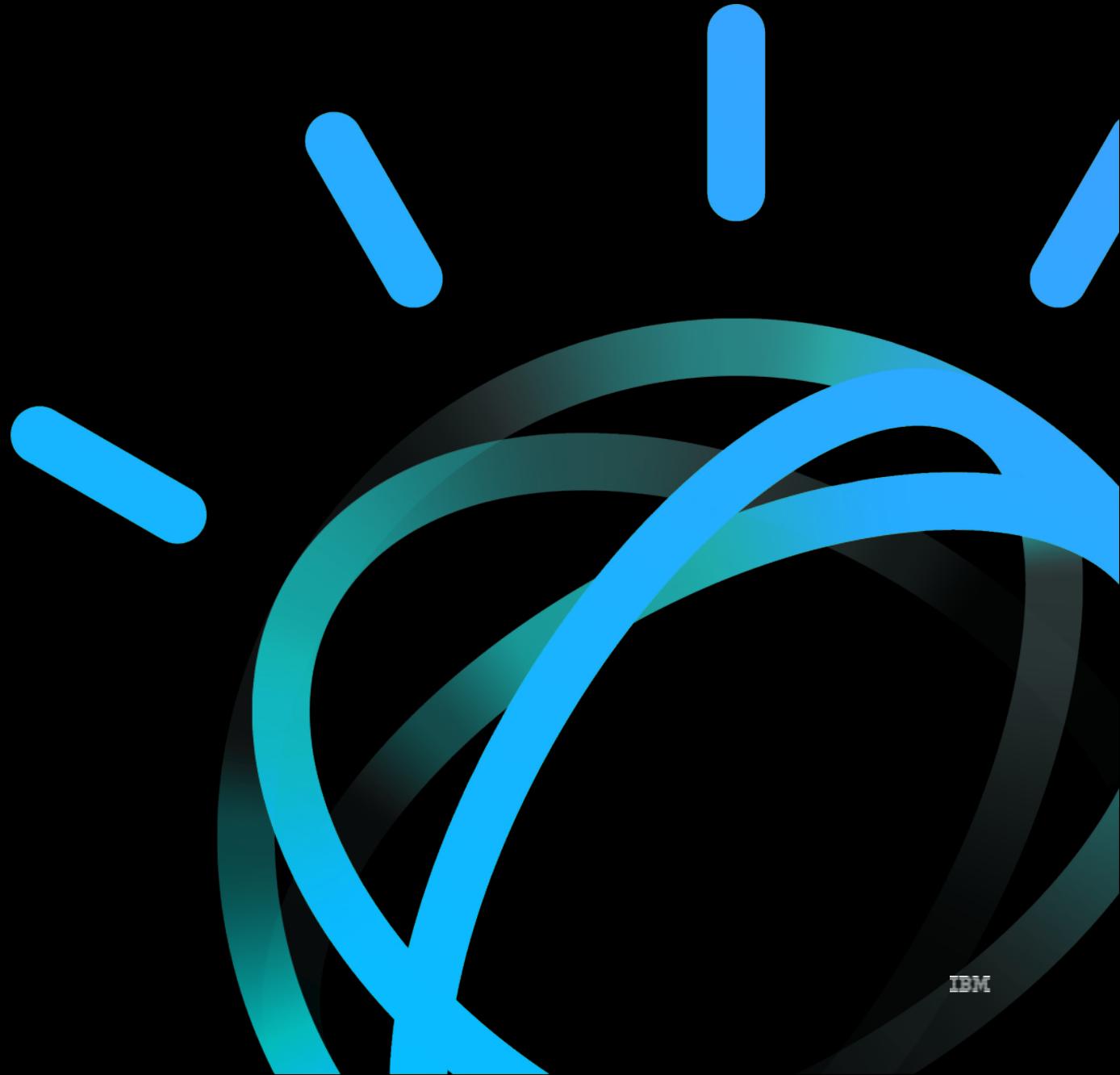


Introducing AutoAI

Upkar Lidder
IBM



Why care about Automation?

Speed AI lifecycle management

“Optimizing pipeline is iterative and not holistic. The pipeline is often optimized for one objective and constraint at a time, which may have severe impact on quality. Typical project consumes **1-6 expert data scientists for 2-12 months.**”

Source: IBM Research estimate, March 2019

Empower Citizen Data Scientists

“Through 2020, the number of citizen data scientists will grow **five** times faster than the number of expert data scientists. Organizations can use citizen data scientists to fill the data science and machine learning **talent gap caused by the shortage and high cost** of data scientists.”

Source: Gartner: [Top 10 strategic technology trends for 2019](#), October 2018

Govern and Monitor AI outcomes

“One magical aspect of software is that it just keeps working... the most common mistake of companies taking their first artificial intelligence (AI) products to market. The **moment** you put a model in production, it starts **degrading.**” Source: Forbes, [Why Machine Learning Models Crash And Burn In Production](#), April. 2019

AutoAI: What we are hearing from our clients

Healthcare insurer



AutoAI is my first step now with any new use case, I can test it out and see immediately if there is signal in the data

Demonstrated with live data that AutoAI outperforms the existing loyalty campaigns designed by their in-house marketing teams.



