Applying AI to Engineering:

A perspective inspired by systems thinking

Daniel Moul, Principal Product Manager

April 2024

New rewards

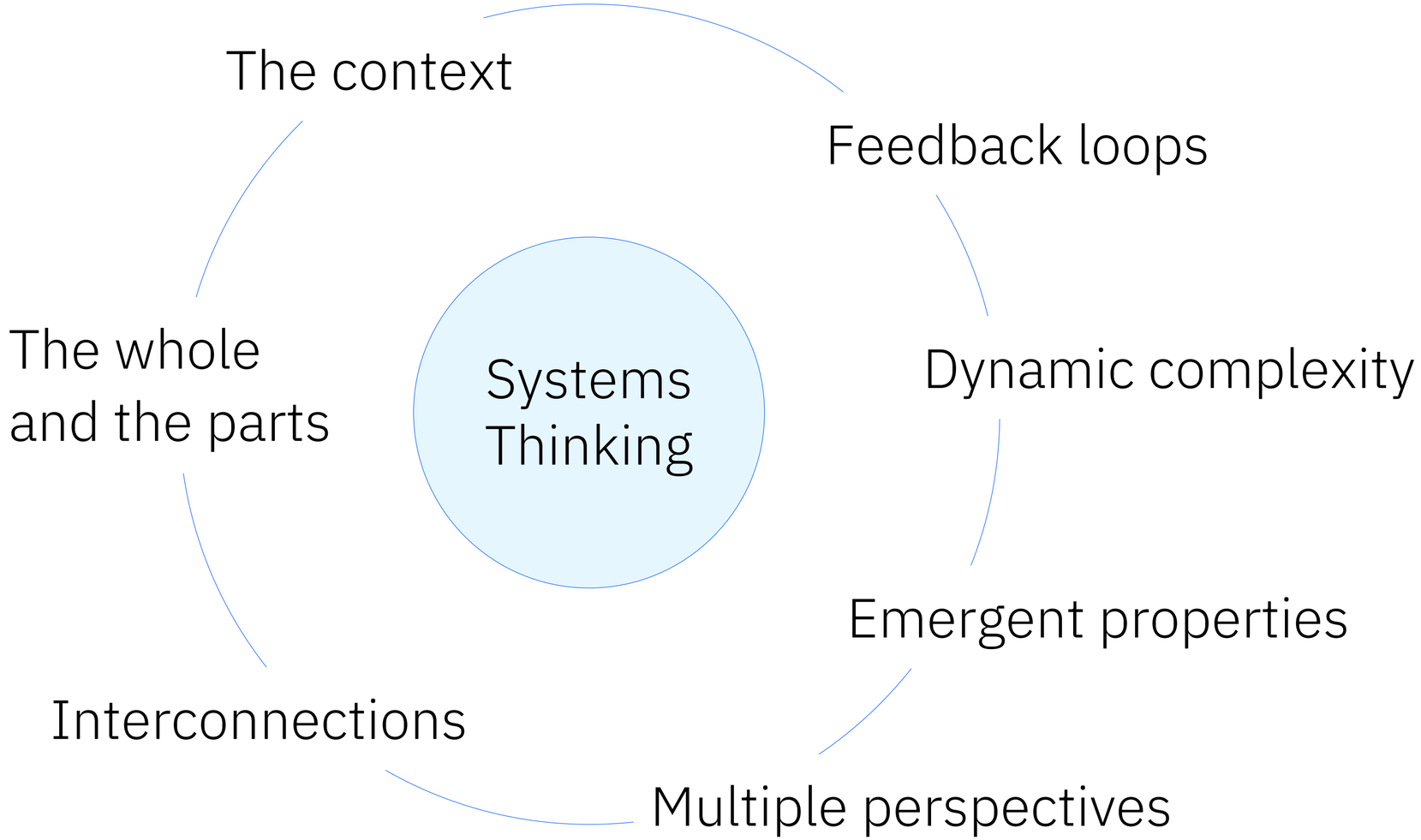
New risks

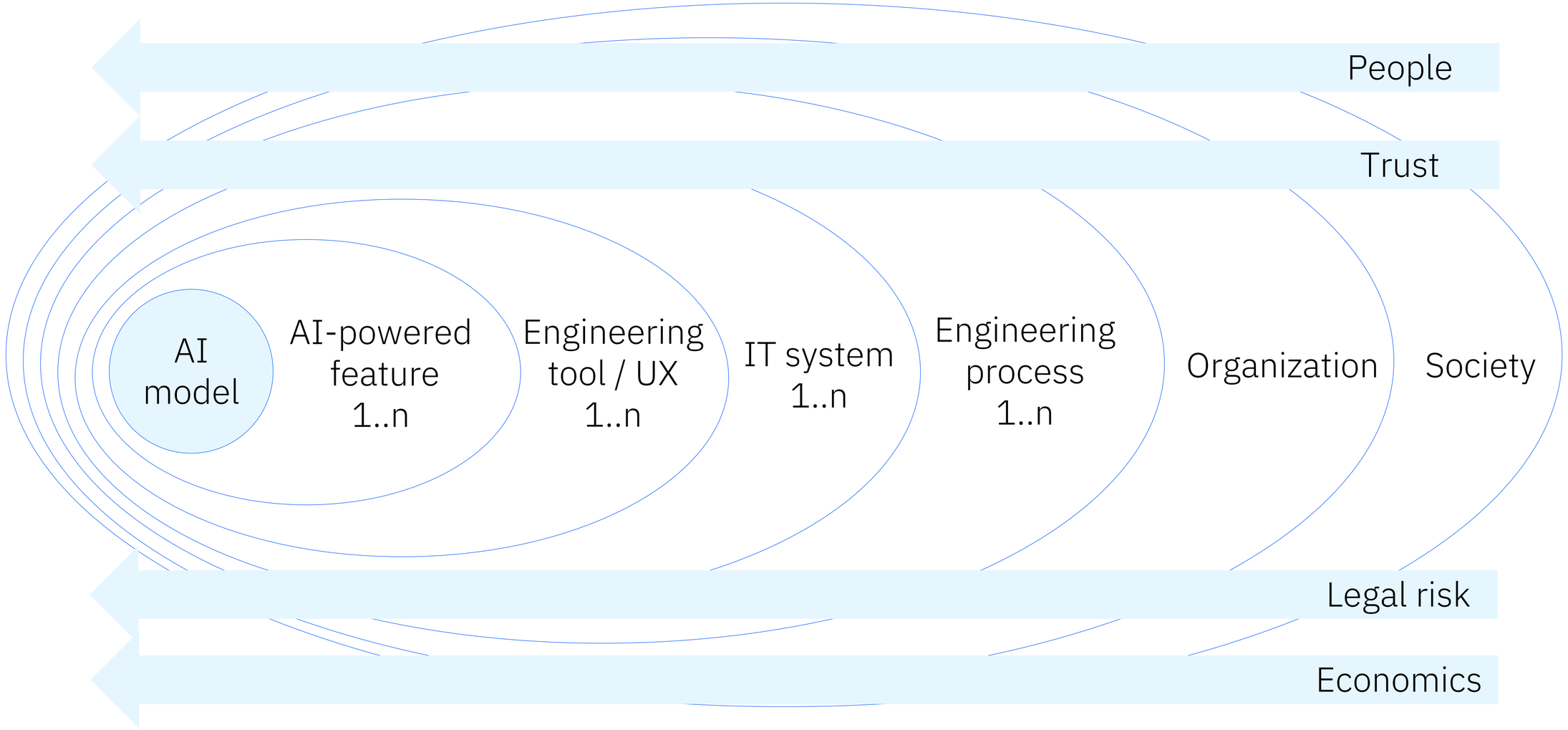
New unknowns

Source: [Theatrvm orbis terrarvm](), 1570
US Library of Congress

Systems Thinking

1

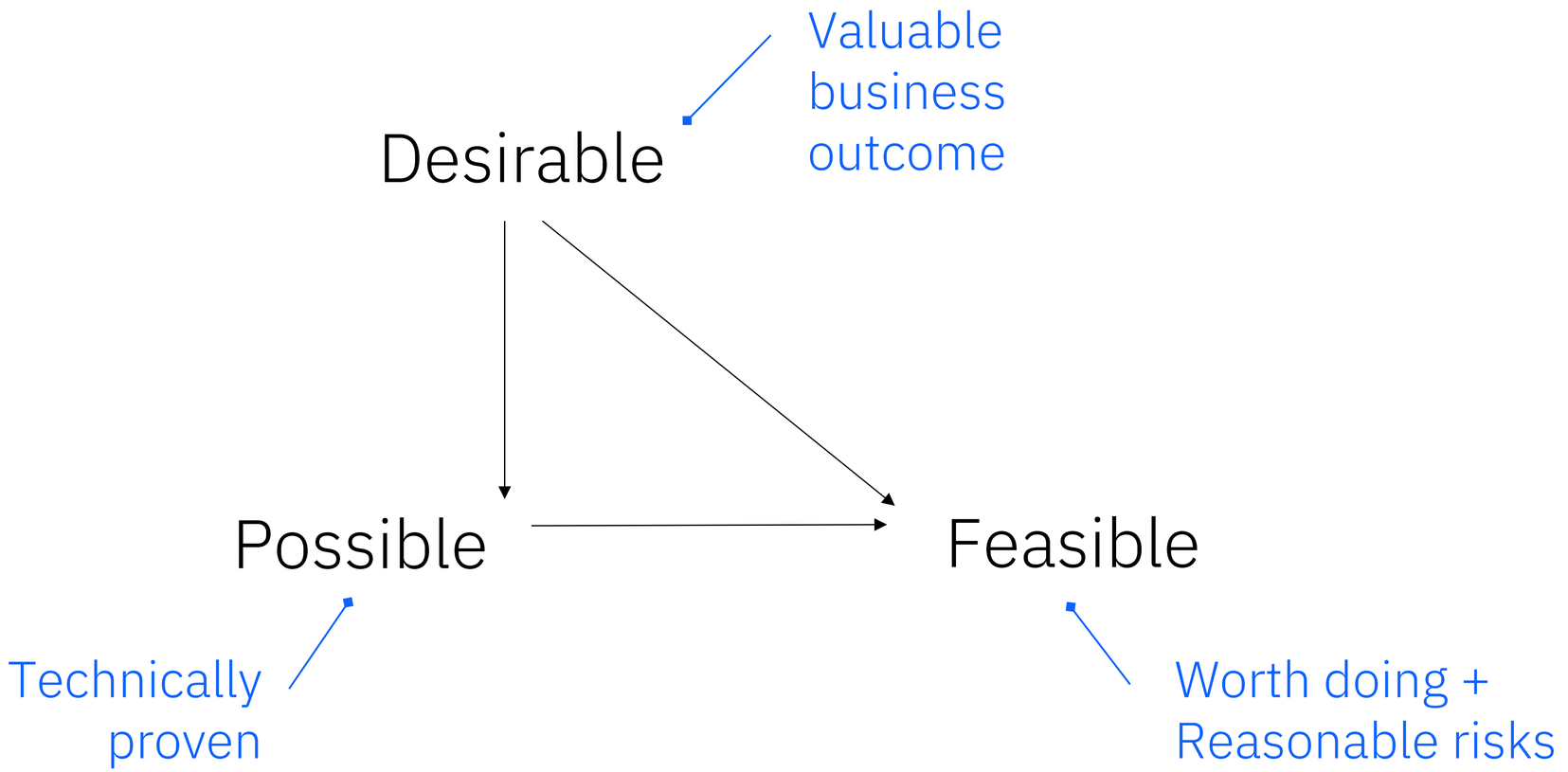

The context
