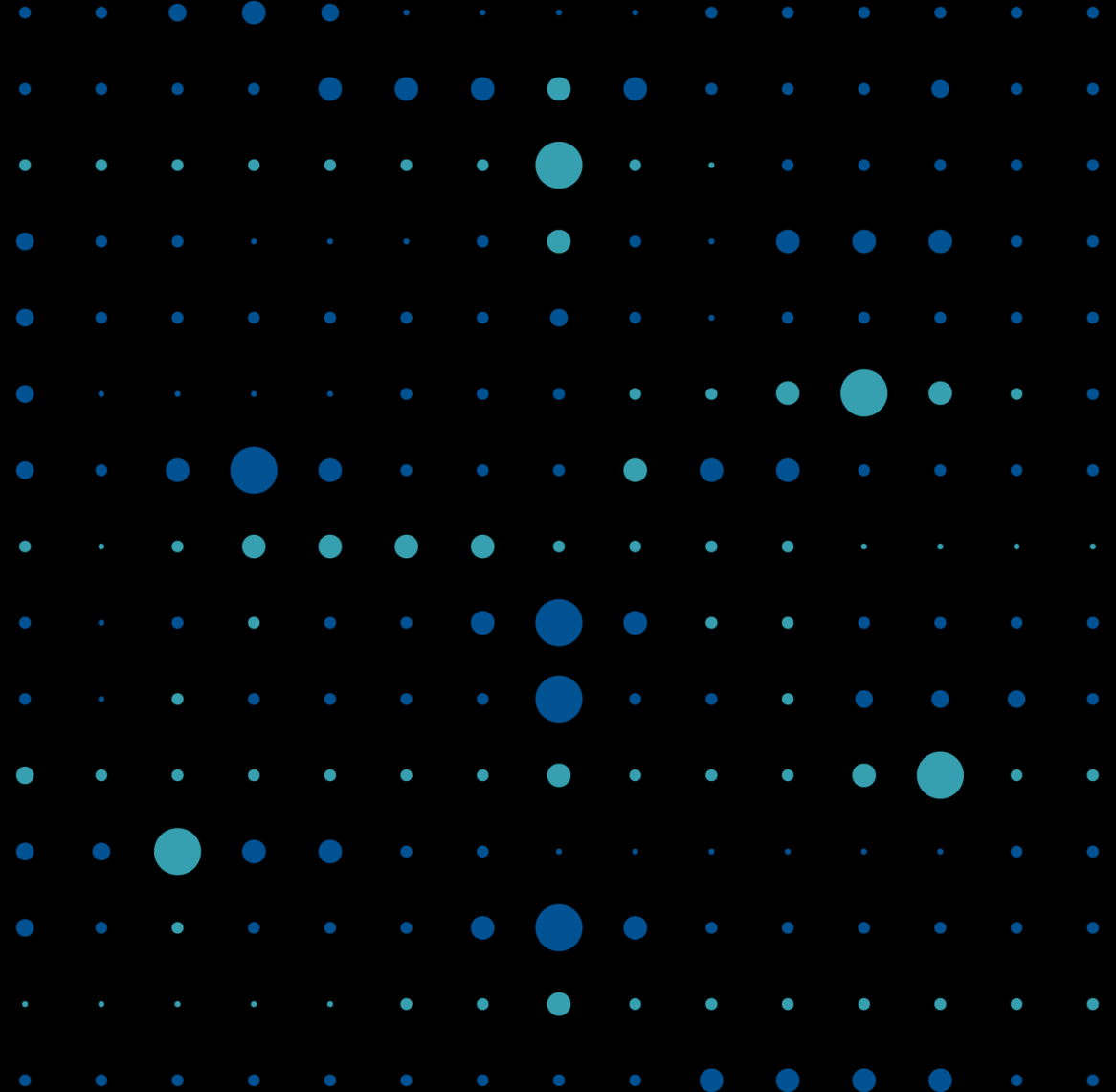


# Core Infra for AI

Core infrastructure in the world of AI

Roman Russev

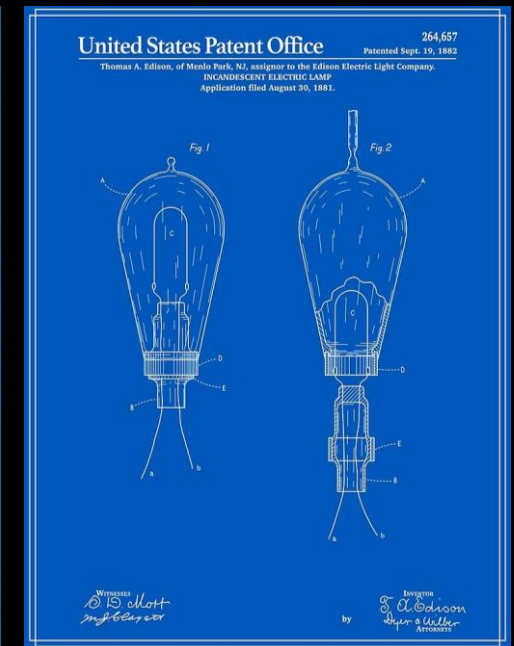
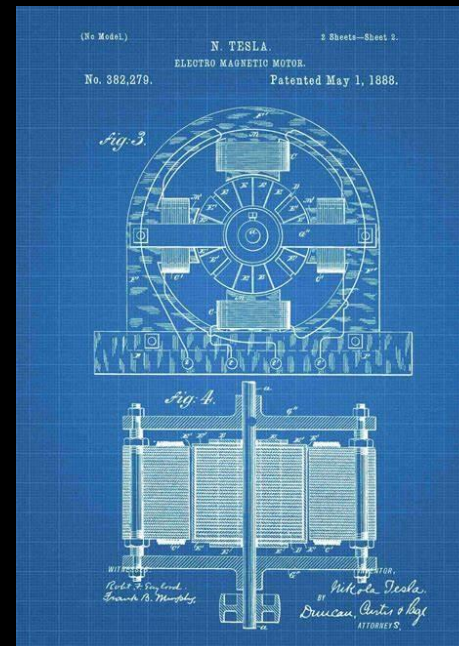
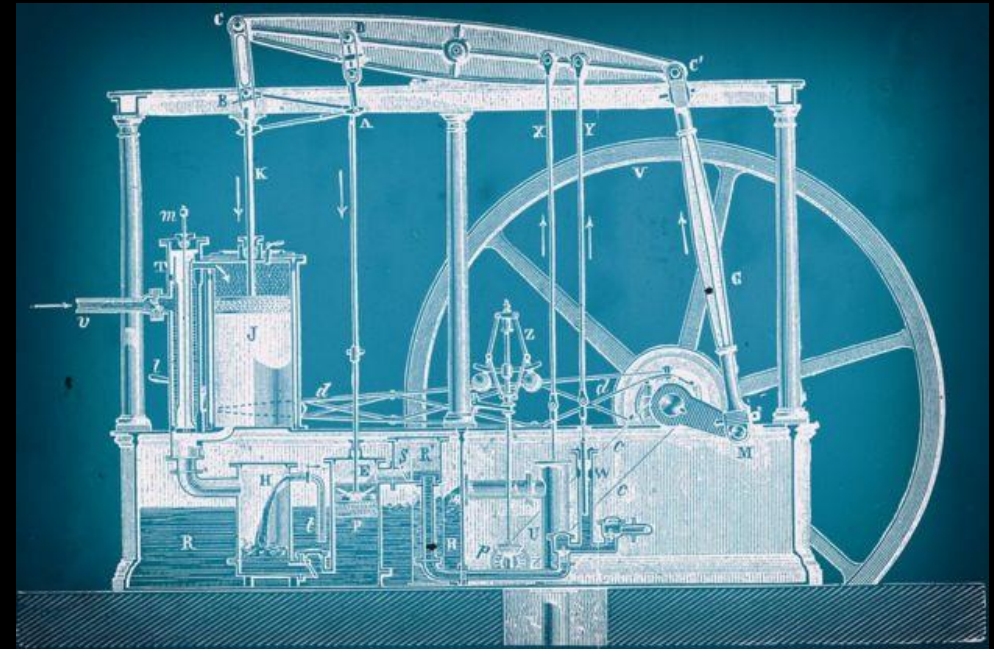
Sr. Partner Solution Architect  
Microsoft  
October, 2023