Airline Leader

International Advertising Giant



The AutoAI function has played an important role in the feature engineering process. We are super excited and want to get deeper dive on the potential of AutoAI

Introducing AutoAI with IBM Watson Studio

Automates AI lifecycle management data prep, feature engineering, and hyper parameter optimization

Makes experiments and neural network search easy

Is explainable, debiased and trusted

Making enterprise AI simple and scale

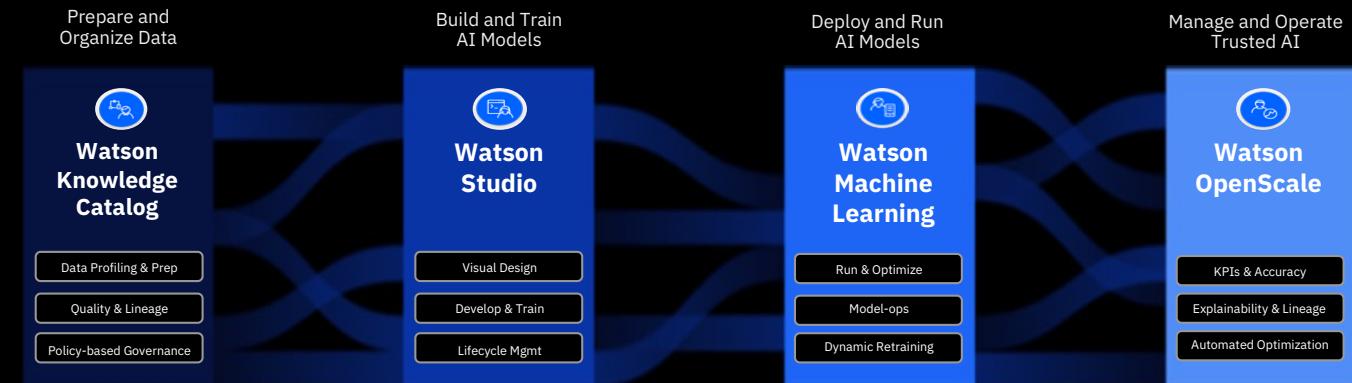
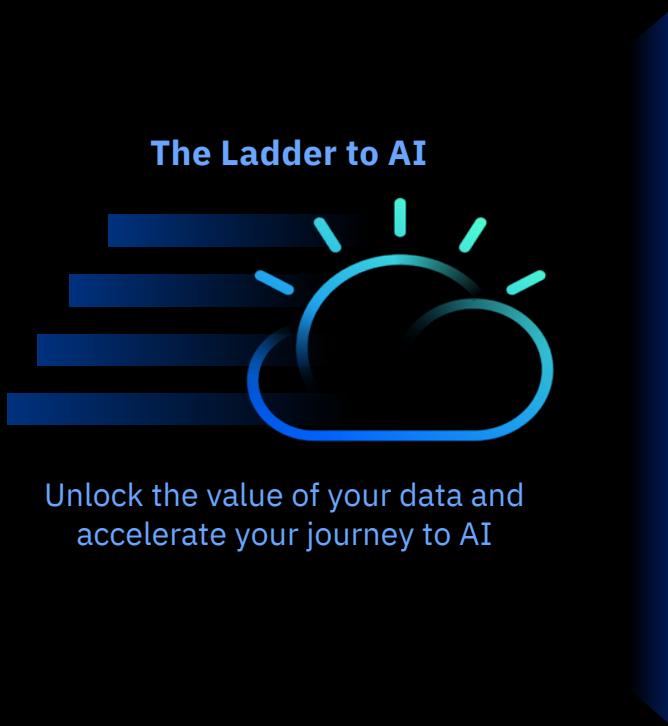
Paving a path toward industrialization of AI



IBM Data and AI portfolio

Delivers the AI Ladder within one unified multicloud platform

Everything you need for enterprise data science and AI



✓ AutoAI Lifecycle Automation – “AI developing AI”



IBM Cloud



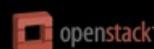
RED HAT
OPENSHIFT



aws



Azure



openstack



Google Cloud



intel



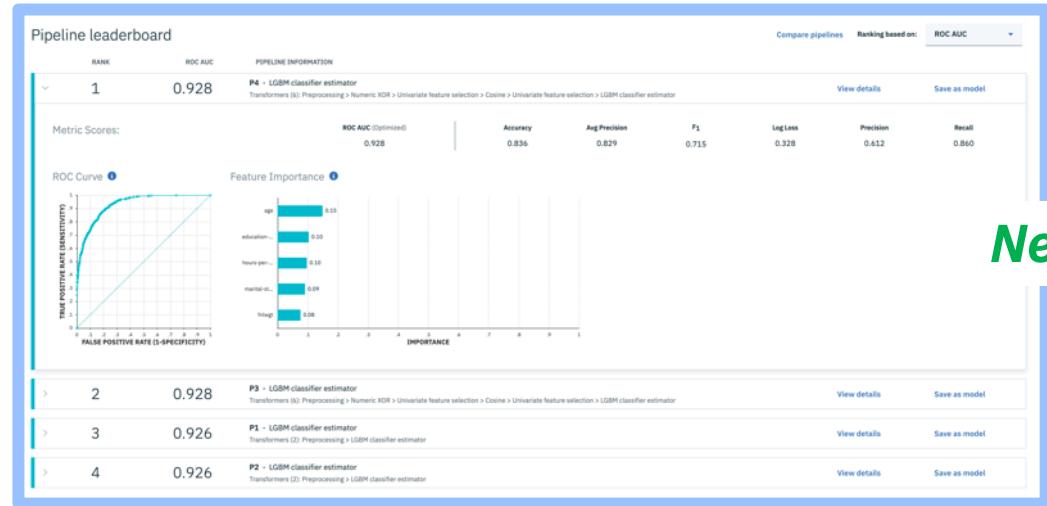
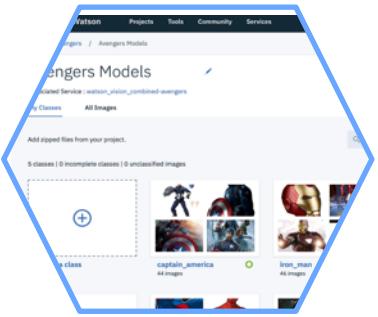
IBM Z

IBM Power Systems

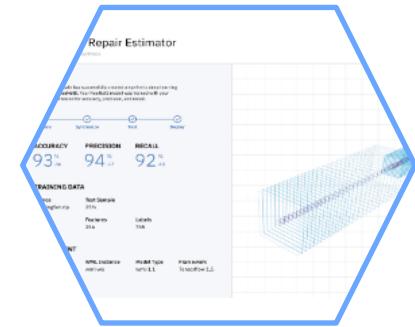


Introducing AutoAI with Watson Studio

IBM's Strategy for Automation of AI Development



New



Transfer Learning

- Transfer knowledge learning in one deep learning system to apply to a different domain
- Featured in Watson Services, available through Watson Studio

AutoAI Experiments | Pipeline optimization

- Auto clean data, engineer features, and complete HPO to find the optimal end to end pipeline
- **New! AutoAI** GA as of May 2019

Neural Network Search

- Just bring data and automatically generate a custom deep neural network through searching the best architectures for the input data
- **NeuNetS** as a feature of Watson Studio, available in Open Beta

What does AutoAI with IBM Watson Studio do?

