Microsoft Azure

Core Infra for AI

Core infrastructure in the world of AI

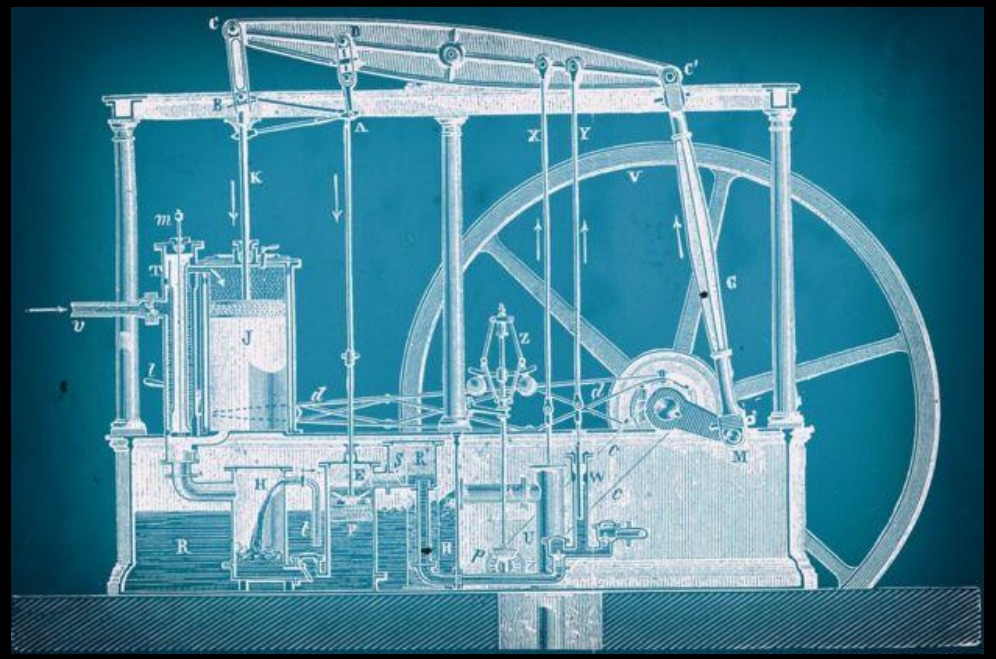
Roman Russev

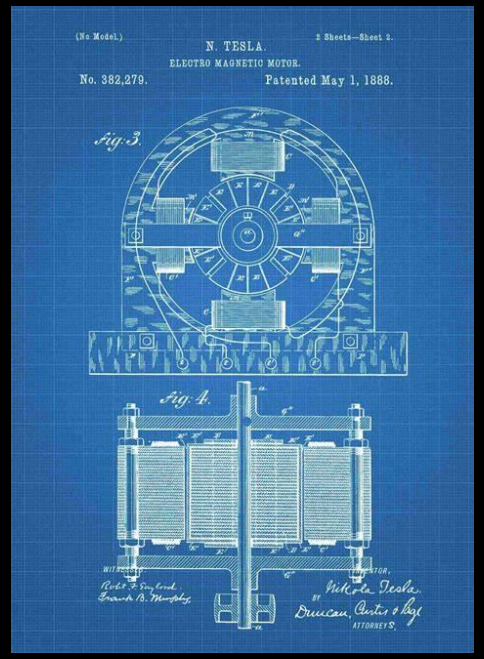
Sr. Partner Solution Architect
Microsoft
October, 2023

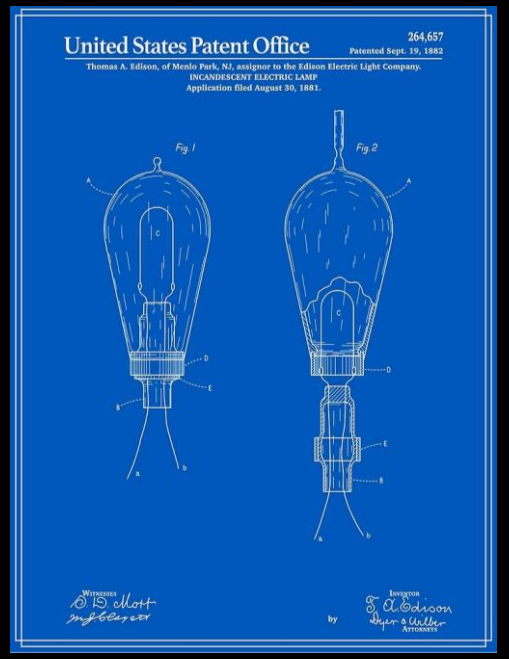
Industrial revolutions … happened before

Some saw them as a threat

Some saw the potential








AI

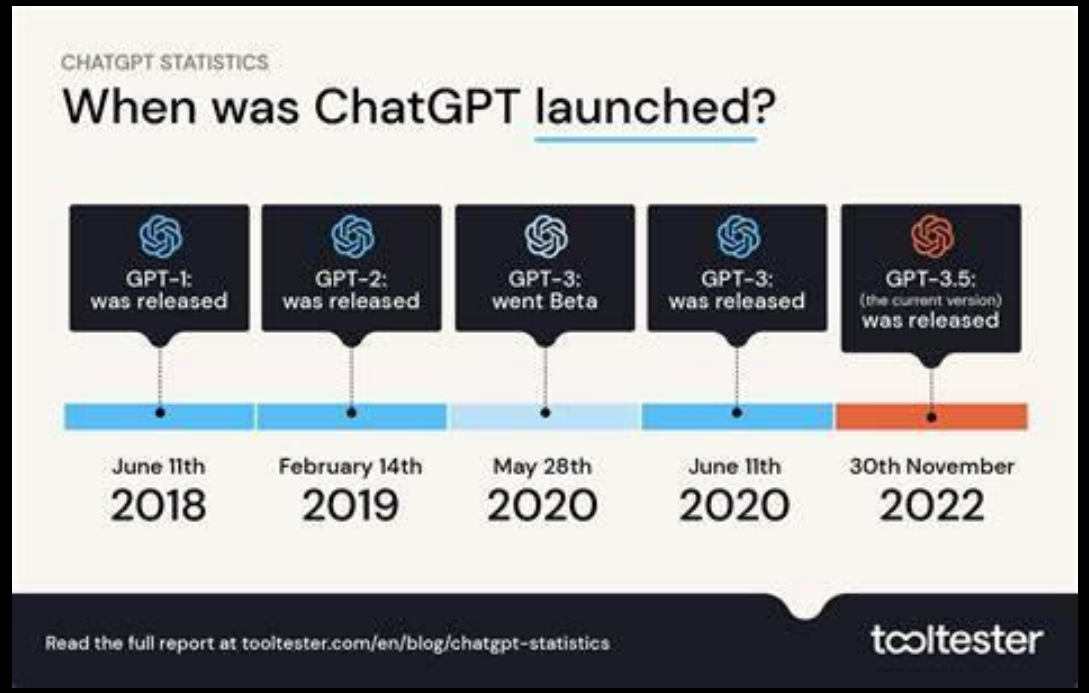
It’s AI time!

