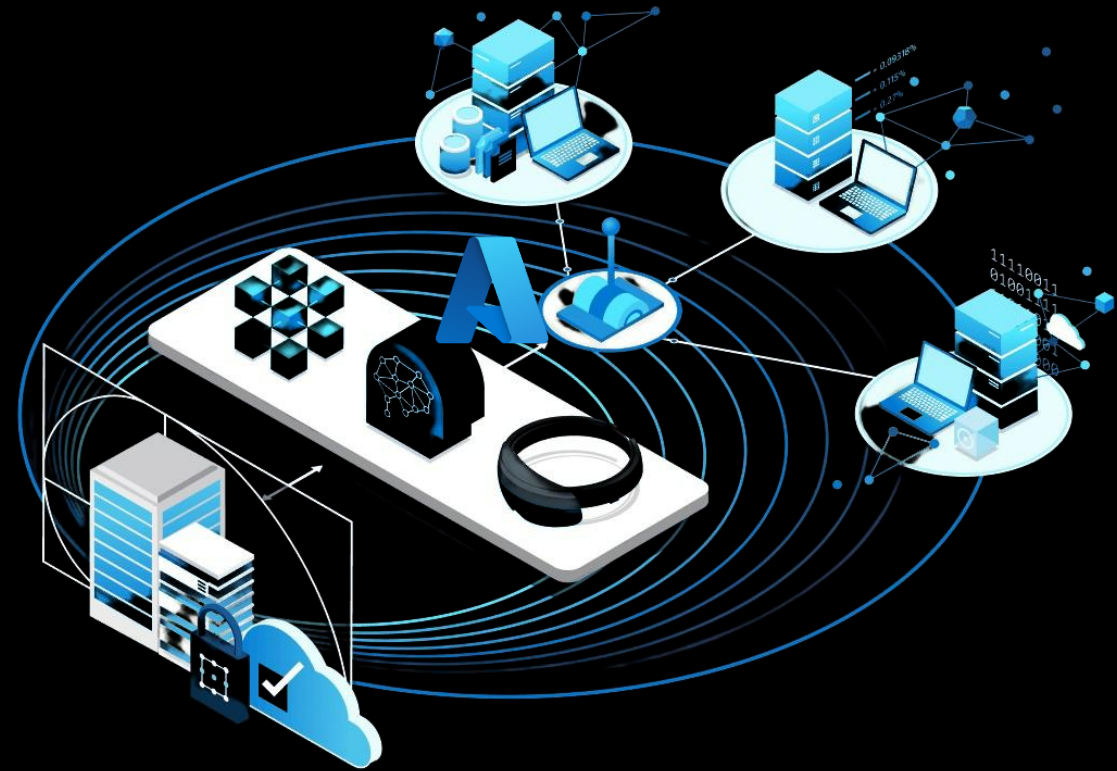


# Cloud & Edge Computing Workshop Series Day 2

Embrace the **future of IT** today

Tom Claes – CSA Azure Core

Merlijn Stoffels – Intern CSA Data & AI



---

## Agenda – Day 2

---

Introduction to Containerization & kubernetes

Refresh on Generative AI & LLMs

Break 10'

