

BOUSSIAS
events

19th IT Directors Forum

Accelerating Digital Business

16 Οκτωβρίου 2024 | Μέγαρο Μουσικής Αθηνών

InTTrust
Information Technology Trust



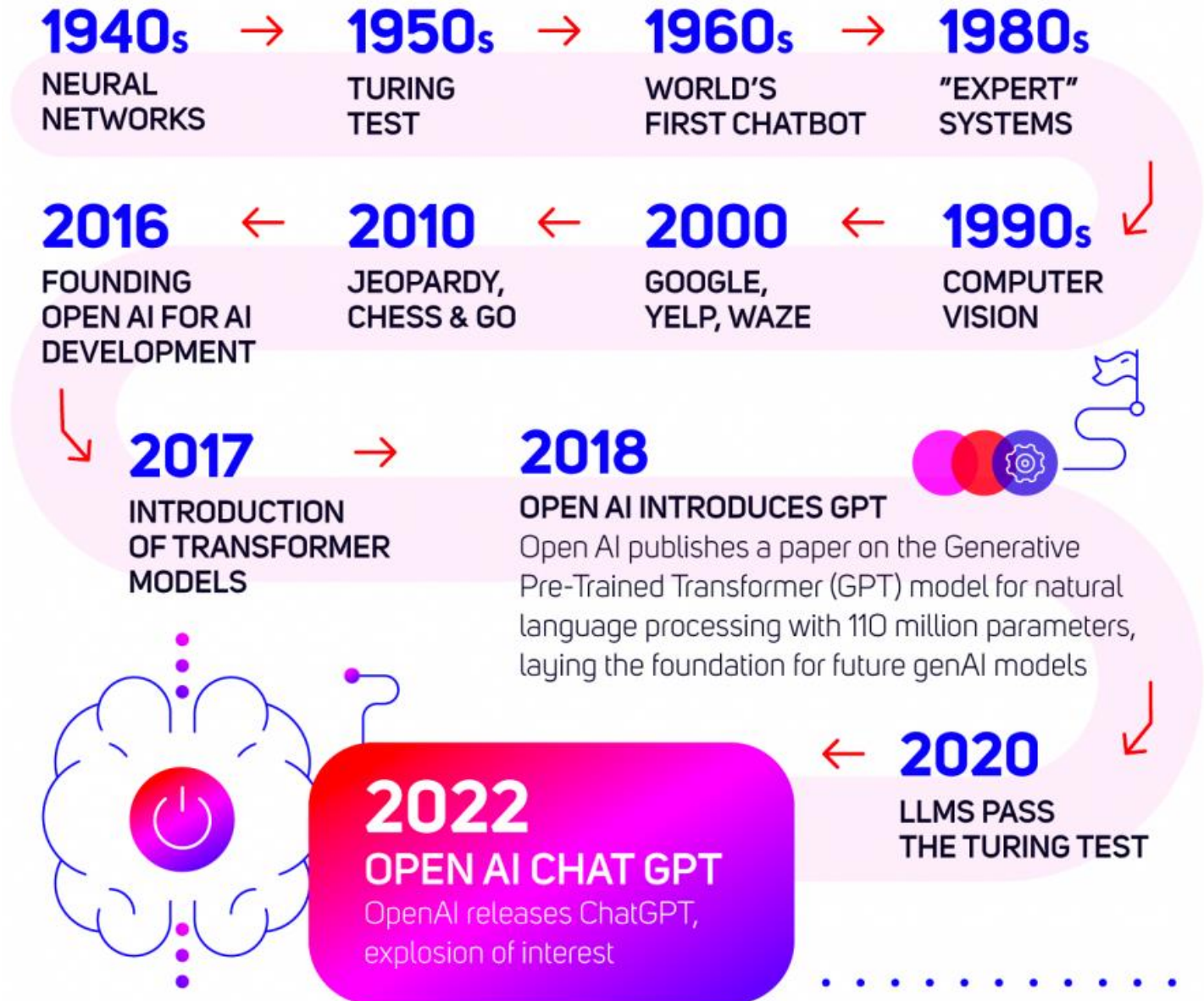
watsonx

Putting Trustworthy AI to Work

Manos Kokolakis, MSc
Customer Success Manager
mkokolakis@intrust.gr



AI evolution ...