AI will significantly increase productivity

AI

It’s AI time!

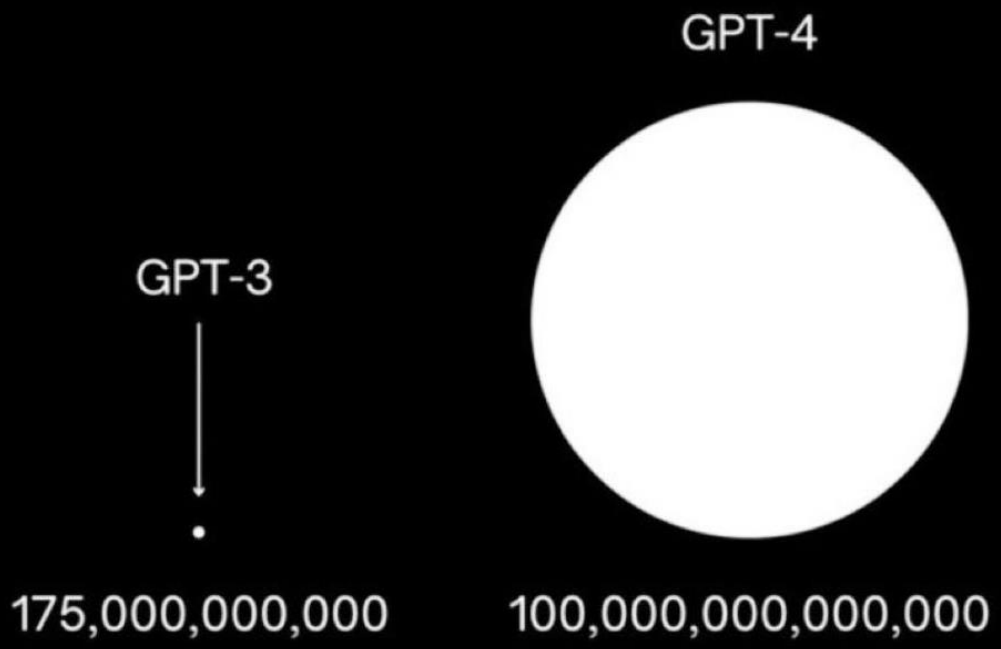
AI matures fast



AI

It’s AI time!