Feedback loops
The whole
and the parts
Systems
Thinking
Dynamic complexity
Emergent properties
Interconnections
Multiple perspectives


People
Trust
AI model
AI-powered feature 1..n
Engineering tool / UX 1..n
IT system 1..n
Engineering process 1..n
Organization
Society
Legal risk
Economics

The AI dynamic

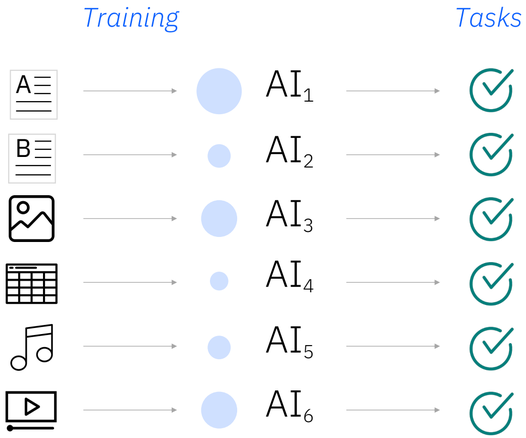
2

Where to focus



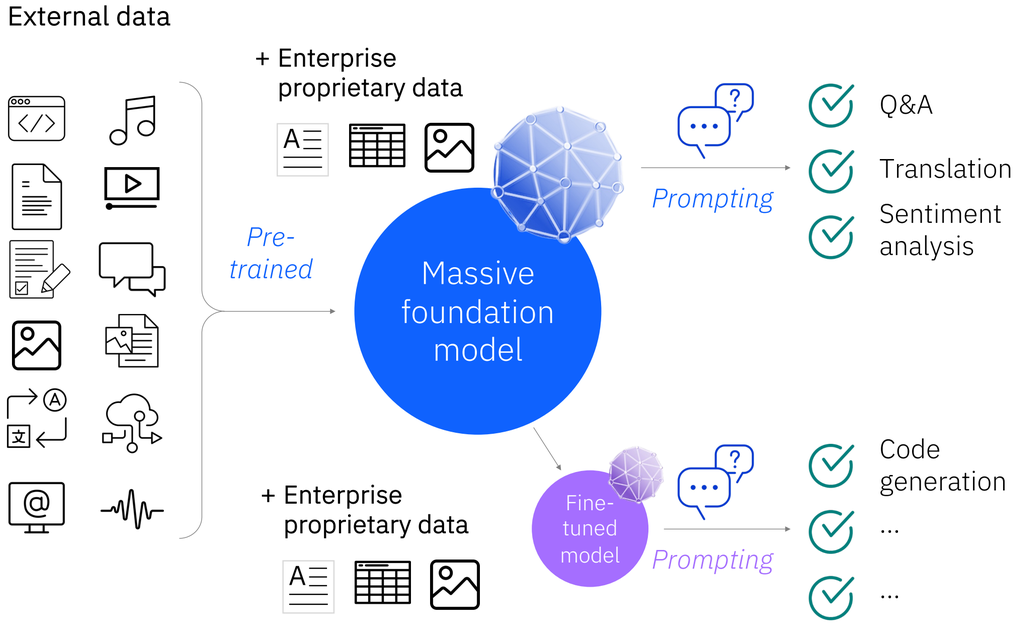
Generative AI establishes a new paradigm for AI capabilities

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

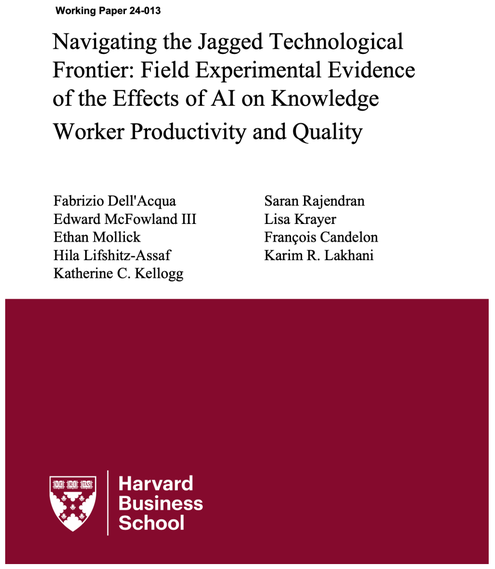
- Summarization
- Conversational knowledge
- Content creation
- Code co-creation

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

What’s possible? Where is the frontier for engineers?


