

watsonx

AI for Business - Ansvarlig AI



Bo Holtemann
Data & AI Governance Leader
IBM
bholte@dk.ibm.com
+4528808188
<https://www.linkedin.com/in/boholtemann/>

Accelerer din AI indsats – få indblik i IBM's AI-plattform watsonx.

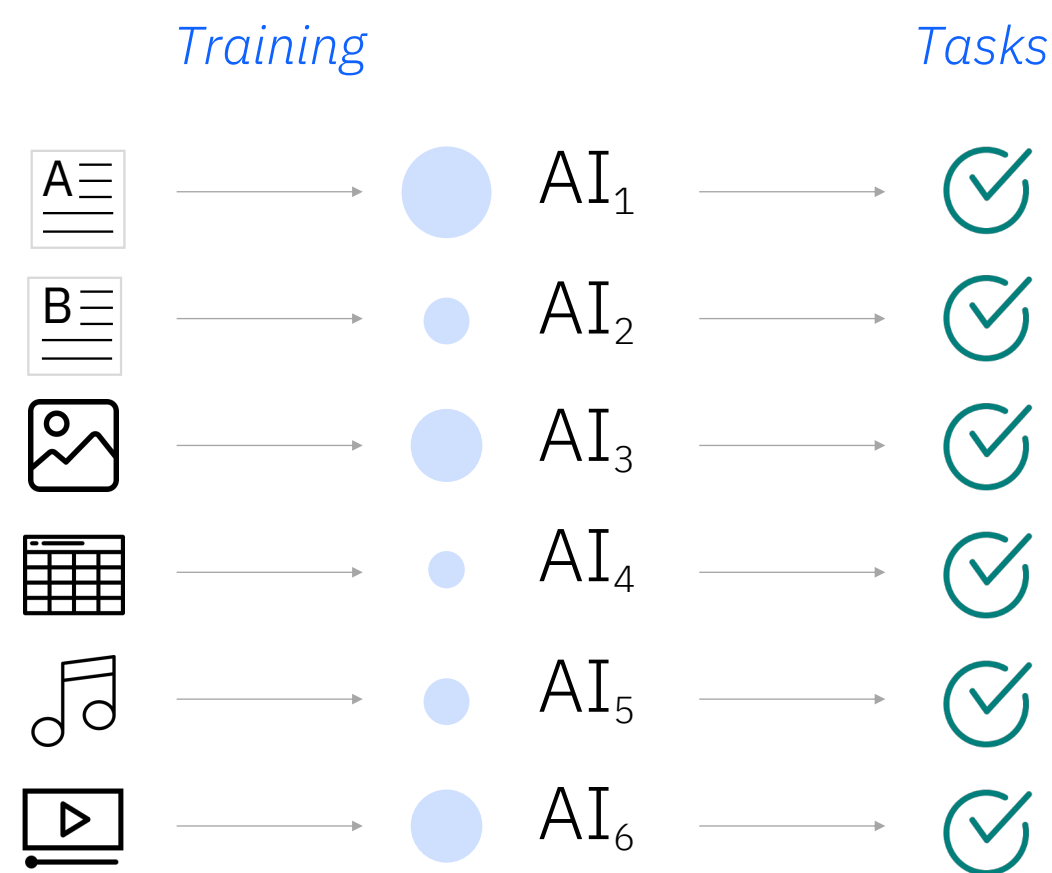
watsonx – IBMs AI platform er seneste løsning fra IBM som tillader virksomheder at deltage i rejsen på ML & Generativ AI i fuld kontrol.

- Din data er din data
- Undgå lock in
- fuld evidens & transparens dokumenteret fra data source til AI output
- At du er klar til EU AI act

Konkrete eksempler på ansvarlig AI ved use cases / kundecases

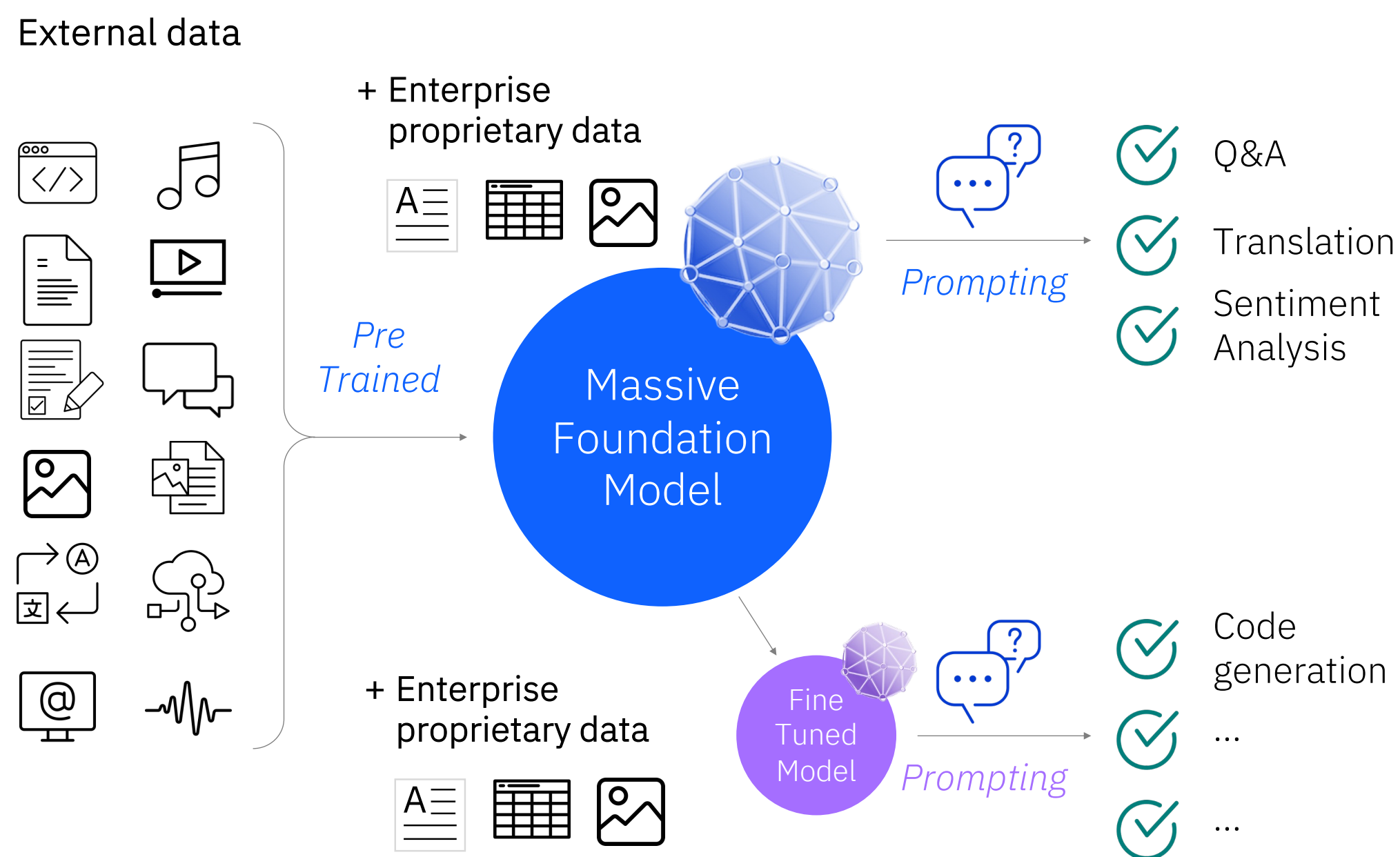
Foundation models establish a new paradigm for AI capabilities

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 and inferencing
- Equal or better accuracy for multiple use cases
- Incremental revenue through better performance

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

Har du som andre bekymringer vedrørende (gælder for traditionel ML og GenAI)?

How was it trained?

- Garbage in -> garbage out
- An enterprise cannot use a foundation model trained with a Wikipedia crawl
- The training material needs to be huge and comprehensive but must also be curated

Can it detect & minimize bias & hallucination?

- How does the platform detect and correct bias?
- How can it prevent hallucination (providing random and untrue answers with absolute aplomb and convictions)?

Is it transparent?

- Open vs black-box
- How to audit, and explain the model and the answers it generates?
- Does the model track drift and bias? And how does it address them?

Does it support regulatory compliance?

- How do foundation models and their usage comply with privacy and government regulations?
- What are the guardrails?
- Who is responsible for an inadvertently exposed PII or a “wrong answer”?

Is it safe?

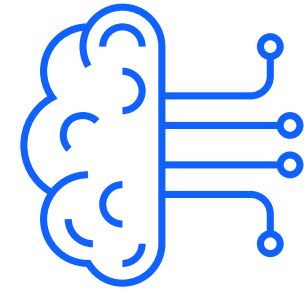
- Who has control over the model, input data, and output data?
- How to ensure that confidential information is not given out?
- How is it monitored?
- What safety features and guardrails are in place?

Can it be customized?

- Hybrid and multicloud?
- Can the model be fine-tuned with clients’ data?
- How can clients update, and extend the model to make it more suitable for their use cases?
- How to integrate with applications? What APIs are in place?

IBM has continuously strived for responsible innovation capable of bringing benefits to everyone and not just a few.

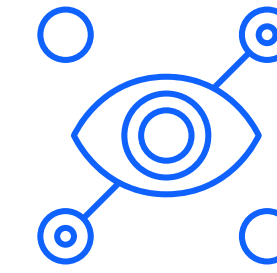
IBM applies the same philosophy to AI through its [principles for trust and transparency](#).



The purpose of AI is to augment human intelligence



Data and insights belong to their creator



New technology, including AI systems, must be transparent and explainable

Open

open source foundation models +

Targeted

Der er ikke en stor model, i stedet mange mindre og fokuserede til dit domain – din industri

Trusted

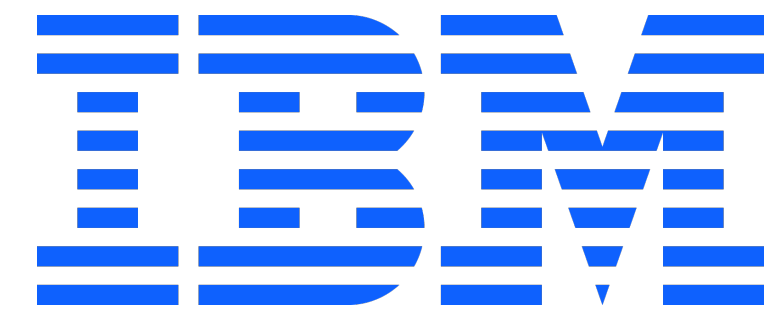
Skab evidens og transparens fra data kilde til AI output

Empowering

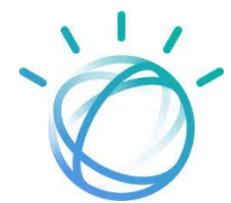
“Airgapped” mulighed for din data i dit miljø (i din cloud eller datacenter

= ANSVARLIG AI,
AI FOR BUSINESS

watsonx



IBM watson storyline + fremtid



2008: IBM Watson brand created to align to innovation in AI

2011: IBM debuts system to compete against long standing Jeopardy champions Ken Jennings and Brad Rutter



2014-2015: : IBM debuts partnership with Mayo Clinic and launches Watson Health for health-related data, assets and AI applications



2017: IBM embarks on creating a standard set of NLP/Speech/Dialog libraries.

2018: Watson Assistant is launched, IBM's conversational AI technology

2019: IBM builds, Project Debater, an AI powered system to debate a human in a live debate competition using deep NLP and generative AI capabilities

2019: IBM introduces standard libraries that will power all "Watson" products going forward. These libraries contain foundation models, Language Model compression, proprietary models and open source components.

2020: IBM infuses Foundational models Into AI applications like Watson Assistant and Watson Discovery

2023: IBM launches WatsonX - an AI built and infused platform for working with, customizing, tuning, deploying and governing foundation models.