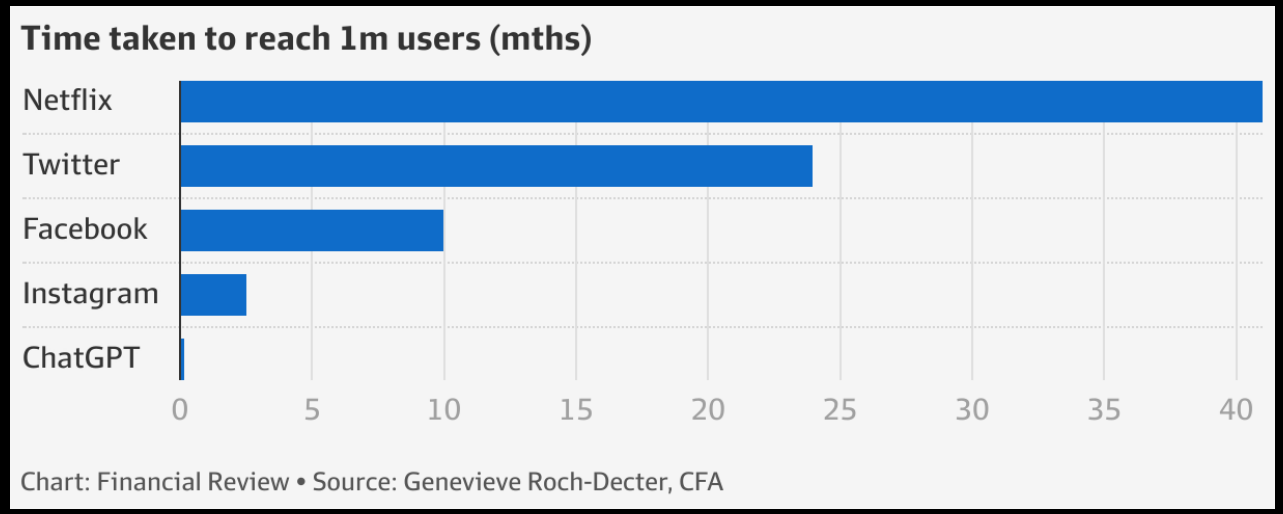
AI matures fast



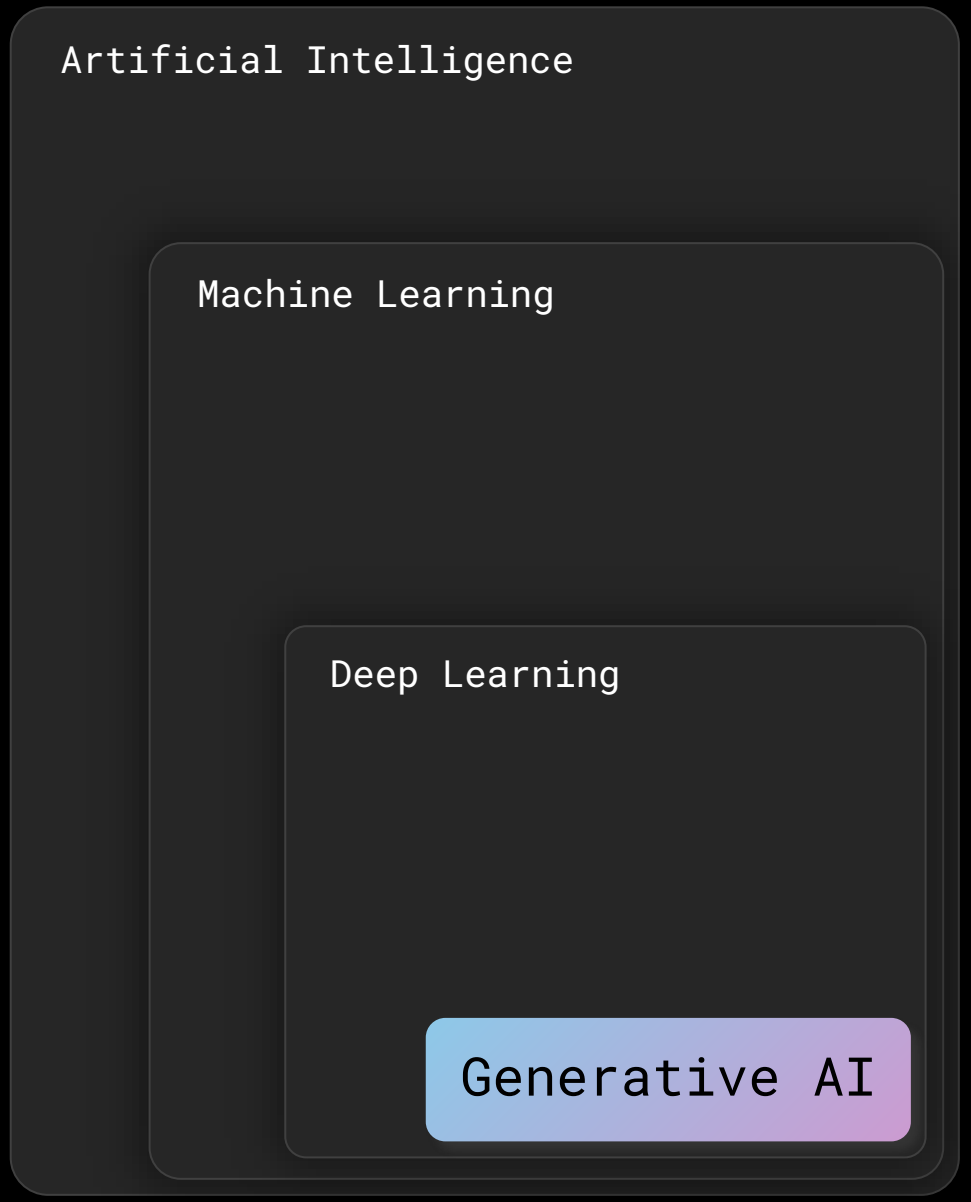
AI

It’s AI time!

AI is one of the fastest growing tech … ever



AI



1956

Artificial Intelligence

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

1997

Machine Learning

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

2017

Deep Learning

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

2021

Generative AI

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

Ensure that artificial general intelligence (AGI) benefits humanity

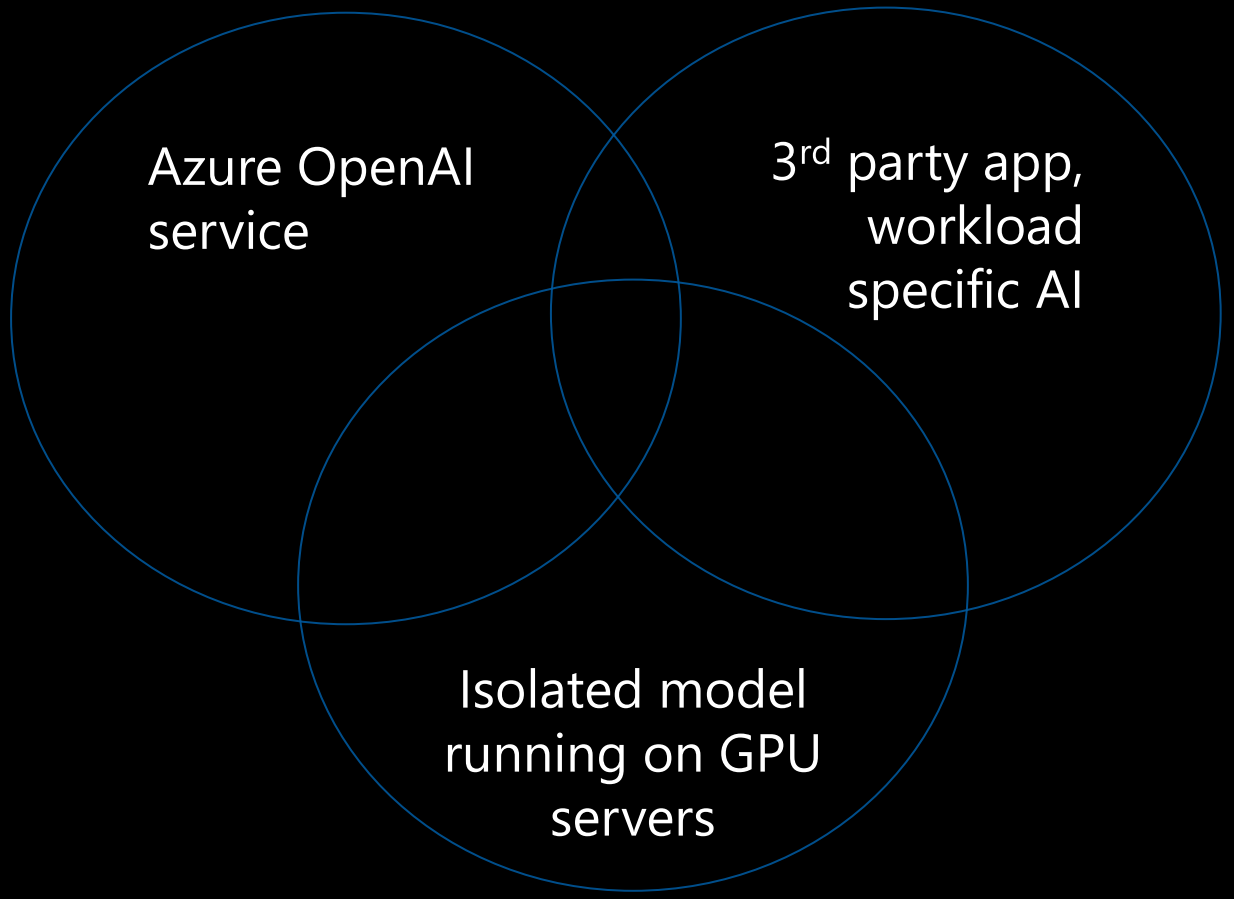
Empower every person and organization on the planet to achieve more

