Advancing the Future of AI

IBM Research AI

February 2021

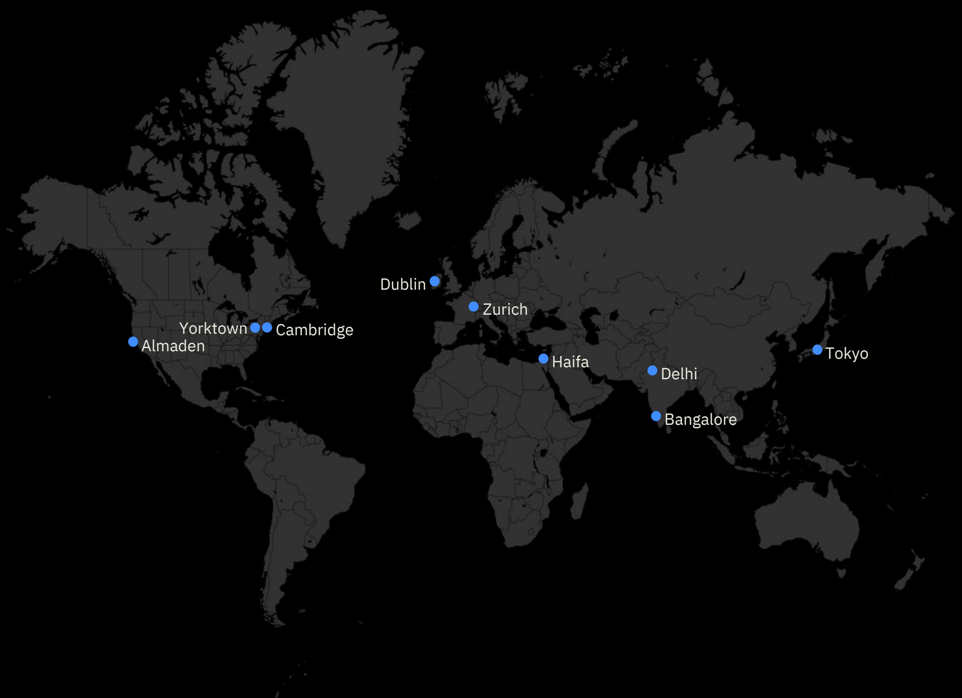
Introduction	Trusted AI	Scaling and Automating AI
		
Dr. John R. Smith, IBM Fellow IBM T. J. Watson Research Center	Dr. Karthikeyan N. Ramamurthy IBM T. J. Watson Research Center	Dr. Lisa Amini, Director IBM Research Cambridge

Introduction

Dr. John R. Smith, IBM Fellow
IBM T. J. Watson Research Center

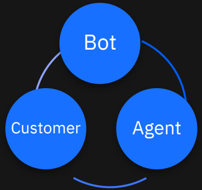
IBM Research AI

- Global footprint with an integrated AI strategy
- Network of academic partnerships in AI



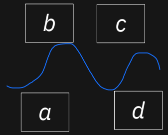
AI Focus Areas include:

Language

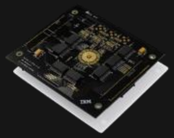


Speech

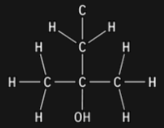
Code



Time series



Hardware



Chemistry

... more

MIT-IBM AI LAB

MIT-IBM AI Lab
By the numbers

50+ Running projects

23 Departments and centers at MIT

$300m 10-year investment to found a joint Lab

150+ Researchers across MIT and IBM

100+ Publications in top academic conferences and journals

Narrow AI

Emerging

Deep learning

Single-task,
single-domain,
high accuracy

Requires large
amounts of data

CPU and GPU

Broad AI

Disruptive and pervasive

Neuro-symbolic AI

Trusted AI capable of learning
with much less data

Automating AI development
and deployment

Reduced-precision and
analog HW

We are here now

General AI

Revolutionary

True neuro-AI

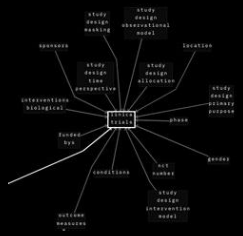
Cross-domain learning
and reasoning

Broad autonomy with
moral reasoning

Wetware?