Industrial revolutions ...  
happened before

Some saw them as a threat

Some saw the potential



AI

It's AI time!

AI will significantly increase productivity



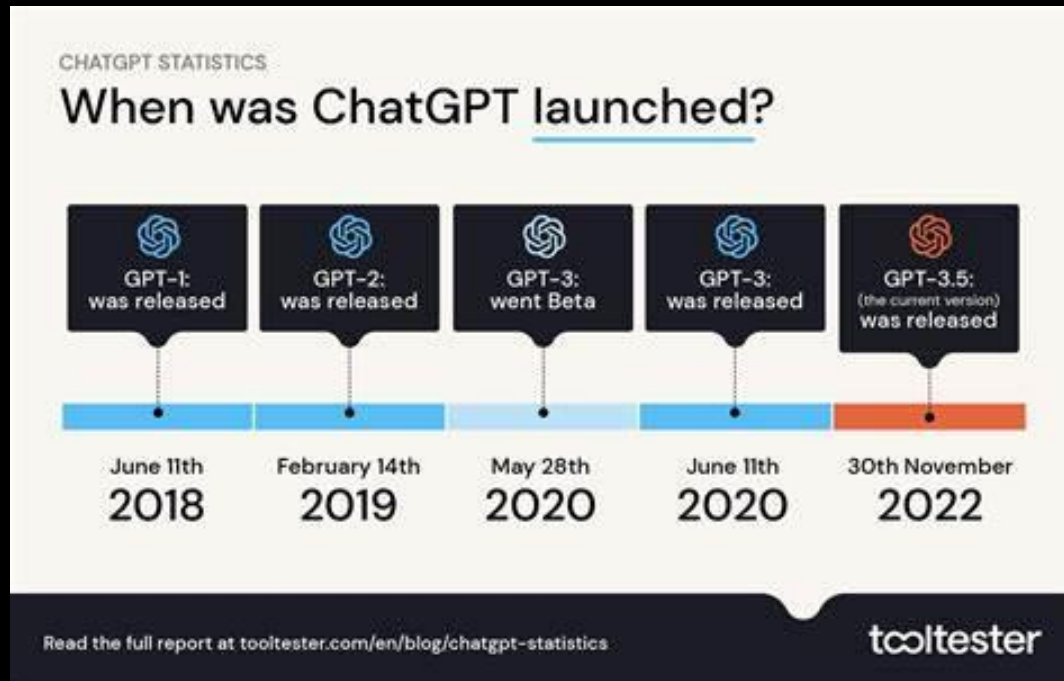
**Sizing the prize**  
PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution  
What's the real value of AI for your business and how can you capitalise?

|   |   |  |
|---|---|--|
| <b>\$15.7tr</b><br>Potential contribution to the global economy by 2030 from AI | <b>+26%</b><br>Up to 26% boost in GDP for local economies from AI by 2030 | <b>~300</b><br>AI use cases identified and rated are captured in our AI Impact Index |
|---|---|--|

AI

It's AI time!

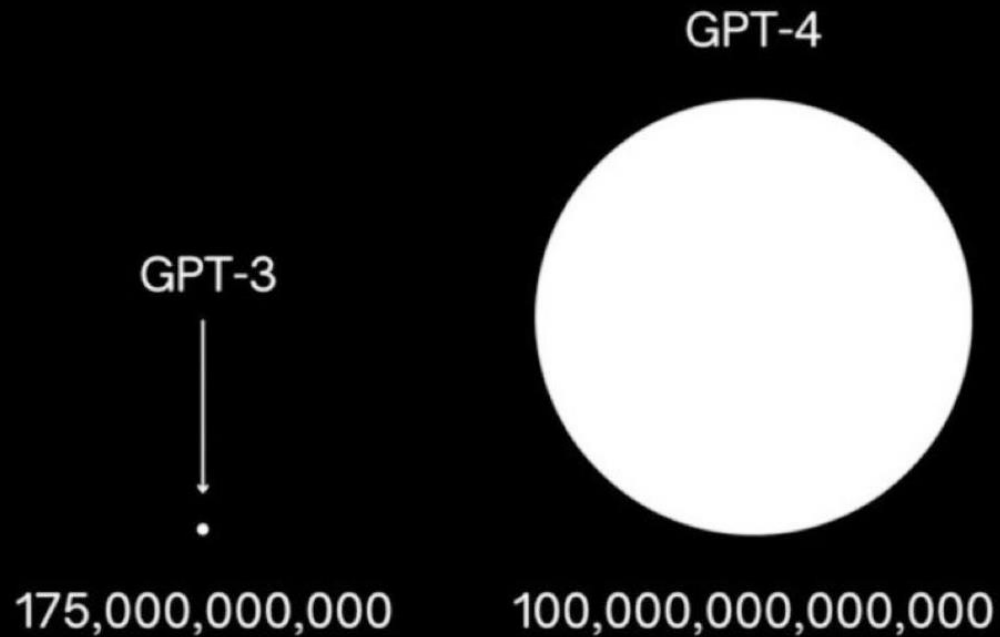
AI matures fast



AI

It's AI time!