---

GPT-3.5 and GPT-4	ChatGPT	Codex	DALL·E 2
Text	Conversation	Code	Images

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 ≠ any AI

Pick any industry

AI will be the new enabler

Entertainment

Public

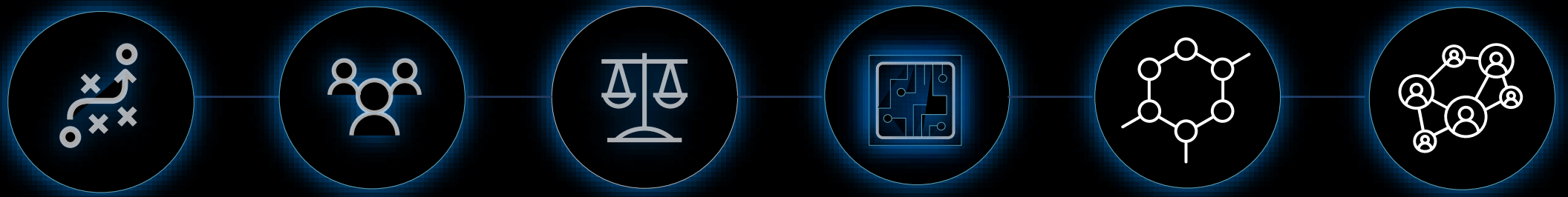
Retail

IT and APP development

Utilities

Manufacturing

Fintech



Get relevant company data understanding

Operations

Security

Product, dev, manufacturing

Customer services

Marketing

Copilot Boosts Productivity

Research shows developers...