AI is making incredible impact

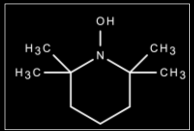
Example Applications (Accelerated discovery)



Deep Search (PDF Corpora)

Knowledge Ingestion and Reasoning

www.research.ibm.com/covid19/deep-search/



RXN for Chemistry

Chemical Reaction Prediction

https://rxn.res.ibm.com

RoboRXN (AI-Driven Synthesis)

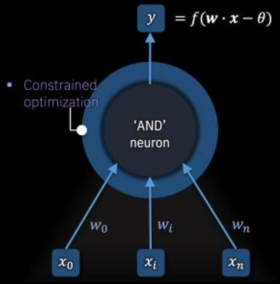
Cloud-based Autonomous Labs

https://rxn.res.ibm.com/rxn/robo-rxn

What's Next (Advances in AI foundations)

Neuro-symbolic Reasoning

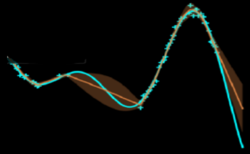
www.research.ibm.com/artificial-intelligence/vision/#neurosymbolic



Logical Neurons

Trusted AI

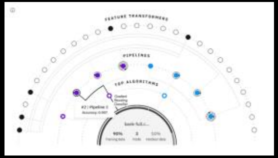
www.research.ibm.com/artificial-intelligence/trusted-ai/



Uncertainty Quantification

Auto AI

https://www.ibm.com/cloud/watson-studio/autoai

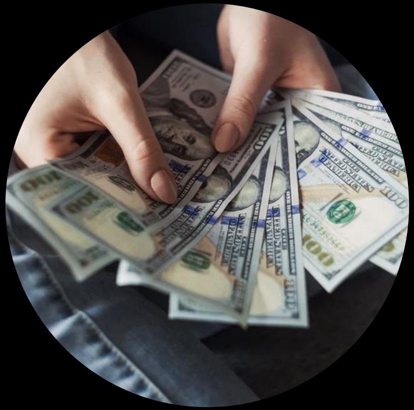


AutoAI

Trusted AI

Dr. Karthikeyan N. Ramamurthy
IBM T. J. Watson Research Center

AI is powering critical workflows and trust is essential



loan
processing

employment

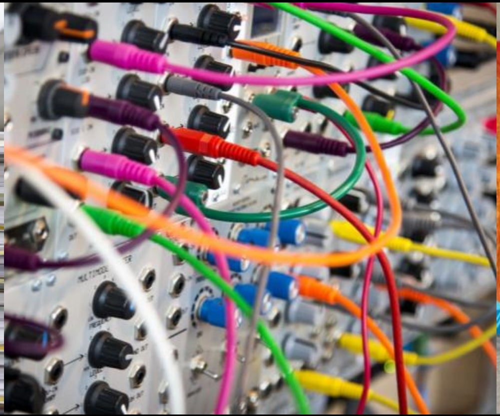
customer
management

quality control

Multiple factors are placing trust in AI as a top client priority

brand reputation

increased regulation



complexity of AI deployments

focus on social justice

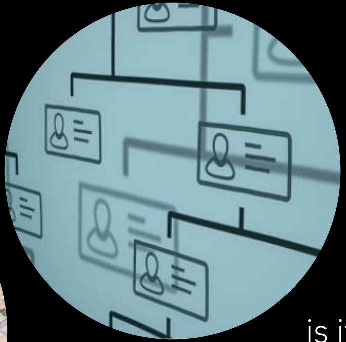
What does it take to trust a decision made by a machine?

is it accurate?

is it fair?

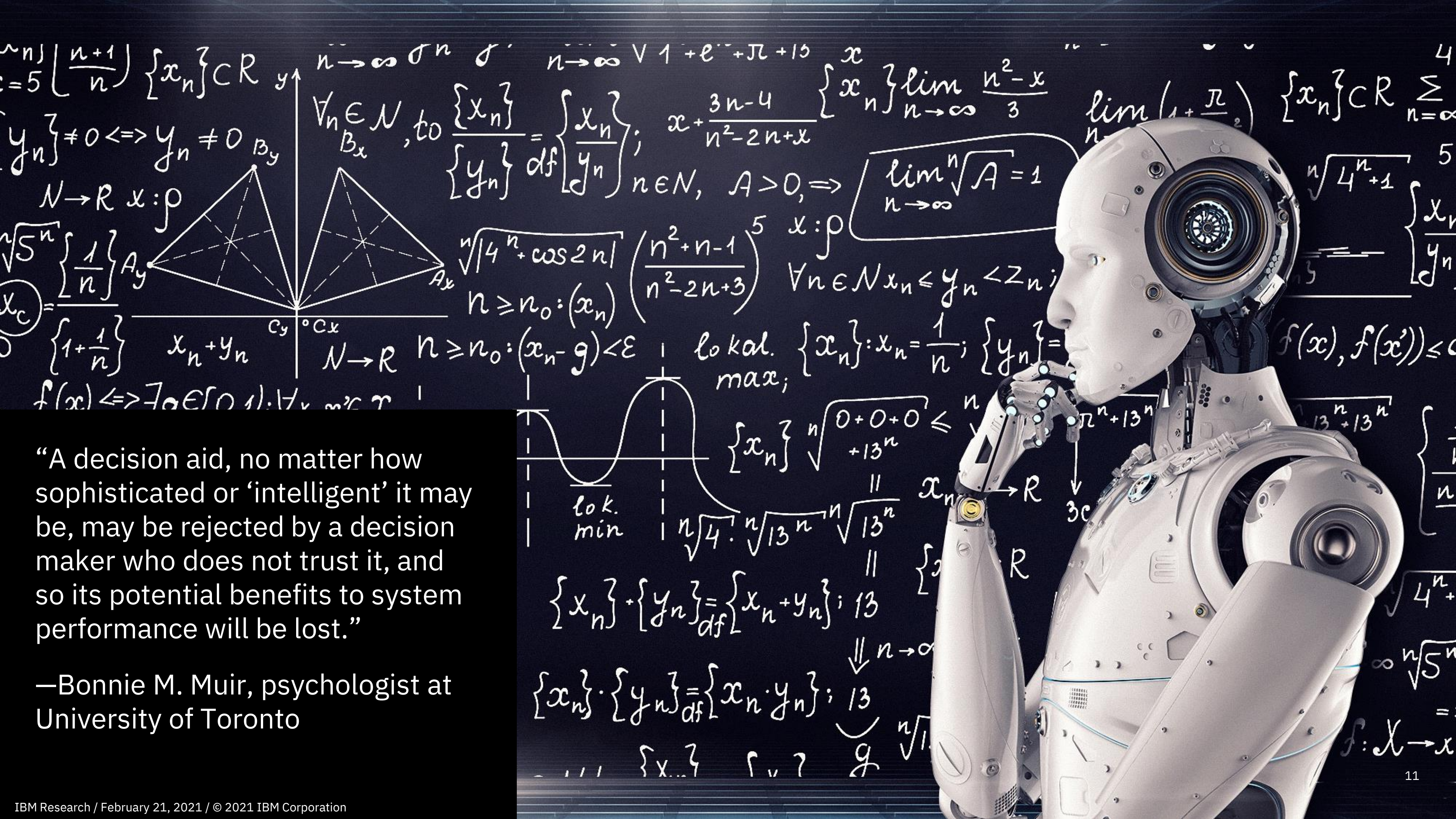
did anyone tamper with it?