GenAI in a nutshell

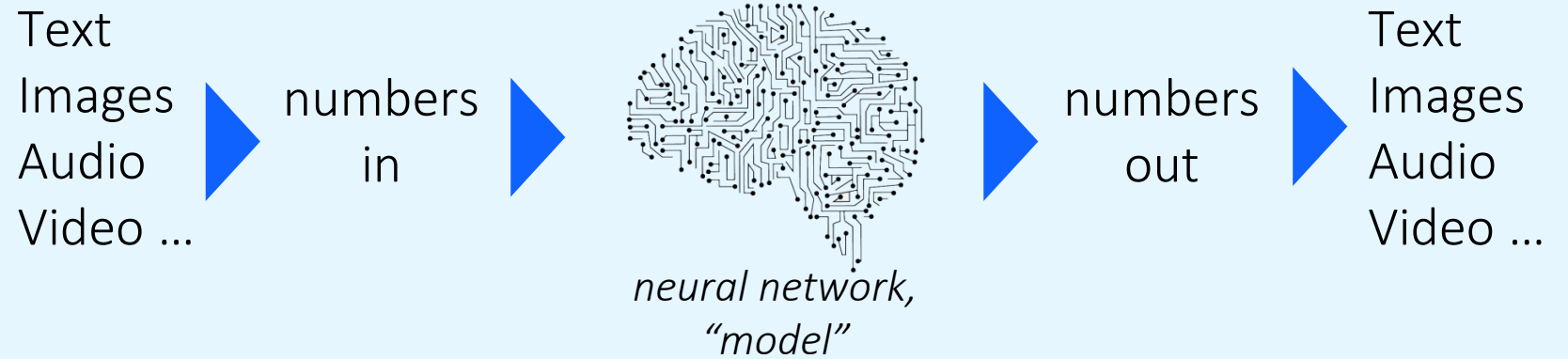
GenAI in a Nutshell

“Generative AI refers to deep-learning models that can generate high-quality text, images, and other content based on the data they were trained on.”

Terminology

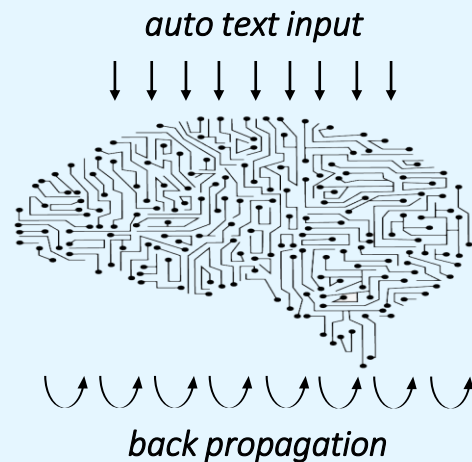
AI, Machine Learning, Deep Learning, Computer vision and more ...

How it works

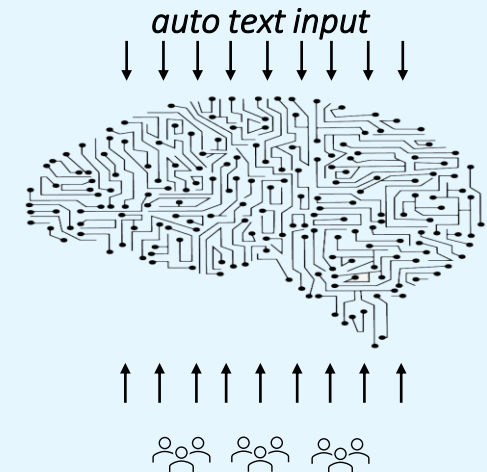


Training a model

1. Unsupervised Generative Pretraining



2. Reinforcement Learning with Human Feedback



Models

Llama
Claude
Bard

GPT 4
Mistral
Meta

vary on...

Speed
Capability
Cost
Architecture

Ease of use
Specialization
Integration
Open/ proprietary ...

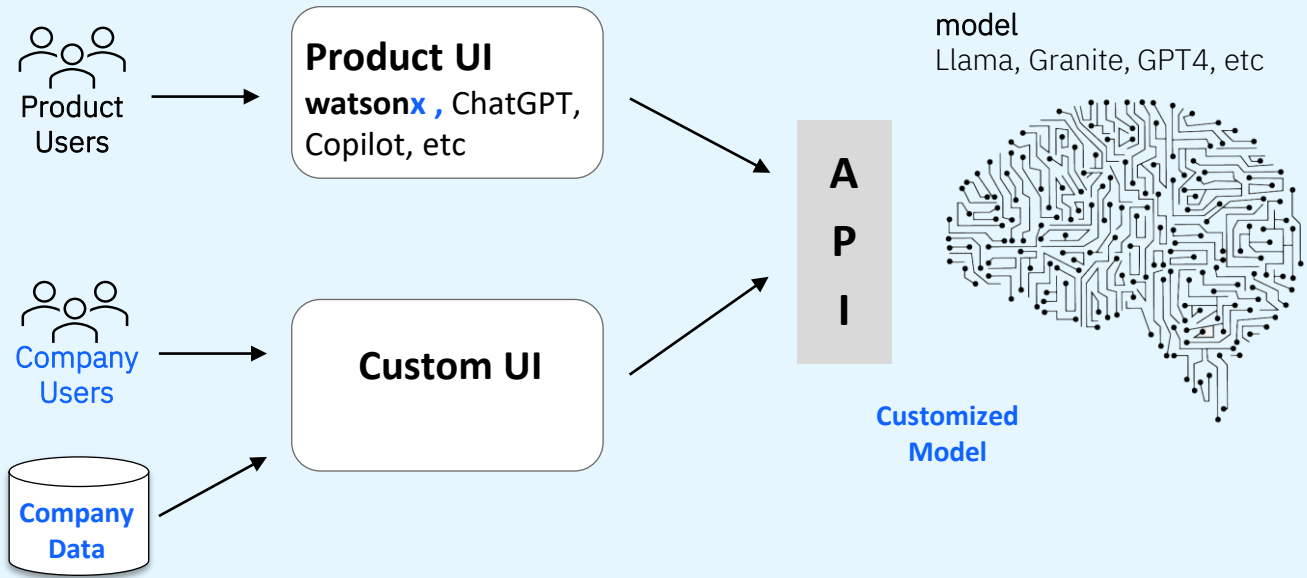
Model types

Text to Text
Speech to Text

Text to Audio
Text to Video

Text to Image
Image to Image

Using AI powered products



GenAI for business

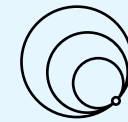
75%

of CEOs say their competitive advantage rests on generative AI¹

“Generative AI models surprise, impress, and scare us, all at the same time.”