- Integrated with **Watson Studio** and **Watson Machine learning**
- Automatically ingest, clean, transform, and model with hyperparameter optimization
- Training feedback visualizations provide real-time results to see model performance
- One-click deployment to Watson Machine Learning

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'My Projects / AutoAI Playpen / Demo with Elena'. On the right, it shows 'greg illia's Account' and various icons. Below the navigation, the main area displays a pipeline titled 'Classifying Titanic Survivors' using the 'titanic_train.csv' source table. The pipeline status is 'Completed 15 minutes ago'. The pipeline diagram shows a flow from 'Read Dataset' through 'Split holdout data', 'Read training data', 'Preprocessing', 'Model Selection', and then branching into three parallel paths for different classifiers: 'Gradient boosting classifier', 'Extra trees classifier', and 'LGBM classifier'. Each path involves 'Hyperparameter Optimization', 'Feature Engineering', and another 'Hyperparameter Optimization' step. The pipeline leaderboard below lists 12 entries, each with a rank, ROC AUC score, pipeline information, and 'View details' and 'Save model' buttons. The top-ranked model is P11 (LGBM classifier estimator) with an ROC AUC of 0.860.

RANK	ROC AUC	PIPELINE INFORMATION	View details	Save model
1	0.860	P11 - LGBM classifier estimator Transformers (8): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > LGBM classifier estimator	View details	Save model
2	0.852	P9 - LGBM classifier estimator Transformers (2): Preprocessing > LGBM classifier estimator	View details	Save model
3	0.852	P10 - LGBM classifier estimator Transformers (2): Preprocessing > LGBM classifier estimator	View details	Save model
4	0.851	P7 - Extra trees classifier estimator Transformers (8): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > Extra trees classifier estimator	View details	Save model
5	0.839	P5 - Extra trees classifier estimator Transformers (2): Preprocessing > Extra trees classifier estimator	View details	Save model
6	0.839	P6 - Extra trees classifier estimator Transformers (2): Preprocessing > Extra trees classifier estimator	View details	Save model
7	0.838	P1 - Gradient boosting classifier estimator Transformers (2): Preprocessing > Gradient boosting classifier estimator	View details	Save model
8	0.838	P2 - Gradient boosting classifier estimator Transformers (2): Preprocessing > Gradient boosting classifier estimator	View details	Save model
9	0.831	P4 - Gradient boosting classifier estimator Transformers (8): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > Gradient boosting classifier estimator	View details	Save model
10	0.826	P3 - Gradient boosting classifier estimator Transformers (8): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > Gradient boosting classifier estimator	View details	Save model
11	0.744	P12 - LGBM classifier estimator Transformers (8): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > LGBM classifier estimator	View details	Save model
12	0.500	P8 - Extra trees classifier estimator Transformers (8): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > Extra trees classifier estimator	View details	Save model

User Benefits of using AutoAI



Build models faster

Automate [data preparation](#) and model development



Jump the skills gap

[No coding](#)? No problem – get started with a couple clicks



Discover more use cases

Supercharge [collaboration](#) with [AI everywhere](#) to disrupt and transform



Find signal from noise

[Auto-feature engineering](#) makes it easy to extract more predictive power from your data



Rank and explore models

Quickly compare [candidate pipelines](#) to find the best model for the job



Ready, set, deploy

Pipelines generated with AutoAI can be deployed to REST APIs with [one click](#)

demo



Defining AutoAI - Difference between AutoAI vs traditional AutoML

AutoAI is the process of automating AI lifecycle management. It involves an end-to-end process:

- Taking advantage of pre-trained models, eg. transfer learning – visual recognition and natural language classification
- Selecting the neural networks
- Optimizing pipeline – data prep, feature engineering, HPO and deployment
- Explainability and debiasing

It is more adaptive, multi-modal and multi-sourced compared to the traditional AutoML approach.

	AutoAI	Traditional AutoML
Transfer learning	✓	
Neural network search	✓	
Data preparation	✓	✓
Advanced data refinery	✓	
Feature engineering	✓	✓
Hyper parameter optimization	✓	✓
One click deployment	✓	
Explainability and de-biasing	✓	
AI lifecycle management	✓	