is it easy to understand?



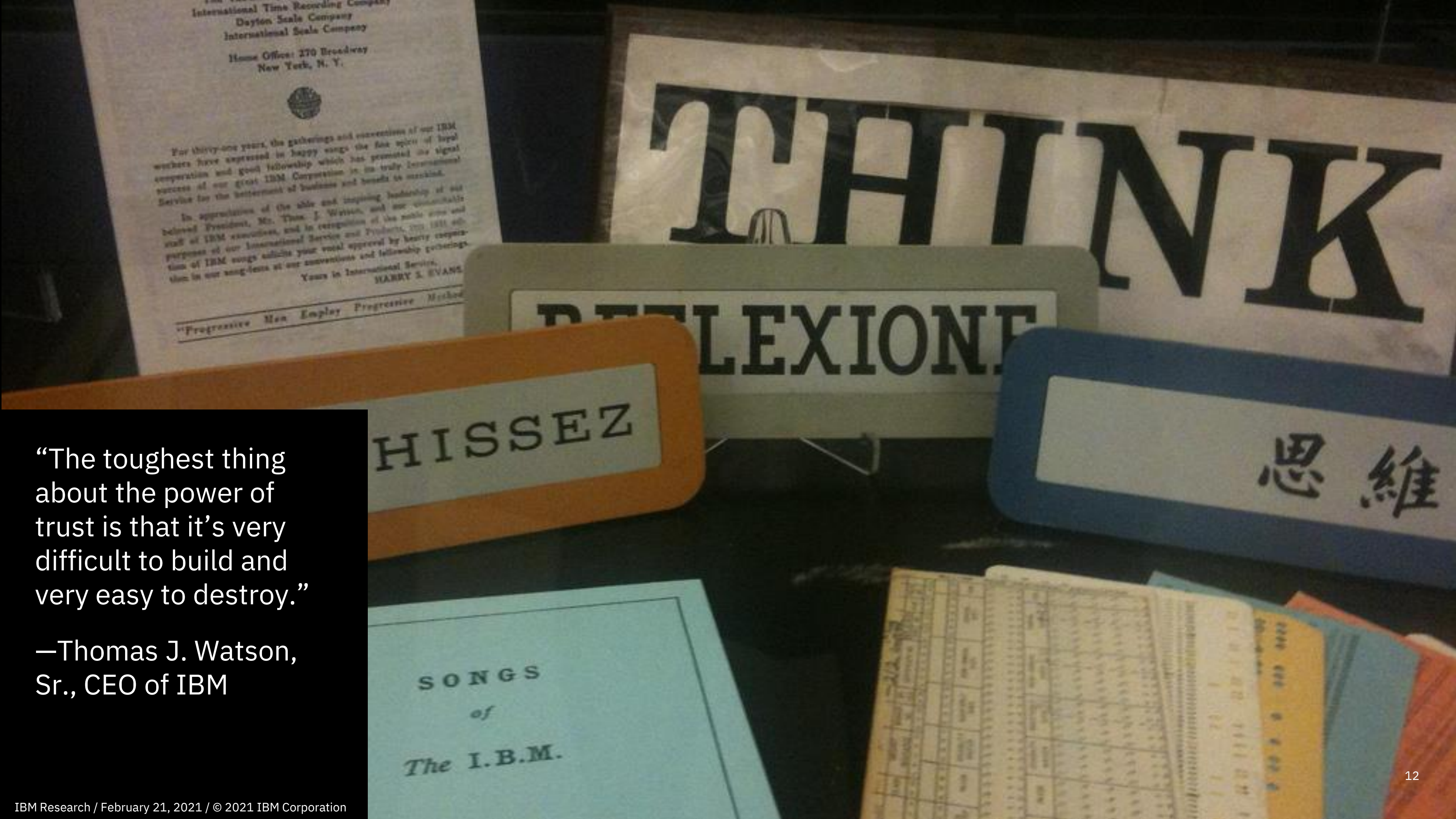
is it accountable?