Working Paper 24-013

Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality

Fabrizio Dell'Acqua
Edward McFowland III
Ethan Mollick
Hila Lifshitz-Assaf
Katherine C. Kellogg

Saran Rajendran
Lisa Krayer
François Candelon
Karim R. Lakhani

Harvard Business School

Paper / Blog

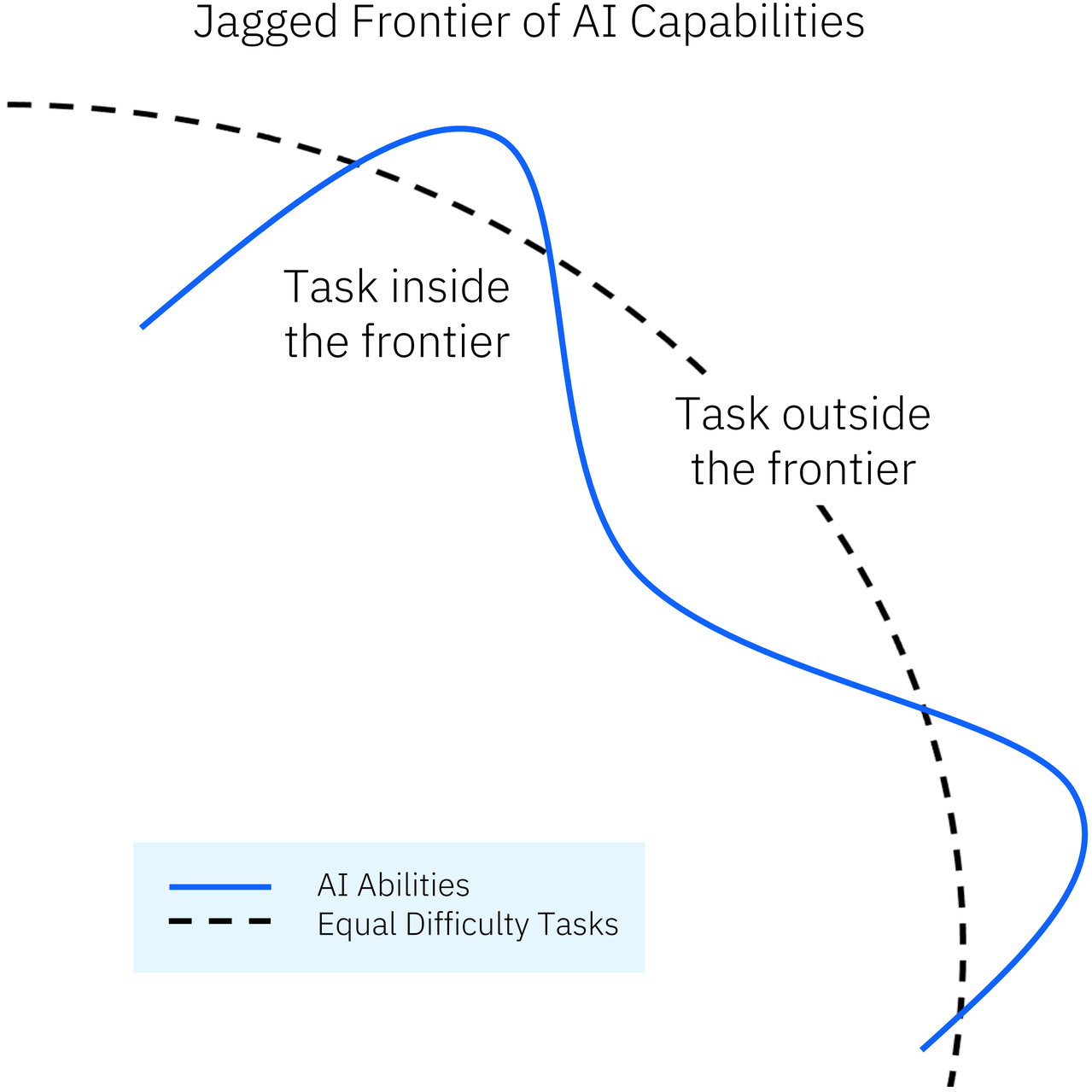


Figure 1: The Jagged AI Frontier, page 27

Feedback loop examples

If we automate use case A …

- Will it short-circuit something important?
- How will our work council respond?
- What level of additional inaccuracy will [our users / our customers / society] accept if it increases their productivity?
- What risks are there to our brand?

Limiting risk …

Legal / ethical / organizational

- Training data provenance
- Intellectual property management (yours and your customers’)
- Bias? Accuracy? Transparency?

→ Governance

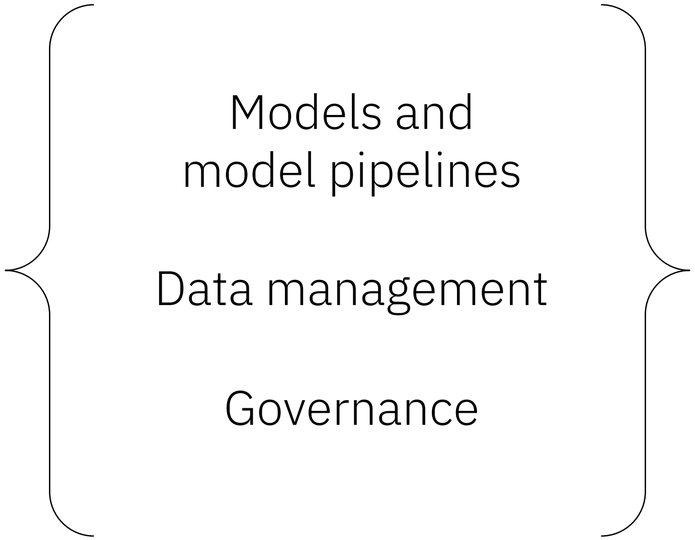
Governance to limit risk

What models are available?