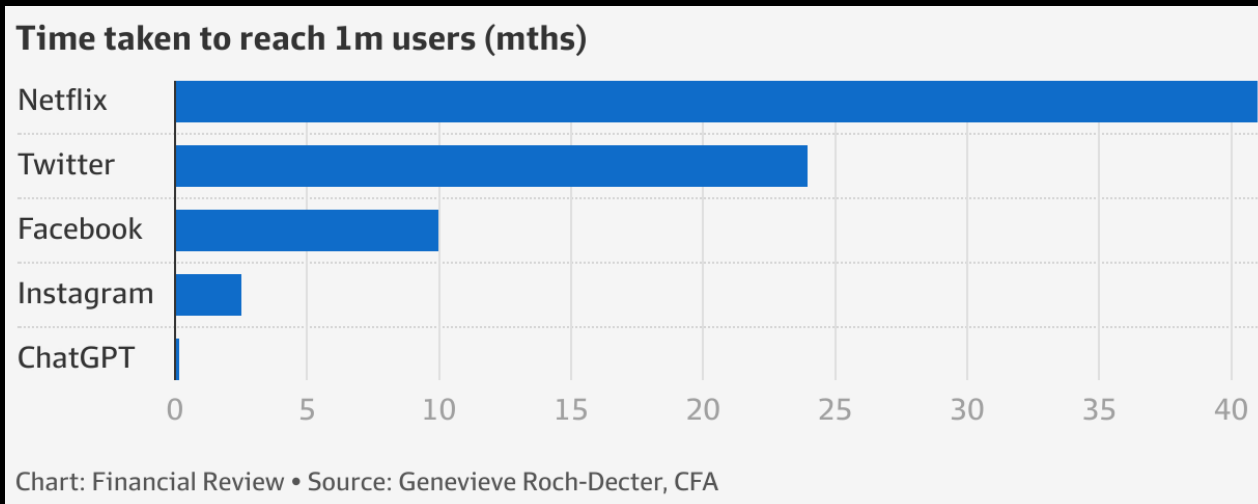
AI matures fast



AI

# It's AI time!

## AI is one of the fastest growing tech ... ever



# AI

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

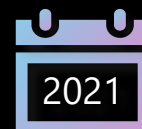
---



## 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





*Ensure that artificial  
general intelligence (AGI)  
benefits humanity*



*Empower every person and  
organization on the planet  
to achieve more*

---

GPT-3.5 and GPT-4

Text

ChatGPT

Conversation

Codex

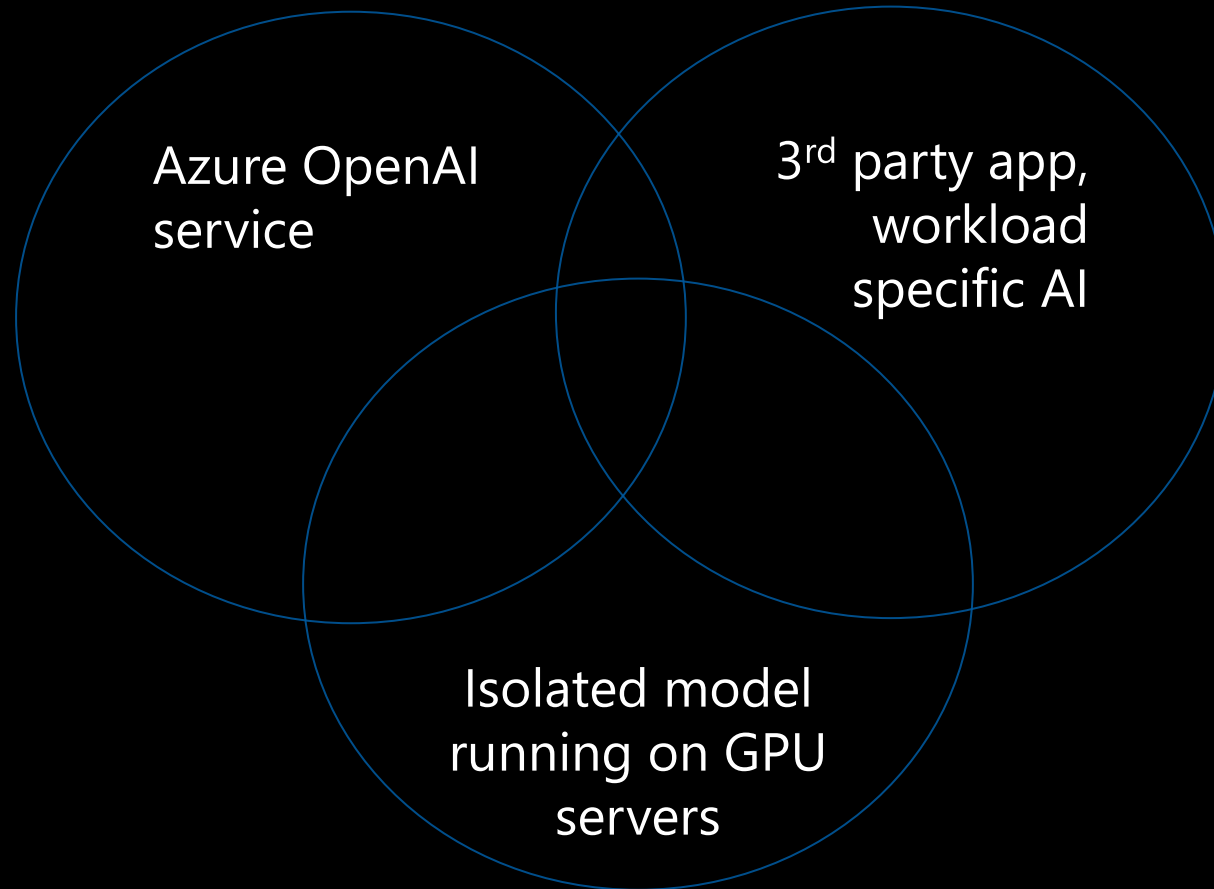
Code

DALL·E 2

Images

# I AI perception

People can be talking about different stuff



Companies will be deploying AI

OpenAI will represent many workloads.

There will be many other AI or AI enriched workloads running in Azure or hybrid infra

Azure OpenAI Service  $\neq$  any AI

# | Pick any industry

AI will be the new enabler

Entertainment

Public

IT and APP development

Retail

Utilities

Manufacturing

Fintech



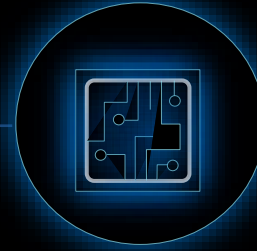
Get relevant  
company data  
understanding



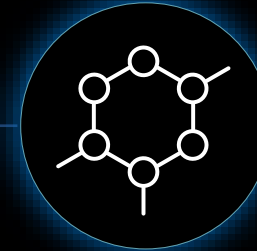
Operations



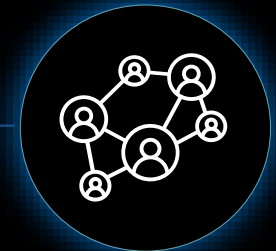
Security



Product, dev,  
manufacturing



Customer  
services



Marketing

# | Copilot Boosts Productivity

Research shows developers...