is it transparent?



"A decision aid, no matter how sophisticated or 'intelligent' it may be, may be rejected by a decision maker who does not trust it, and so its potential benefits to system performance will be lost."

—Bonnie M. Muir, psychologist at University of Toronto



“The toughest thing about the power of trust is that it’s very difficult to build and very easy to destroy.”

—Thomas J. Watson, Sr., CEO of IBM

Trust

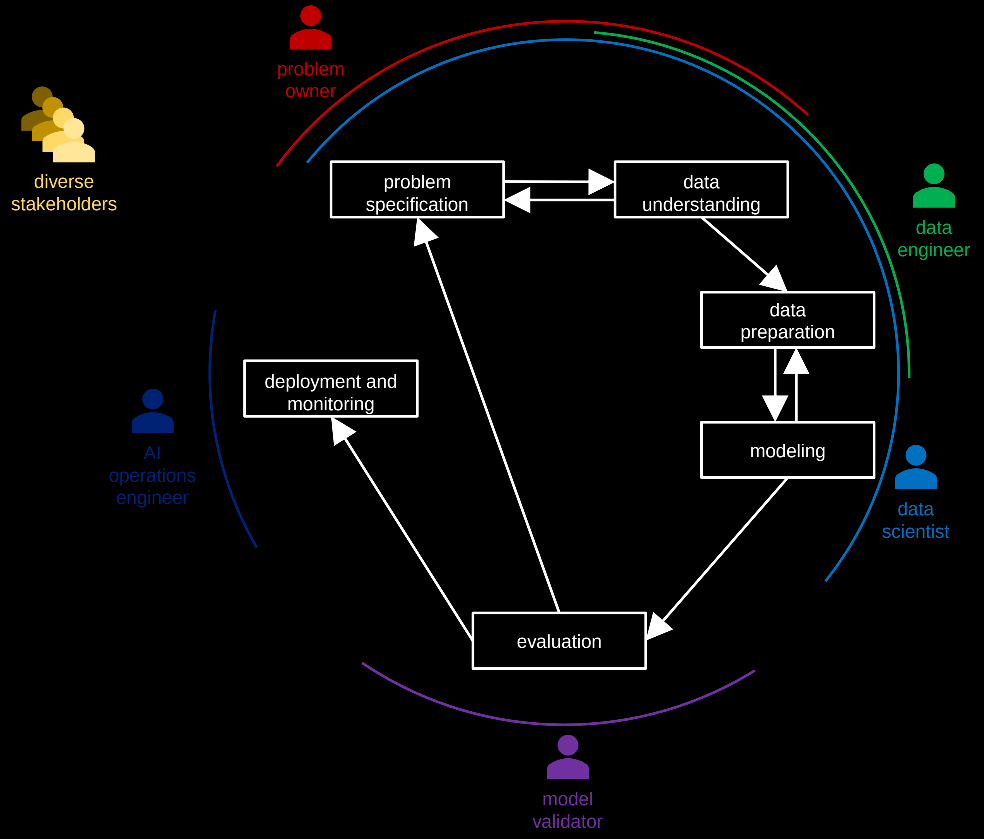
⬇ ⬆

Verify

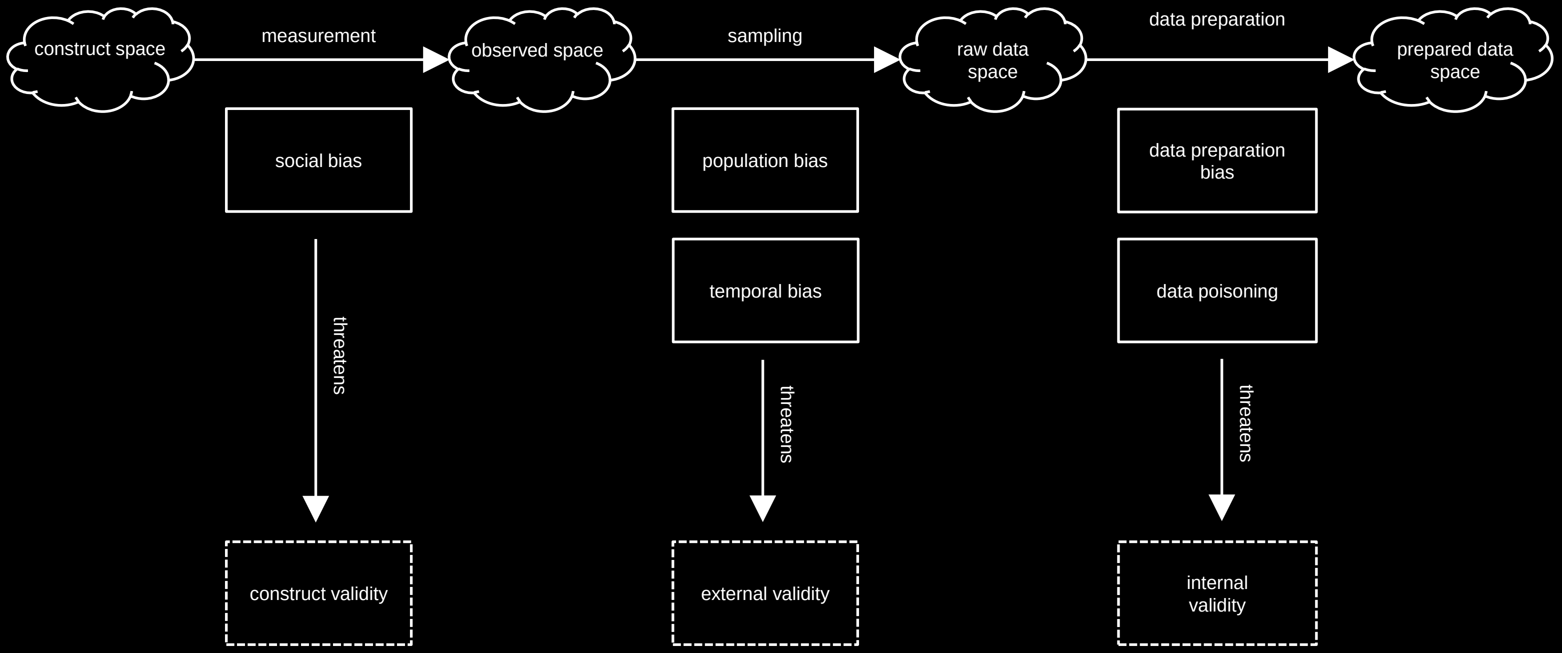