What teams may use which data, which models and for which purposes?

What work was done to ensure model performs OK?

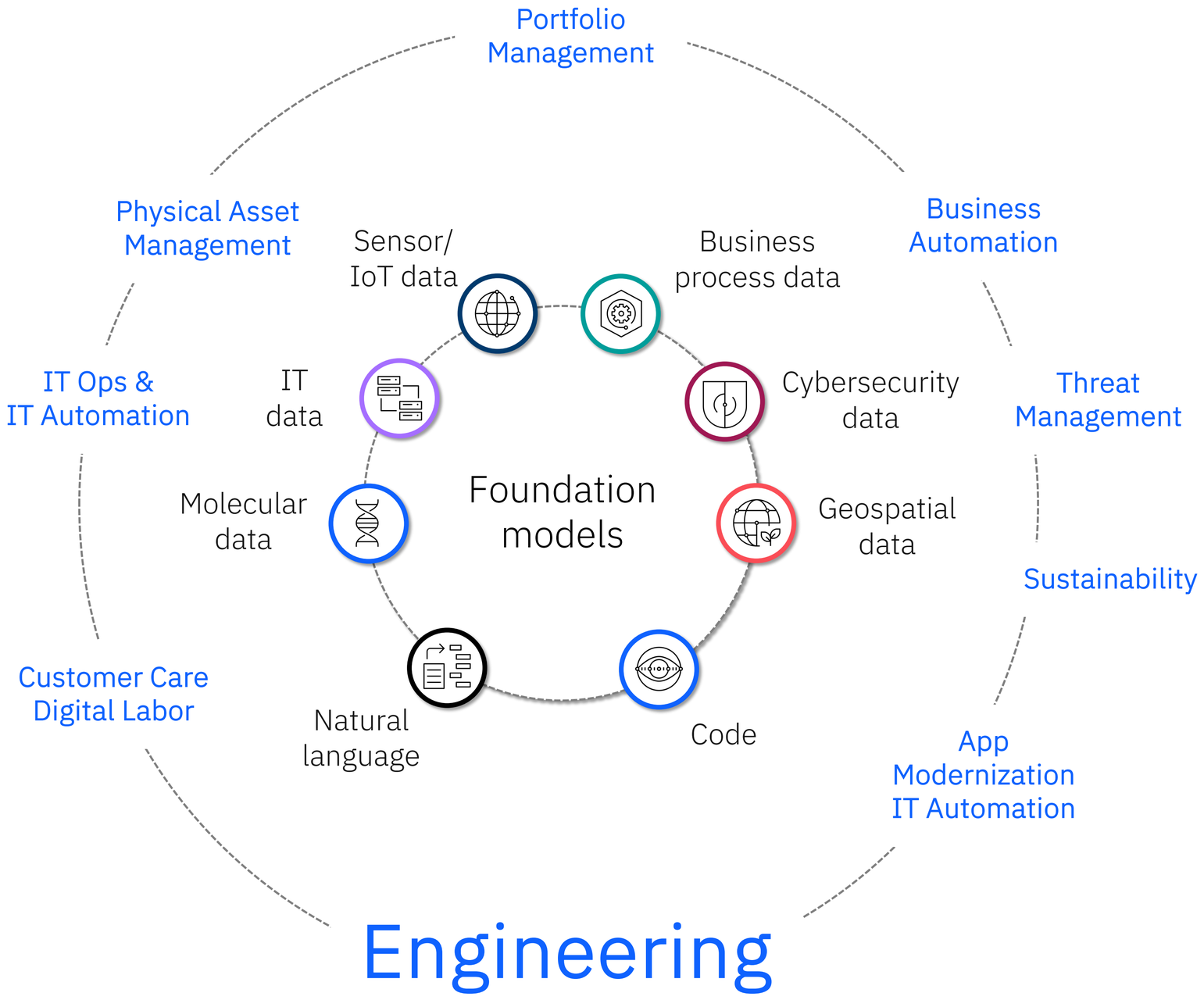
What development process are teams following?