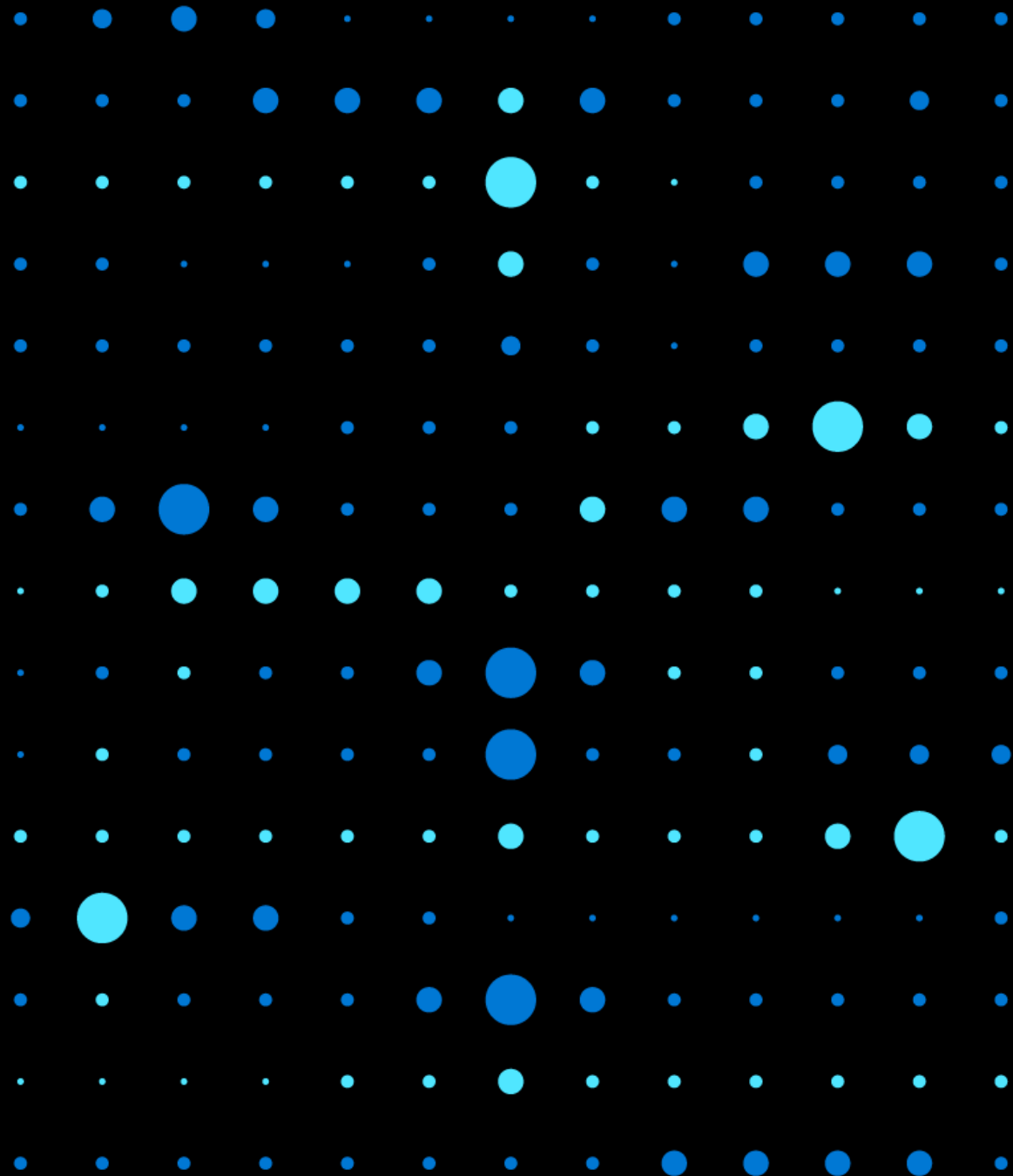
Hands-on project

Wrap up

Quiz (to win some prizes) 😊

Q&A

# Containerization and Kubernetes



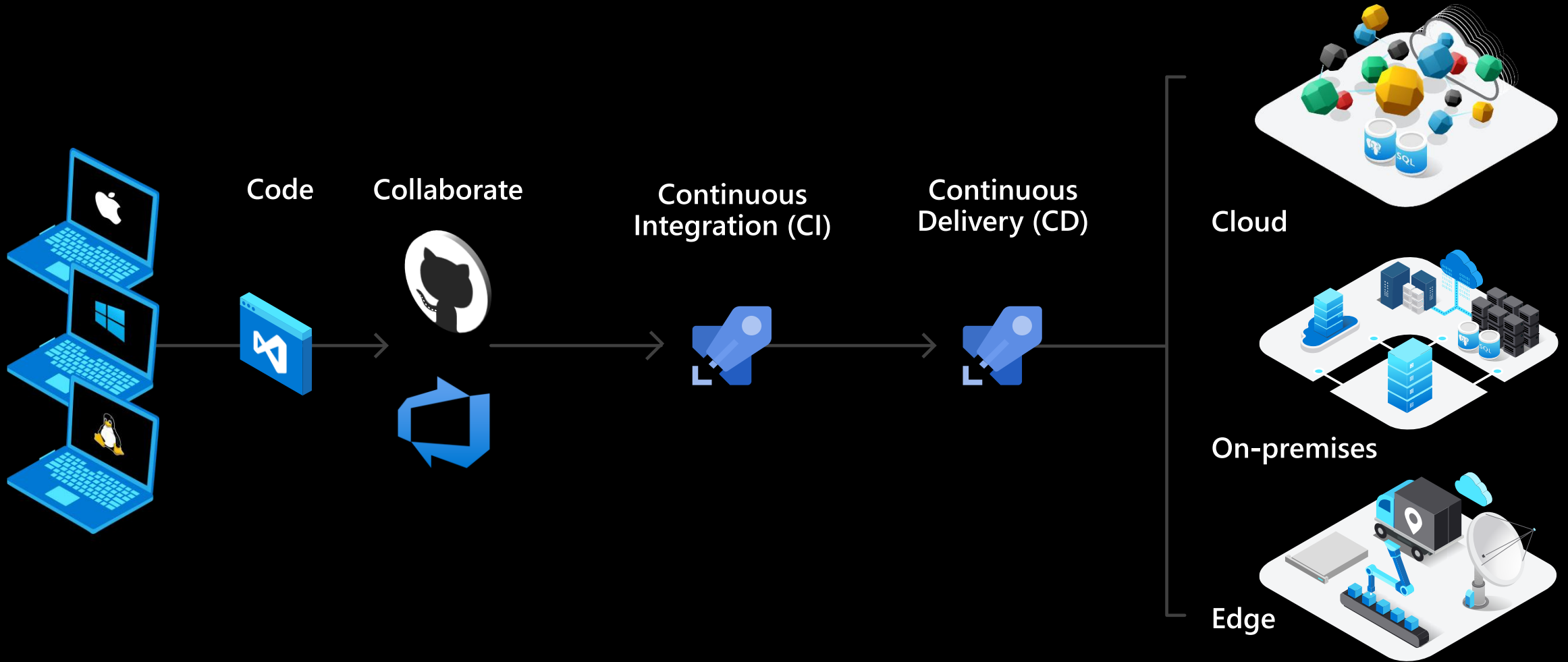
# Empower everyone to innovate

Scale innovation amplifying the tech talent across your organization

- Increase agility across your organization
- Build apps without compromises
- Enable extensibility for developers



# Code to Cloud



# How to host application on the cloud



Control

Productivity



## VMs

Azure Virtual Machines

## Containers

Azure Kubernetes Service | Azure  
Red Hat OpenShift

## PaaS/Serverless

App Service | Functions |  
Spring Apps | Container Apps

## Low code

Power Apps | AI Builder  
Azure Services

## Managed databases

Azure SQL Database | PostgreSQL | MySQL | Cosmos DB

## DevOps

GitHub | Azure DevOps | VS Code



Azure | Azure Arc | Azure Stack



# What is “cloud native?”

*Cloud-native architecture and technologies are an approach to designing, constructing, and operating workloads that are built in the cloud and take full advantage of the cloud computing model.*

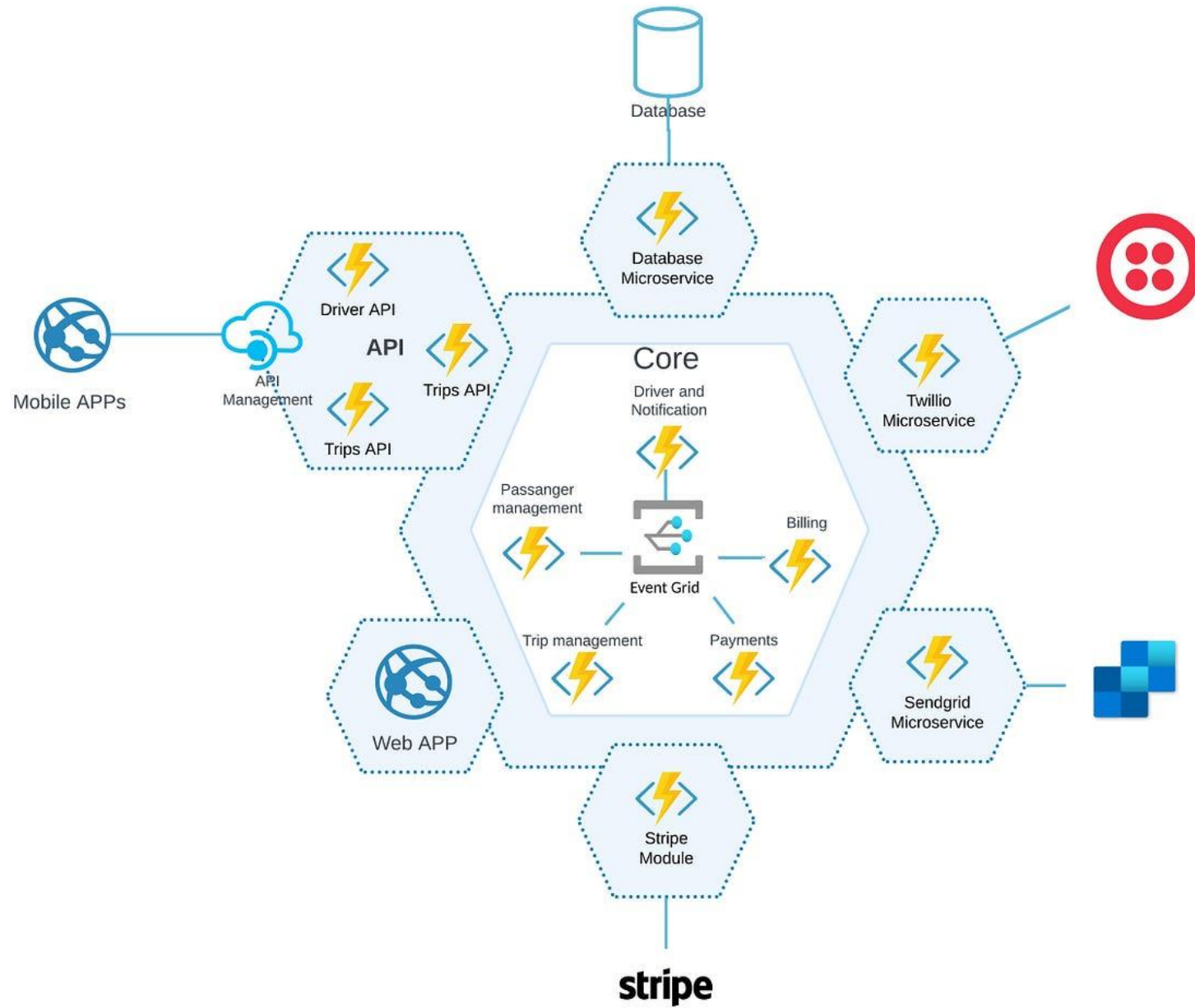


Lift-and-shift enables cloud benefits, while “cloud native” is about fully optimizing those workloads to access full advantages across developers, teams, and organizations