No shortcuts

No shortcuts in problem specification

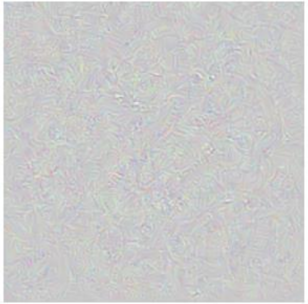
Take advice from a panel of diverse voices

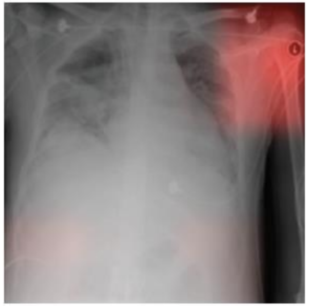


No shortcuts
in data understanding and preparation



No shortcuts in modeling





Article: Super Bowl 50

Paragraph: "*Peython Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had a jersey number 37 in Champ Bowl XXXIV.*"

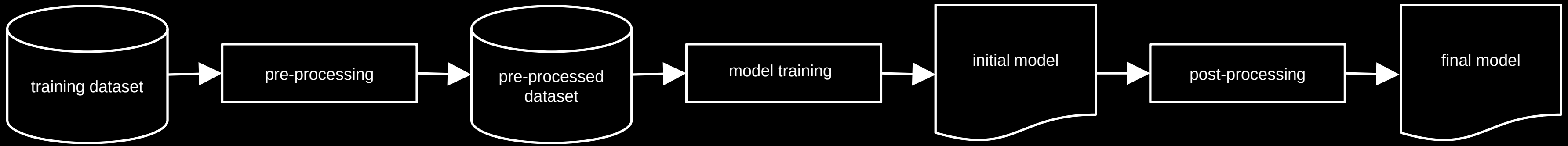
Question: "*What is the name of the quarterback who was 38 in Super Bowl XXXIII?*"