Get Started Today

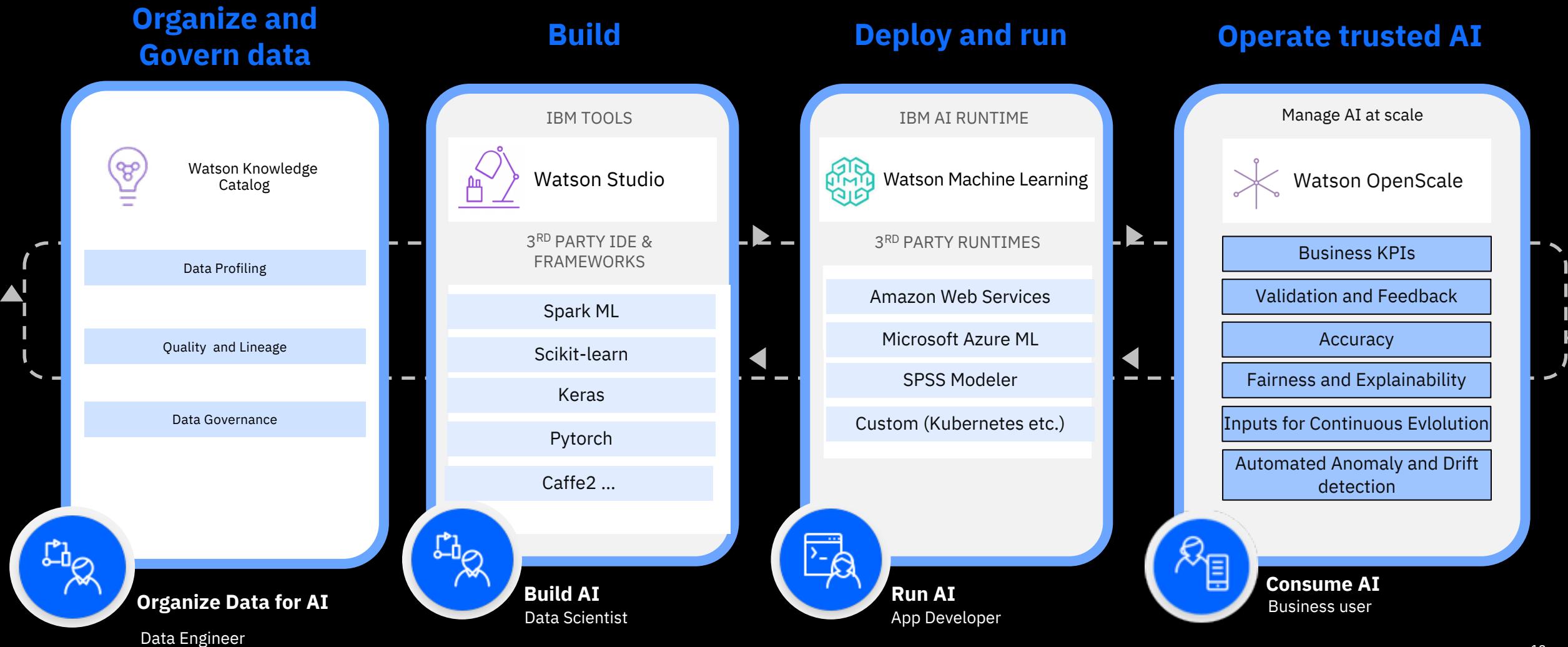


AutoAI:

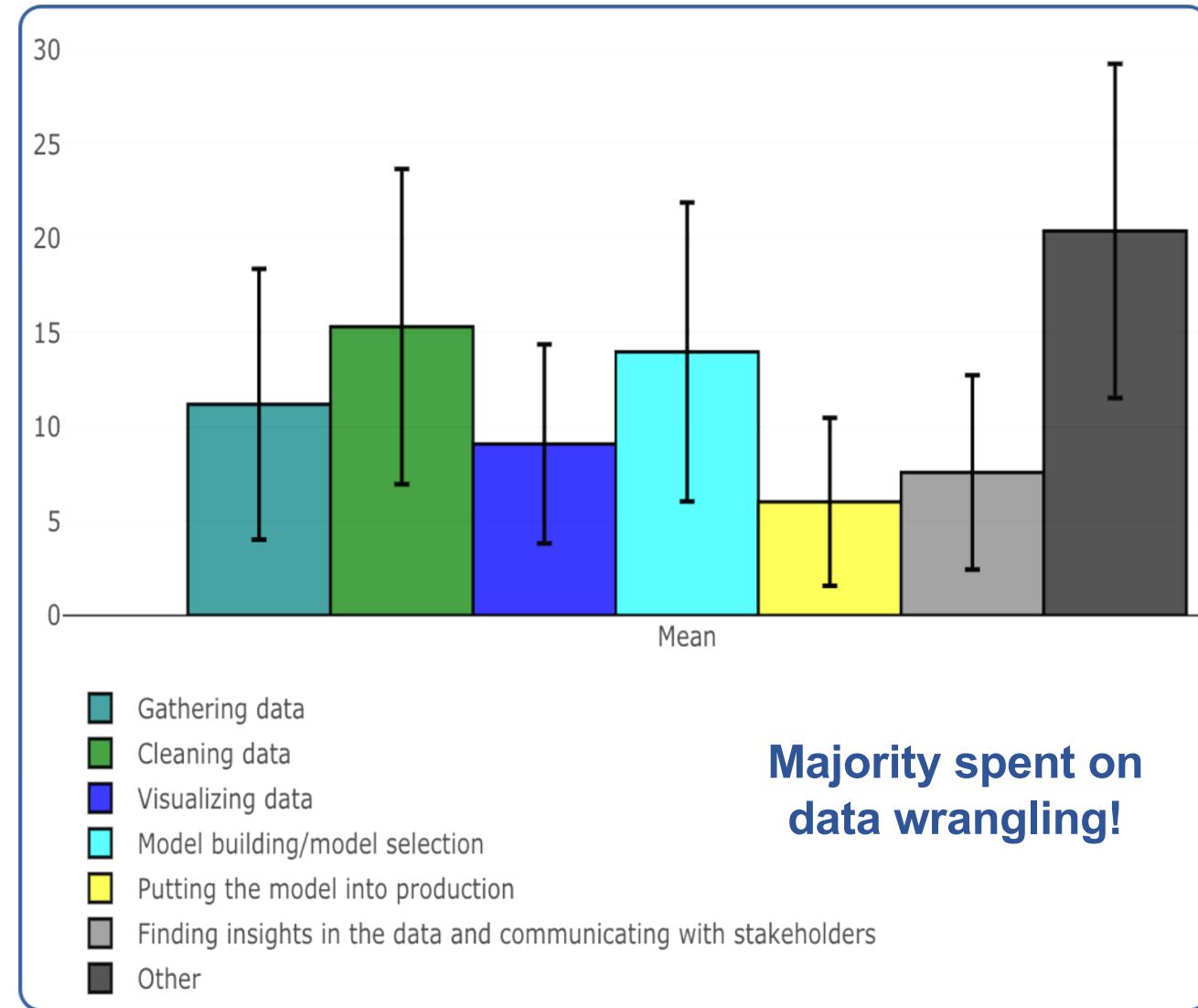
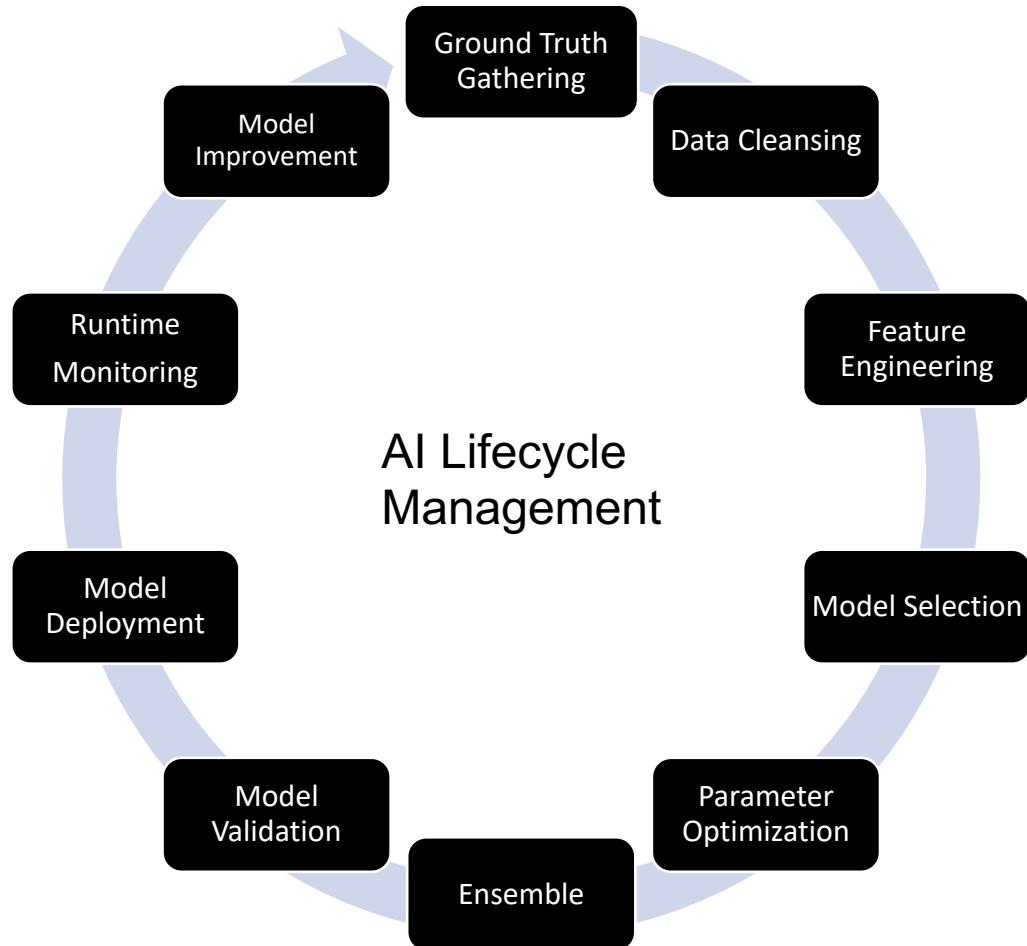
- Accelerate
- Automate
- Optimize