2024: Governance and trust permeate AI

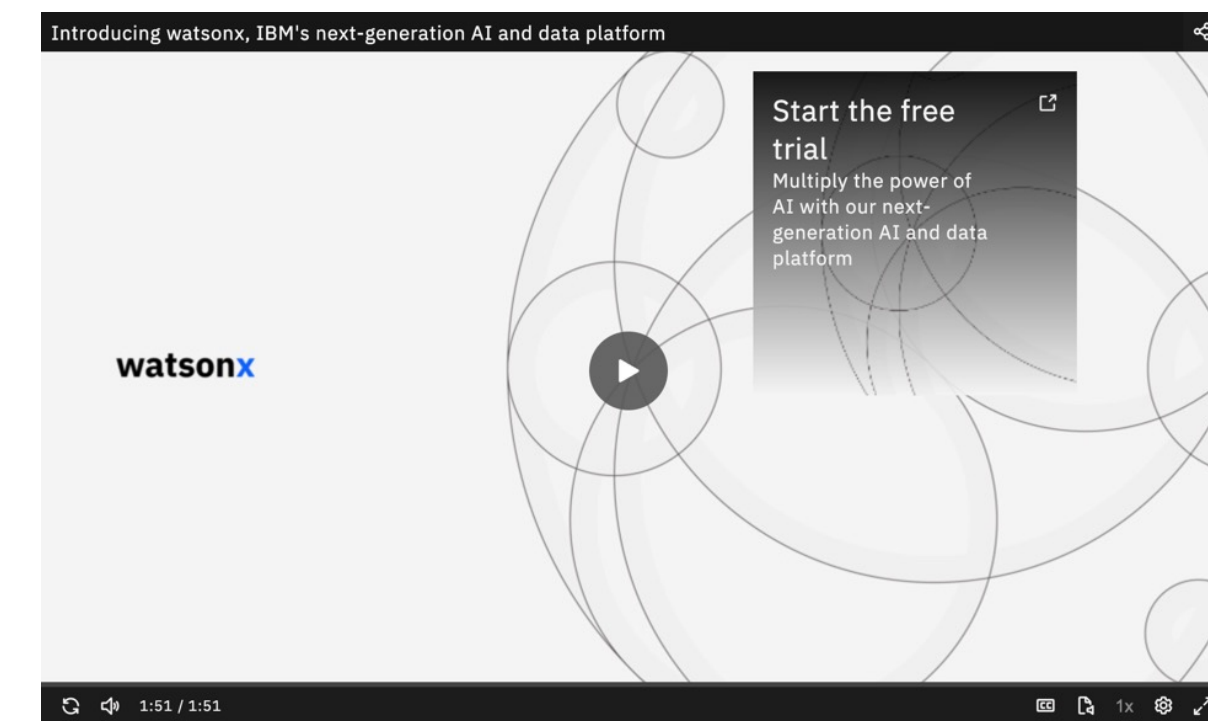
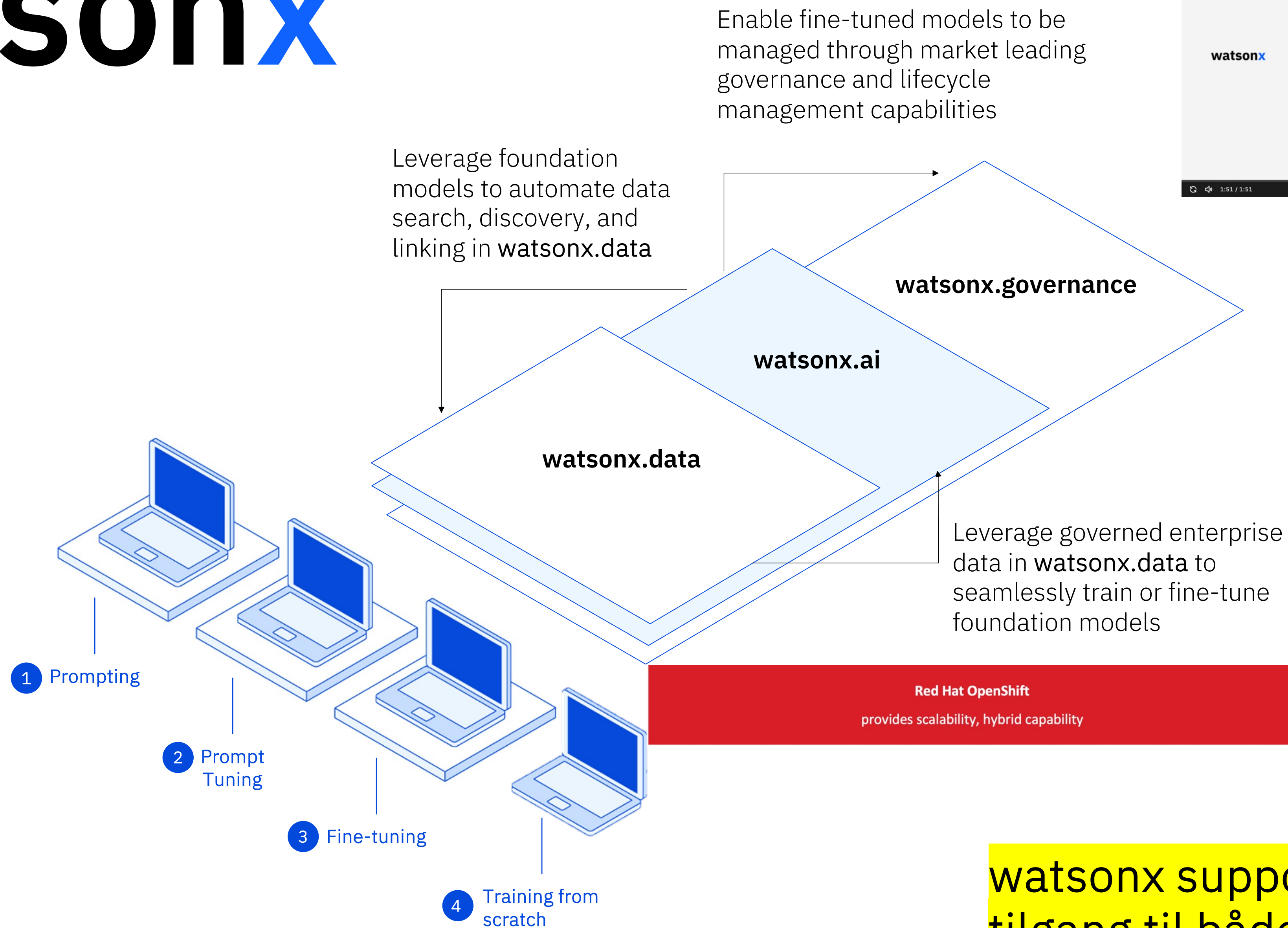
2025: AI becomes more energy and cost efficient

2027: Foundation models in production scale uniquely

2029: Trustworthy and explainable AI starts to reason

2030: Fully multi-modal AI gives enterprises unprecedented scale

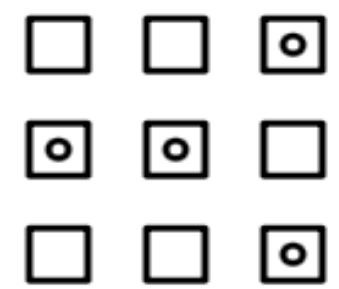
watsonx



3 selvstændige komponenter og pre-integrerede hvis ønsket

watsonx supporterer samme tilgang til både traditionel ML og generative AI

Generative AI capabilities



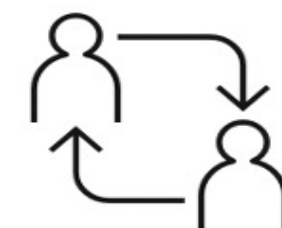
Foundation model library



Prompt Lab



Tuning Studio*

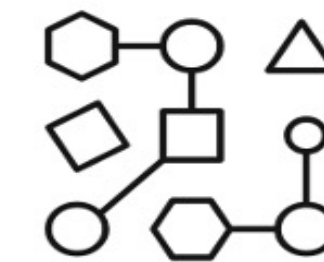


Team collaboration and data preparation

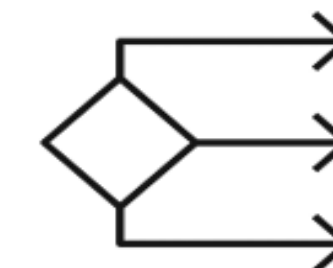
watsonx.ai

Train, validate, tune
and deploy AI models

A proven studio for machine learning



ModelOps

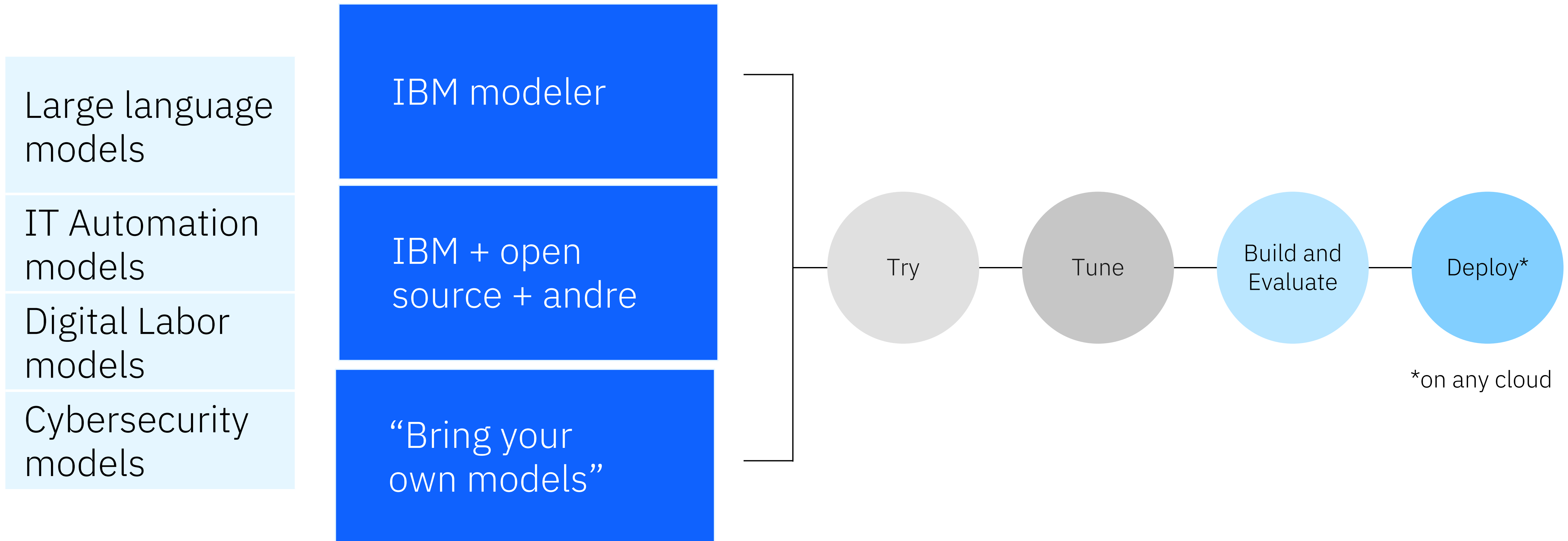


Automated development



Decision optimization

foundation models - multi-model - *multi-cloud - ingen lock-in



watsonx.ai Foundation Model Library


IBM Foundation Models

Slate (encoder-only) NLP models

Granite (decoder-only)

154 million params

Multilingual distilled




Efficient Domain and Task Specialization

Models Coming Soon:

- Finance
- Cybersecurity
- Legal, etc.