GitHub

88%

are more productive

74%

can focus on more satisfying work

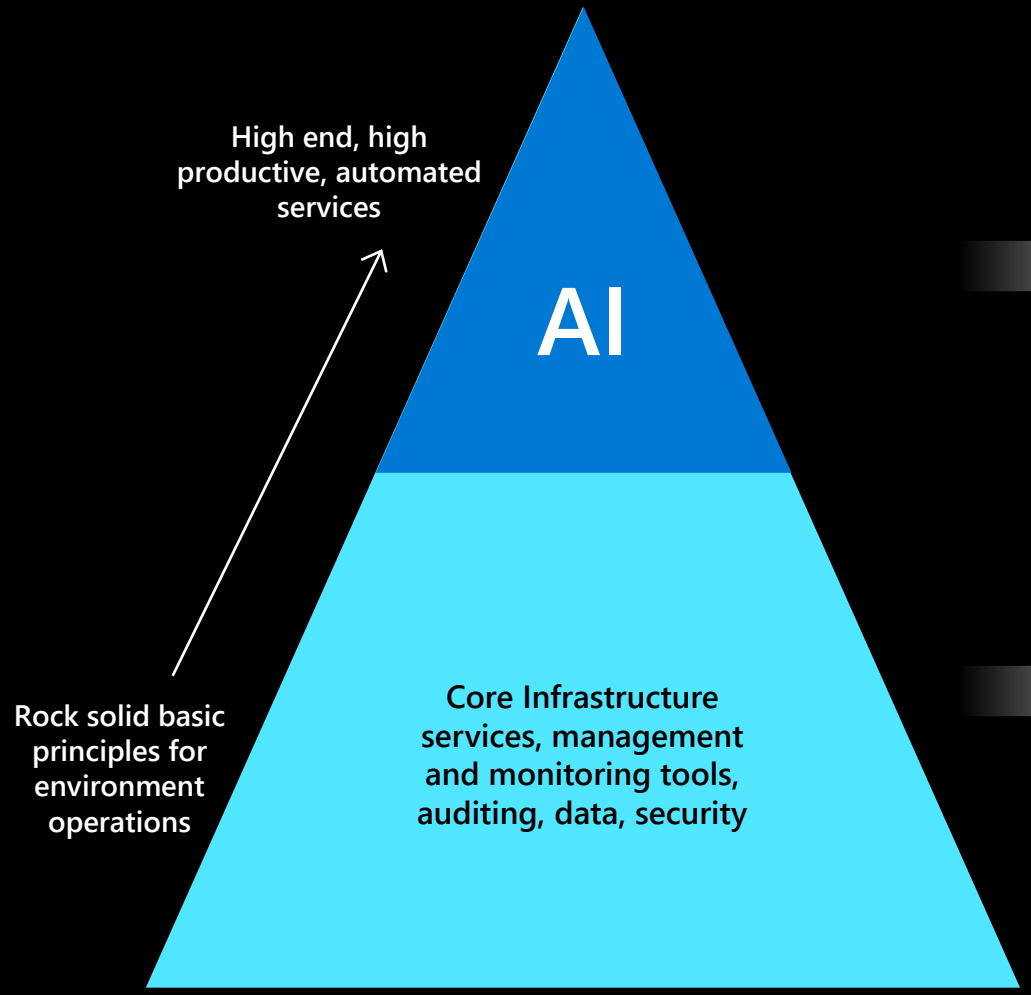
77%

spend less time searching for information or examples

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

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

Microsoft 365

Microsoft Dynamics 365

Partner Solutions

Application Platform

AI Builder

Power BI

Power Apps

Power Automate

Power Virtual Agents

Business Users

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

Developers & Data Scientists

ML Platform

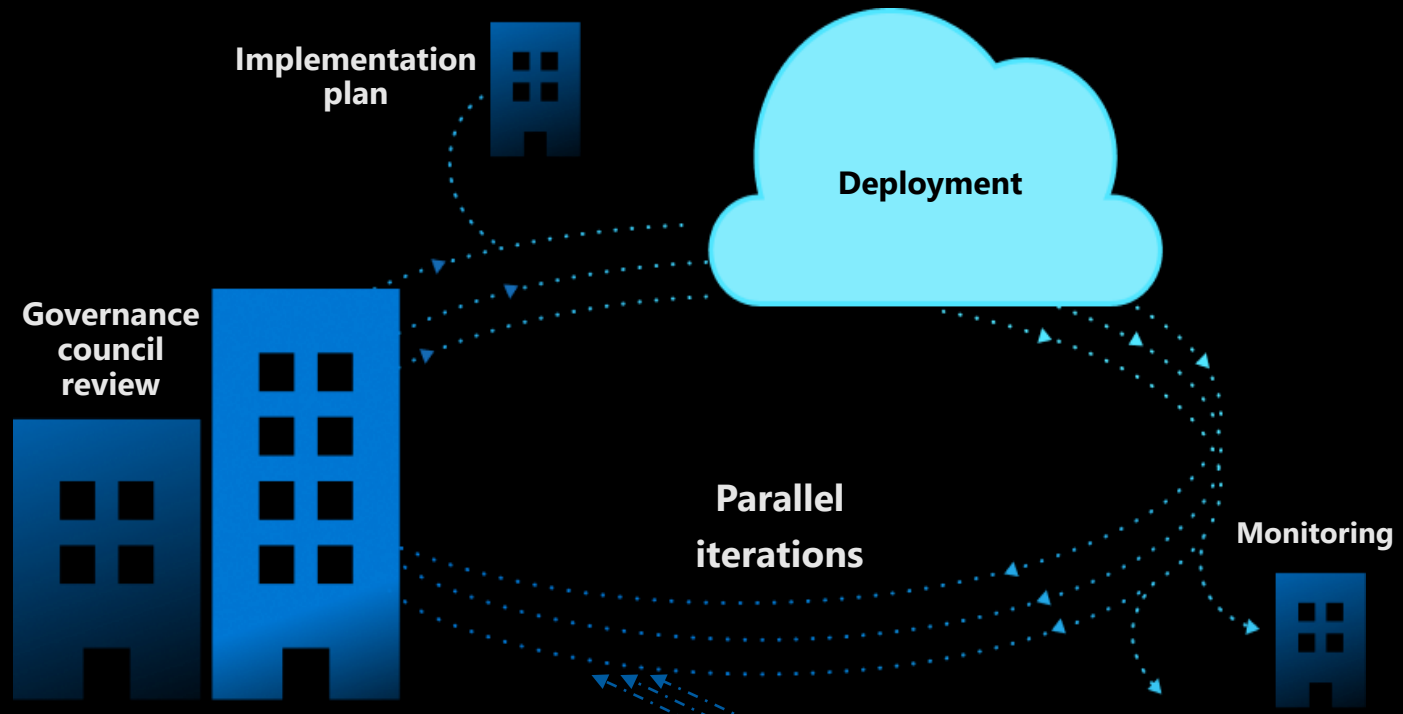
Azure Machine Learning

Approaching infrastructure lifecycle

Azure / hybrid governance and operations

is not equal

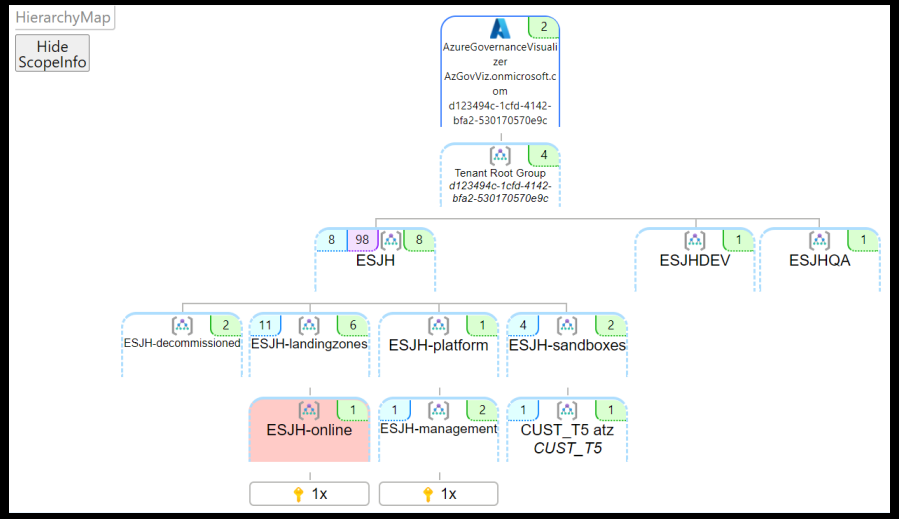
to DC governance / operations

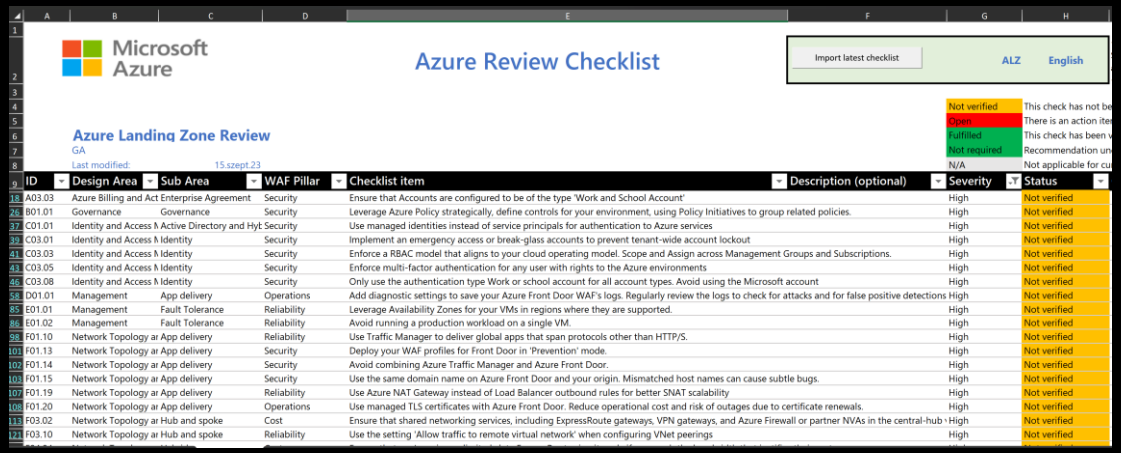


Linear approach

Environment assessment	Transition design	MVP deployment	Workloads transition planning
AzGovViz FT Checklists Advisor ...	ALZ Brownfield DLZ Open AI LZ ...	ALZ Templates AZ Policy ...	Workloads eval And migration plan ...