**88%**

are more  
productive

**74%**

can focus on more  
satisfying work

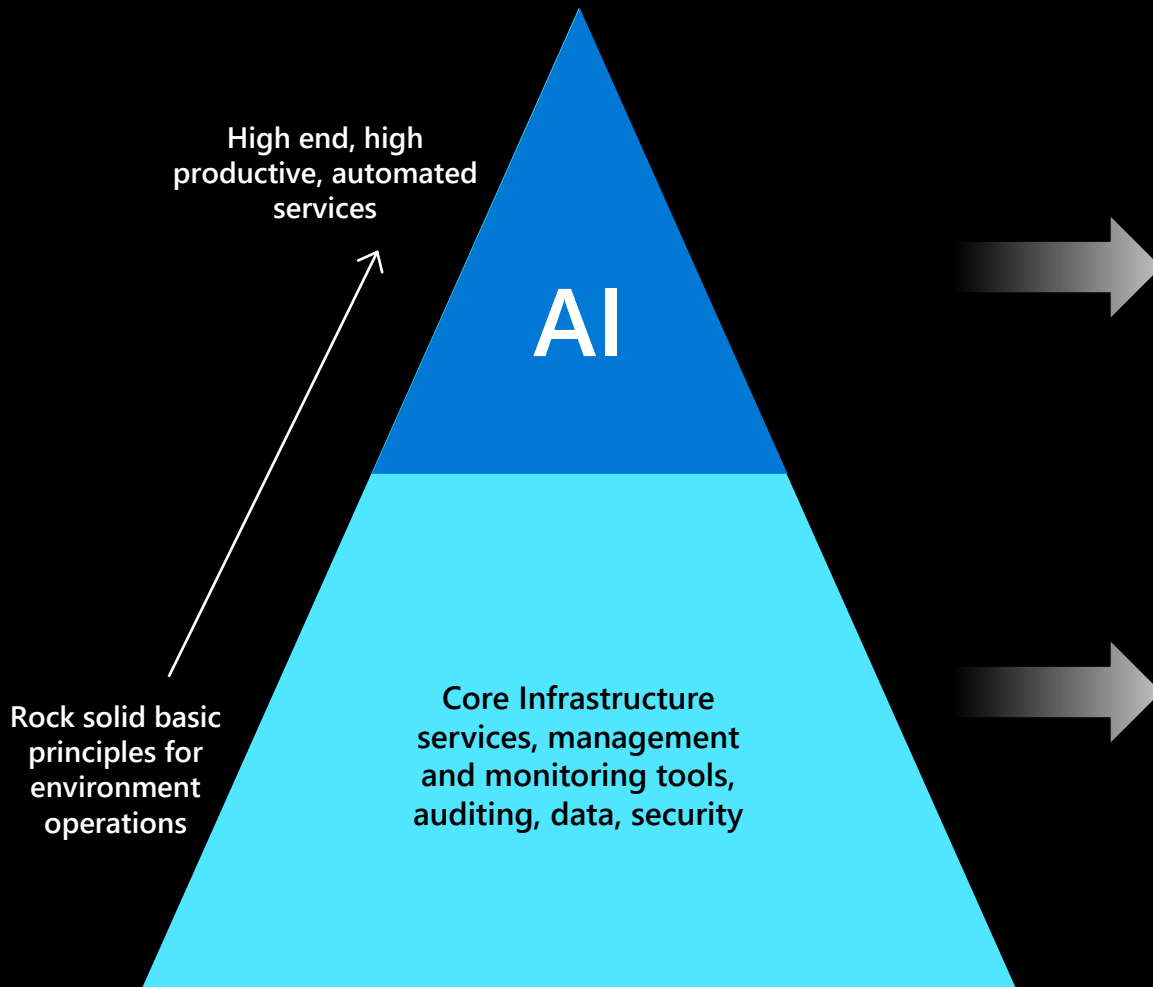
**77%**

spend less time searching  
for information or examples

# I Core infrastructure perception

- Core infrastructure is done and clear
- Everybody has established proper core infra, security and governance
- It is a product, widely adopted and well understood
- How to deploy cloud or hybrid workloads is well documented, guidance is present and everybody is following it
- We don't need additional push
- New, high added value services already think about that stuff, somehow
- Companies already did it with 365 deployment

# I Solution is as good as its weakest part



## High value for companies and customers

- AI usage cannot be defined as only Azure OpenAI Service
- We will enter a massive period of experimentation
- AI will, in time, engage with most of the IT and data landscape
- AI or AI enhanced solutions will be part of the most customer facing or internal solutions
- AI will directly support today deployment of advanced core infra services

## Sets operational margins for AI

- All basic infra areas – identity, networking, policies driven environment etc, create secure and controlled environment for AI usage
- Security ensures the safety and isolation of AI workloads
- Implementation of basic principles in CAF, WAF ensures well designed and managed environment
- Governance secures the long-term operational effectiveness in ever changing environment
- Enables secure areas for experimentation
- Secures data access and distribution
- ALZ brownfield – there is a need on how to implement on the go

# I Can you see any similarities with your environment?

- Well deployed cloud infra requires serious managerial decision
- Most parts are done in the old (on-prem) way. Cloud is operated like DC
- Policy driven environment is rare
- Identity deployment is usually not deployed as complex identity. No RBAC
- Security advanced features are not deployed (defender is good enough)
- ALZ is not deployed or wrongly conceptuated
- Governance is not present
- Data security, classification or data LZ is not deployed

# Risks

experimenting with AI with no infrastructure boundaries

access to internal data from (by process) not accessible data in a very easy way

security breaches (secure score)

data breaches

fast environment degrading (when no governance)

limited spent control

understanding risks does not decrease the value of AI but underlines the need to verify/implement proper base infrastructure and security



# Opportunities

speedup infrastructure, security & data correct deployment

increase Azure solutions proper design and adoption

use the AI experimentation period to setup infrastructure the right way

adopt sandboxing and enable experimentation in secure way

use AI to inflict usage, monitoring and governance of the cloud in effective and secure way