Original Prediction: John Elway

Prediction under adversary: Jeff Dean

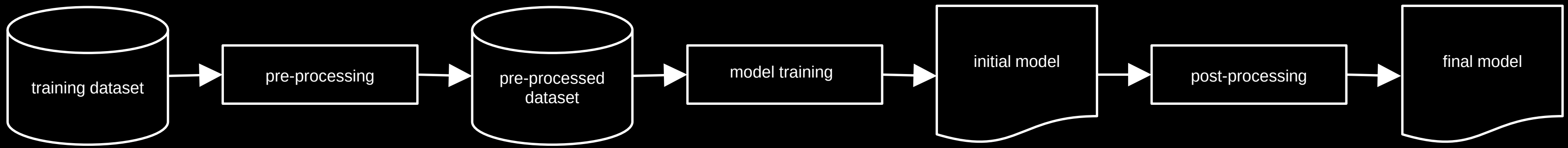
Task for DNN	Caption image	Recognise object	Recognise pneumonia	Answer question
Problem	Describes green hillside as grazing sheep	Hallucinates teapot if certain patterns are present	Fails on scans from new hospitals	Changes answer if irrelevant information is added
Shortcut	Uses background to recognise primary object	Uses features irrecognisable to humans	Looks at hospital token, not lung	Only looks at last sentence and ignores context

No shortcuts in modeling



No shortcuts in modeling

general robustness
adversarial robustness
fairness
explainability

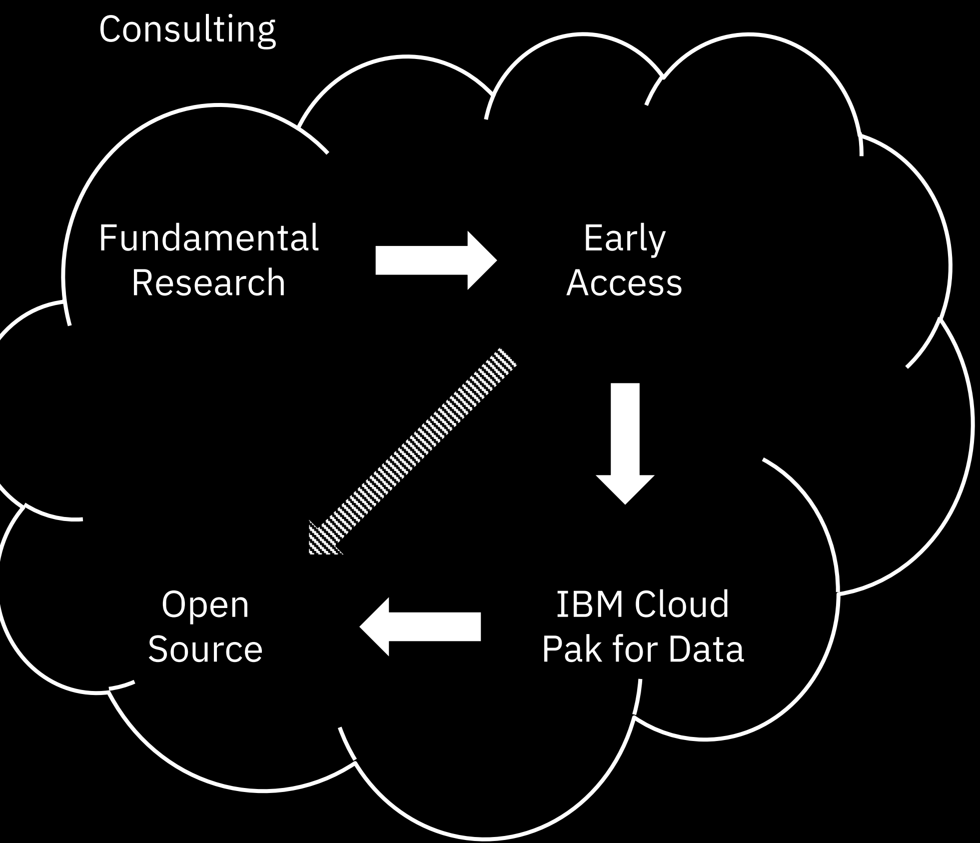


data augmentation
data sanitization
bias mitigation pre-processing
disentangled representations

invariant risk minimization
gradient shaping
bias mitigation in-processing
directly interpretable models

post hoc correction
post hoc defense
bias mitigation post-processing
post hoc explanations

Various routes to serve clients



- Trust capabilities in IBM Cloud Pak for Data
  - Explainability
  - Fairness
  - Drift detection
- Open source toolkits
  - AI Explainabilty 360
  - AI Fairness 360
  - Adversarial Robustness Toolbox
- Early access to enhanced editions of toolkits

Overview of trustworthy AI topics

domain expertise: lending, workforce, health

broader impacts: ecosystem, government affairs

topics and assets should not be viewed only through a technical lens; these are social issues too

AI FactSheets 360	consulting	human-computer interaction
transparency	**specification elicitation**	

transparent documentation and eliciting societal values and preferences from policymakers are critical for AI governance

FACT	DART	FreaAI	VerifAI	Uncertainty Quantification 360
testing				**uncertainty quantification**