Experience AutoAI: <https://www.ibm.com/demos/collection/IBM-Watson-Studio-AutoAI/>

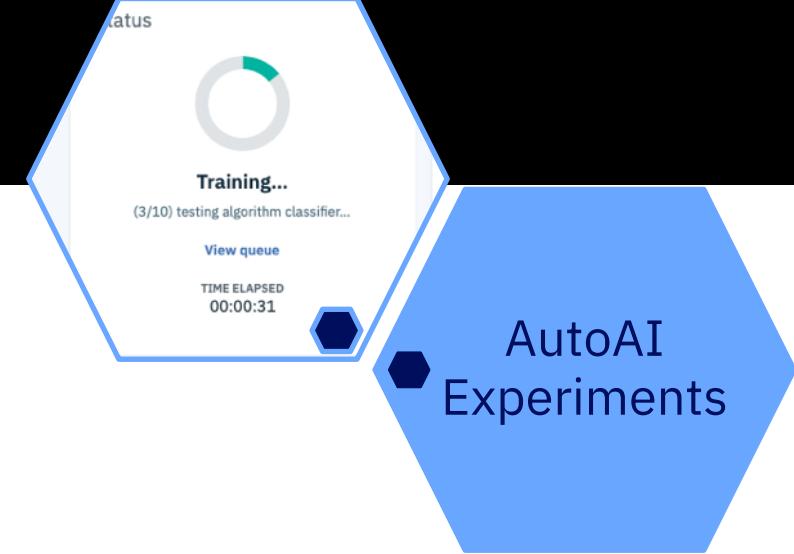
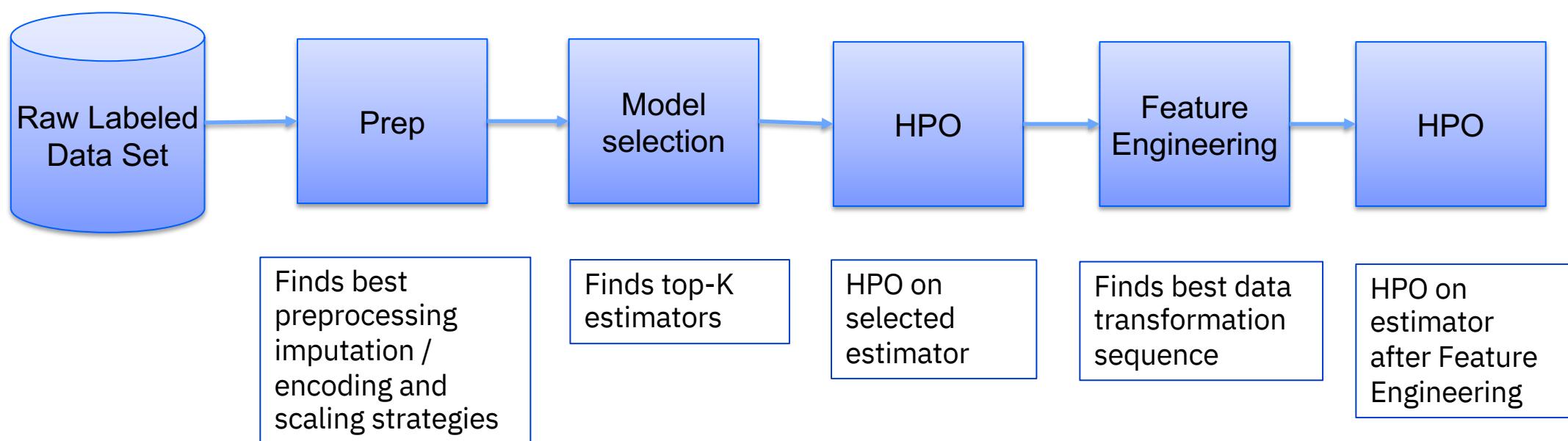
Watson OpenScale along with Watson Studio and WML enables enterprises to operationalize AI across the enterprise