Open-Source Large Language Models



flan-ul2-20b	gpt-neox-20b	mt0-xxl-3b	flan-t5-xxl-11b	mpt-7b-instruct2
20 billion params	20 billion params	13 billion params	11 billion params	7 billion params
encoder/decoder	decoder only	encoder/decoder	encoder/decoder	decoder only

3rd Party Large Language Models

Llama2-chat-70b	Starcode-15.5b
70 billion params	15.5 billion params
decoder only	decoder only

Model variety to cover enterprise use cases
and compliance requirements

Language Tasks

Q&A

Generate

Extract

Summarize

Classify

Model responds to a question in natural language

Model generates content in natural language

Model extracts entities, facts, and info. from text


Model creates summaries of natural language

Model classifies text (e.g., sentiment, group)

Coding Tasks

Code Gen

Model generating code from a natural language prompt



watsonx.ai IBM Foundation Models – data

“rensning” process = **tillid til data**

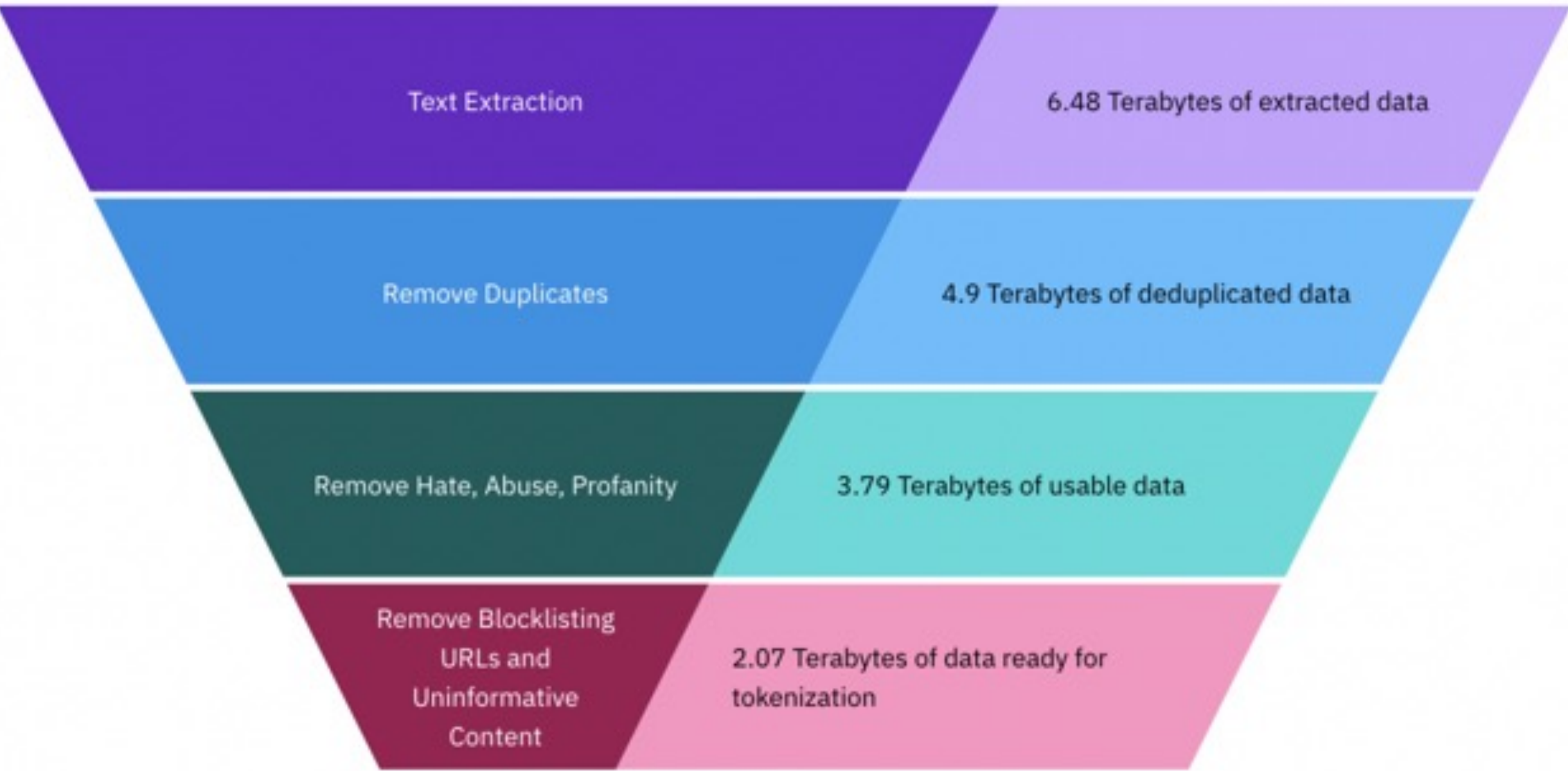


Fig. 2. Summary governance statistics on IBM’s curated pre-training dataset at the time of granite.13b’s training.

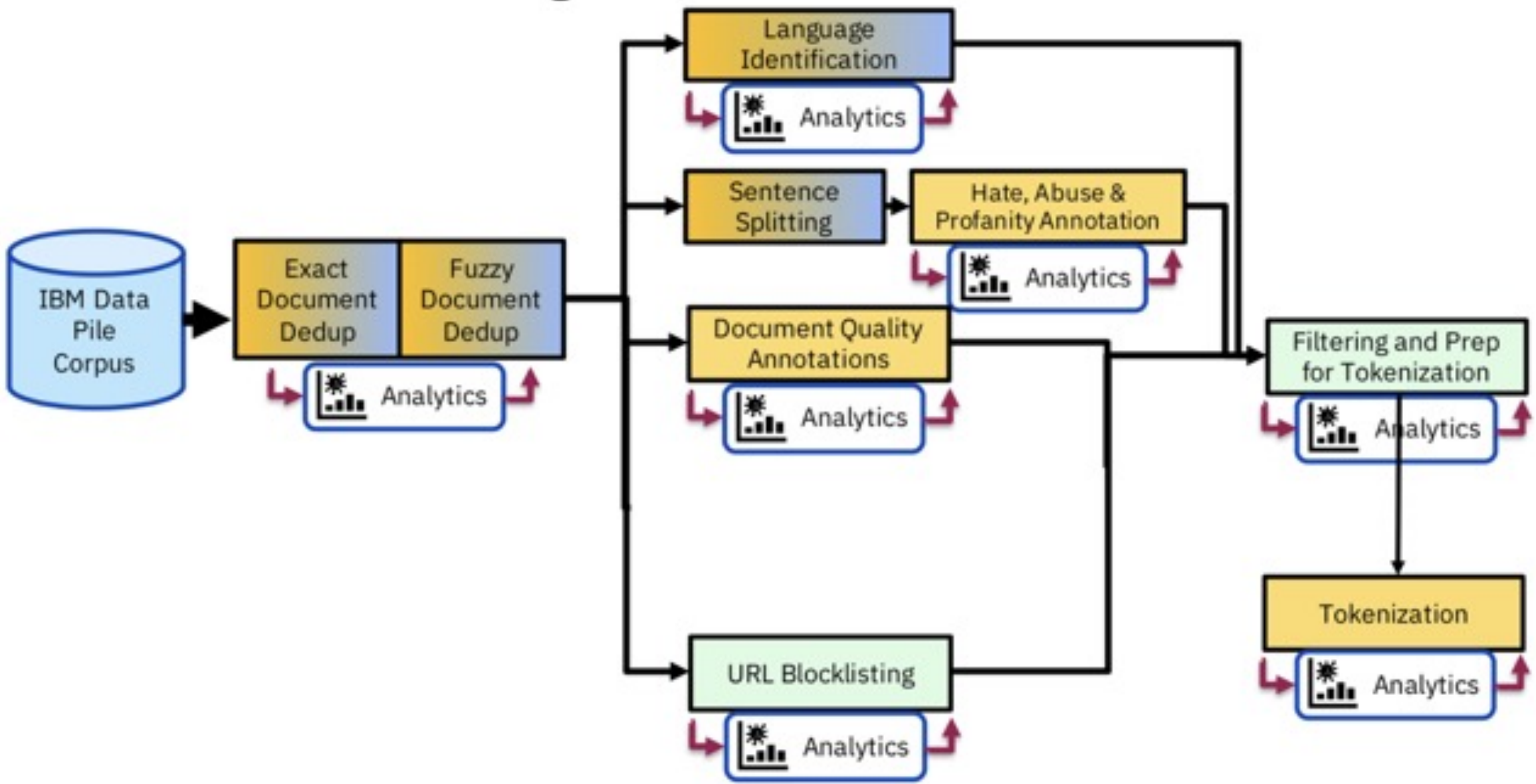


Fig. 3. IBM’s Data pre-processing pipeline.

Transparent Pre-Training on IBM’s trusted Data Lake

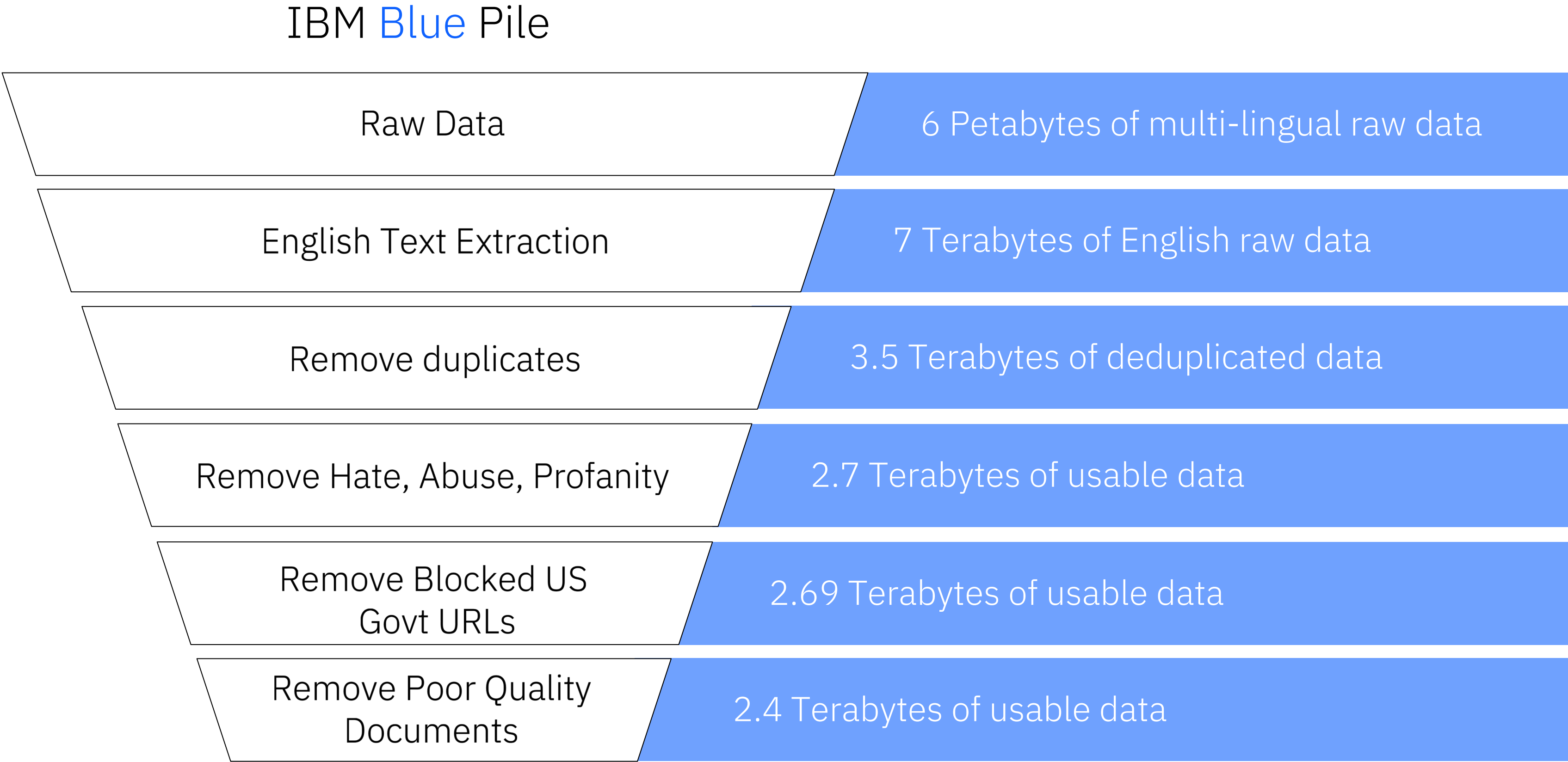
- One of the largest repositories of enterprise-relevant training data
- Verified legal and safety reviews by IBM
- Full, auditable data lineage available for any IBM Model