Gonzalo Gortázar
CEO, CaixaBank

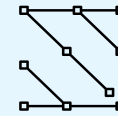
However, the following barriers prevent organizations from benefiting from AI



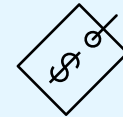
Integration and scaling challenges



AI governance



Data complexity



High price



Limited skill and expertise



Lack of AI model development tools

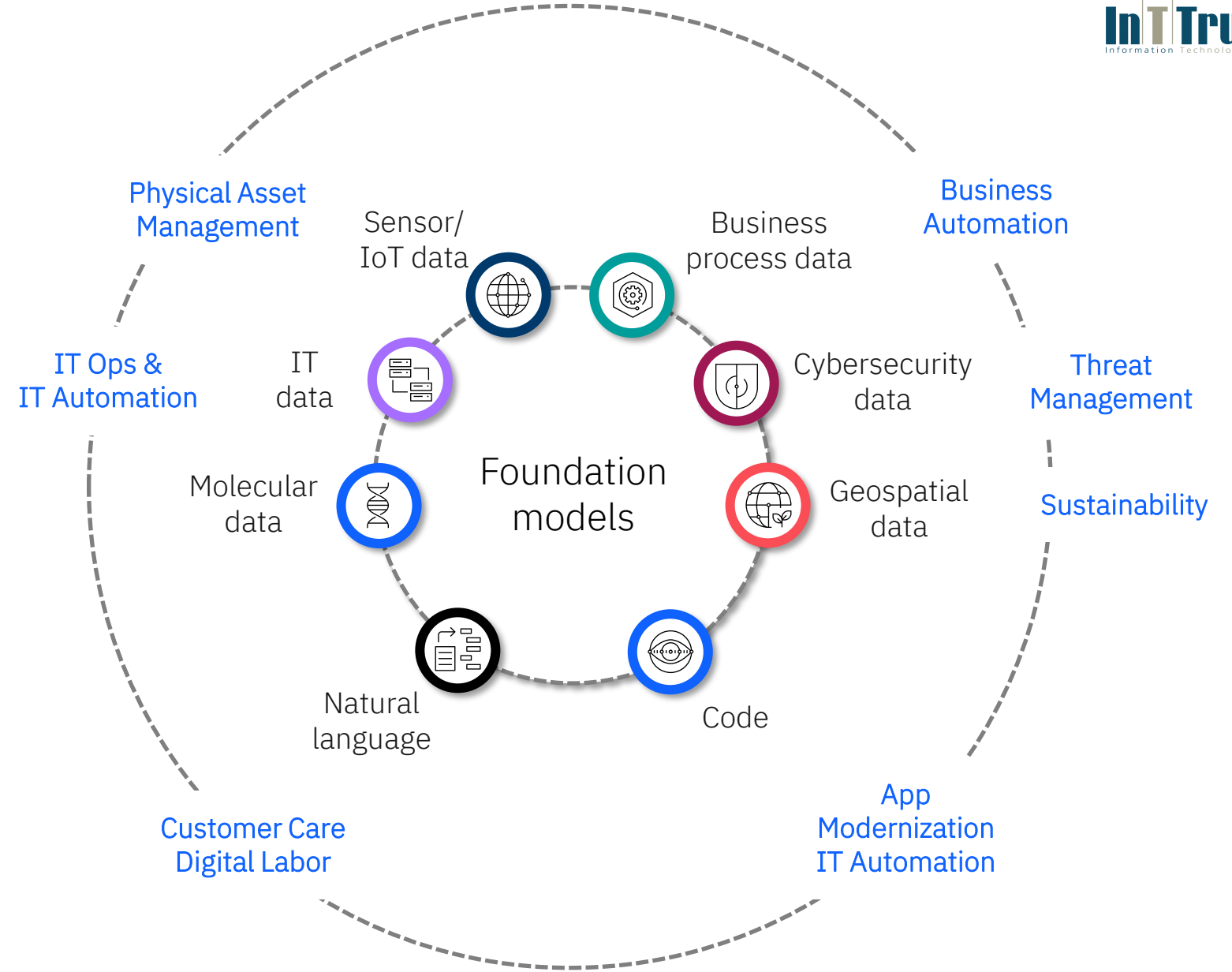
+ AI → AI +

*Reinventing how work gets done across
business domains and industries*

AI for Business

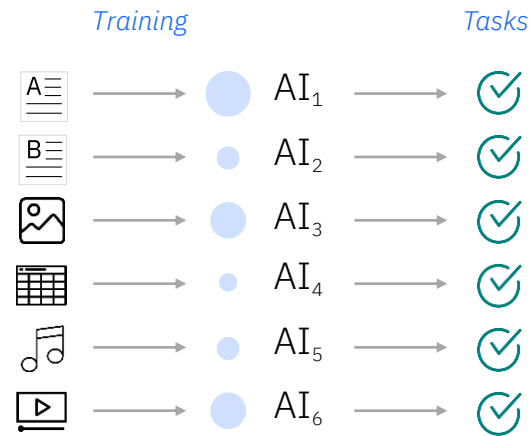
- Enterprise focus: Data privacy, compliance, scalability, and integration.
- Customization: Tailored models for industry-specific applications.
- Security: Ensuring AI models are secure, reliable, and free from bias.
- Cost Efficiency: Optimizing resources while providing actionable insights.