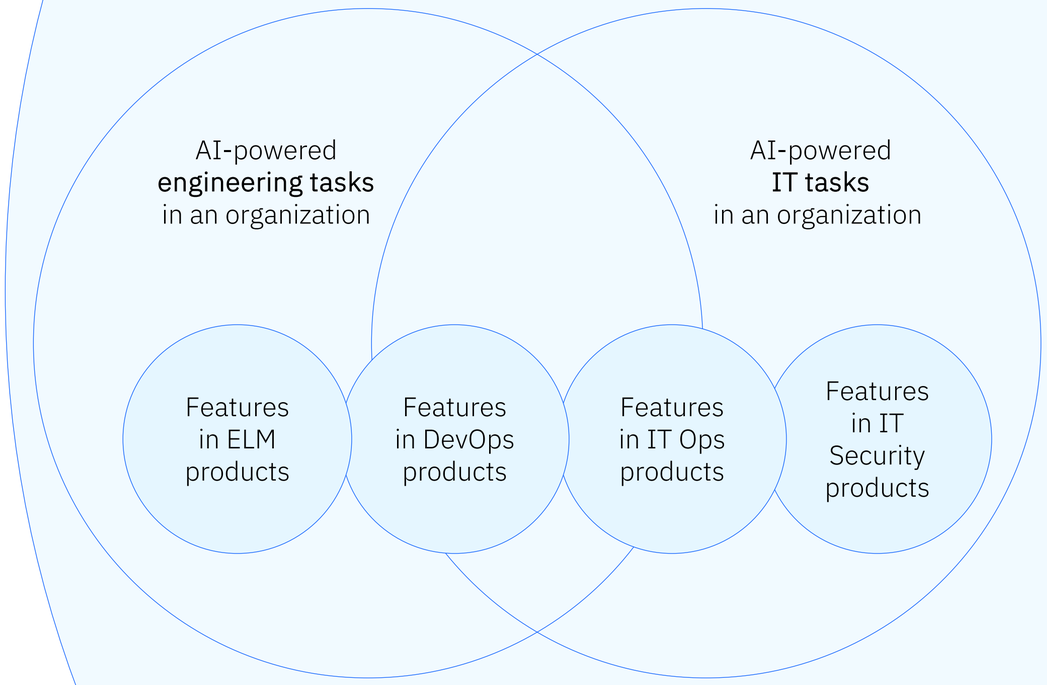
Go bigger

3

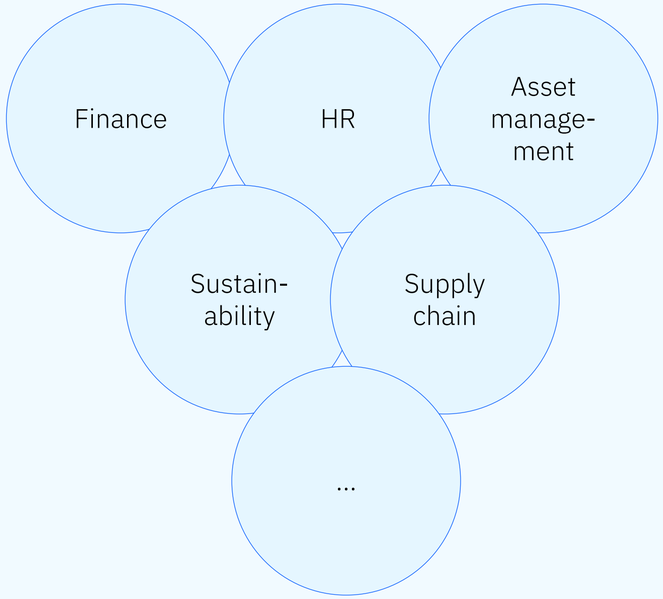
Opportunity to unlock business advantage with foundation models trained across the breadth of enterprise data