[IBM Granite Foundation Model whitepaper](#)

[Foundation models: Opportunities, risks and mitigations](#)

Key Differentiators

- Superior performance at a smaller size and lower TCO
- Trained on highly governed, cleansed and de-duplicated data, with full, auditable data lineage
- AI Factsheets available for any IBM Model
- Legal Indemnification for 3rd Party Copyright Claims



Tokenized to 1+ Trillion Tokens for Training

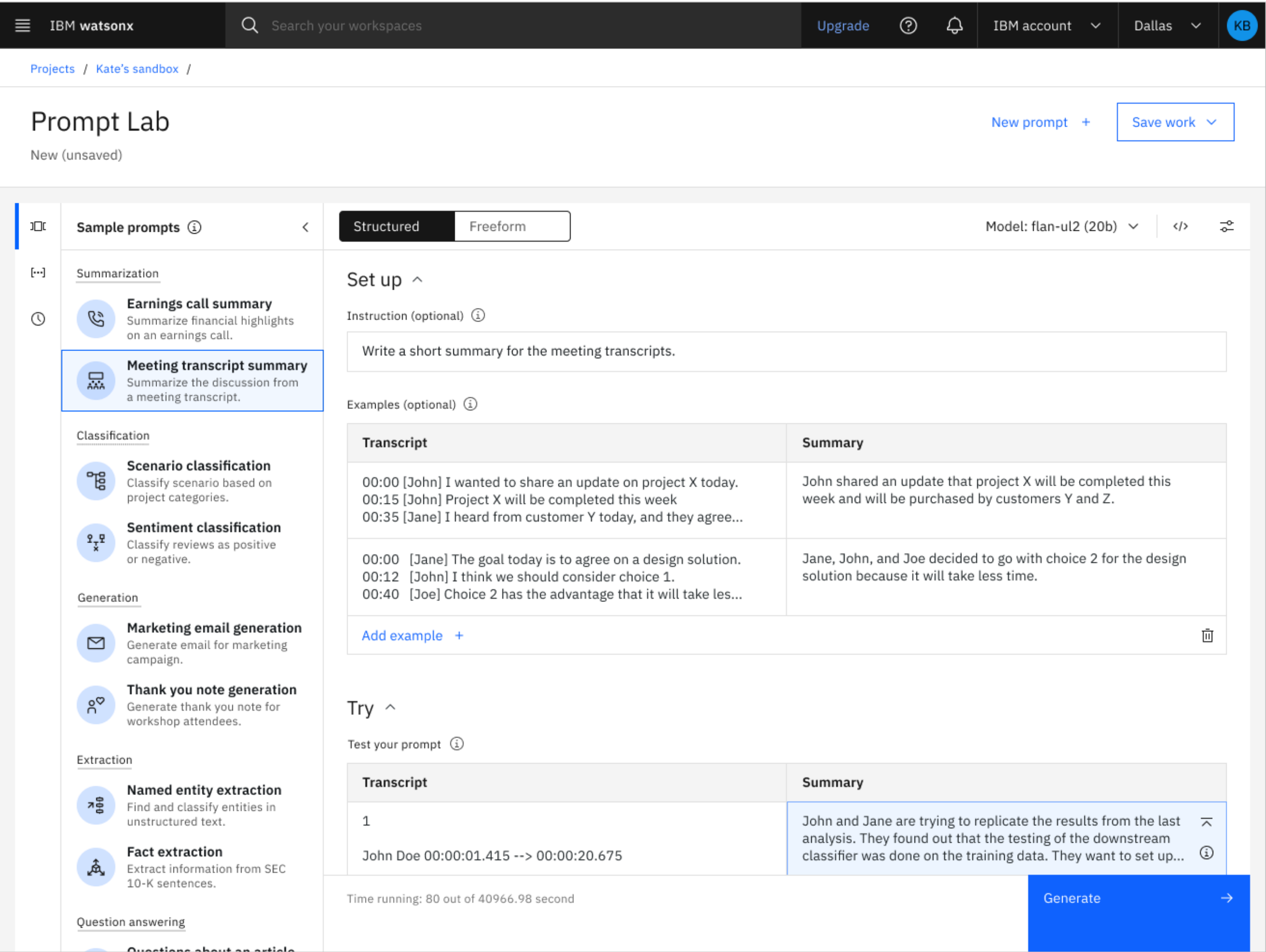
Experiment with foundation models
and build prompts

Interactive prompt
builder

Includes prompt examples
for various use cases
and tasks

Experiment with different
prompts, save and reuse
older prompts, use different
models and vary different
parameters

Experiment with zero-shot,
one-shot, or few-shot
prompting to get the
best results



Experiment with
prompt engineering

Choice of foundation models
to use based on task
requirements

Prevent the model from
generating repeating phrases

Number of min and max
new tokens in the response

Stop sequences – specifies
sequences whose appearances
should stop the model

watsonx.ai: Tuning Studio*

Tune your foundation models with labeled data

Prompt tuning

Efficient, low-cost way of adapting an AI foundation model to new downstream tasks

Tune the prompts with no changes to the underlying base model or weights

Unlike prompt engineering, prompt tuning allows clients to further train the model with focused, business data

Task support in the Tuning Studio

Models support a range of Language Tasks: Q&A, Generate, Extract, Summarize, Classify

Requires a small set of labelled data to perform specialized tasks

Can achieve close to fine-tuning results without model modification, at a lower cost to run

The screenshot displays the IBM watsonx Tuning Studio interface. At the top, there's a navigation bar with the IBM watsonx logo, a search bar, and user account information. The main content area is titled 'Tune foundation models with labeled data' and includes a subtitle 'Start your custom tune by selecting a tuning method, foundation model, and use case. [Learn more](#)'. Below this, a sidebar on the left lists the steps: 'Set up' (active), 'Add training data', 'Edit parameters', and 'Review and tune'. The main panel, titled 'Set up your tune', contains three dropdown menus: 'Tuning method' set to 'Prompt tuning', 'Foundation model' set to 'flan-ul2 (20b)', and 'Use case' set to 'Summarization'. Each dropdown has a brief explanatory text below it. At the bottom of the panel, there are three buttons: 'Cancel', 'Back', and 'Next'.

*Coming soon, available post-GA

watsonx.ai: Data Science and MLOps

Build machine learning models automatically in the studio

Model training and development

Build experiments quickly and enhance training by optimizing pipelines and identifying the right combination of data

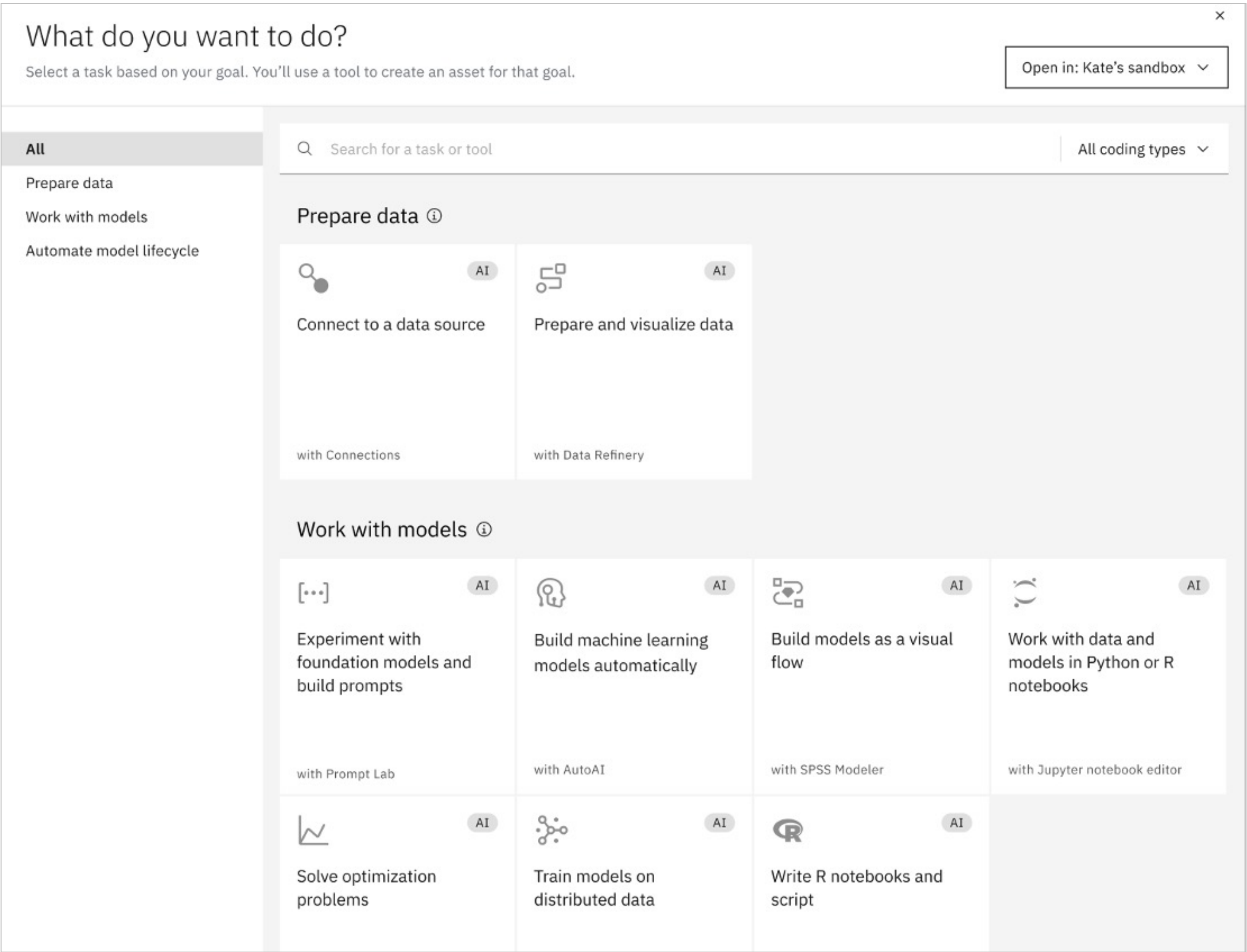
AutoAI, including preparing data for machine learning and generating and ranking candidate model pipelines

Use predictions to optimize decisions, create and edit models in Python, in OPL or with natural language

Integrated visual modeling

Prepare data quickly and develop models visually to help visualize and analyze enterprise data to identify patterns and trends, explore opportunities, and make informed, insightful business decisions

- Uncover correlations
- Insight for hypotheses
- Find relationships and connections within the data