Environment assessment





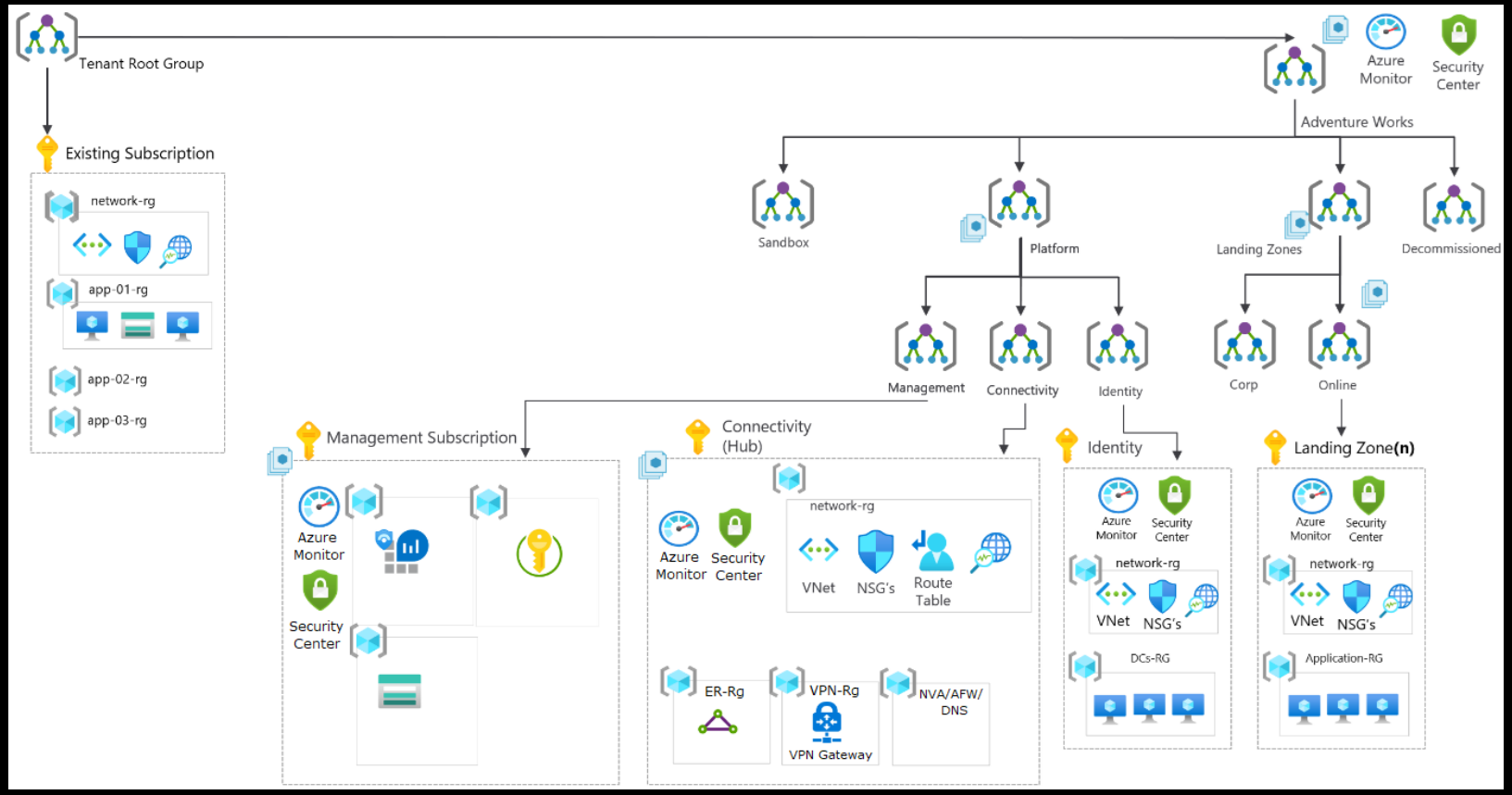
Companies mostly have some Azure environment…