AI opportunities all over the place



AI-powered enterprise



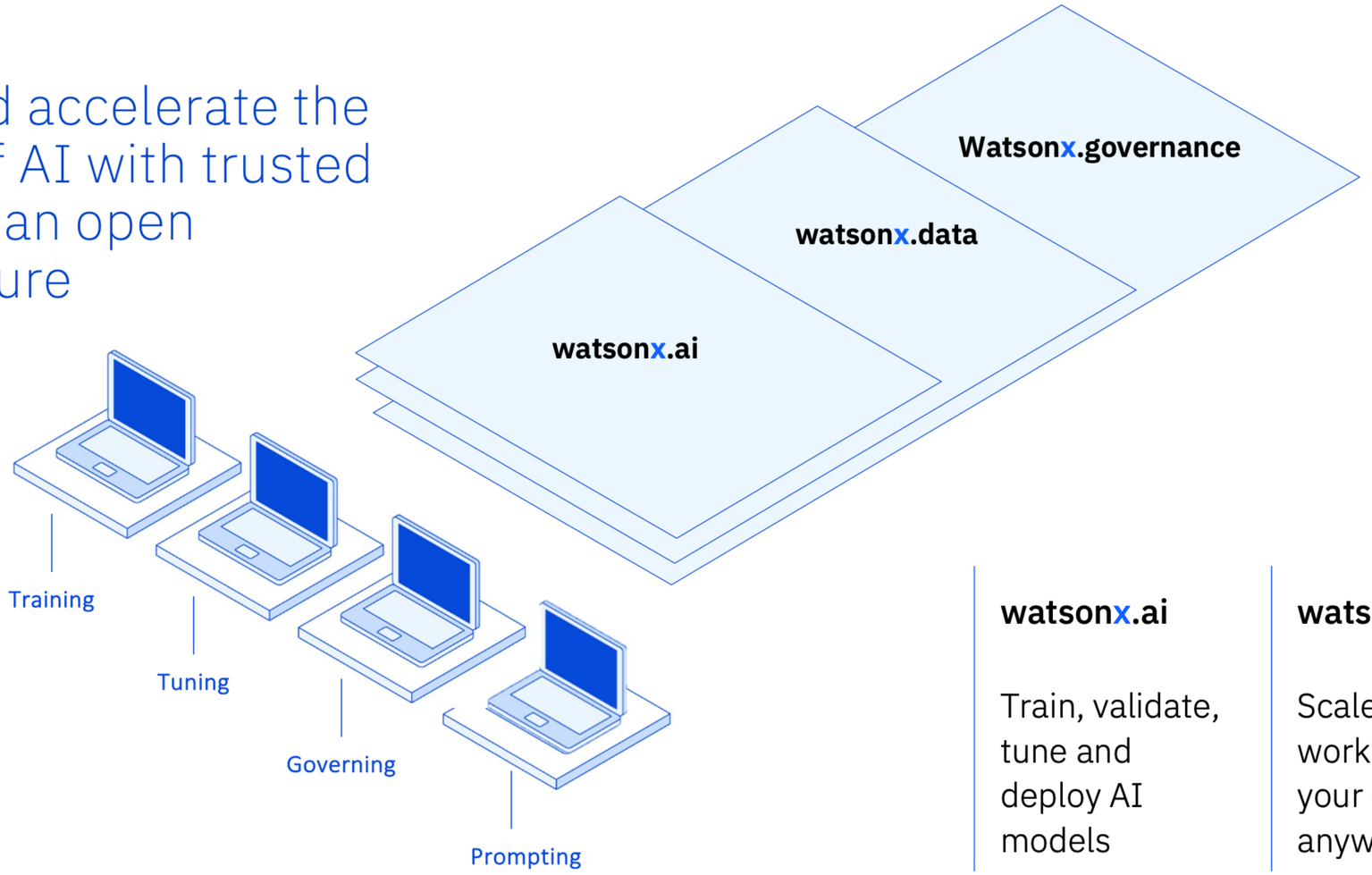
watsonx

4

watsonx

The platform for AI and data

Scale and accelerate the impact of AI with trusted data and an open architecture



watsonx.ai

Train, validate, tune and deploy AI models

watsonx.data

Scale AI workloads, for all your data, anywhere

watsonx.governance

Enable responsible, transparent and explainable AI workflows

IBM watsonx.ai Foundation Models Library: from open source, IBM, partners, your own

starcoder-15.5b

The StarCoder models are 15.5B parameter models that can generate code from natural language descriptions.

Foundation model by BigCode

mt0-xxl-13b

An instruction-tuned iteration on mT5.

Foundation model by BigScience

gpt-neox-20b

A 20 billion parameter autoregressive language model trained on the Pile.

Foundation model by EleutherAI

flan-t5-xl-3b

A pretrained T5 - an encoder-decoder model pre-trained on a mixture of supervised / unsupervised tasks converted into a text-to-text format.

Foundation model by Google

flan-t5-xxl-11b

flan-t5-xxl is an 11 billion parameter model based on the Flan-T5 family.

Foundation model by Google

flan-ul2-20b

flan-ul2 is an encoder decoder model based on the T5 architecture and instruction-tuned using the Fine-tuned Language Net.

Foundation model by Google

mixtral-8x7b-instruct-v01-q

Mixtral-8-7b-instruct-v01-gptq model is made with AutoGPTQ, which mainly leverages the quantization technique to 'compress' the model weights from...

Foundation model by Mistral AI, tuned by IBM

granite-13b-chat-v1

The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks.

Foundation model by IBM

granite-13b-chat-v2

The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks.

Foundation model by IBM

granite-13b-instruct-v1

The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks.

Foundation model by IBM

granite-13b-instruct-v2

The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks.

Foundation model by IBM

mpt-7b-instruct2

MPT-7B is a decoder-style transformer pretrained from scratch on 1T tokens of English text and code. This model was trained by IBM.

Foundation model by Mosaic, tuned by IBM

llama-2-13b-chat

Llama-2-13b-chat is an auto-regressive language model that uses an optimized transformer architecture.

Foundation model by Meta

llama-2-70b-chat

Llama-2-70b-chat is an auto-regressive language model that uses an optimized transformer architecture.

Foundation model by Meta

Model in deprecation

Why watsonx?

1. Hybrid models: open source, IBM, partners, your own
2. Manage IP ownership / IP risk
3. On-prem or on cloud: your choice
4. Whole-lifecycle governance
5. Cost of inference