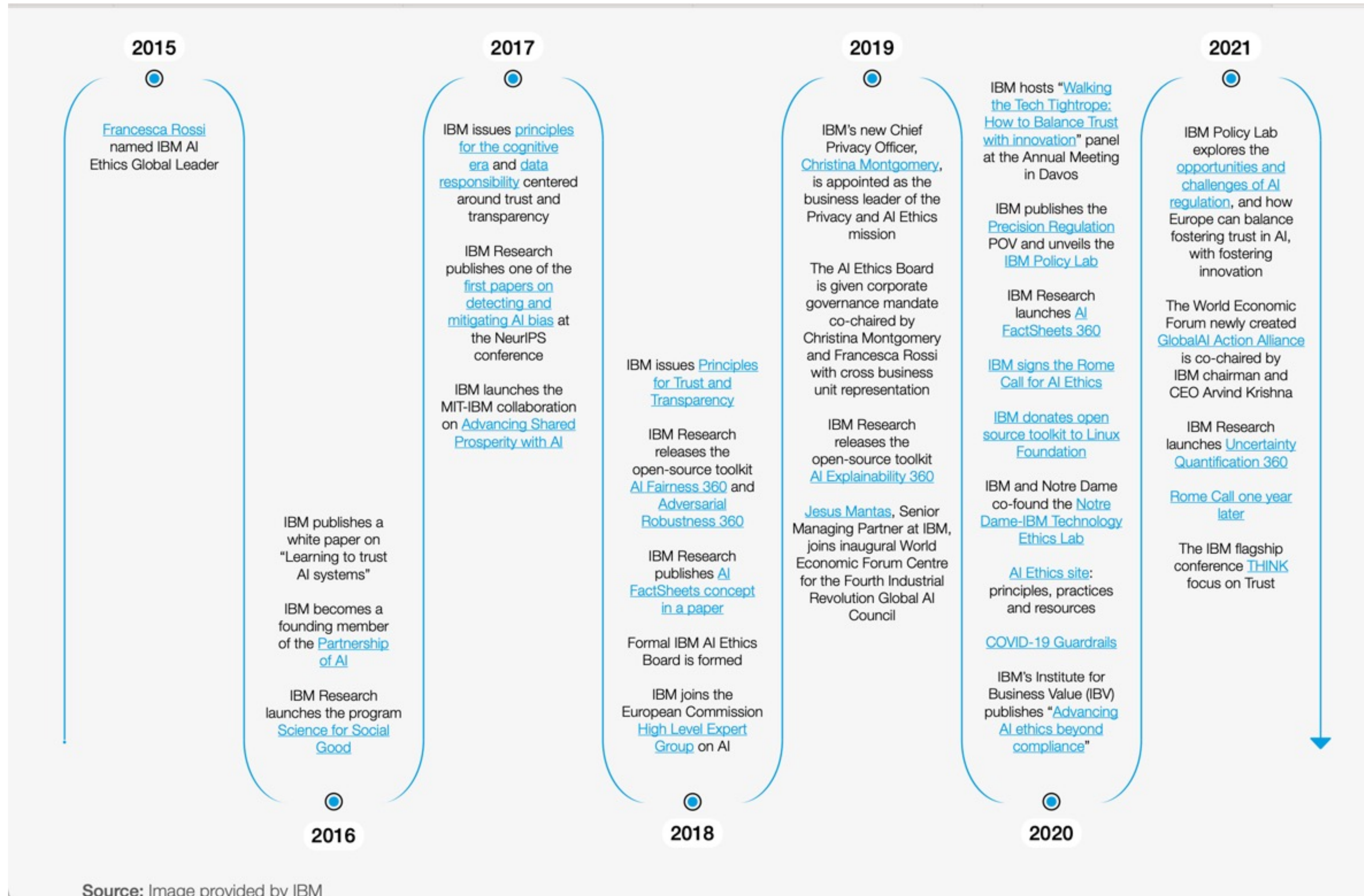


watsonx.governance

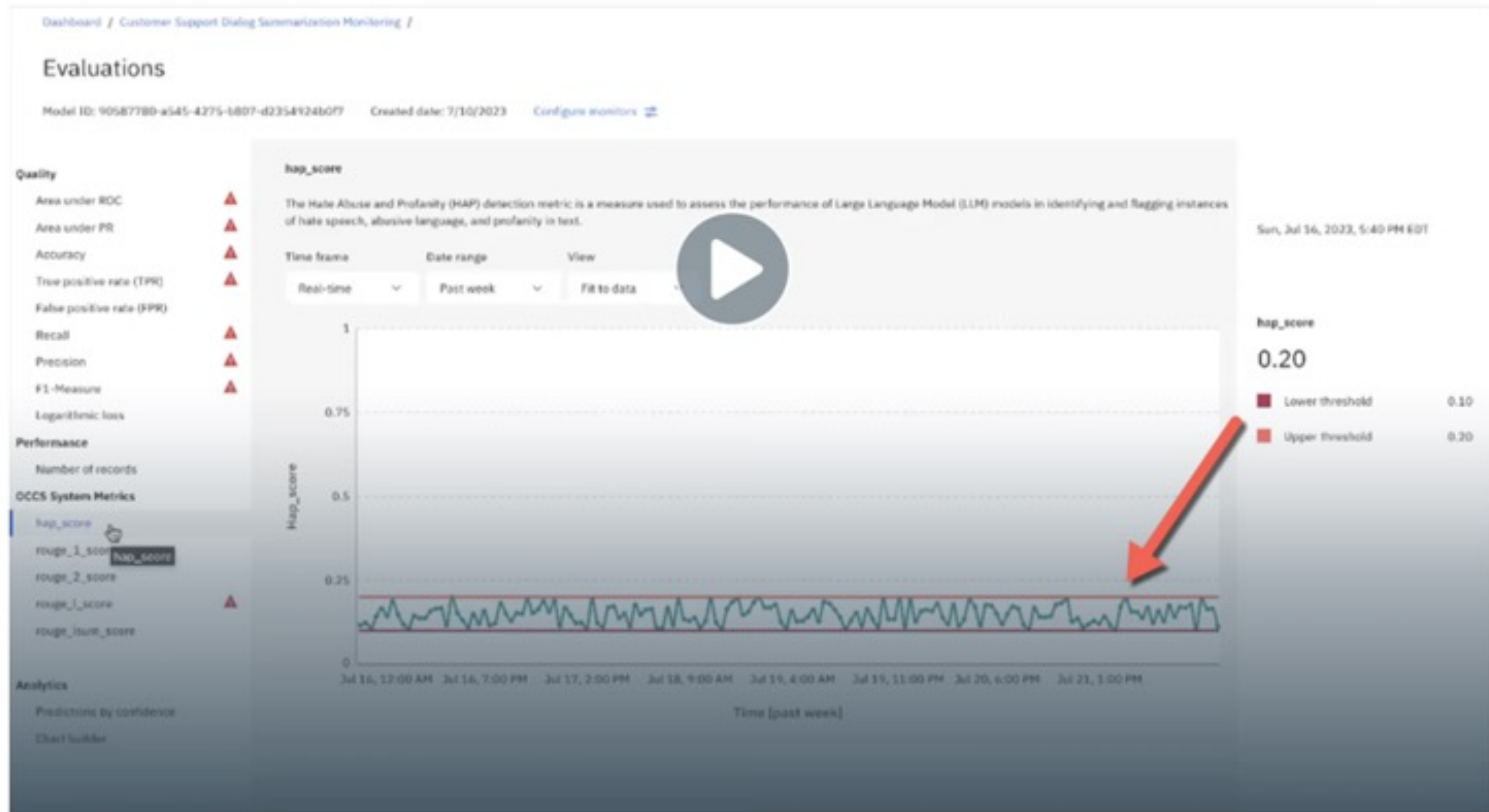
Enable responsible, transparent and explainable data and AI workflows

AI governance time travel

watsonx



Overvåg – rapporter & reager på brud på “hat, abuse & profanity”



De 3 grundstene i “Ansvarlig AI”

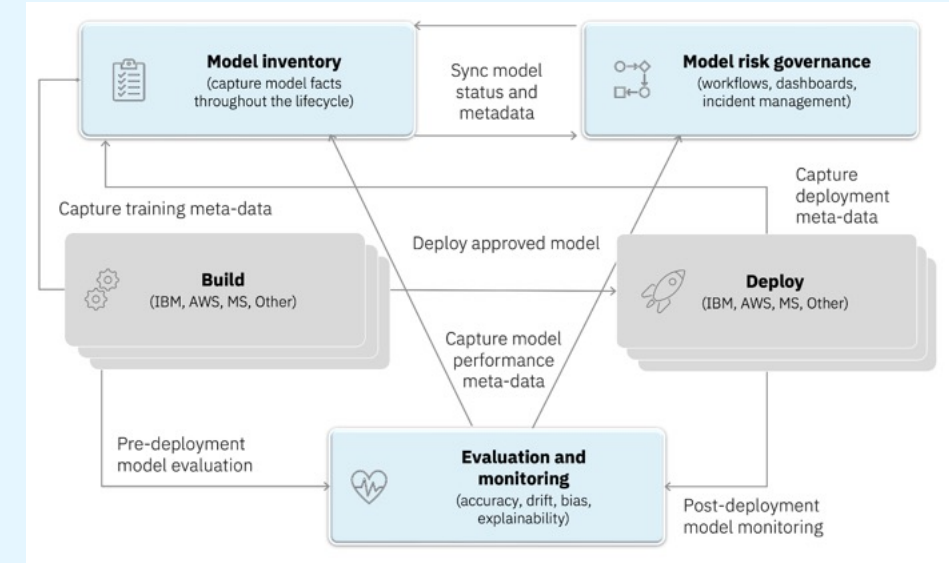


5.december
2023

IBM AI Governance

- Model governance
- Model documentation
- Model evaluation and monitoring

Predictive ML Focus

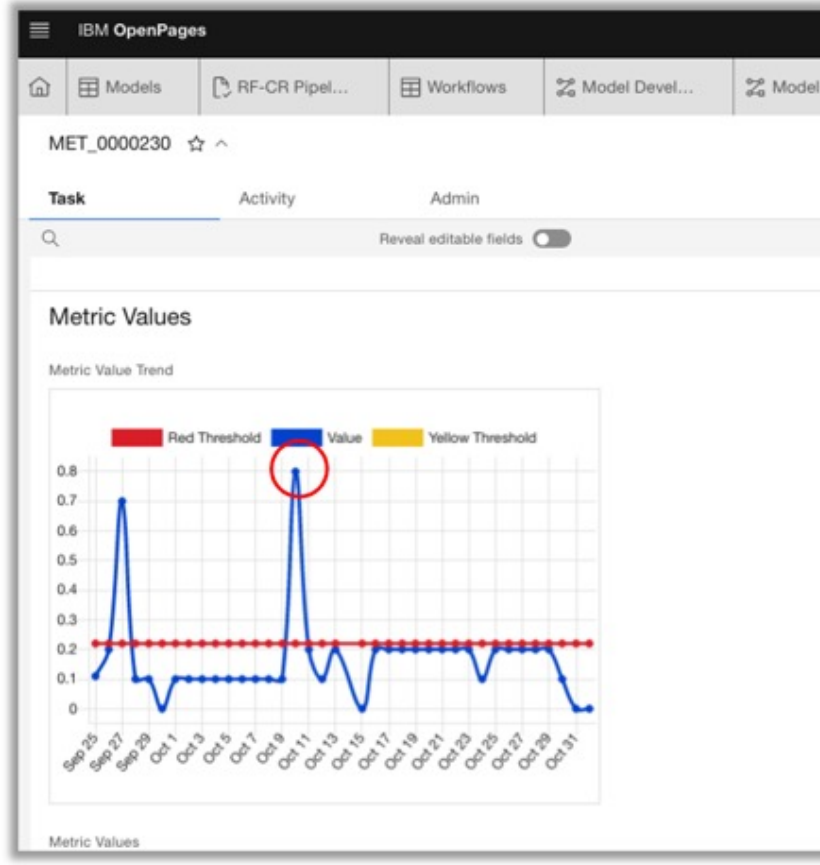


watsonx.governance

Extend the best practices of AI Governance from Predictive ML to Generative AI while monitoring and mitigating the new and amplified risks from models, users, datasets, and regulations.