reduce costs

allow faster and secure future deployment

# Azure AI

## Applications



Partner Solutions

## Application Platform

AI Builder



Power BI



Power Apps



Power Automate



Power Virtual Agents

## Scenario-Based Services

Applied AI Services



Bot Service



Cognitive Search



Form Recognizer



Video Indexer



Metrics Advisor



Immersive Reader

## Customizable AI Models

Cognitive Services



Vision



Speech



Language



Decision

Azure OpenAI Service

## ML Platform



Azure Machine Learning



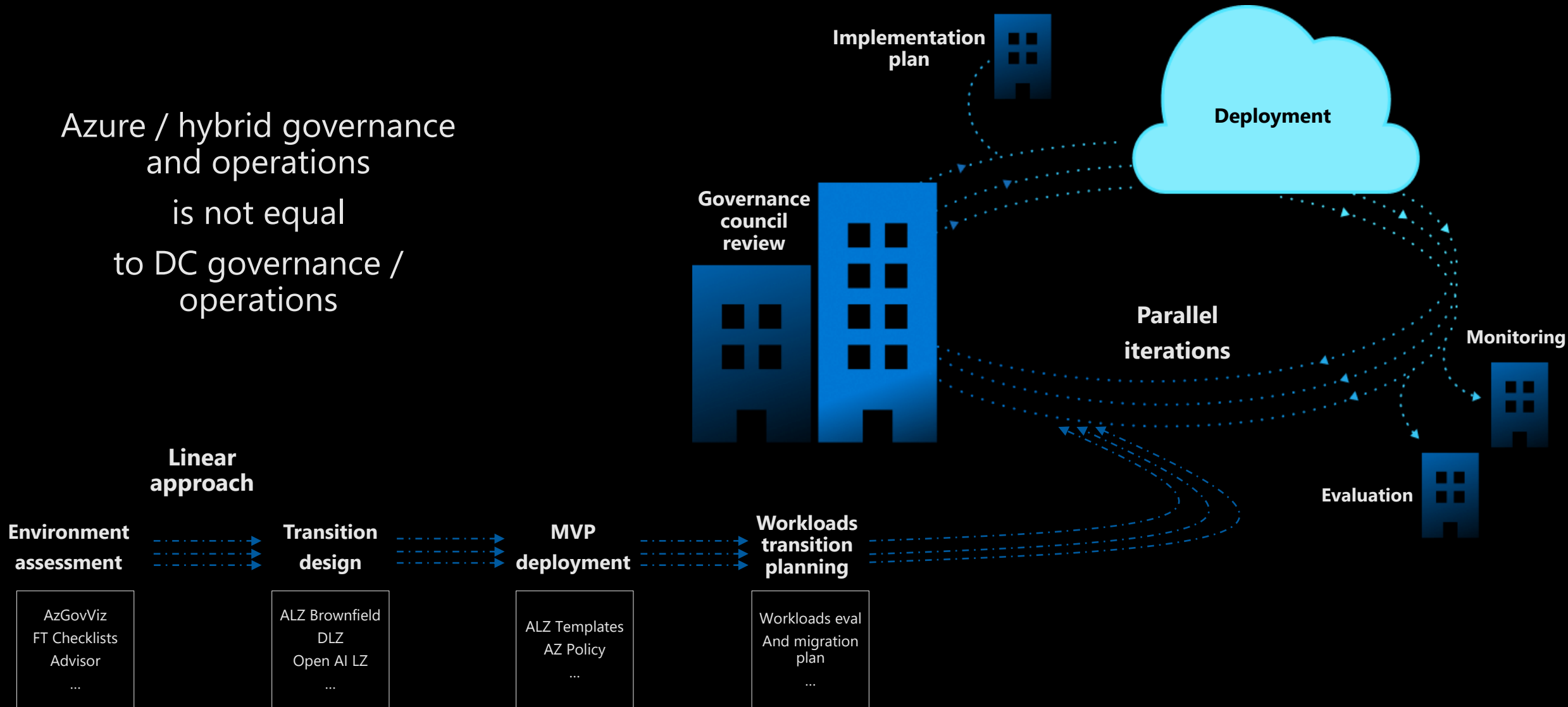
Business Users



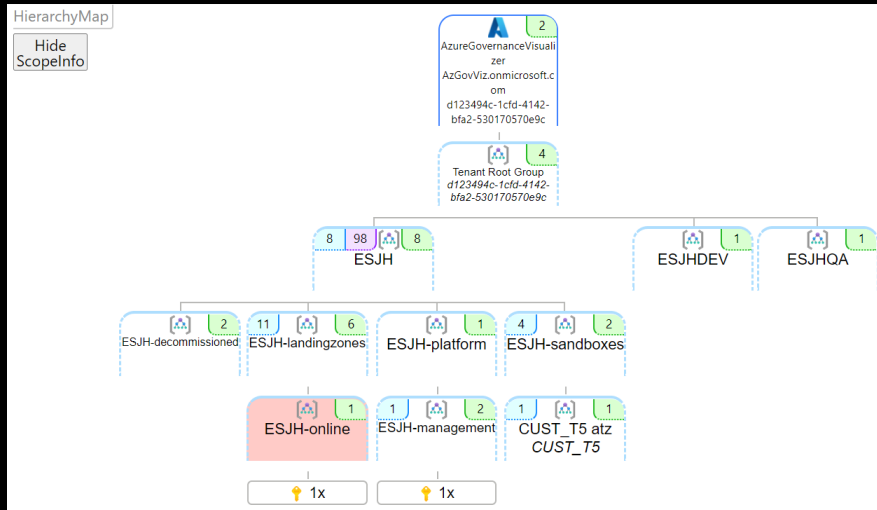
Developers & Data Scientists

# | Approaching infrastructure lifecycle

Azure / hybrid governance and operations is not equal to DC governance / operations



# Environment assessment



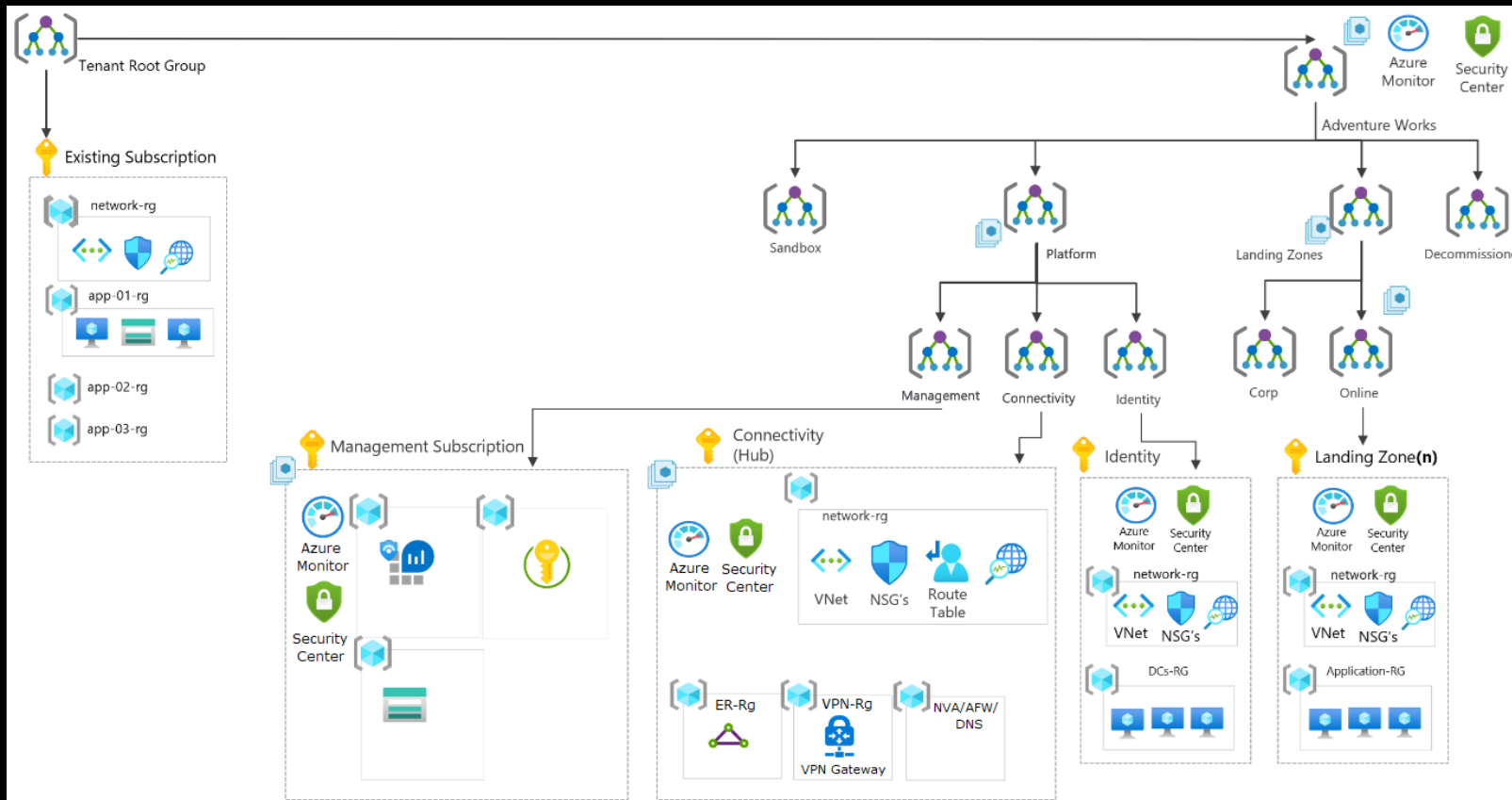
Companies mostly have some Azure environment...