Case for AI Automation: AI Workflow's Bigger & More Complex



How does it work?



IBM Watson Studio



Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business

Analyze any data, no matter where it lives

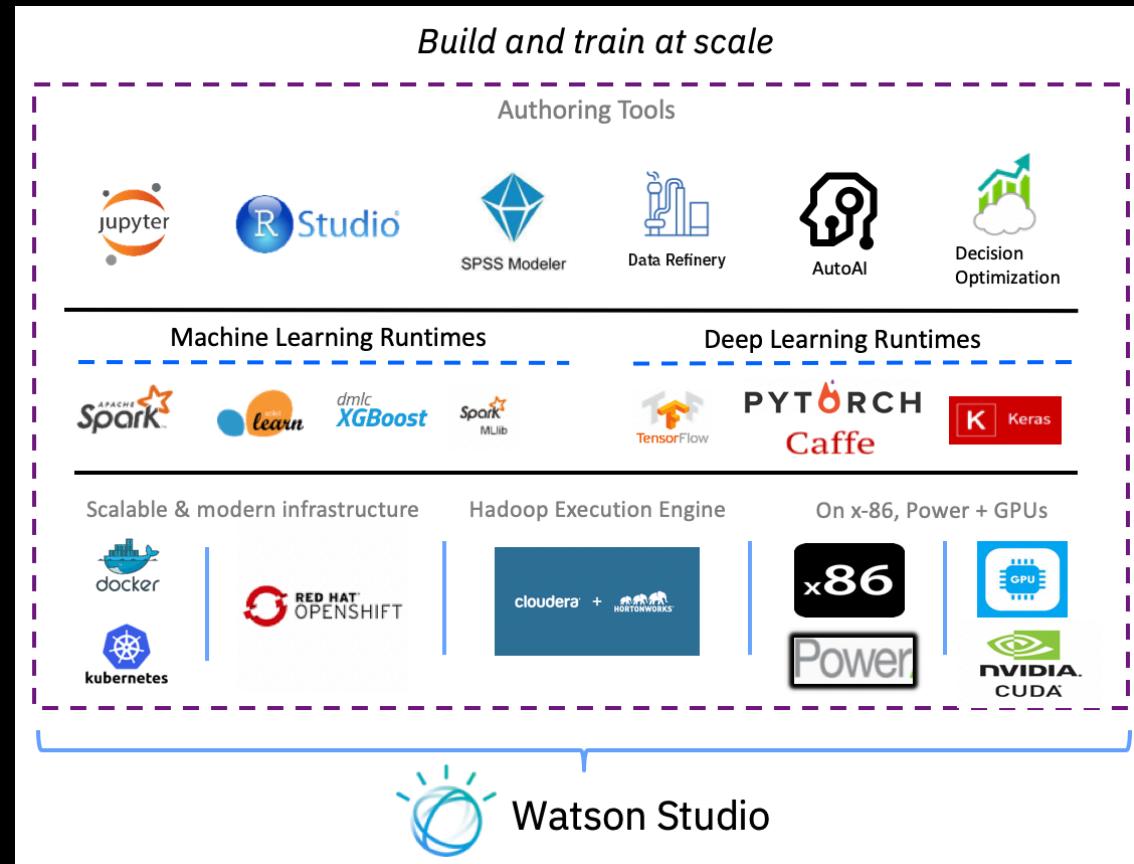
Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale



IBM Watson Studio

Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business

► **Analyze any data, no matter where it lives**

Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale

The screenshot shows the 'New connection' screen in the IBM Watson Studio interface. It is divided into two main sections: 'IBM services' and 'Third-party services'.
IBM services:

- BigInsights HDFS
- Cognos Analytics
- Db2 Big SQL
- Db2 on Cloud
- PureData for Analytics

Third-party services:

- Amazon Redshift
- Dropbox
- Hortonworks HDFS
- Microsoft SQL Server

- Cloud Object Storage
- Compose for MySQL
- Db2 for i
- Db2 Warehouse

- Amazon S3
- FTP
- Looker
- MySQL

- IBM Services like **Cognos & DB2**
- 3rd Party Services like **Amazon S3, Hadoop, & Microsoft SQL Server**
- We have **Public Cloud, Private Cloud, & Desktop/Server** deployment options

IBM Watson Studio

Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business

Analyze any data, no matter where it lives

Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale

The screenshot shows a Jupyter Notebook titled "Train and deploy a heart disease prediction model using XGBoost and IBM Watson Machine Learning APIs". The notebook includes a diagram of a heart on a circuit board, a brief description of the task, and learning goals related to loading CSV files and preparing data.

Super charged Jupyter Notebooks & R Studio as most popular IDEs for data scientists well integrated with data connectors and rich set of default environments

The screenshot shows the SPSS Modeler Lab interface with a complex data flow diagram. On the left, there's a sidebar with various modeling nodes like Association Rules, Auto Classifier, and Bayesian Network. The main area displays a network of nodes connected by lines, representing a data mining process.