• Dansk 2022 customer case med Deloitte: Håndtering af angreb på kreditmodel

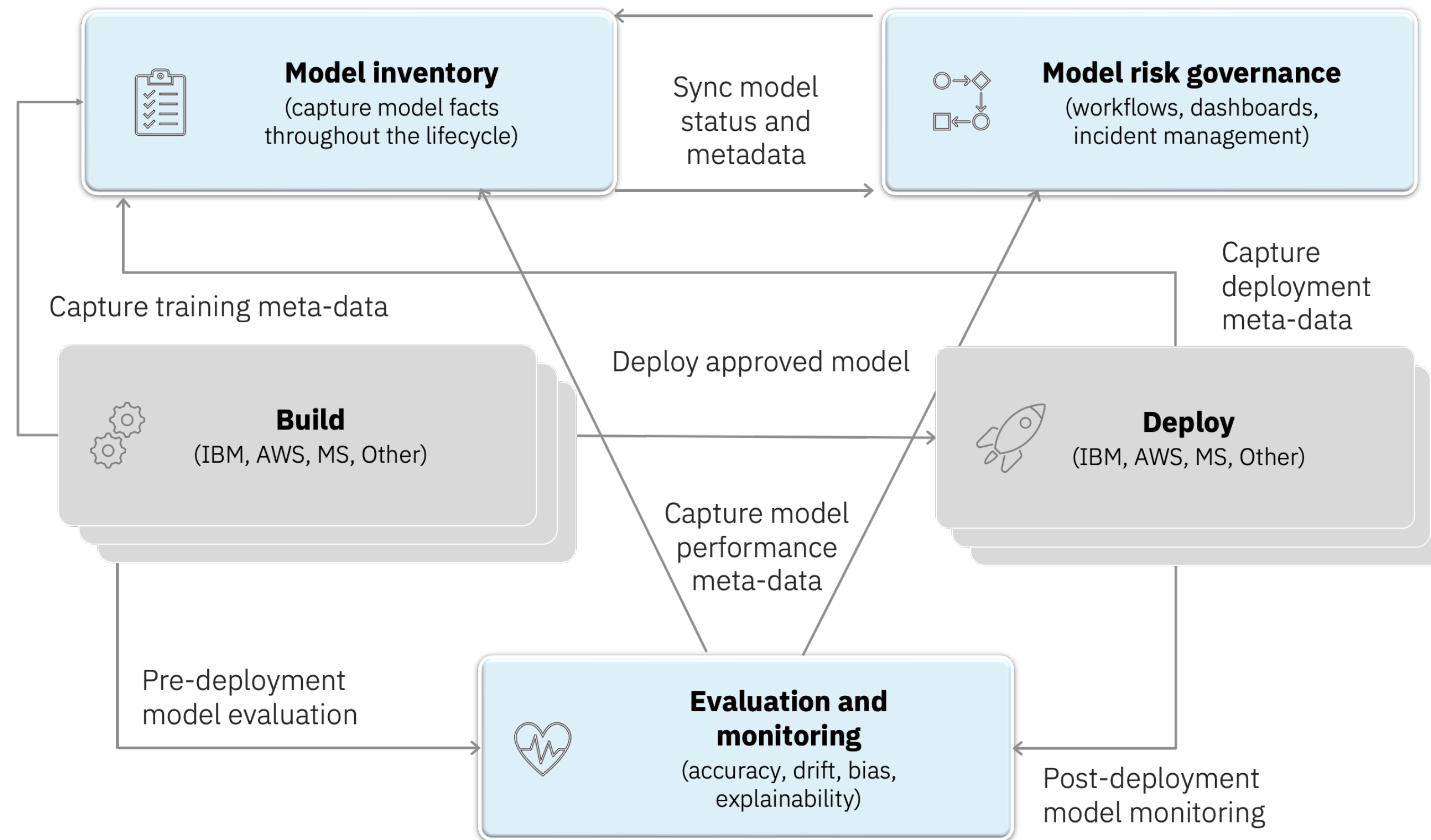
“ What should I do if there is a *breach*?



Expand AI Governance capabilities with net new features for Predictive ML and a complete new experience and functionality for governance of Generative AI

Manage and monitor

- Automate AI workflows to increase model transparency and explainability
- Capture the origin of data sets, model metadata and pipelines
- Support the governance of models built and deployed using 3rd party tools
- Use integrated collaboration and communication tools



Confidently scale

- Increase predictive accuracy, proactively identify and mitigate bias and drift
- Decrease time to model deployment using automated processes and collaborative tools
- Optimize data scientist talent, minimize costly human errors and speed time to model deployment
- Drive transparent processes and explainable analytic results

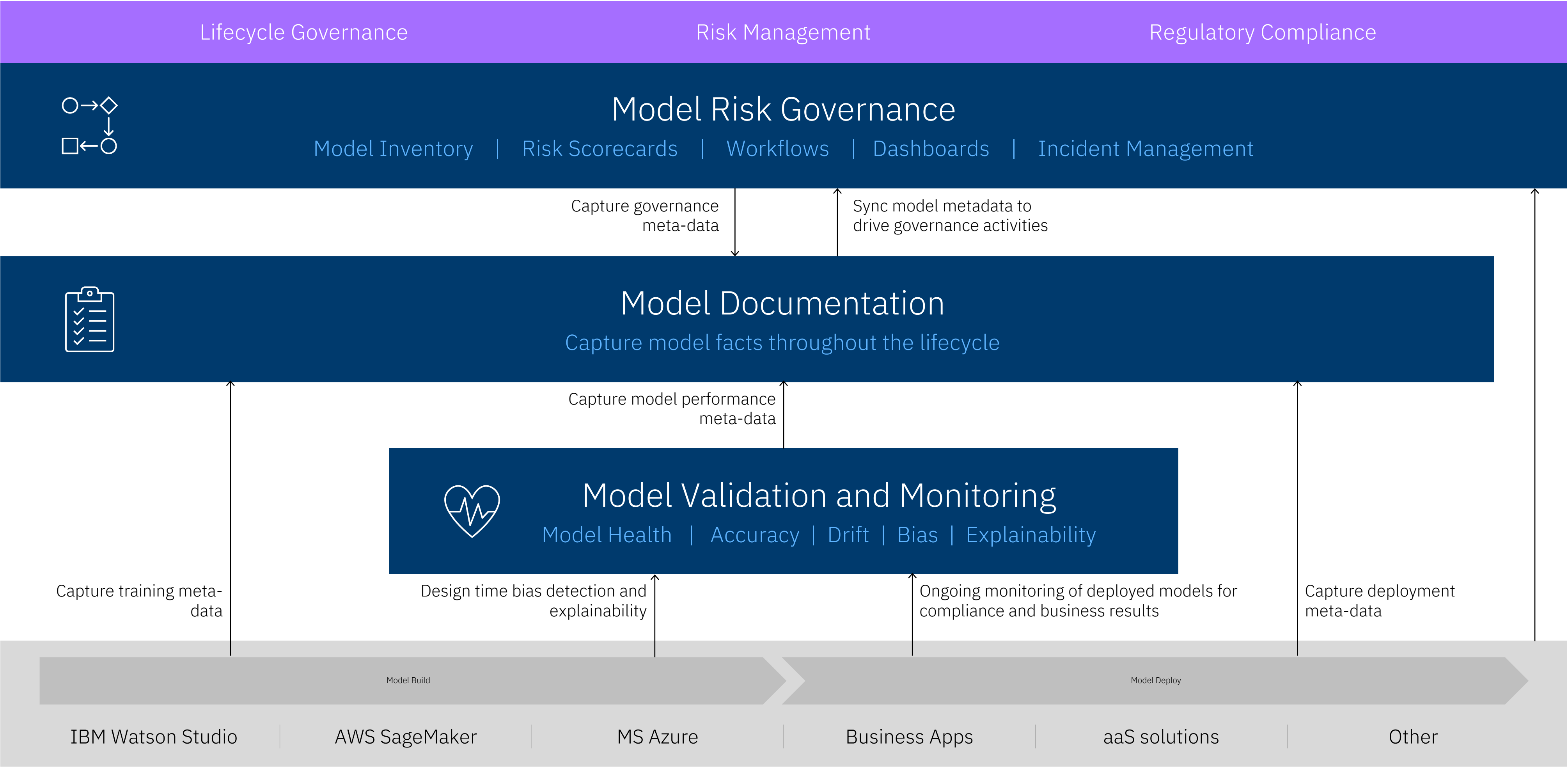
End-to-end toolkit for AI governance across the entire model lifecycle to enable responsible, transparent, and explainable AI workflows.



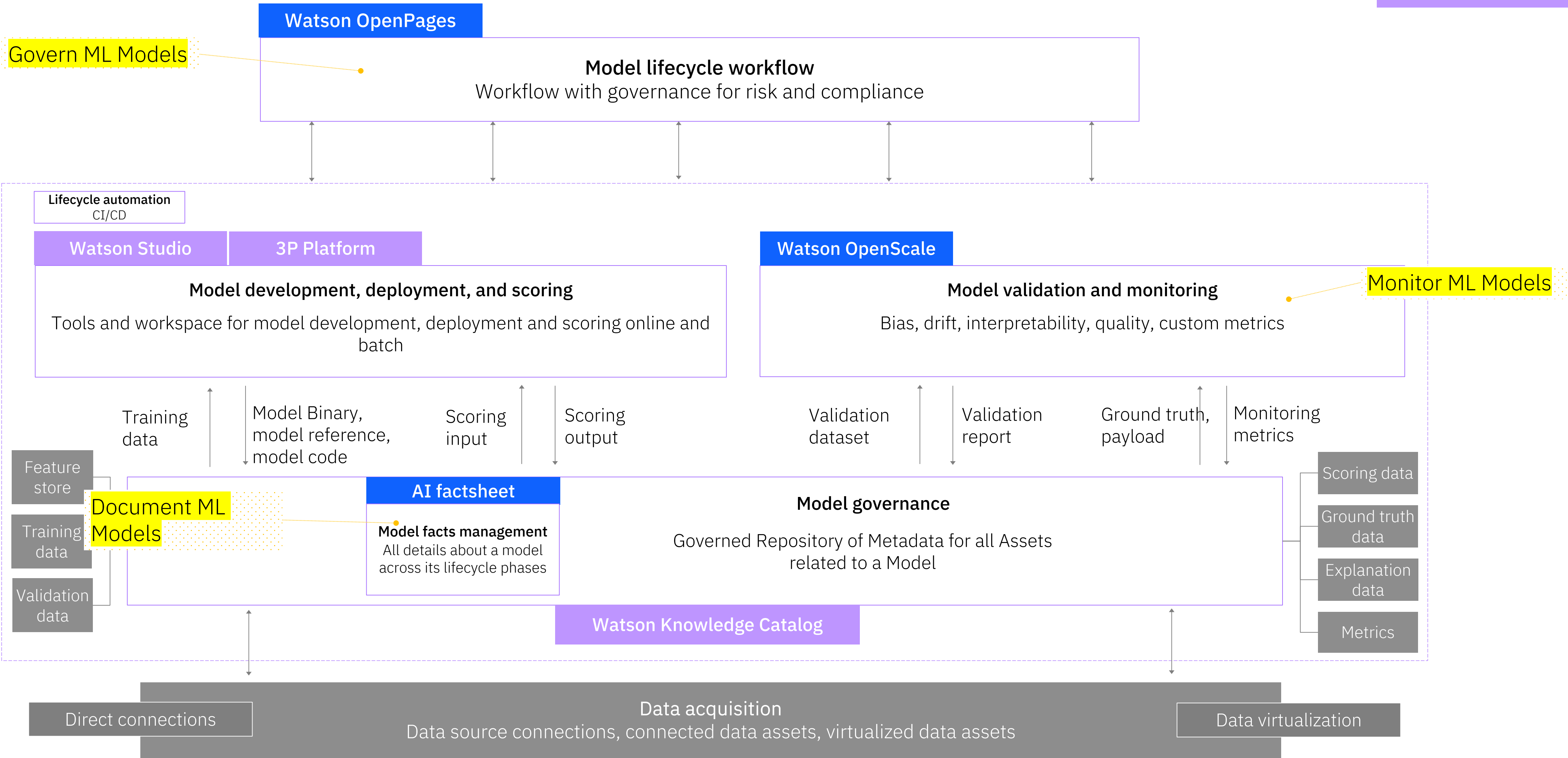
- Model Owners
- Model Validators
- Audit Teams
- Compliance Teams
- Risk Management Teams
- Data Privacy Teams
- Principal Data Scientists