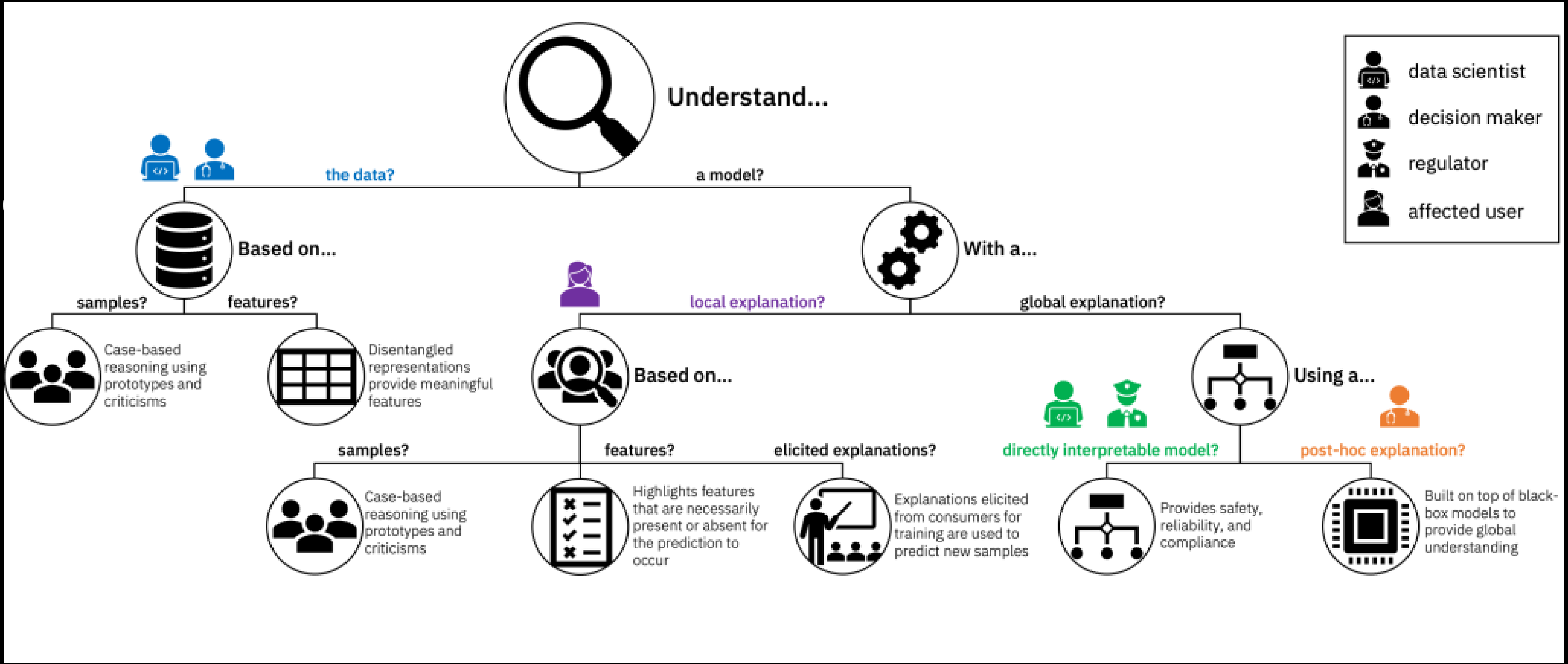
all elements of trust should be tested and reported with error bars

AI Explainability 360	AI Fairness 360	Adversarial Robustness 360	Trusted Generation 360
explainability	**fairness**	**robustness**	**generation**

causality and constraint-aware learning are important in their own right and also as foundations for the pillars of trust above

Causallib	Diffprivlib
causal modeling	**constraint-aware learning**

Explanations



Open Source AI Explainability 360

Supporting diverse and rich explanations.