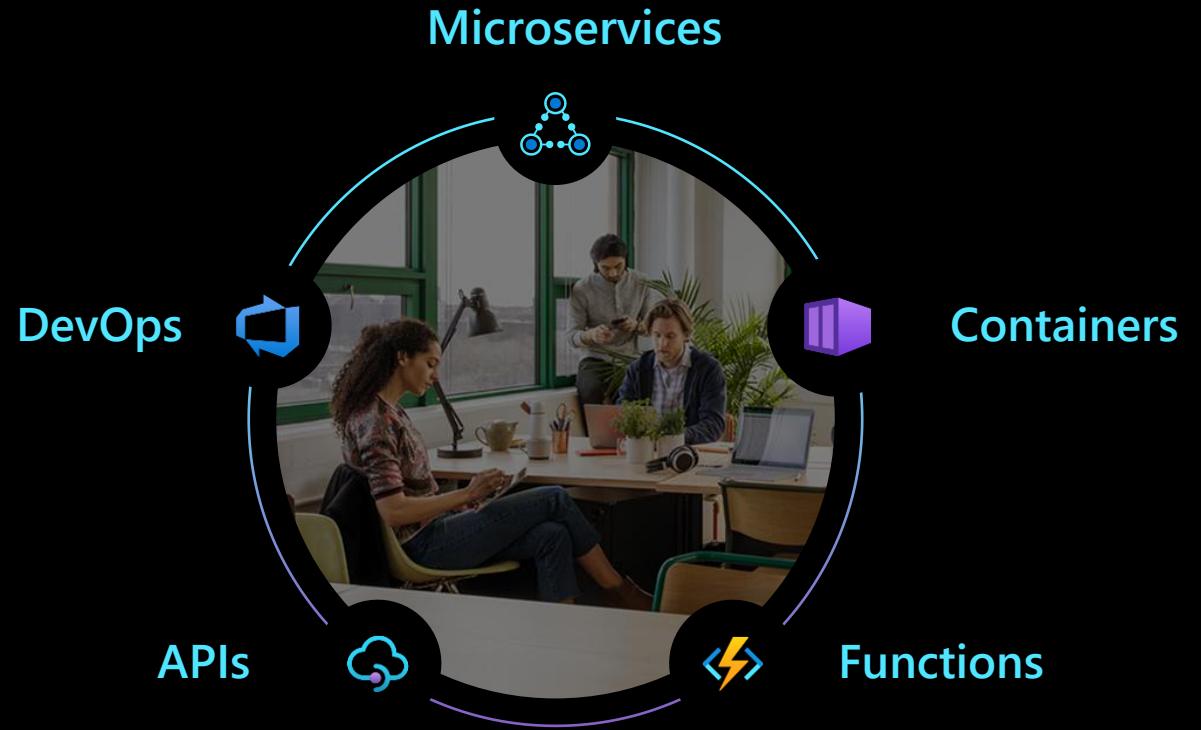
# Microservices

- Architectural framework
- Decentralized management and development
- Technology & language agnostic
- Scalability & elasticity