Multiple approach vectors enable environment analysis in depth

- resource organization
- Azure policy status
- Secure score details analysis
- Identity approach incl. RBAC
- Isolation status and networking
- Data landscape
- Governance approach

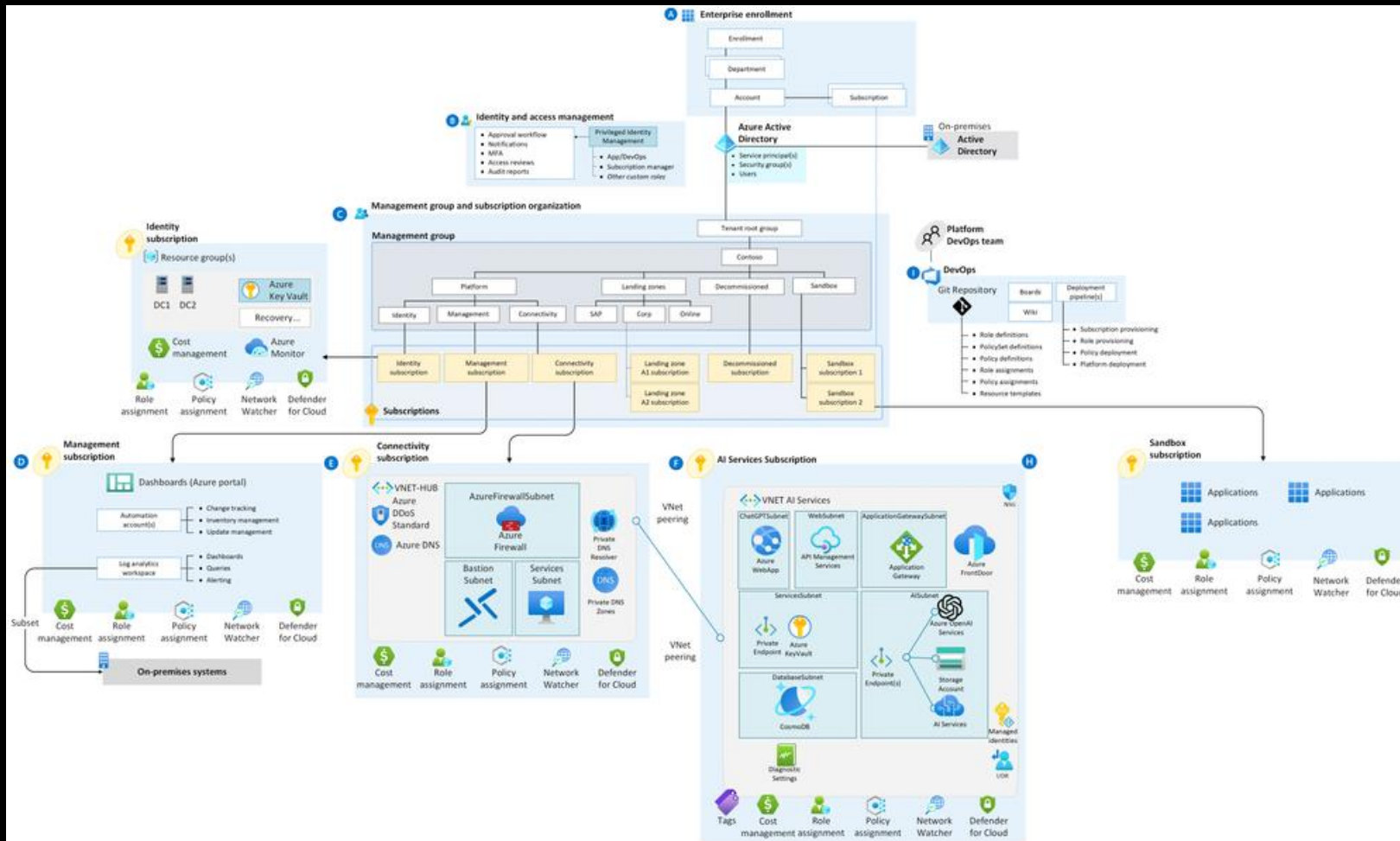
| ID     | Design Area           | Sub Area                    | WAF Pillar  | Checklist item  | Description (optional) | Severity | Status       |
|--------|-----------------------|-----------------------------|-------------|---|------------------------|----------|--------------|
| A03.03 | Azure Billing and Act | Enterprise Agreement        | Security    | Ensure that Accounts are configured to be of the type 'Work and School Account'   |                        | High     | Not verified |
| B01.01 | Governance            | Governance                  | Security    | Leverage Azure Policy strategically, define controls for your environment, using Policy Initiatives to group related policies.                      |                        | High     | Not verified |
| C01.01 | Identity and Access   | Active Directory and Hybrid | Security    | Use managed identities instead of service principals for authentication to Azure services.  |                        | High     | Not verified |
| C03.01 | Identity and Access   | Identity                    | Security    | Implement an emergency access or break-glass accounts to prevent tenant-wide account lockout.   |                        | High     | Not verified |
| C03.03 | Identity and Access   | Identity                    | Security    | Enforce a RBAC model that aligns to your cloud operating model. Scope and Assign across Management Groups and Subscriptions.                        |                        | High     | Not verified |
| C03.05 | Identity and Access   | Identity                    | Security    | Enforce multi-factor authentication for any user with rights to the Azure environments.   |                        | High     | Not verified |
| C03.08 | Identity and Access   | Identity                    | Security    | Only use the authentication type Work or school account for all account types. Avoid using the Microsoft account.                                   |                        | High     | Not verified |
| D01.01 | Management            | App delivery                | Operations  | Add diagnostic settings to save your Azure Front Door WAF's logs. Regularly review the logs to check for attacks and for false positive detections. |                        | High     | Not verified |
| E01.01 | Management            | Fault Tolerance             | Reliability | Leverage Availability Zones for your VMs in regions where they are supported.   |                        | High     | Not verified |
| E01.02 | Management            | Fault Tolerance             | Reliability | Avoid running a production workload on a single VM.   |                        | High     | Not verified |
| F01.10 | Network Topology      | App delivery                | Reliability | Use Traffic Manager to deliver global apps that span protocols other than HTTP/S.   |                        | High     | Not verified |
| F01.13 | Network Topology      | App delivery                | Security    | Deploy your WAF profiles for Front Door in 'Prevention' mode.   |                        | High     | Not verified |
| F01.14 | Network Topology      | App delivery                | Security    | Avoid combining Azure Traffic Manager and Azure Front Door.   |                        | High     | Not verified |
| F01.15 | Network Topology      | App delivery                | Security    | Use the same domain name on Azure Front Door and your origin. Mismatched host names can cause subtle bugs.  |                        | High     | Not verified |
| F01.19 | Network Topology      | App delivery                | Reliability | Use Azure NAT Gateway instead of Load Balancer outbound rules for better SNAT scalability.  |                        | High     | Not verified |
| F01.20 | Network Topology      | App delivery                | Operations  | Use managed TLS certificates with Azure Front Door. Reduce operational cost and risk of outages due to certificate renewals.                        |                        | High     | Not verified |
| F03.02 | Network Topology      | at Hub and spoke            | Cost        | Ensure that shared networking services, including ExpressRoute gateways, VPN gateways, and Azure Firewall or partner NVAs in the central-hub.       |                        | High     | Not verified |
| F03.10 | Network Topology      | at Hub and spoke            | Reliability | Use the setting 'Allow traffic to remote virtual network' when configuring VNet peering.  |                        | High     | Not verified |