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

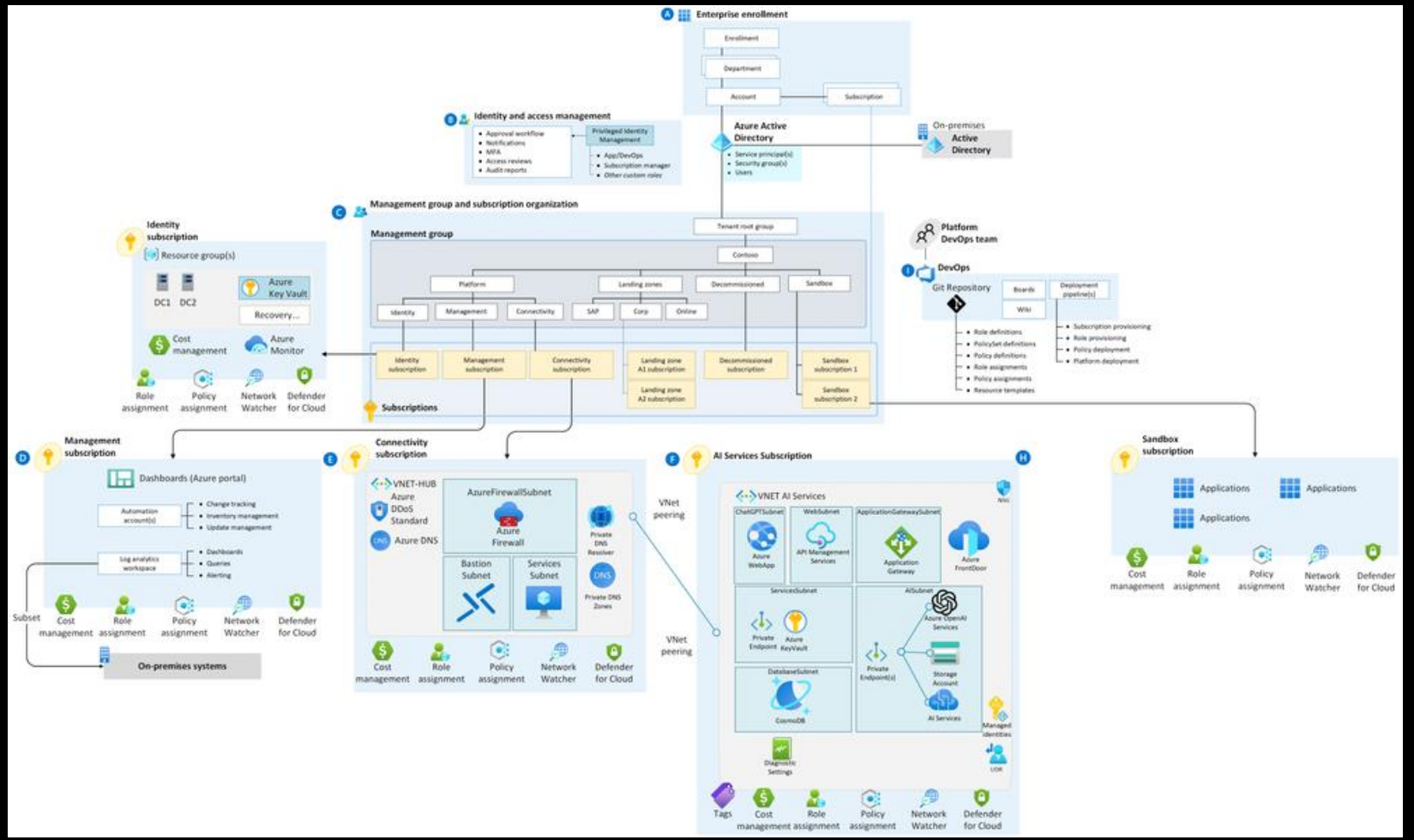
…

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



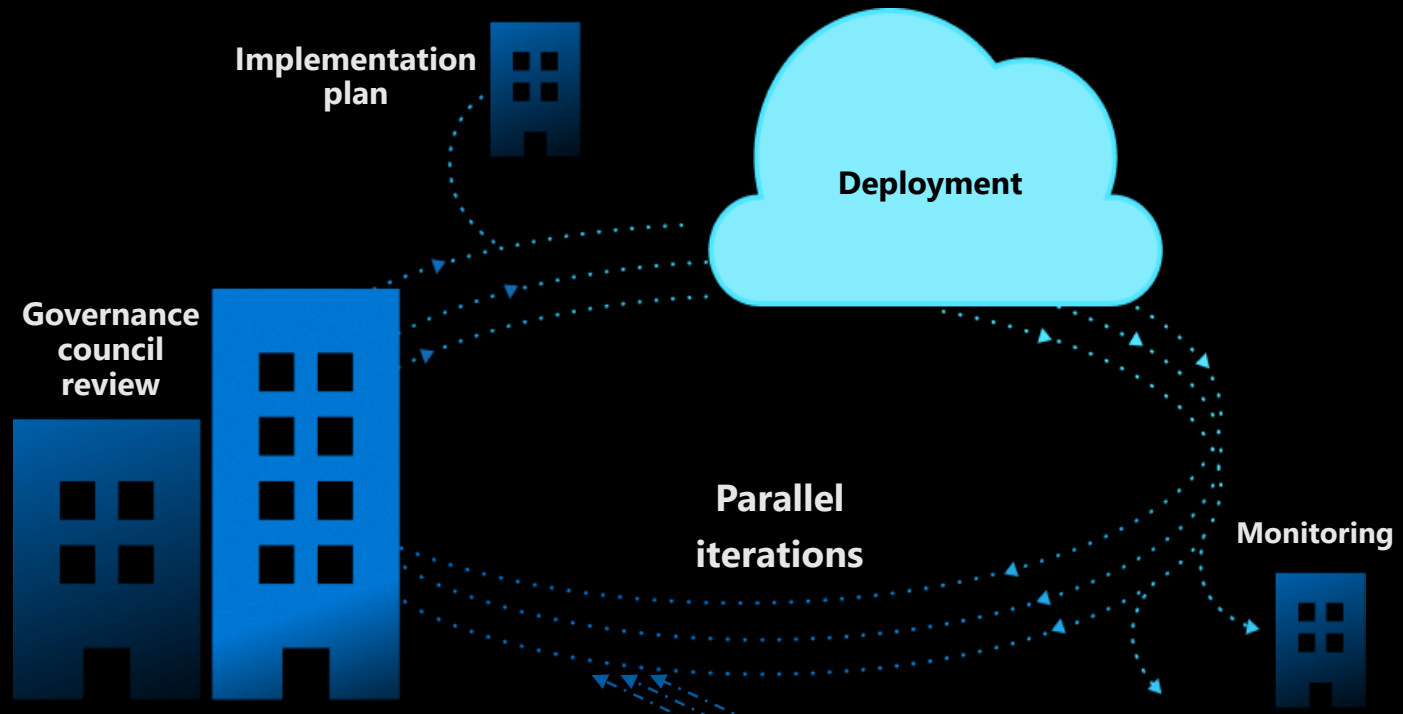
- based on ALZ variant
- apply accelerators
- apply additional extensions
- deploy parallel to production environment