# Build new intelligent cloud-native applications

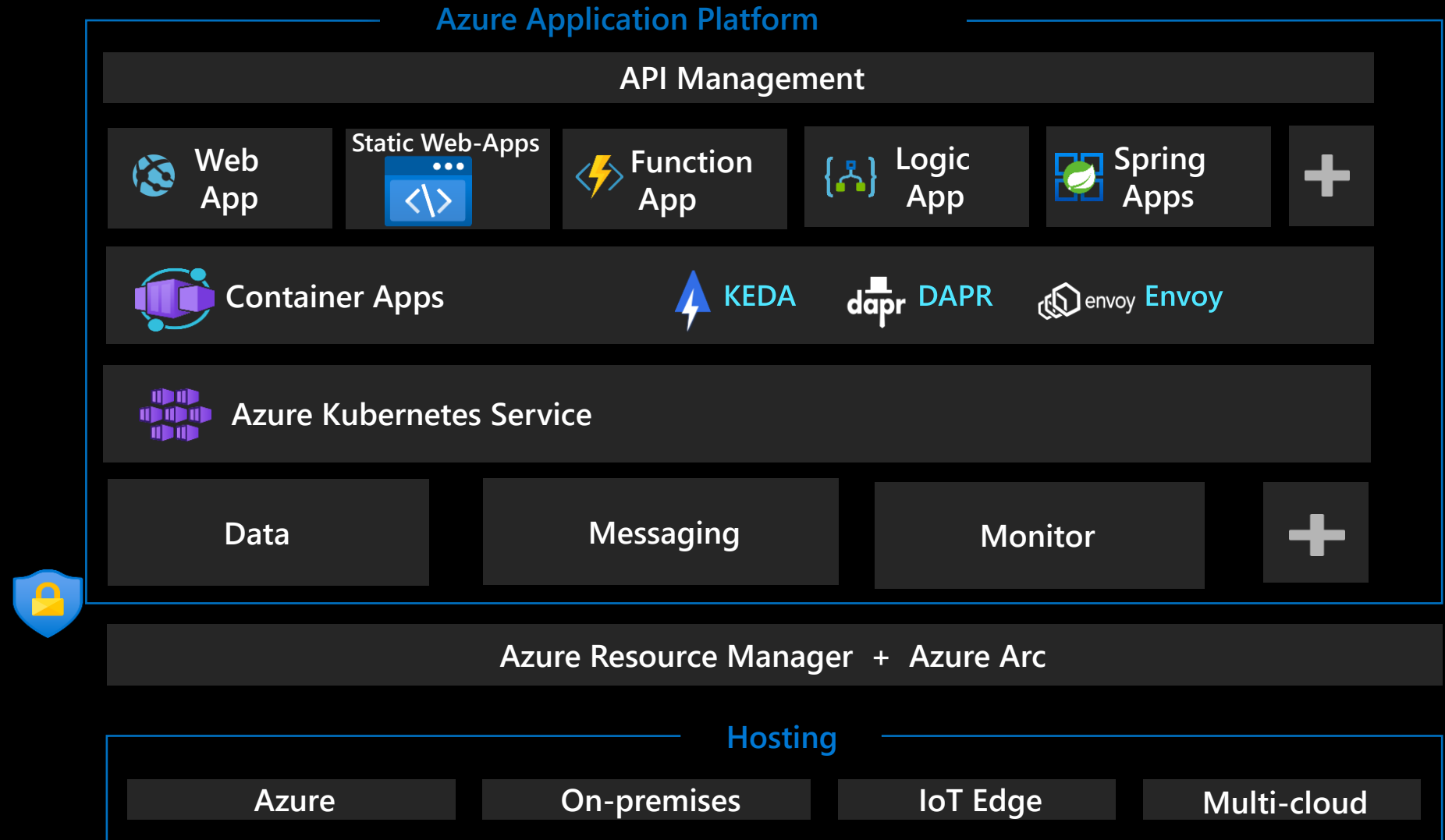


AI-powered UI and application logic

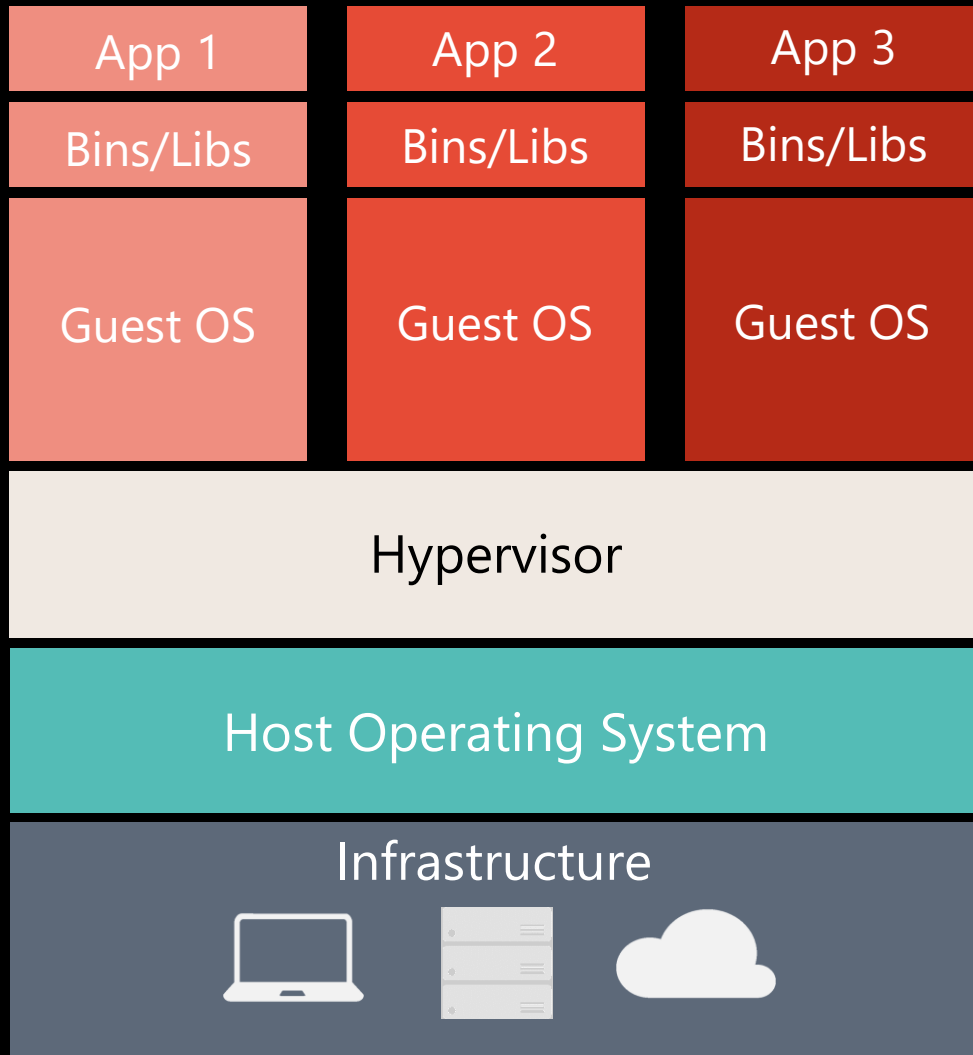


Fully managed data, analytics and insights

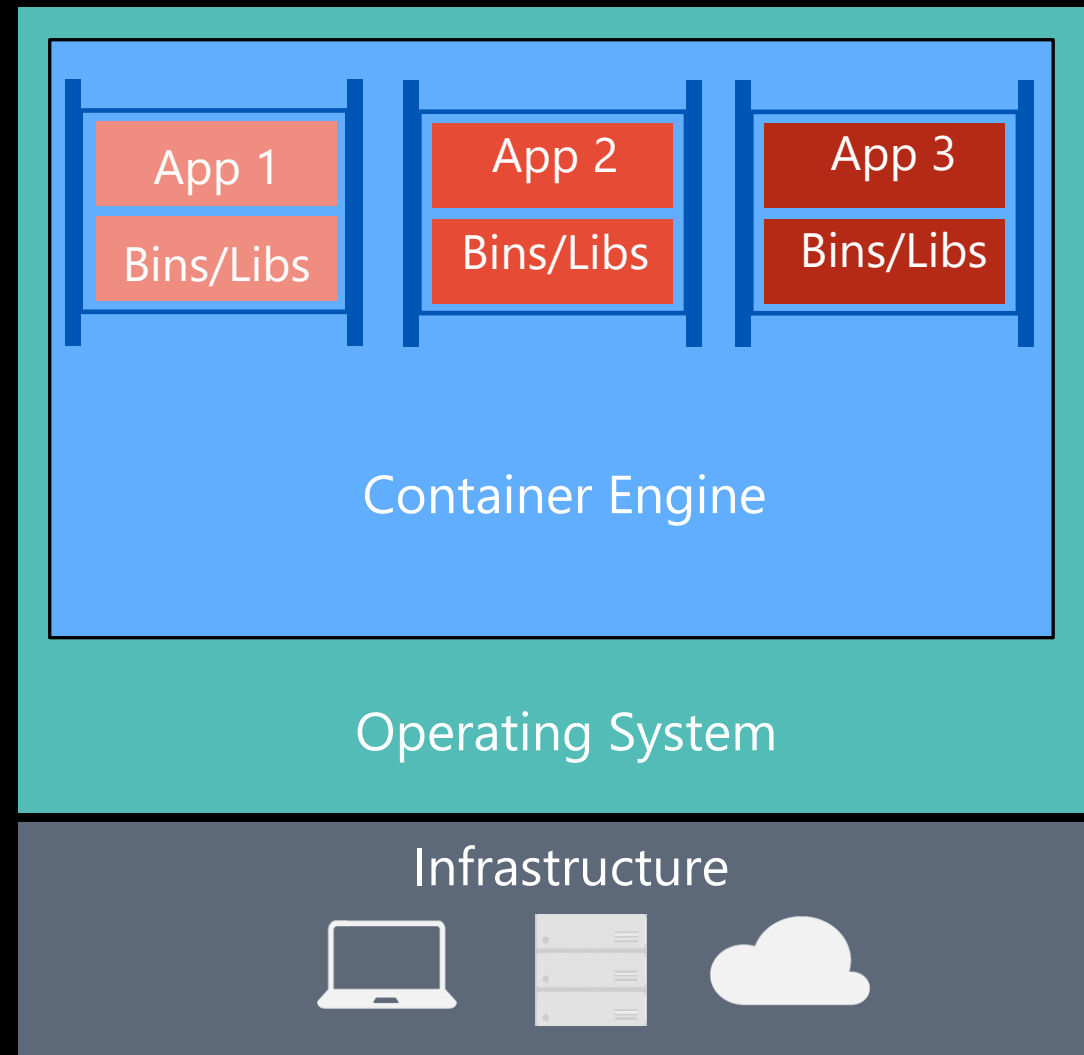
# Azure's Application Platform



# Virtual Machines



# Containers

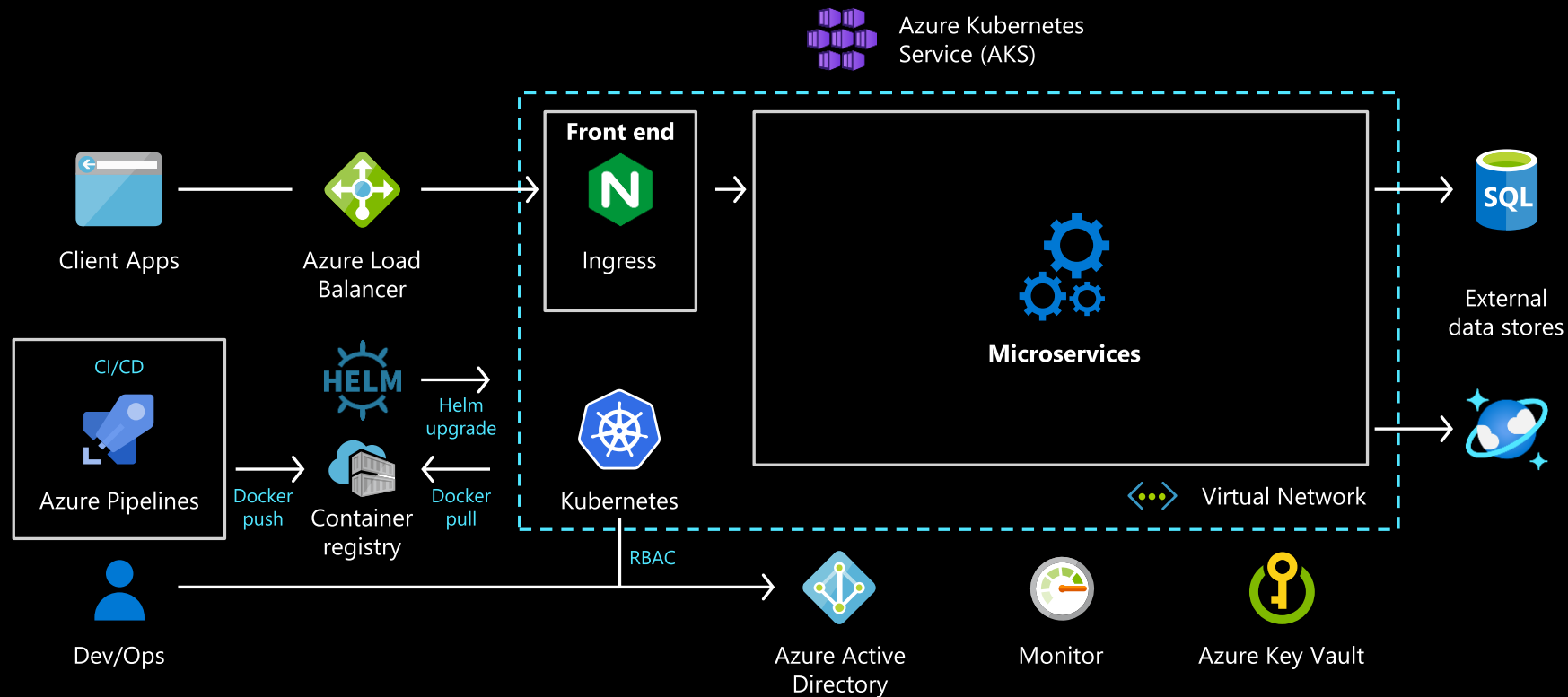


# Why containers?

- Resource Efficiency
- Portability
- Faster Startup Times
- Simplified Management
- Immutability and Version Control
- Agile development of applications
- Microservices architecture