Visual tools such as SPSS Modeler, Data Refinery, & AutoAI for non coders to analyze data and build models

IBM Watson Studio

Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business

Analyze any data, no matter where it lives

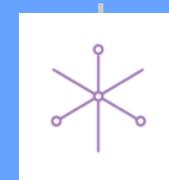
Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale



Seamless integration with:

- Watson Knowledge Catalog brings governance & catalog management
- Watson Machine Learning deploys, manages, & automatically re-trains your models
- Watson OpenScale offers unmatched levels of trust in automation

Bring Watson Studio to wherever your data lives.

Cloud

Data science as a team sport on IBM's secure cloud.

Desktop

Most popular data science tools as an easy desktop installation.

Local/Private

Secure site-wide data science behind your firewall.

Watson Machine Learning

IBM Watson Machine Learning

Embed Machine Learning and Deep Learning
in your Business

Deploy and Manage Models

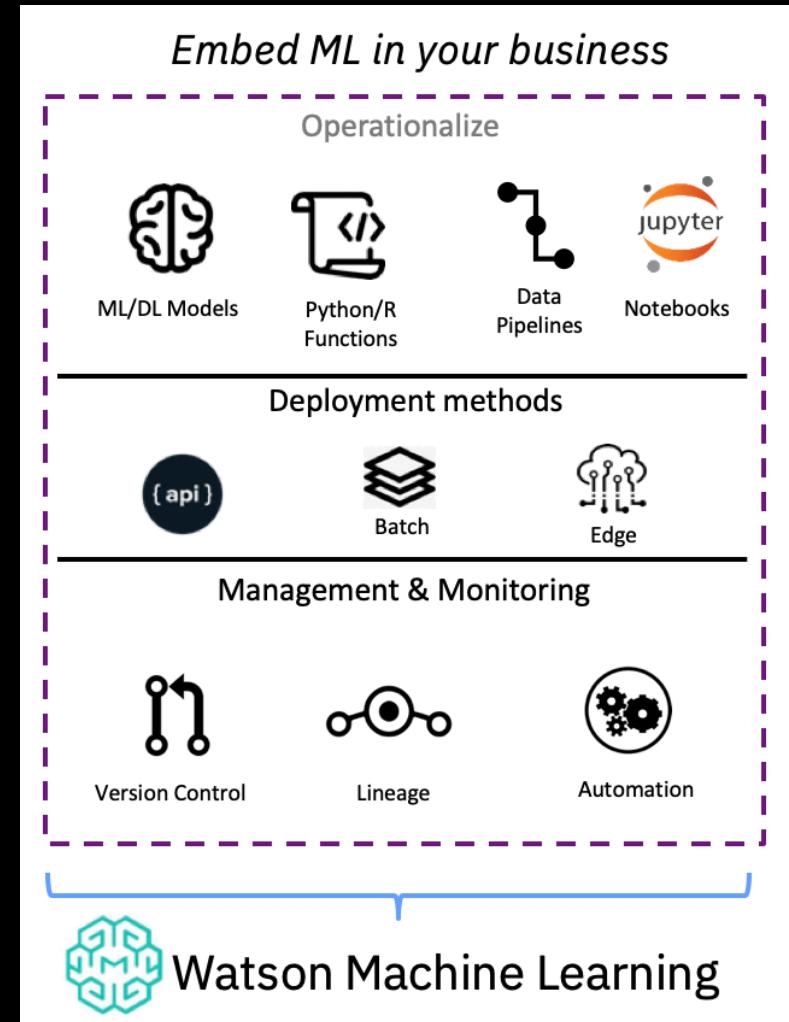
Move models to production, in an easy, secure, and compliant way

Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

Accelerate Compute Intensive Workloads

Distribute your deep learning training and Hadoop/Spark workloads with multi-tenant job scheduling



IBM Watson Machine Learning

Embed Machine Learning & Deep Learning in your Business

Deploy and Manage Models

Move models to production, in an easy, secure, and compliant way

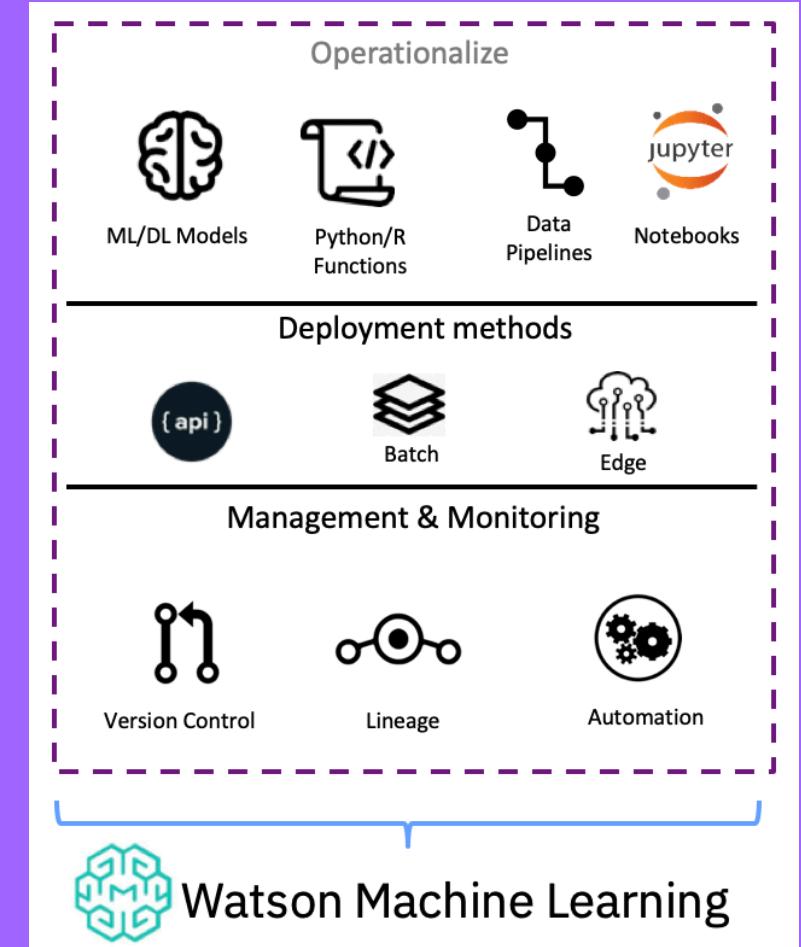
Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