- Data Engineers
- (Citizen) Data Scientists
- MLOps
- ML Engineer



Reference Arkitektur for IBM AI Governance



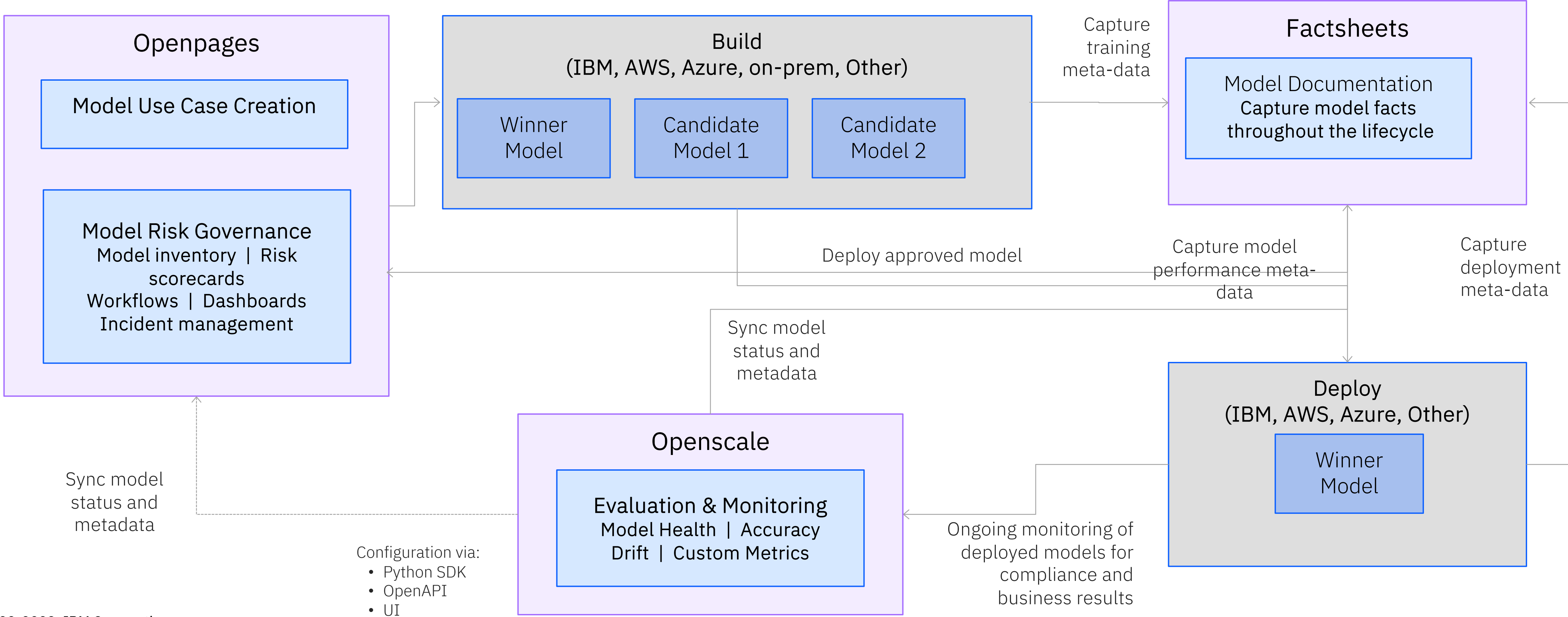
Model development flow Model Use Case

- Model Owners
- Model Validators
- Audit Teams
- Compliance Teams
- Risk Management Teams
- Data Privacy Teams
- Principal Data Scientists

- Data Engineers
- (Citizen) Data Scientists
- AI Engineers
- MLOps

- Soon to come for LLMs:
- Text Classification
 - Entity Extraction
 - Question and Answering
 - Content Generation
 - Text Summarization