IBM Engineering Lifecycle Management

5

Gen AI offers many opportunities along the ‘V’, for example…

Digital thread / trace analysis

Plan Assist

Assess requirements from customer

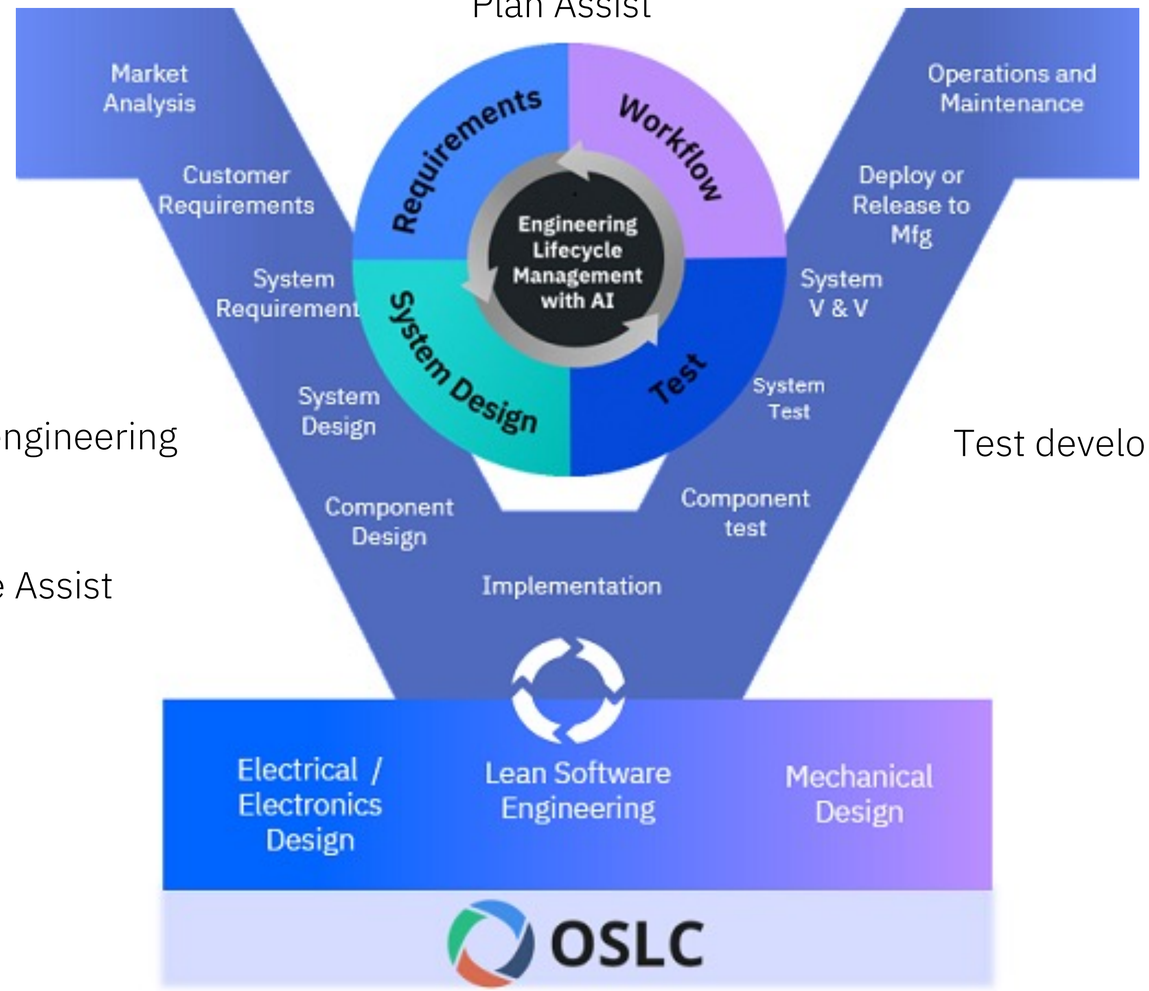
Find duplicate and conflicting requirements

Requirements quality

Accelerate systems engineering model development

Code Assist

Test development



Some scenarios are more interesting than others

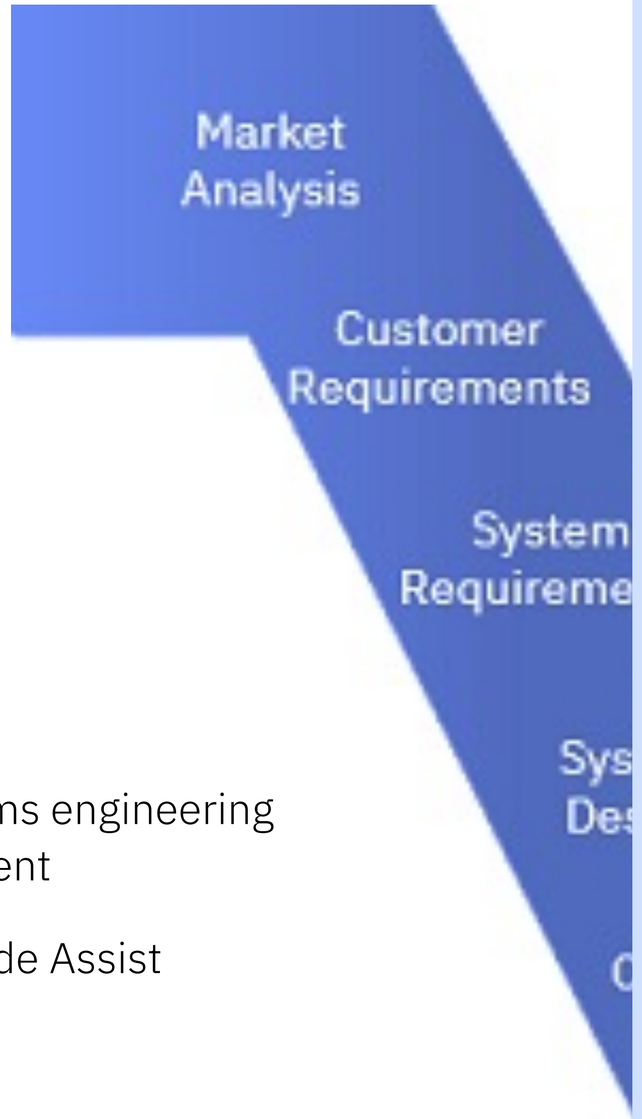
Assess requirements from customer

Find duplicate and conflicting requirements

Requirements quality

Accelerate systems engineering model development

Code Assist



Why these scenarios?

Unbounded problems

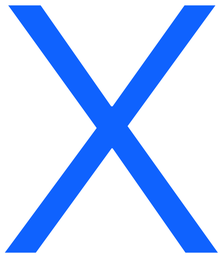
“Shift left” acceleration

Q&A

6

Put AI to work with **watsonx**.

Learn more



Learn more

https://www.ibm.com/watsonx

https://www.ibm.com/products/engineering-lifecycle-management

Thank you