Opportunity to unlock business advantage with foundation models and LLMs trained across the breadth of enterprise data



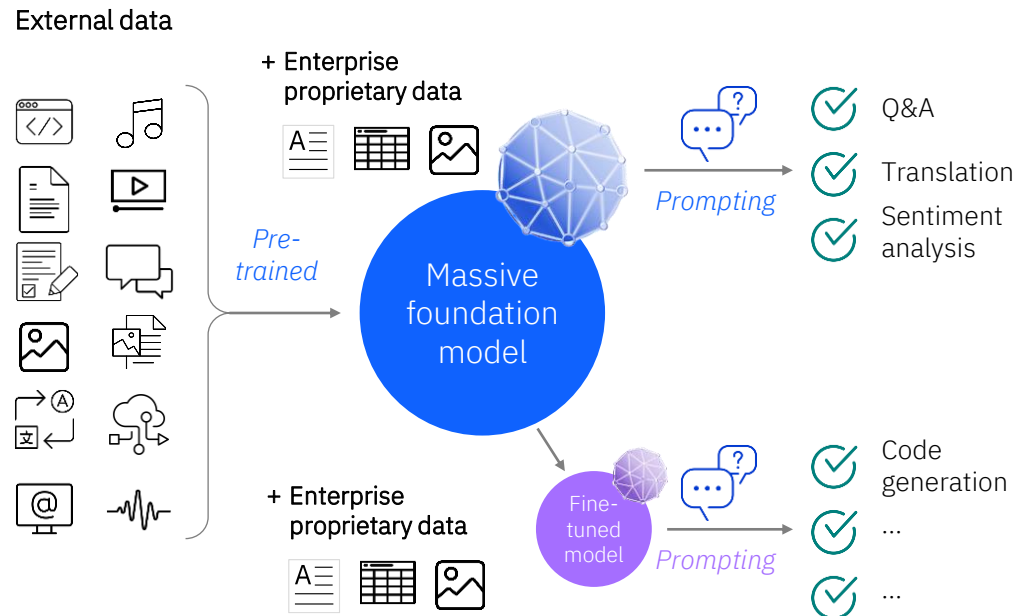
Generative AI establishes a new paradigm for AI capabilities

Traditional AI models



- Individual siloed models
- Require task specific training
- Lots of human supervised training

Foundation models



- Massive multi-tasking model
- Adaptable with minimized training
- Pre-trained unsupervised learning

Enhanced capabilities

- Summarization
- Conversational knowledge
- Content creation
- Code co-creation

Key advantages

- Lower upfront costs through less labeling
- Faster deployment through fine tuning
- Equal or better accuracy for multiple use cases
- Incremental revenue through better performance

up to **70% reduction** in certain NLP tasks

watsonx *putting AI to work*

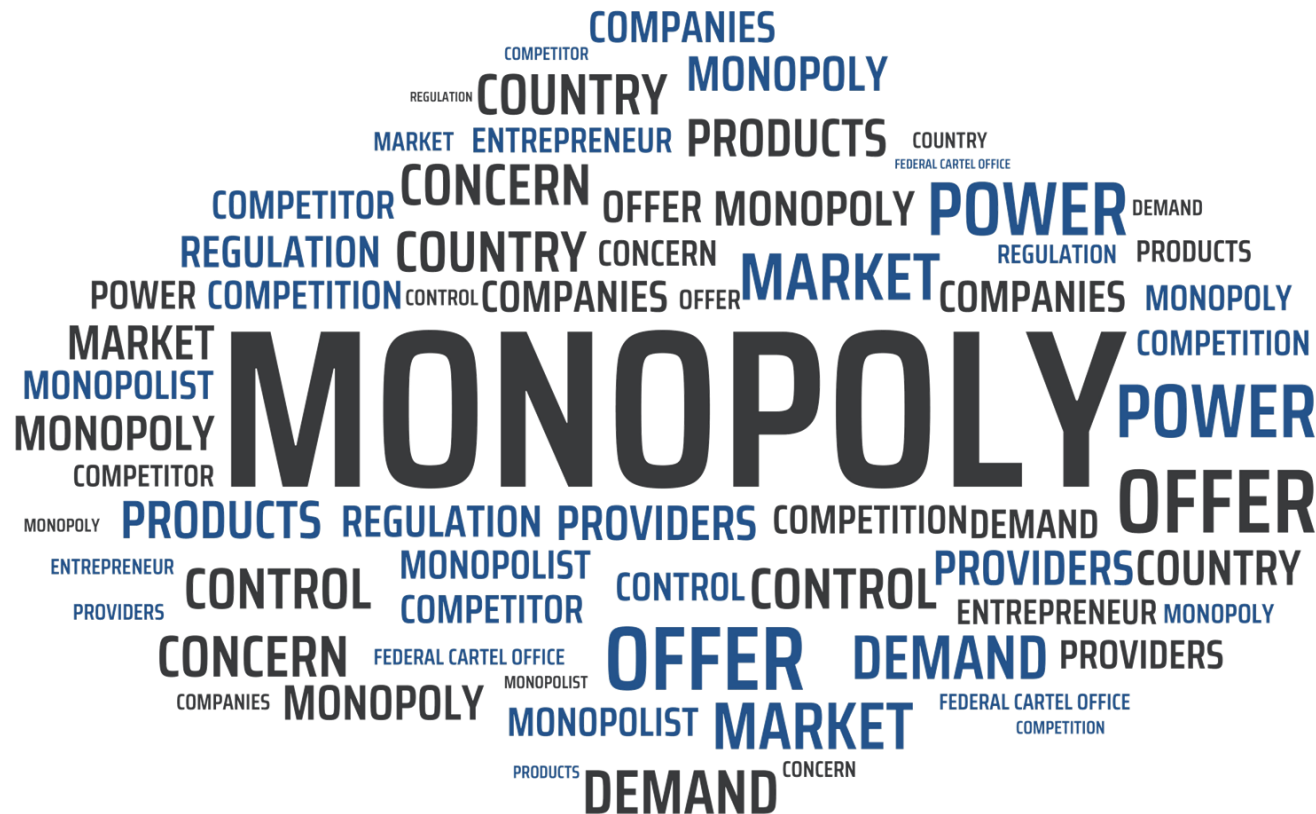
*A suite designed to help businesses harness AI securely and at scale.
Offers data privacy management, explainable AI, customizable
models, easy integration with tools.*