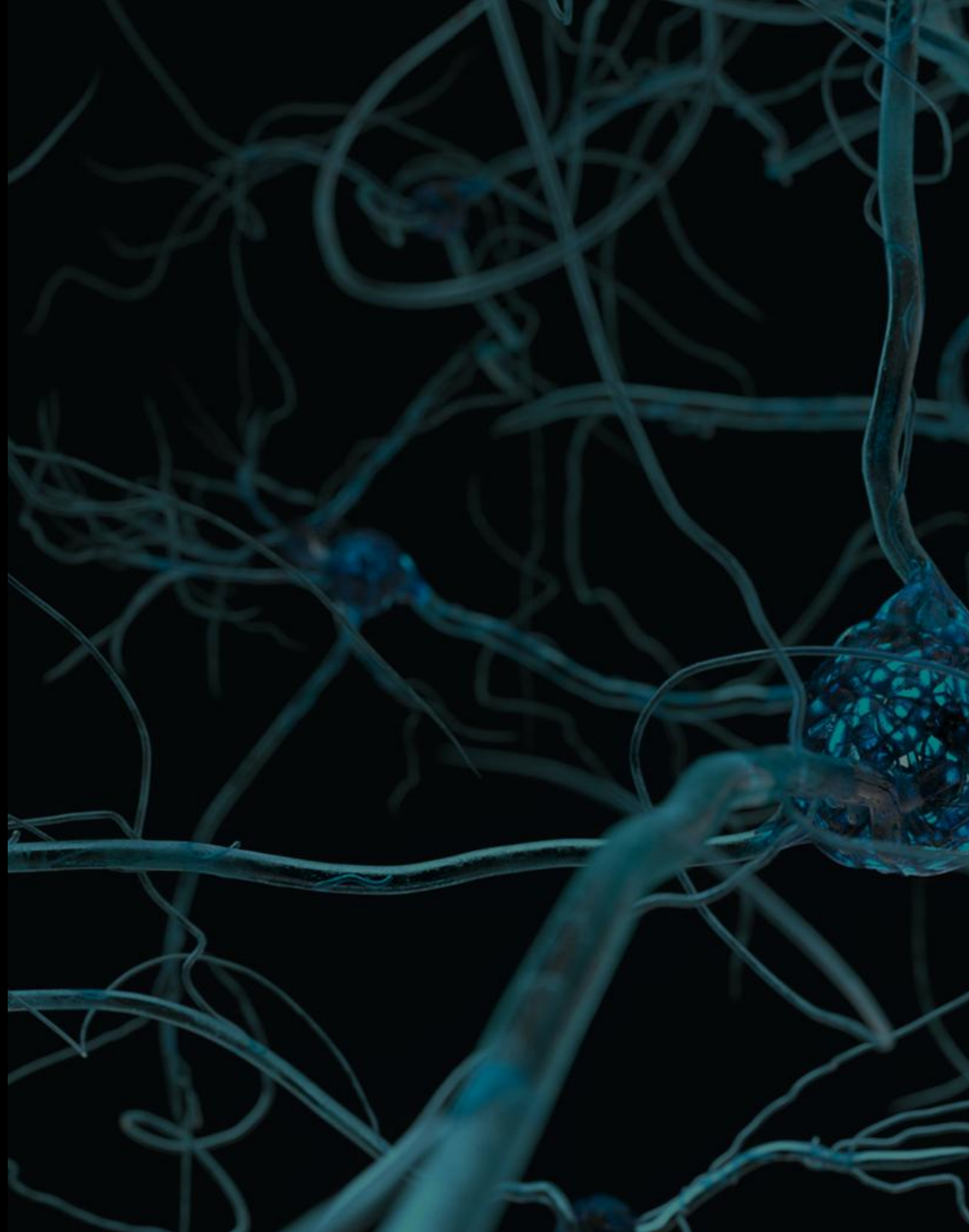
# Modern application pattern

## Kubernetes



# GenAI & Large Language Models (LLMs)

Tom Claes

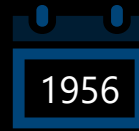


# Artificial Intelligence

## Machine Learning

### Deep Learning

#### Generative AI



## Artificial Intelligence

the field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence

---



## Machine Learning

subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions

---



## Deep Learning

a machine learning technique in which layers of neural networks are used to process data and make decisions

---



## Generative AI

Create new written, visual, and auditory content given prompts or existing data.



# Generative AI

GPT

SORA

Dall-E

Stable-diffusion

Vall-E

Llama 2

Bard

PaLM

Midjourney

...



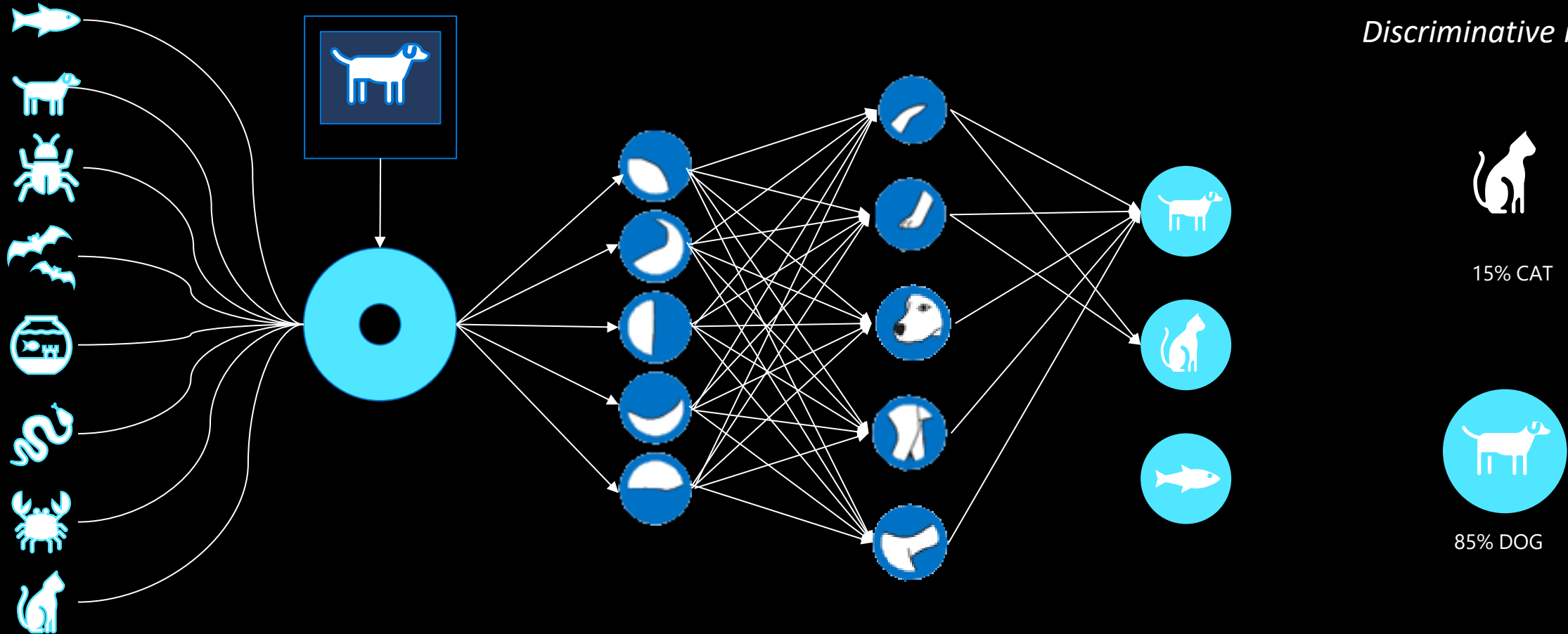
## Generative AI

Create new written, visual, and auditory content given prompts or existing data.