Approaching infrastructure lifecycle

Azure / hybrid governance and operations

is not equal

to DC governance / operations

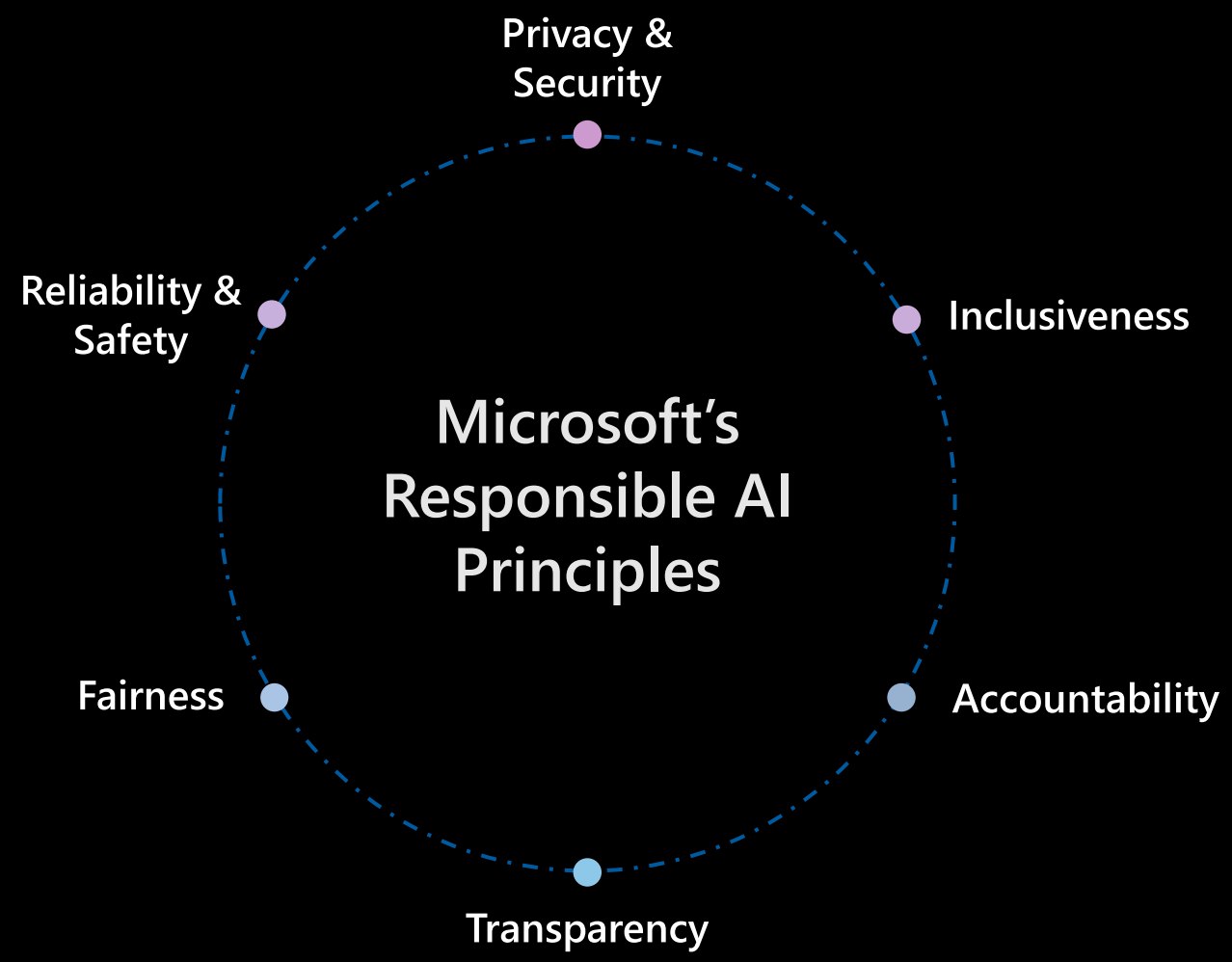


Linear approach

Environment assessment	Transition design	MVP deployment	Workloads transition planning
AzGovViz FT Checklists Advisor ...	ALZ Brownfield DLZ Open AI LZ ...	ALZ Templates AZ Policy ...	Workloads eval And migration plan ...

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



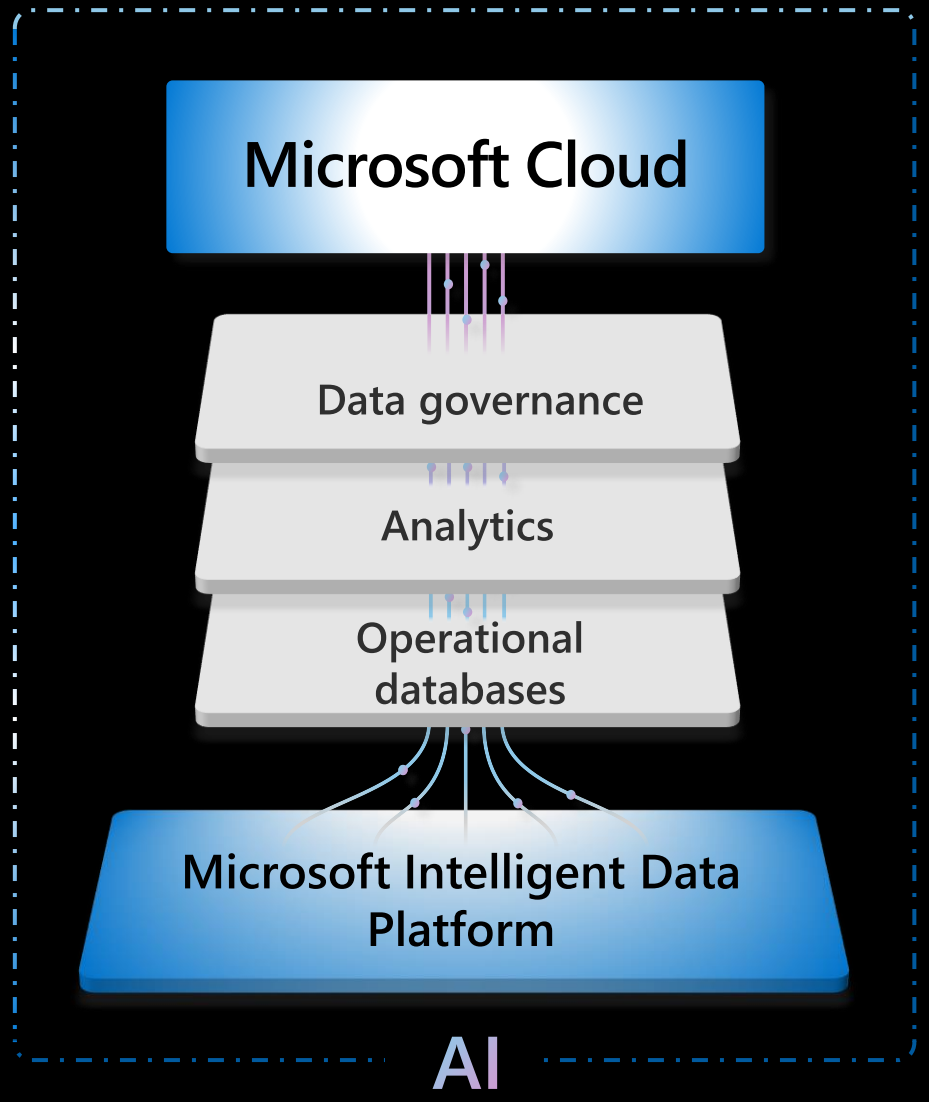
Building blocks to enact principles

- 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