Accelerate Compute Intensive Workloads

Distribute your deep learning Training and Hadoop/Spark workloads with multi-tenant job scheduling

Flexible deployment capabilities



IBM Watson Machine Learning

Embed Machine Learning & Deep Learning in your Business

Deploy and Manage Models

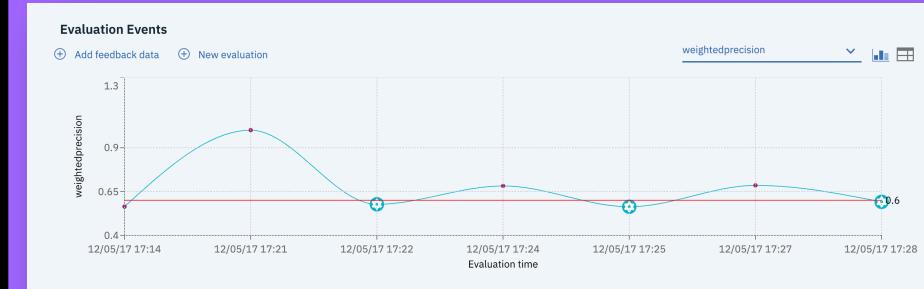
Move models to production, in an easy, secure, and compliant way

Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

Accelerate Compute Intensive Workloads

Distribute your deep learning Training and Hadoop/Spark workloads with multi-tenant job scheduling



- Configure learning systems
- Flexible management experience
 - User Interface
 - CLI
 - APIs
 - Python SDK

IBM Watson Machine Learning

Embed Machine Learning & Deep Learning in your Business

Deploy and Manage Models

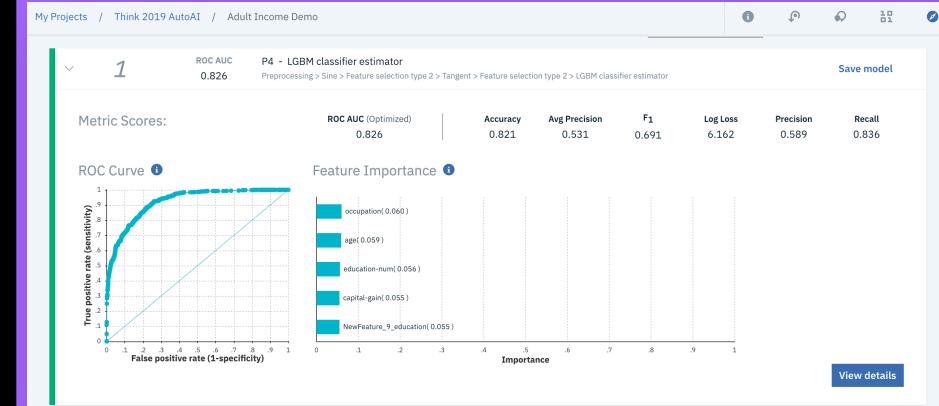
Move models to production, in an easy, secure, and compliant way

Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

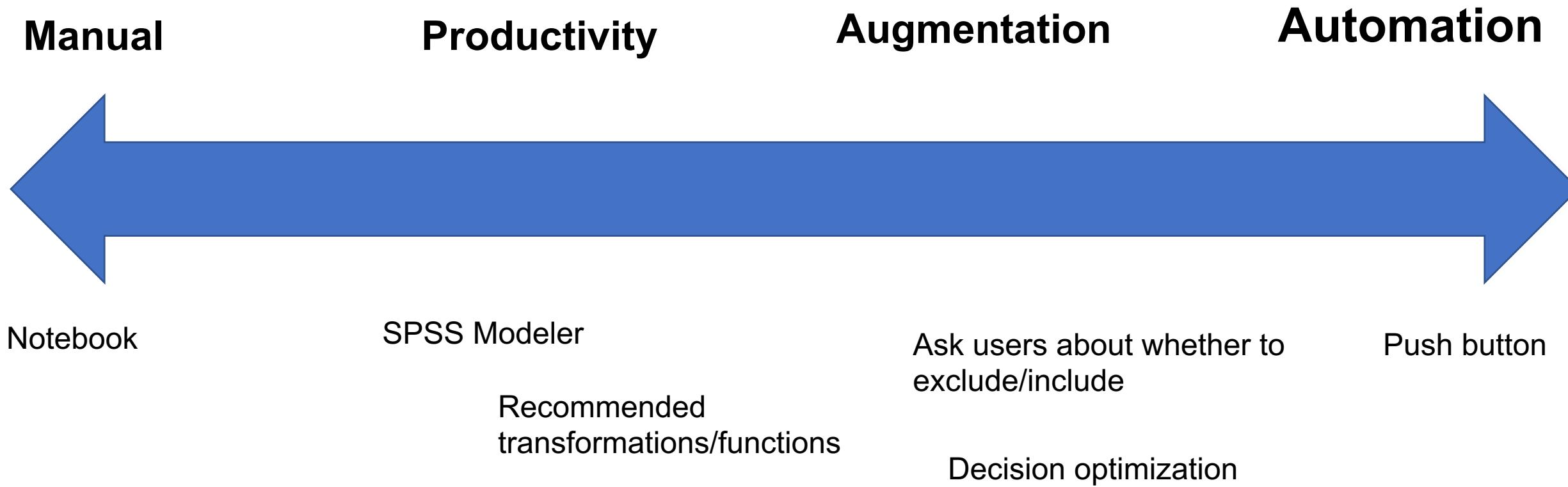
Accelerate Compute Intensive Workloads

Distribute your deep learning Training and Hadoop/Spark workloads with multi-tenant job scheduling



- **Visually monitor training jobs**
 - Experiments
 - AutoAI
- **WML Accelerator for optimized Spark and GPU jobs**
- **Integration with Hadoop for in-place batch training and scoring**

Addressing Data Science Automation-Productivity Spectrum



IBM