- **Text to text**
- Text to image
- Text to speech
- Text to audio
- Image to text
- Image to image
- Text to video
- Image to video
- ...

# Neural Networks & Deep Learning

*Discriminative model*



## TRAINING

During the training phase, a neural network is fed thousands of **labeled images** of various animals, learning to classify them.

## INPUT

An unlabeled image is shown to the **pretrained network**.

## FIRST LAYER

The neurons respond to different simple shapes, like edges.

## HIDDEN LAYER

Neurons respond to more complex structures.

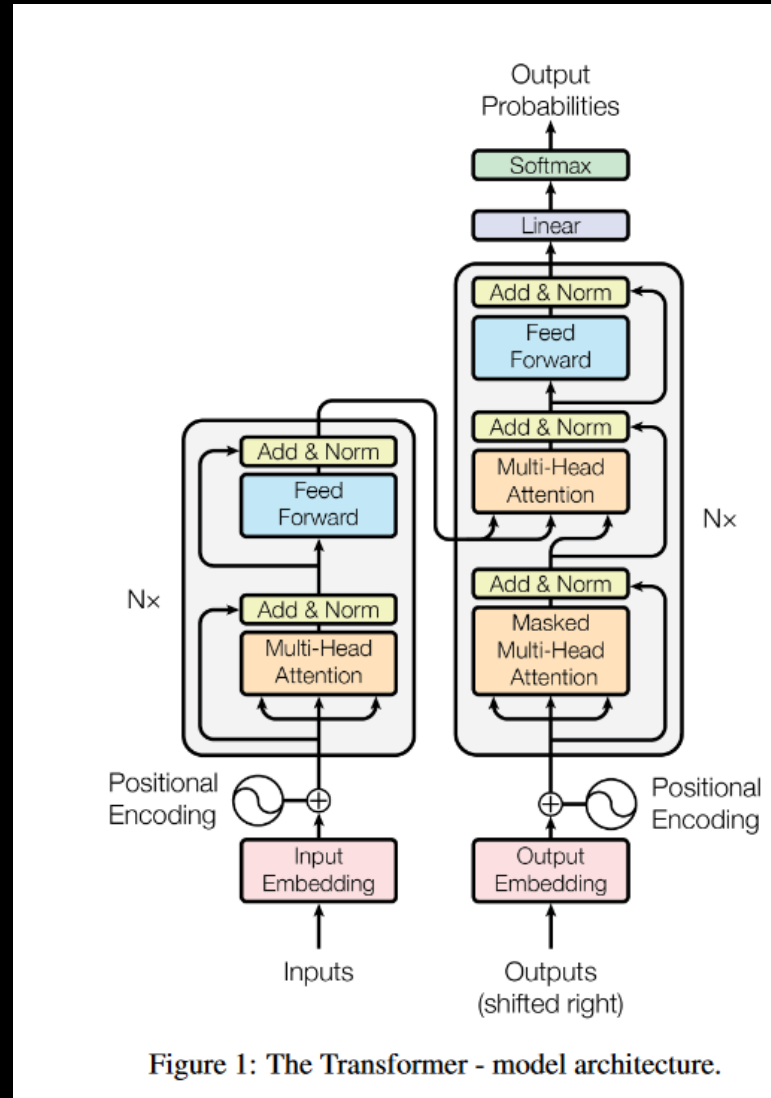
## TOP LAYER

Neurons respond to highly complex, abstract concepts that we would identify as different animals.

## OUTPUT

The network predicts what the object most likely is, based on its training.

# Large Language Models



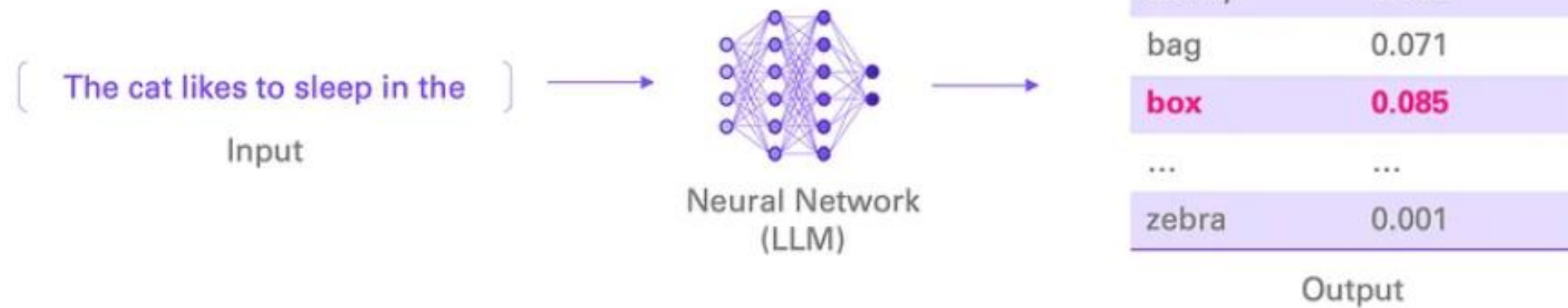
Must read paper: [Attention is all you need](#)

# Large Language Models

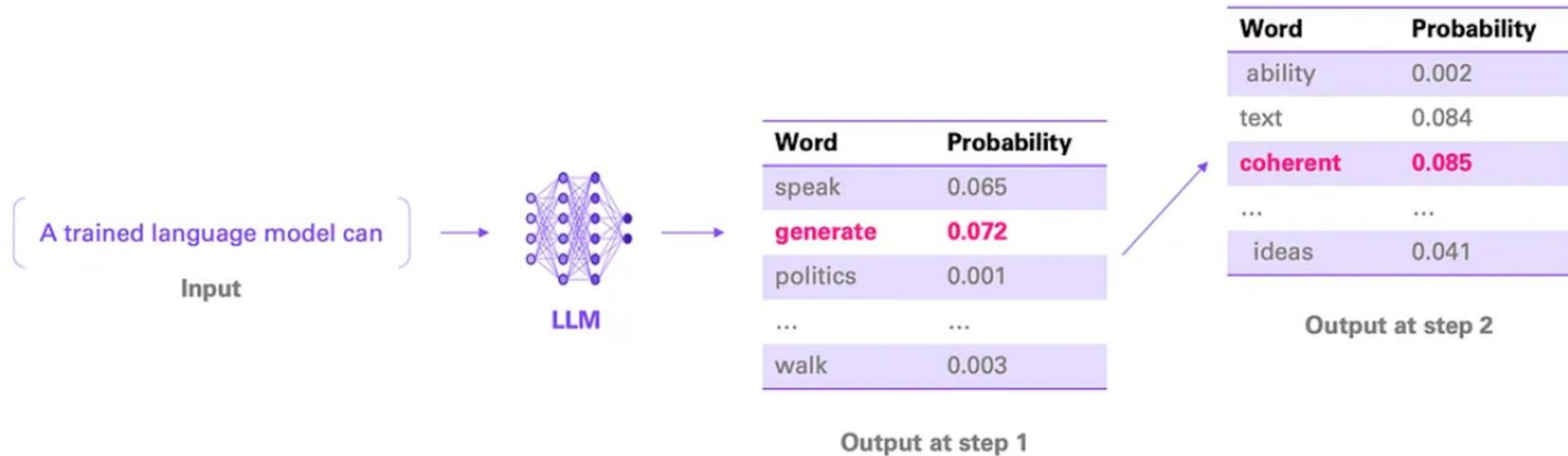
**Large** – 100s of GBs of RAM required to load