Why watsonx ?

- **Enterprise-Grade**, secure, scalable, and compliant AI solution.
- **Customizable**, tailored to meet specific industry and business needs.
- **Trustworthy**, robust governance and explainability for transparency and ethical AI use.
- **Future-Proof**, IBM's ongoing innovations in AI and cloud technologies.
- **Strengths**, enterprise focus, strong data governance, hybrid cloud support, open, custom and mixed models.
- **Deployment**, hybrid, on-premise, and cloud-based

A differentiated approach to AI foundation models

There is no monopoly.
No one model will rule them all



watsonx model strategy

Multi-model

One model doesn't fit all use cases. IBM offers **IBM**-developed, **open-source**, **third-party**, Original Equipment Manufacturer (**OEM**) models, and **Custom** Foundation Models.

Bigger is not always better. Specialized models can outperform general-purpose models with lower infrastructure requirements.

Multi-cloud

Deploy anywhere. IBM provides the flexibility to deploy models on the platform of choice.

IBM's model strategy: multiple models and open

IBM models

- granite-3b-code-instruct
- granite-7b-lab
- granite-8b-code-instruct
- granite-13b-chat-v2
- granite-13b-instruct-v2
- granite-20b-code-instruct
- granite-34b-code-instruct
- granite-20b-multilingual

[IBM is open-sourcing](#)

[Granite code models](#)¹

Open/Third Party

- mt0-xxl-13b
- codellama-34b-instruct-hf
- flan-t5-xl-3b
- flan-t5-xxl-11b
- flan-ul2-20b
- merlinite-7b
- mixtral-8x7b-instruct-v01
- llama-2-13b-chat
- llama-2-70b-chat
- llama-3-70b-instruct
- llama-3-8b-instruct

OEM models

- Mistral Large

To come:

- Mistral Small
- Mistral Embedding
- Codestral
- Other Mistral models

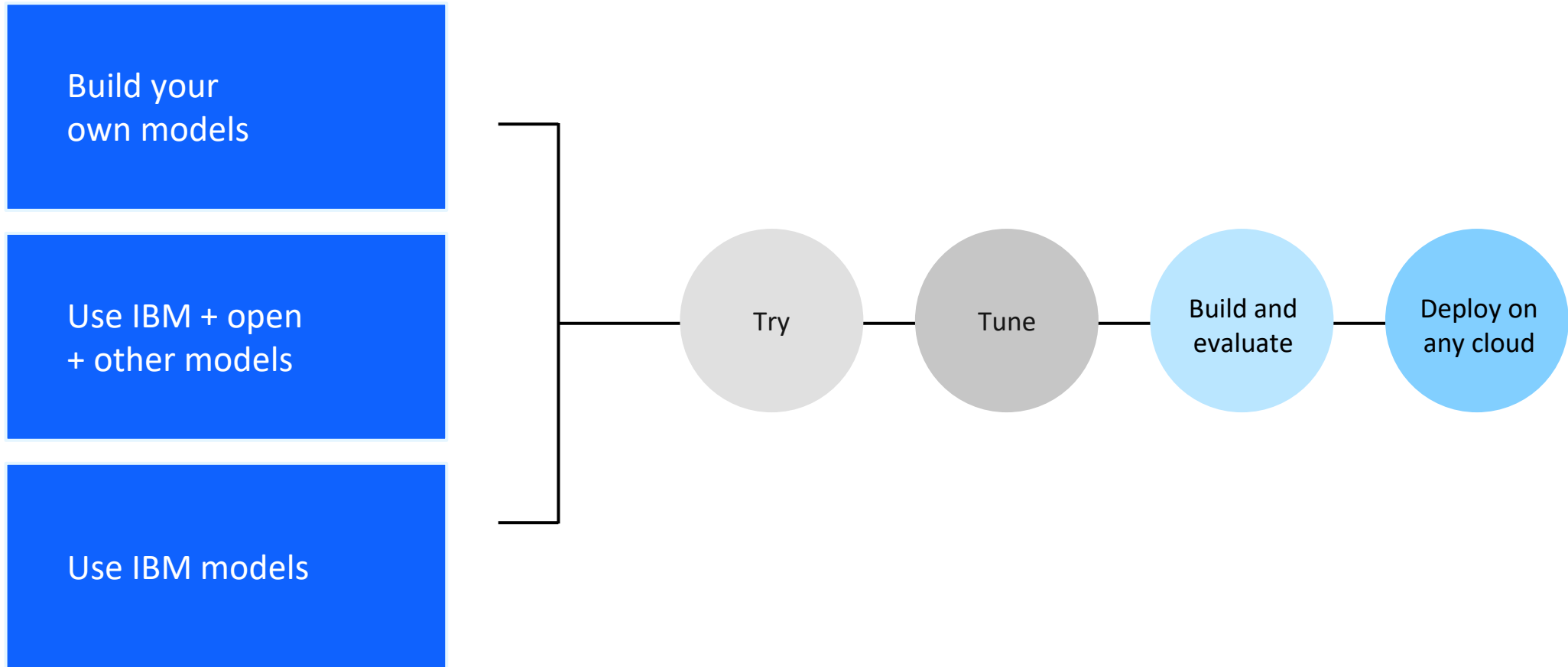
Custom models

Clients can import and deploy their own or third-party foundation models (LLM or code) if they meet the supported model architecture [criteria](#) and type.

The architecture list includes Gpt, Llama, Mistral, Mixtral, T5, Codegen, and more.

This list is constantly adding new models coming online and older ones being deprecated.

Open AI based on foundation models that are multi-model on multcloud



The platform for AI and data

watsonx

watsonx.ai

*Train, validate, tune,
and deploy AI models*

foundation models for
language, code, and
generative tasks

A next generation enterprise [studio for AI builders](#) to train, validate, tune, and deploy both traditional machine learning and new generative AI capabilities powered by [foundation models](#). It enables you to build AI applications in a fraction of the time with a fraction of the data.

watsonx.data

*Scale AI workloads, for
all your data, anywhere*

data management solutions,
from preparation to
governance.

Fit-for-purpose [data store](#), built on an open lakehouse architecture, supported by querying, governance and open data formats to access and share data.