- Configuration via:
- Python SDK
 - OpenAPI
 - UI



Eksempel på AI Governance workflow - AI Governance er en team sport

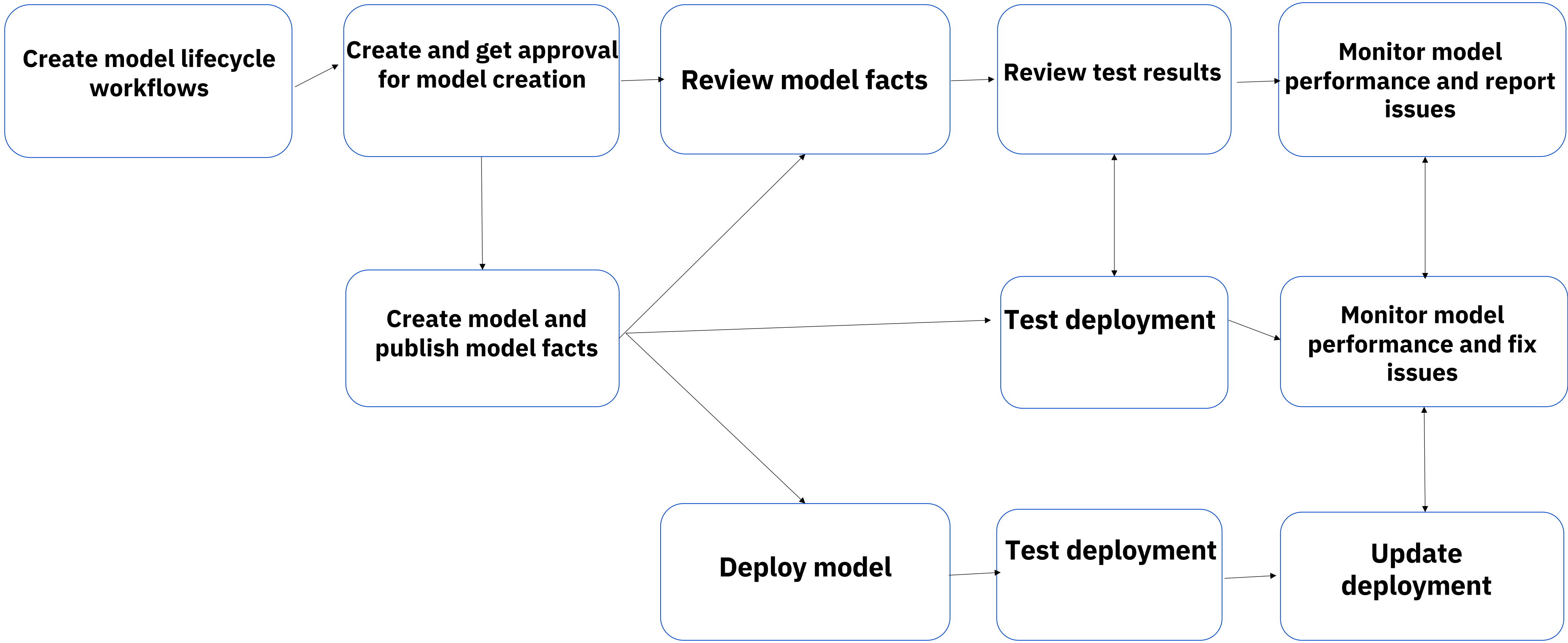
Model Risk Governance team



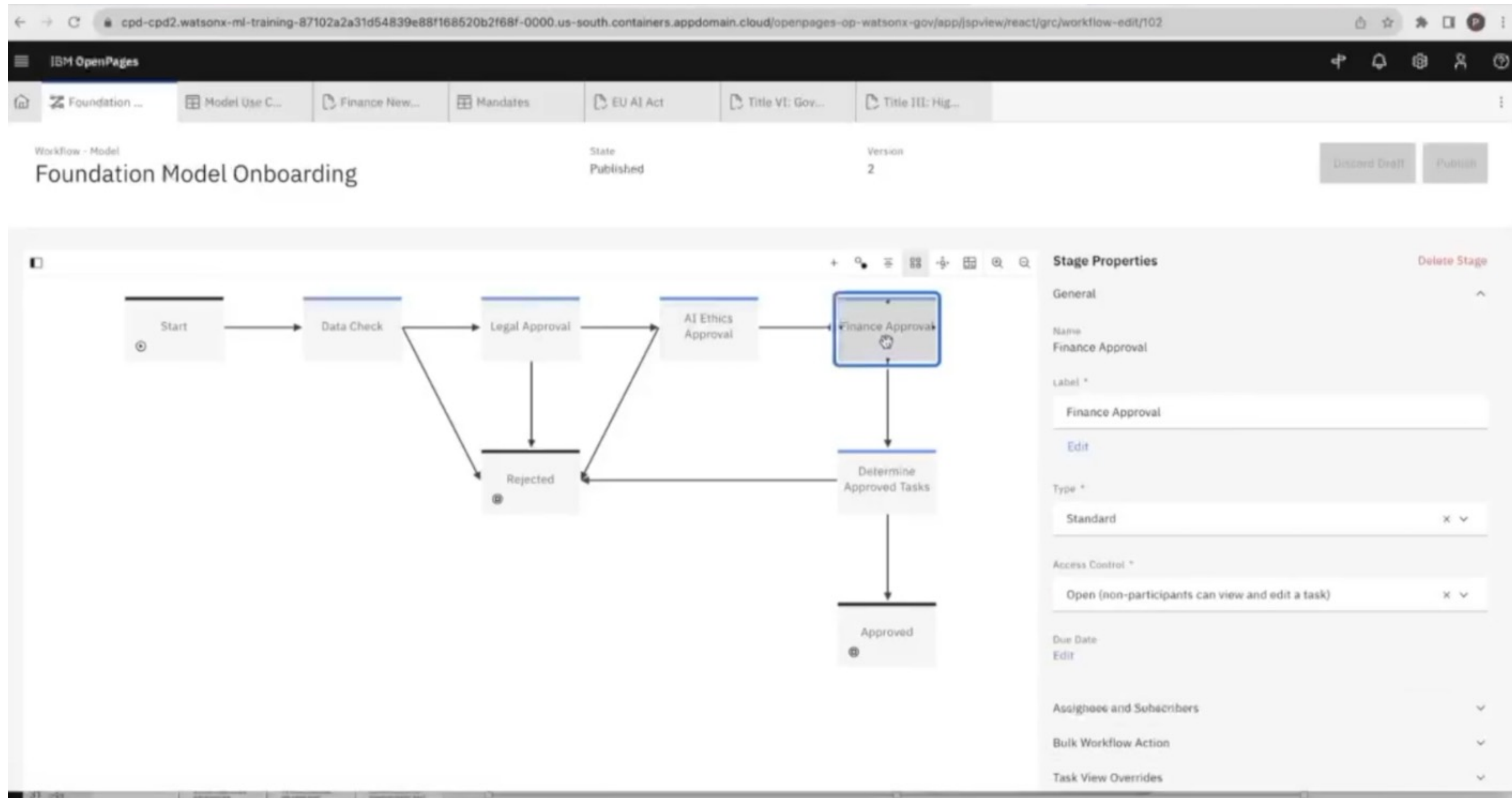
Data science team



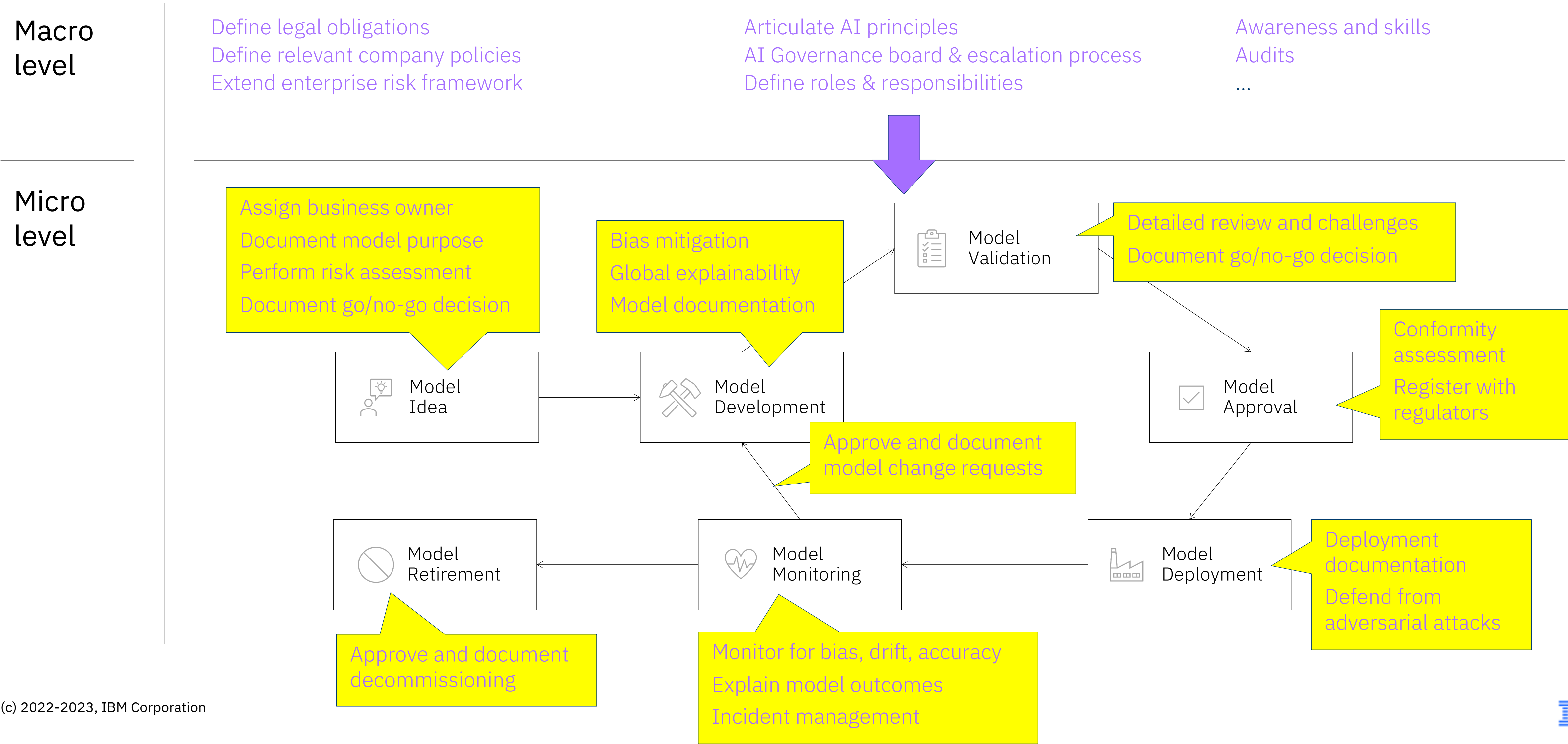
MLOps team



Eksempel på process styring af AI governance



I praksis, AI governance indeholder



Sneek peak: EU AI Act “compliance

IBM OpenPages with Watson

MOD-00002

MOD-00002 ☆ ^

Action ▾

Task

Activity

Admin

Reveal editable fields

* Modified Required *

Name

MOD-00002

Description

Loan Automation Use Case

Model Status

Under Development

Compliance Status

Compliant

Model or Non-Model

Model

Additional Description

Assists with automating the process of issuing a loan to an approved applicant

Model Details

Machine Learning Model

Yes

Model Category

Operations

Model Type

Binary Classification

Monitored with Watson OpenScale

Yes

Measurement Type

Other

Basel Model

No

Model general view

A model is a method, system or approach that processes input data into quantitative estimates, patterns or predictions.

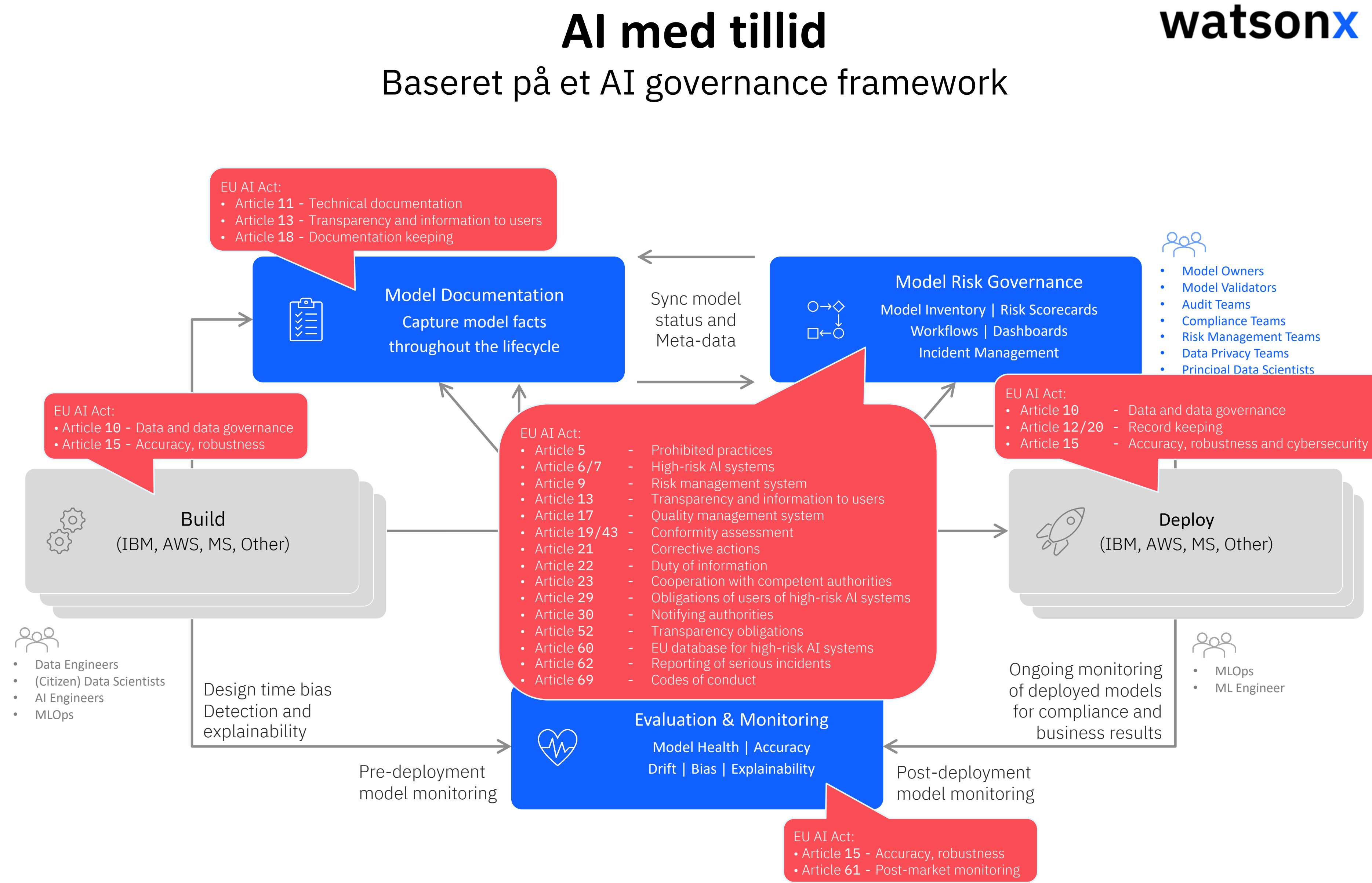
Select an action to validate

All Key Items (1)

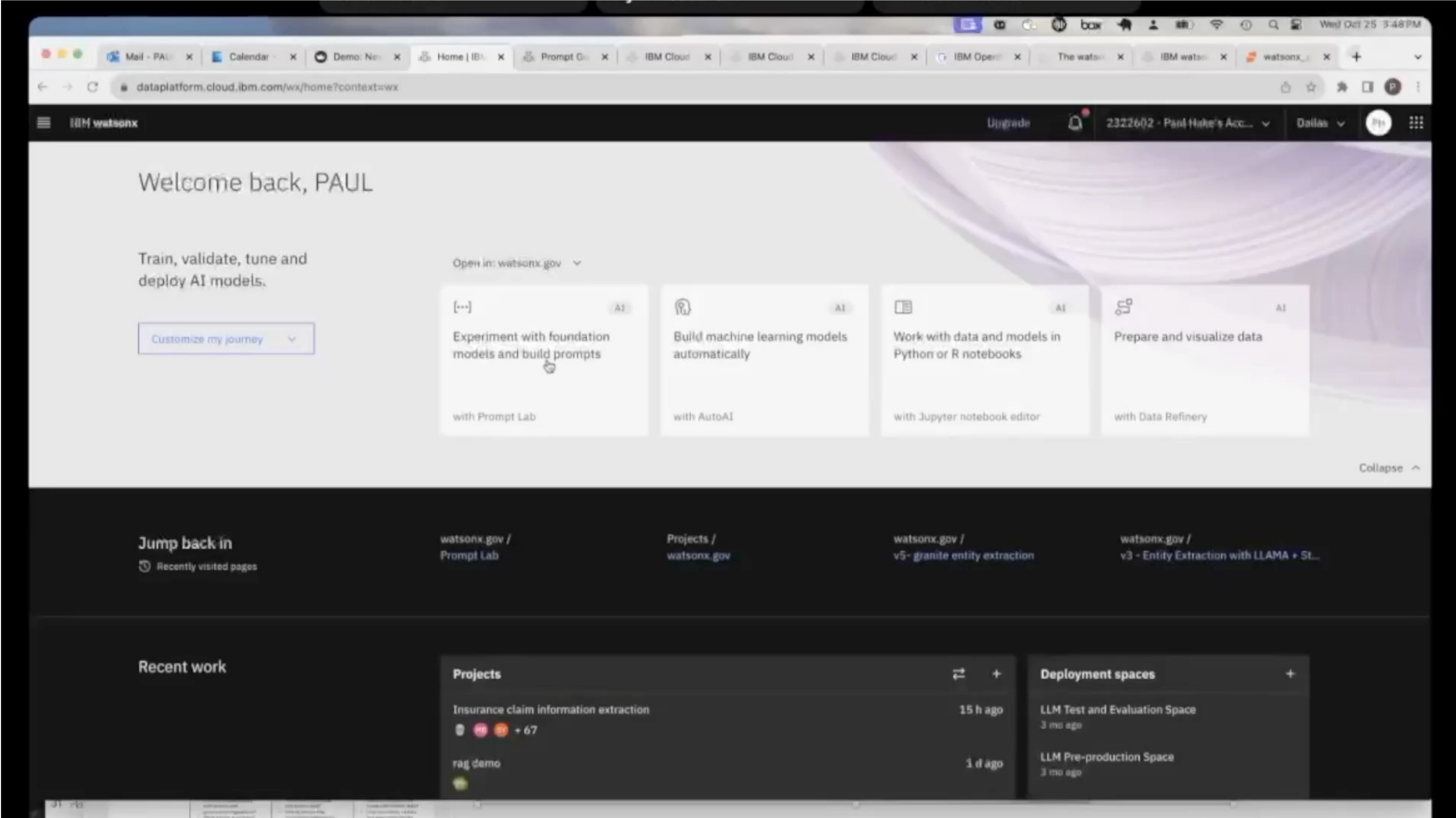
Model Status

Headstart til EU AI Act

- Enterprise fokuseret AI governance - klar til EU AI act
- Klar til at adressere Risk definitioner og implikatione



GenAI Governance demo



watsonx

AI for Business -
Ansvarlig AI



Bo Holtemann
Data & AI Governance Leader
IBM
bholte@dk.ibm.com
+4528808188
<https://www.linkedin.com/in/boholtemann/>

TAK