**Language** – built for natural language processing

**Model** – pre trained and ready to use



# Generate text by predicting one word at the time



# What does **Generative Pre-trained Transformer (GPT)** mean

```
graph TD; A[What does Generative Pre-trained Transformer (GPT) mean] -- pink arrow --> B[Generative]; A -- purple arrow --> C[Pre-trained]; A -- orange arrow --> D[Transformer];
```

## **Generative**

Means “**next word prediction.**”

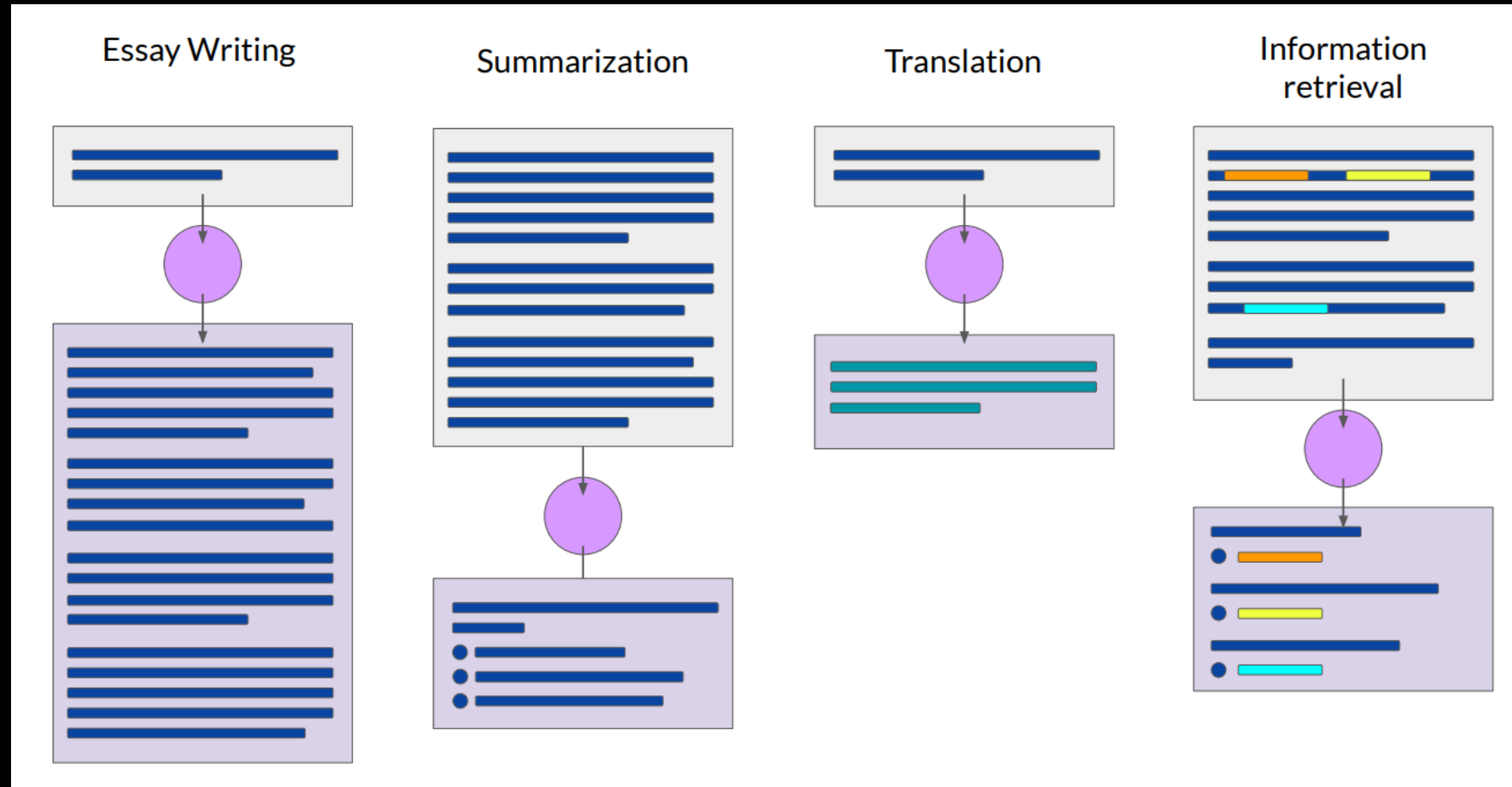
## **Pre-trained**

The LLM is pretrained on massive amounts of text from the internet and other sources.

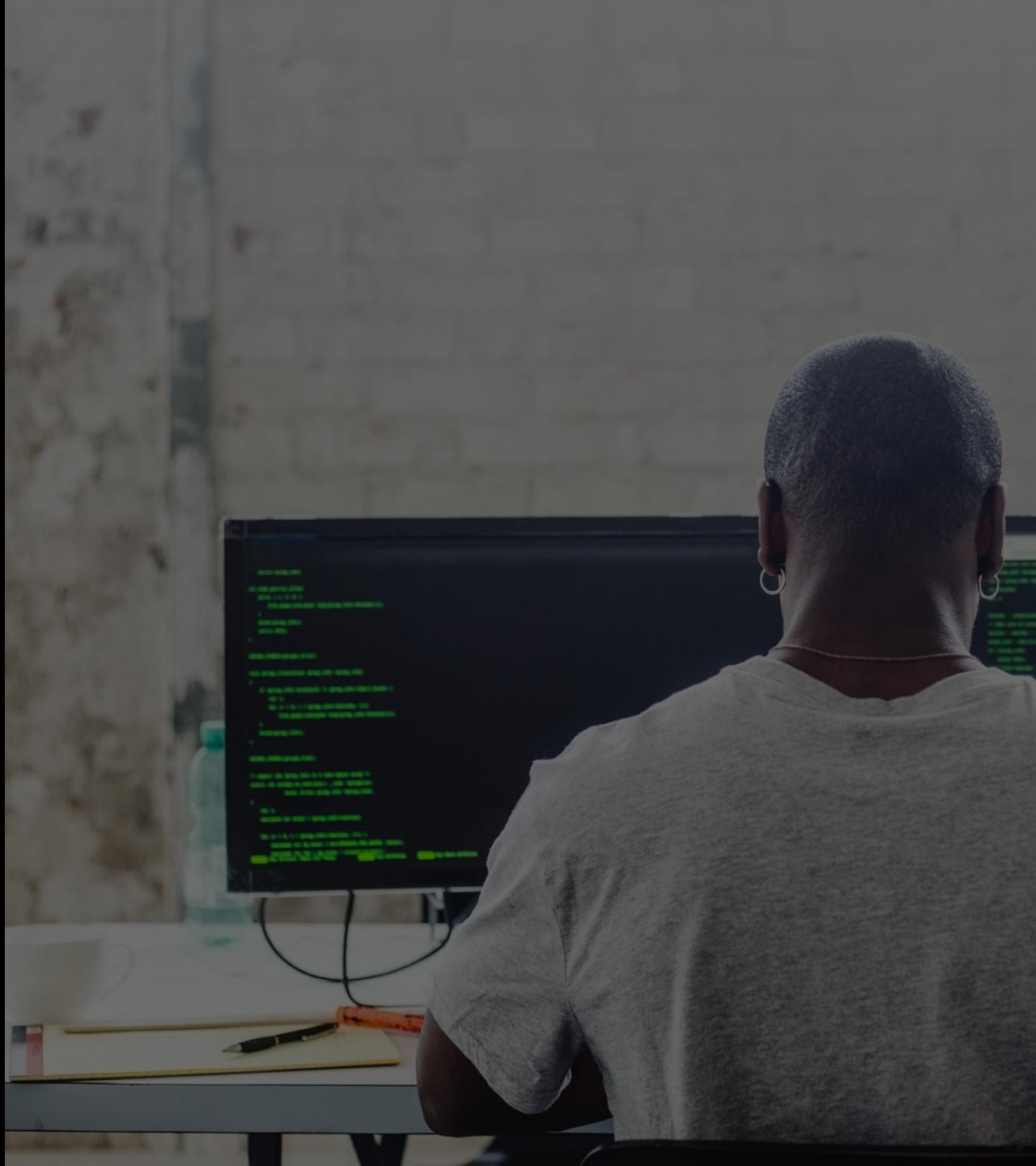
## **Transformer**

The neural network architecture used (introduced in 2017).