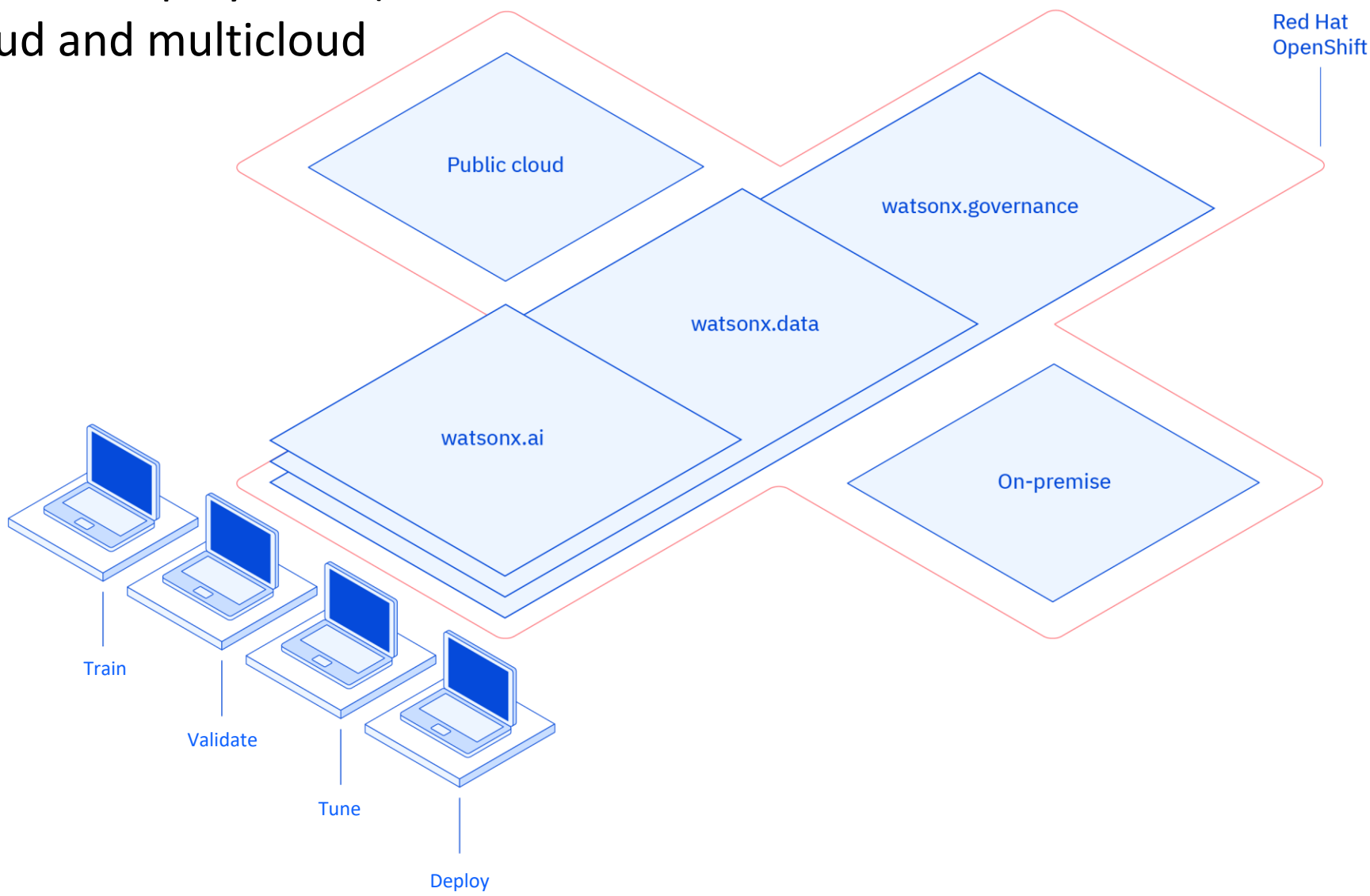
watsonx.governance

*Accelerate responsible,
transparent, and explainable
AI workflows*

transparency, accountability,
and compliance.

End-to-end toolkit for AI [governance across](#) the entire model lifecycle to enable responsible, transparent, and explainable AI workflows.

watsonx: With flexible deployment (on-premise, public cloud and multicloud deployment)



Why IBM's AI?

Open

IBM's AI is based on the best open technologies available

Trusted

IBM's AI is transparent, responsible, and governed

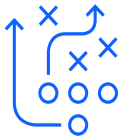
Targeted

IBM's AI is designed for enterprise and targeted at business domains

Empowering

IBM's AI is for value creators, not just users

Three ways to get started with watsonx today



Free trial

Core watsonx.ai and watsonx.data features to start building AI models and accessing data across your organization.

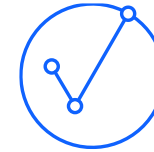
Available now



Client briefing

Discussion and custom demonstration of watsonx point of view and capabilities. Understand where generative AI can be leveraged now for impact in your business.

2-4 hours
onsite or virtual



Pilot program

Engagement with an InTTrust and IBM multi-disciplinary team to jointly innovate and rapidly prove the business value of generative AI solutions using watsonx.

1-4 weeks

Watson who ?

IBM Watson is a computer system capable of answering questions posed in natural language. It was developed as a part of IBM's DeepQA project by a research team, led by principal investigator David Ferrucci.

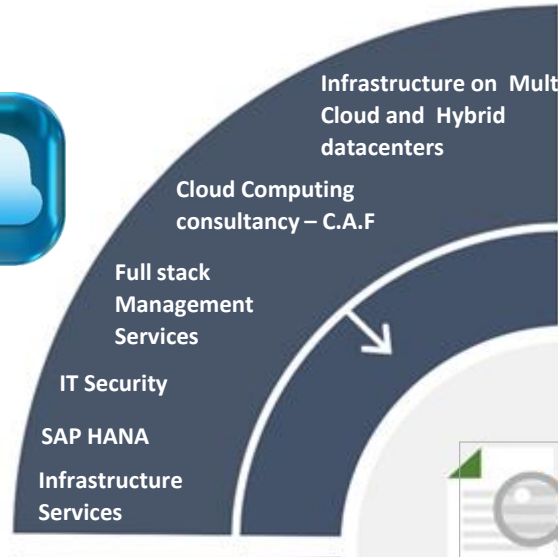


Watson was named after IBM's founder and first CEO, industrialist Thomas J. Watson.

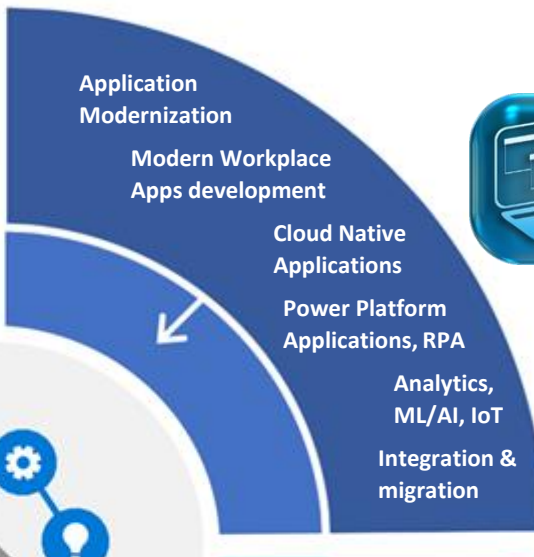
[Wikipedia](#)