# Transition design



- evaluate app dependencies
- understand ownership and access rights
- identify non movable resources
- AI scope definition
- apply learnings to combined AI template (ALZ + DLZ + other accelerators)
- set MVP boundaries, deployment stages and stage contents

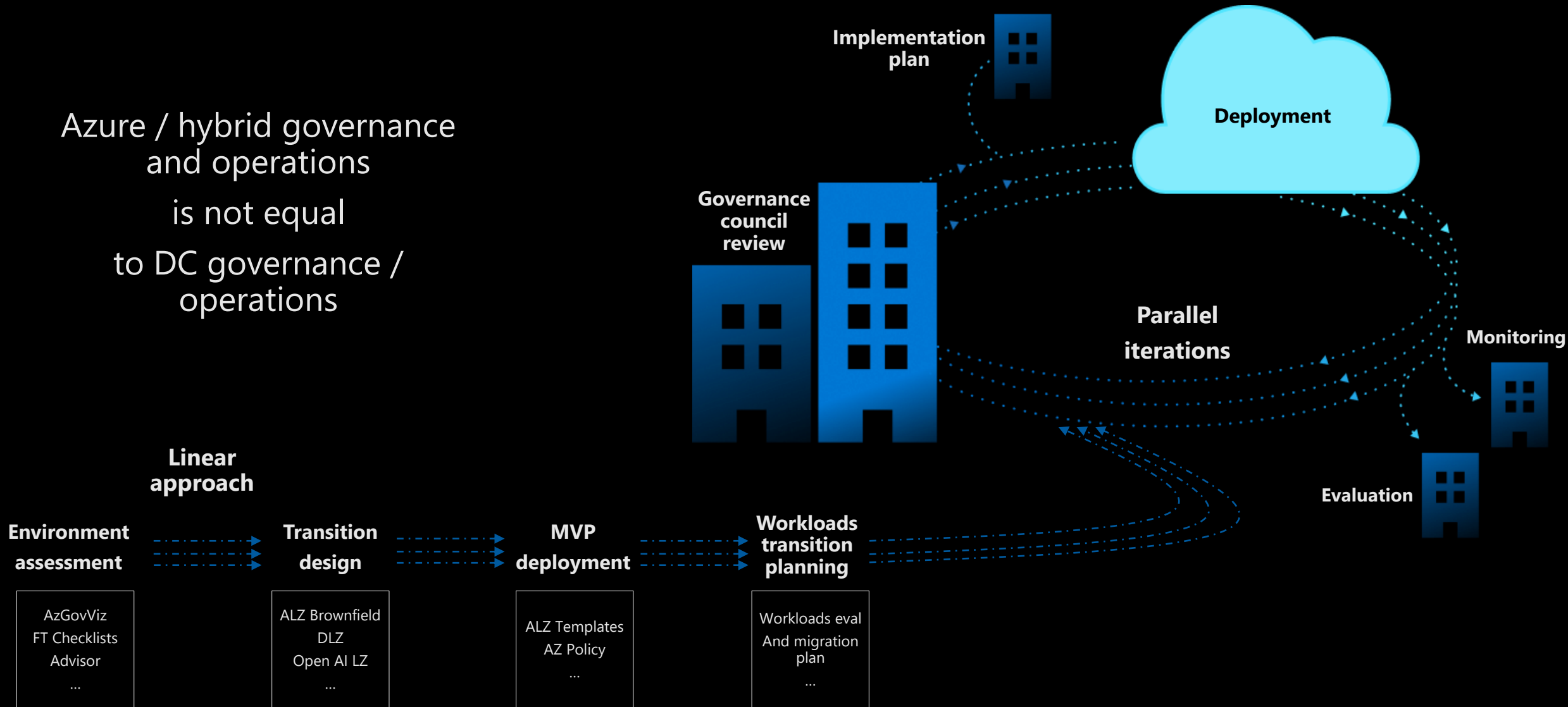
# MVP deployment



- based on ALZ variant
- apply accelerators
- apply additional extensions
- deploy parallel to production environment