## Tasks & Use cases



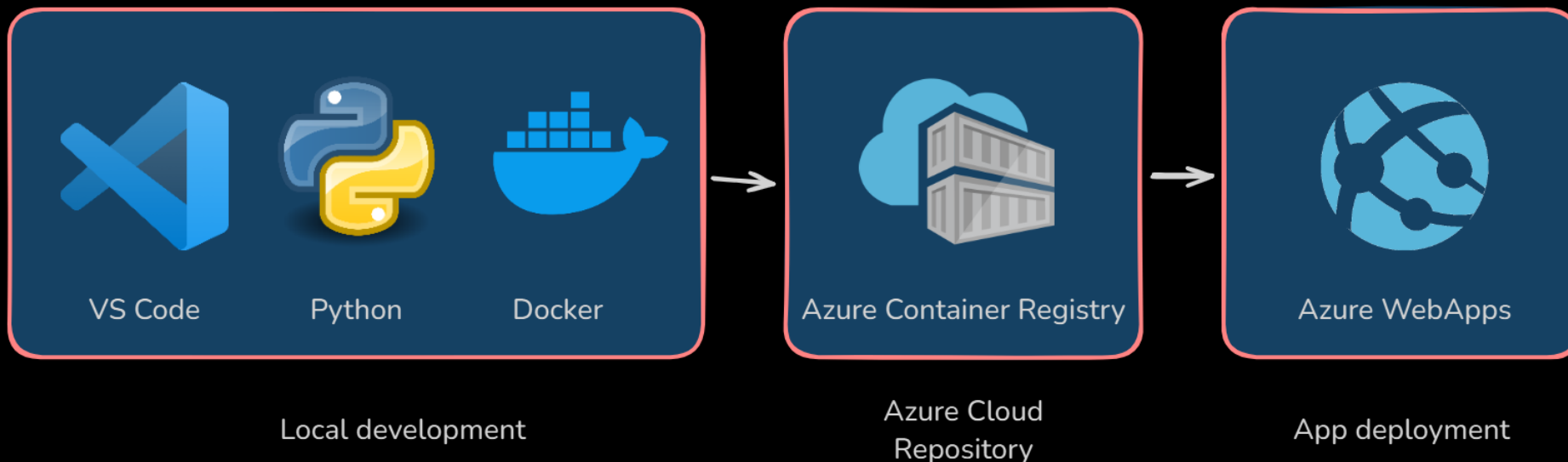
# Hands-on demo





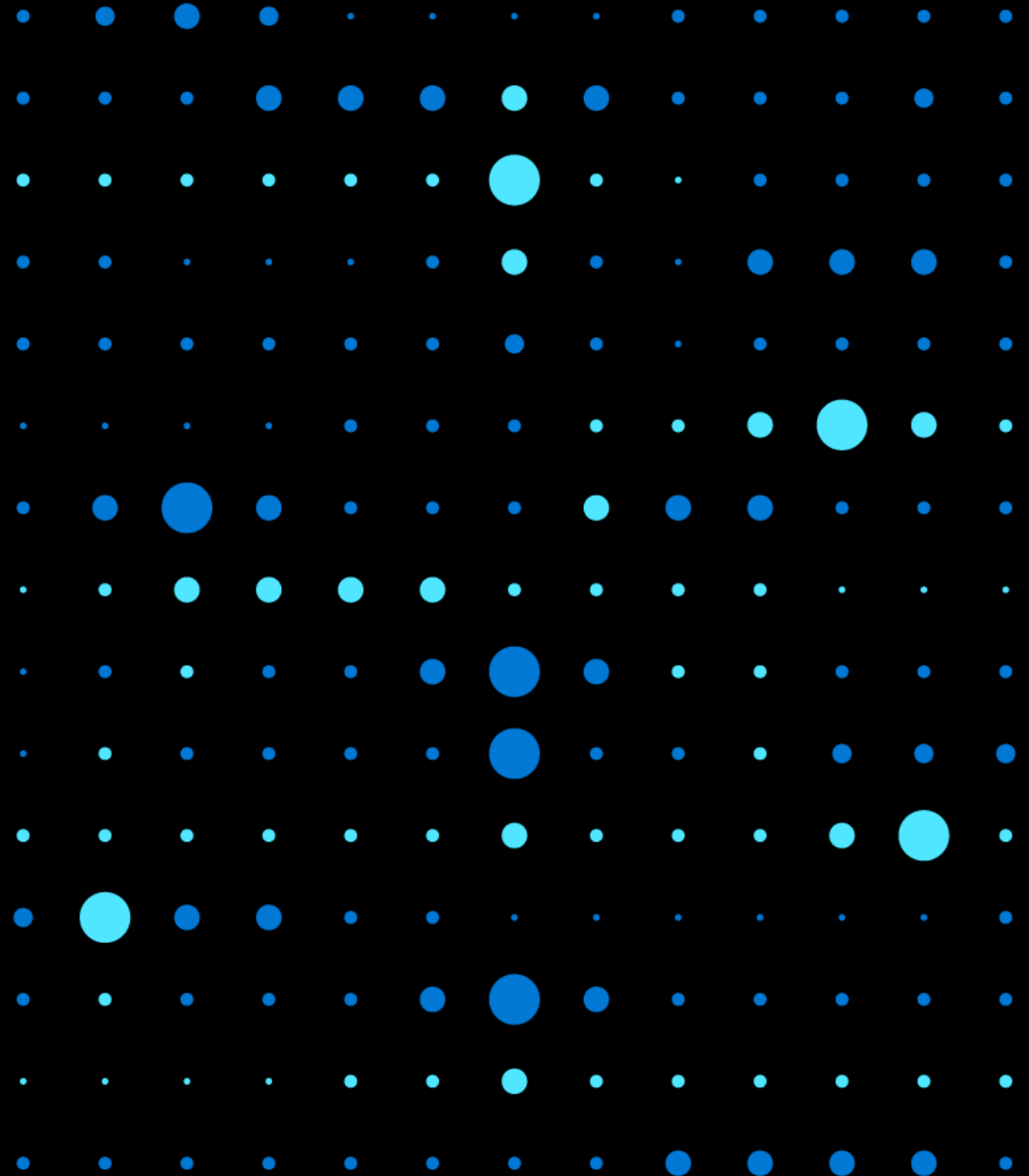
# Hands-on demo

1. Prepare Python Flask application with Azure OpenAI endpoint
2. Create Dockerfile & build Docker image
3. Create Azure Container Registry
4. Push image to ACR
5. Deploy WebApp based on the image

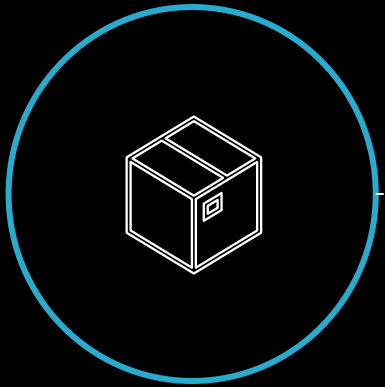


<https://github.com/claestom/emergent-azureopenai-webapp>

# Quick demo



# Recap of the day



Containerization &  
AKS



GenAI & LLMs



Hands-on demo



Containerized LLM  
demo

Questions?





# Cloud & Edge Computing

## Thank you!

Tom Claes – CSA Azure Core

Merlijn Stoffels – Intern CSA Data & AI