Portfolio of Services

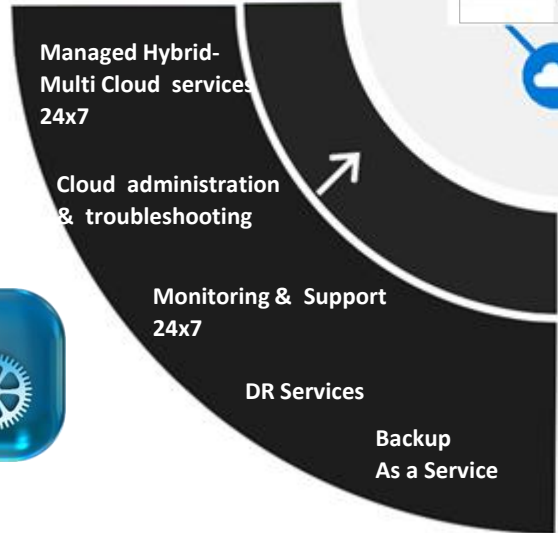
Dynamic Infrastructure Services



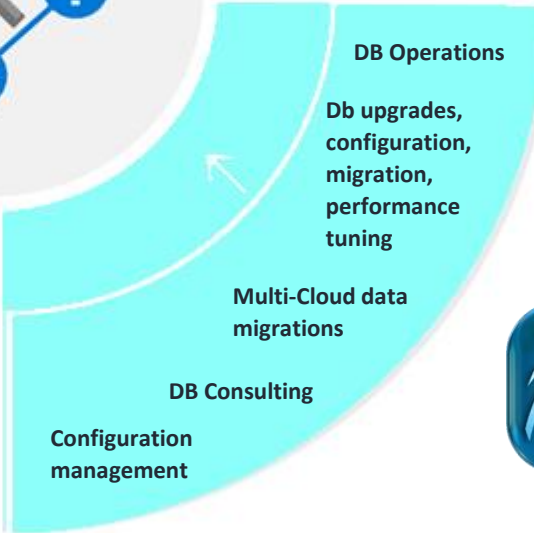
Application Development Services



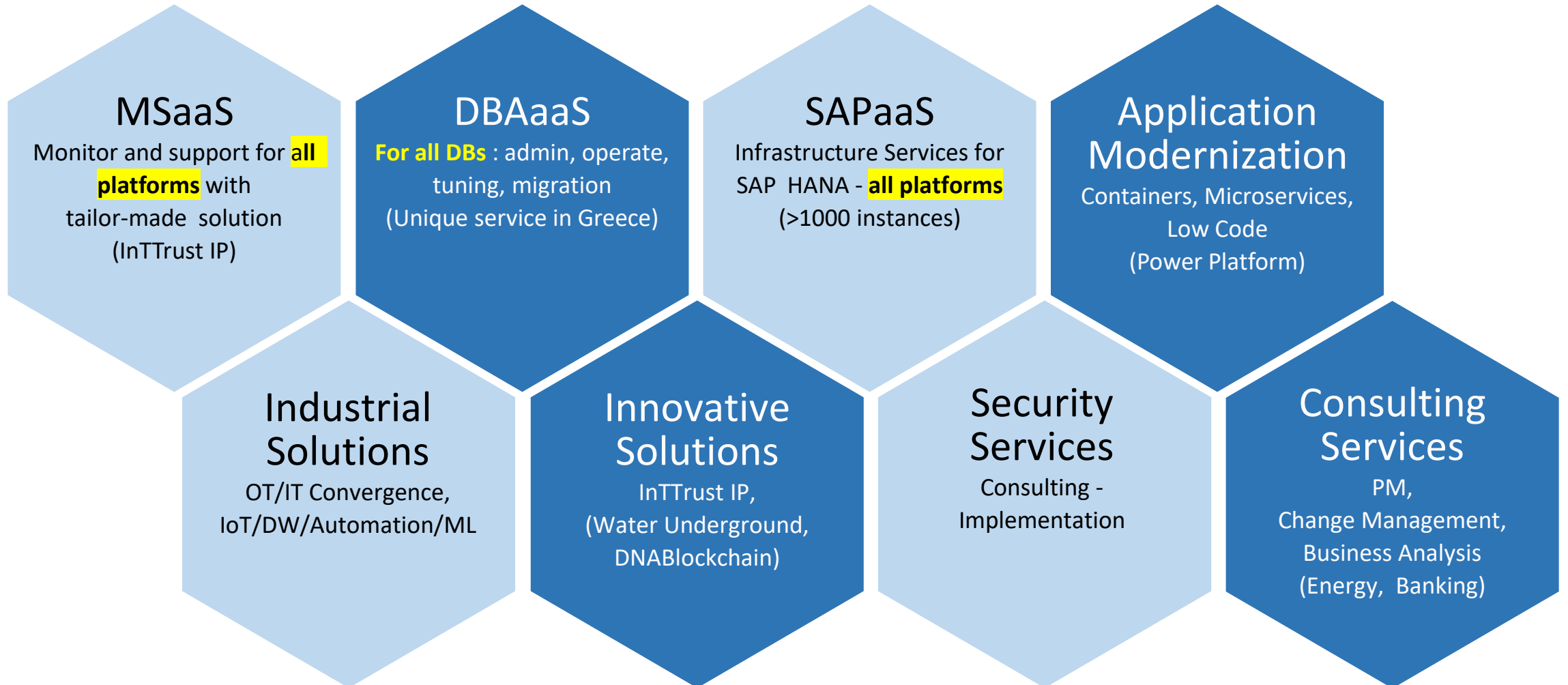
Managed Services



DBA Services



Key differentiators





Thank you.

Manos Kokolakis, MSc
Customer Success Manager
mkokolakis@intrust.gr