# | Approaching infrastructure lifecycle

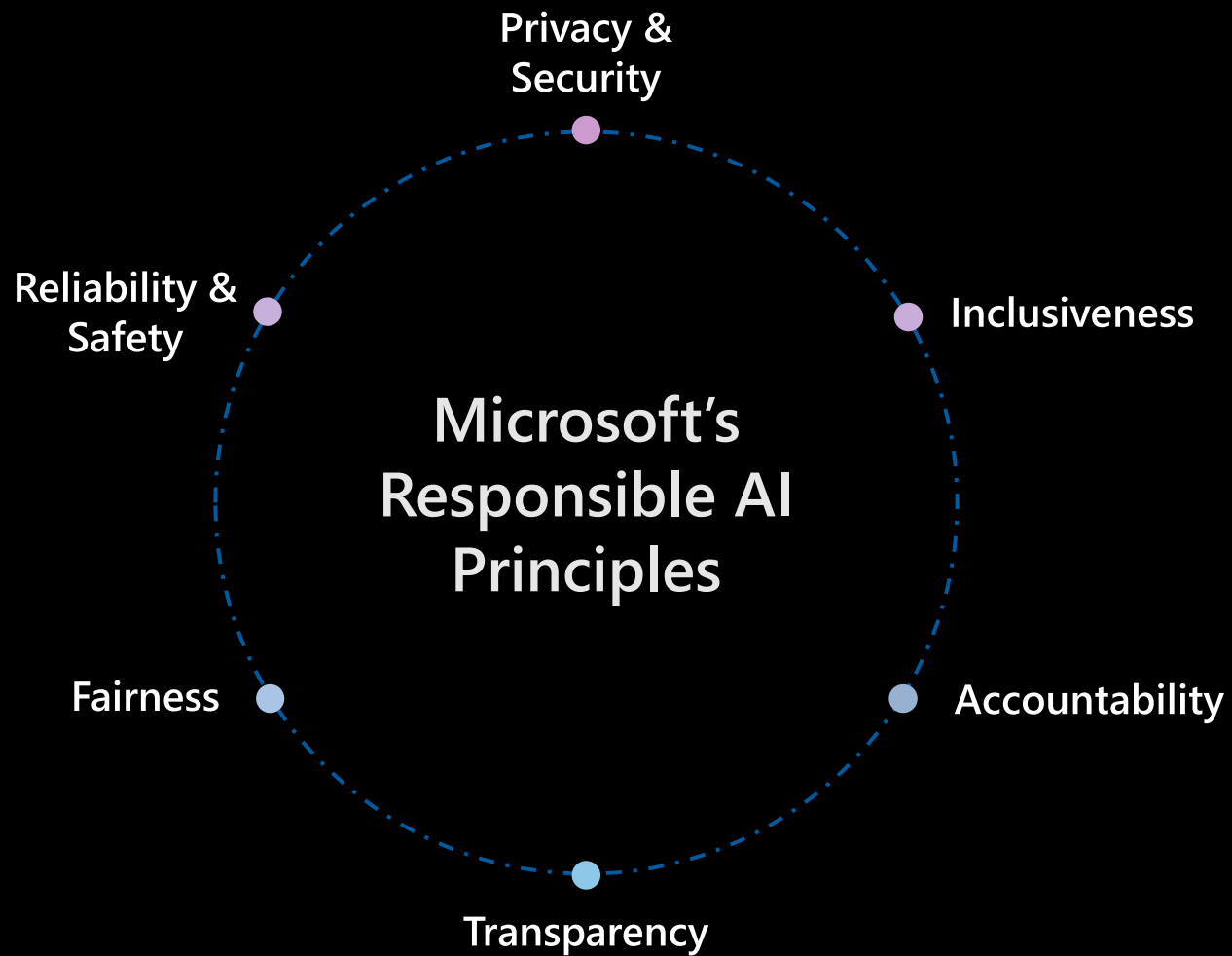
Azure / hybrid governance and operations is not equal to DC governance / operations



# I Where you see your environment in 1 year?

- Deployed cloud / hybrid infra includes BDM req, follows CAF and WAF.
- Infrastructure is Policy driven
- Identity is a complex unique identity
- Security, zero trust based advanced features are deployed
- ALZ is well conceptuated for future extension, follows business strategy, provides secure production, test and sandboxed environments, is cost optimized, includes automation and supports operations
- Governance cycle is regular and effective, SecOps is present, Cost Opt
- Data security, classification is present





## Building blocks to enact principles



Tools and processes



Training and practices



Rules



Governance

# | Microsoft Azure Cloud

## Runs on trust

Your data is your data

---

Data is stored encrypted in *your Azure subscription*

Your data from any fine-tuning is not used to train the foundation AI models

---

Azure OpenAI Service provisioned in *your Azure subscription*

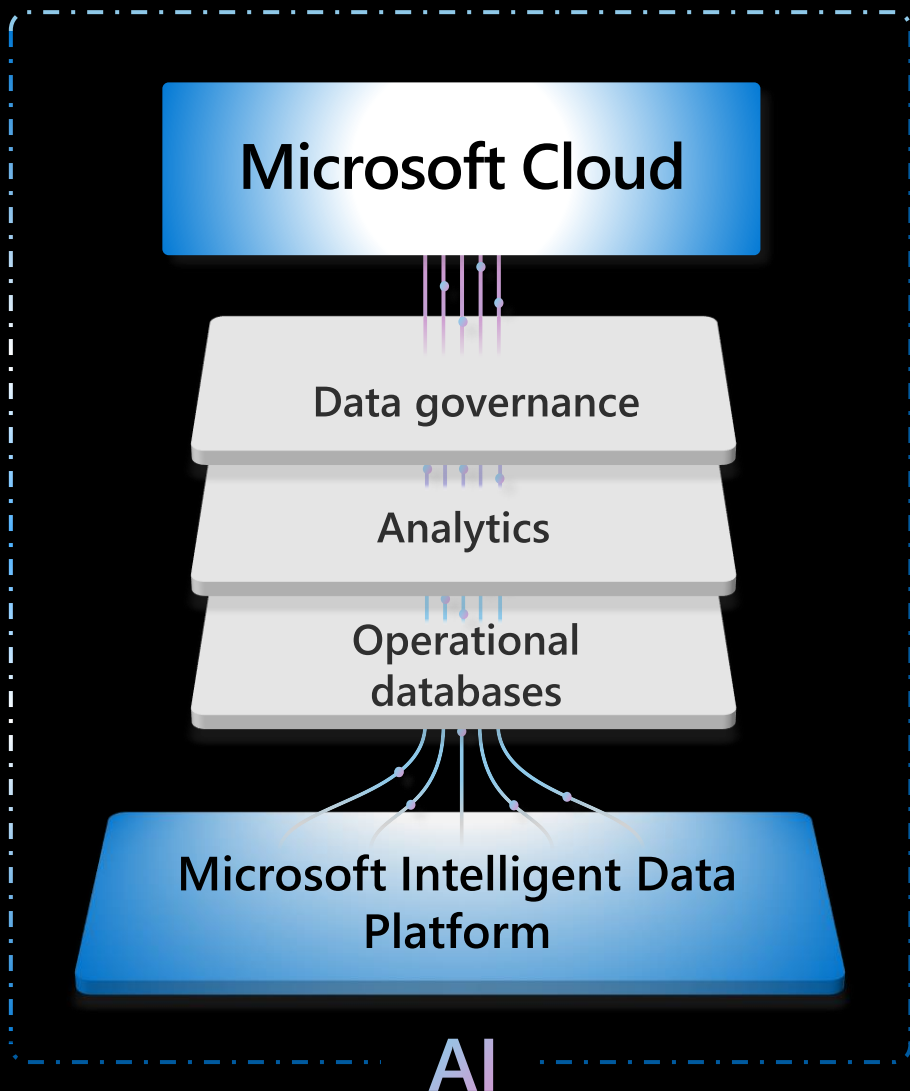
Model fine tuning stays in *your Azure subscription* and never moves into the foundation AI models

Your data is protected by the most comprehensive enterprise compliance and security controls

Encrypted with Customer Managed Keys

Private Virtual Networks, Role Based Access Control

Soc2, ISO, HIPPA, CSA STAR Compliant



Deployed within your Azure subscription, secured by you, accessed only by you, and tied to your datasets and applications



Enterprise-grade security with role-based access control (RBAC) and authentication



Secure networking through private endpoints and VNETs

## Next steps

Define clear business expectations

Launch parallel streams

- Infrastructure optimization

- AI evaluation

Initiate regular governance cycle

Use documentation, learning and templates

- self assessment scripts and guidance for partners

- checklists for AI optimized ALZ

- AI optimized ALZ guidance

- Microsoft learning

Q & AI