http://github.com/ibm/aix360

http://aix360.mybluemix.net

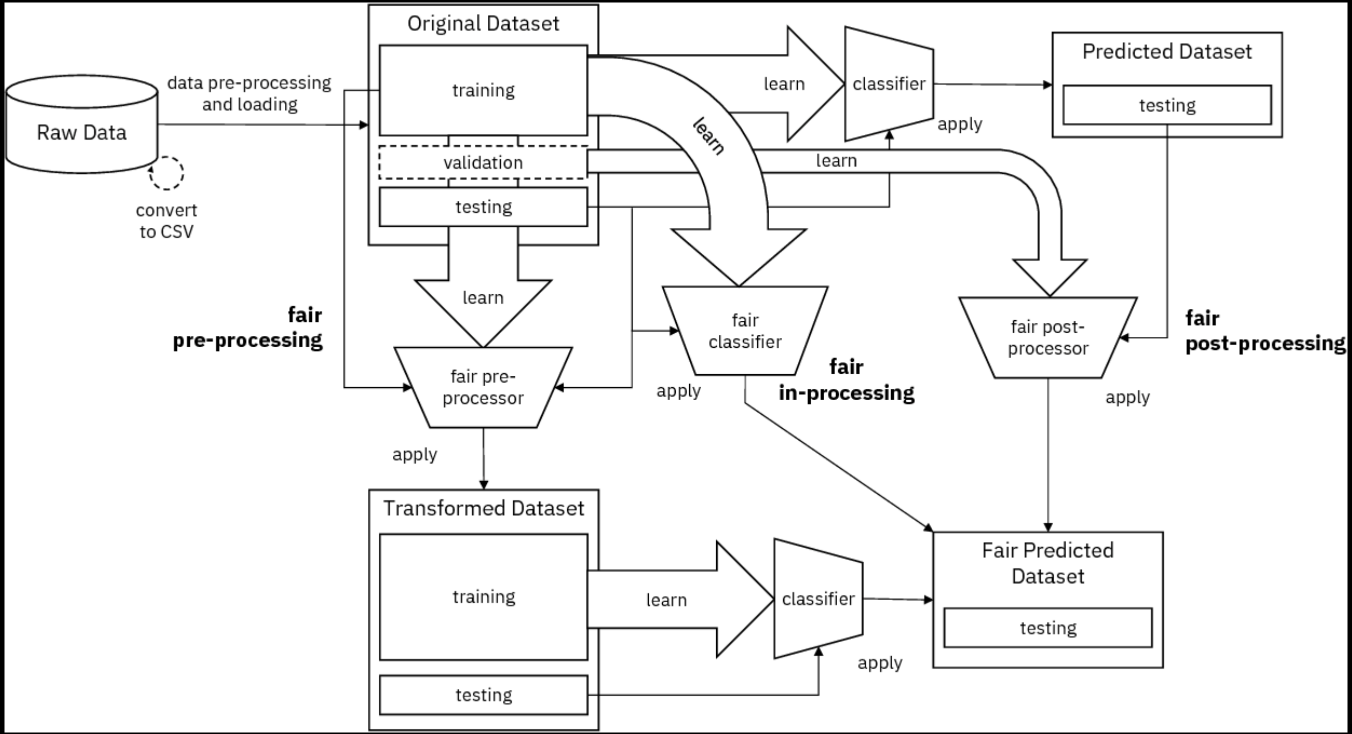
- 8 unique techniques from IBM Research
  - data vs. model
  - global vs. local
  - direct vs. post hoc
- LIME and SHAP
- 2 explainability metrics
- Extensive industry tutorials to educate users and practitioners
- Interactive demo

AI Explainability Enhanced Edition

Early access offering

- New state-of-the-art explanation methods from IBM Research
- Interactive explanation
- Better support for text
- Better support for regression

Fairness



Original Dataset
training
validation
testing
Raw Data
data pre-processing
and loading
convert
to CSV
learn
classifier
apply
Predicted Dataset
testing
learn
learn
learn
fair
pre-processing
fair pre-
processor
apply
fair
classifier
apply
fair
in-processing
fair post-
processor
apply
fair
post-processing
Transformed Dataset
training
testing
learn
classifier
apply
Fair Predicted
Dataset
testing

Open Source AI Fairness 360

The most comprehensive toolkit for handling bias in machine learning.

http://github.com/ibm/aif360

http://aif360.mybluemix.net

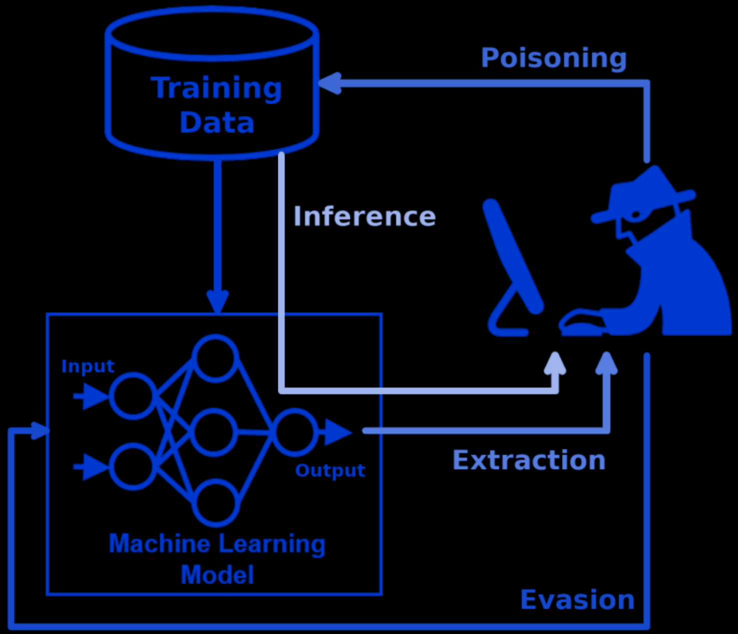
- Comprehensive set of fairness metrics
  - Group fairness
  - Individual fairness
- 12 state-of-the-art bias mitigation algorithms
  - Pre-processing
  - In-processing
  - Post-processing
- Extensive industry tutorials to educate users and practitioners
- Interactive demo

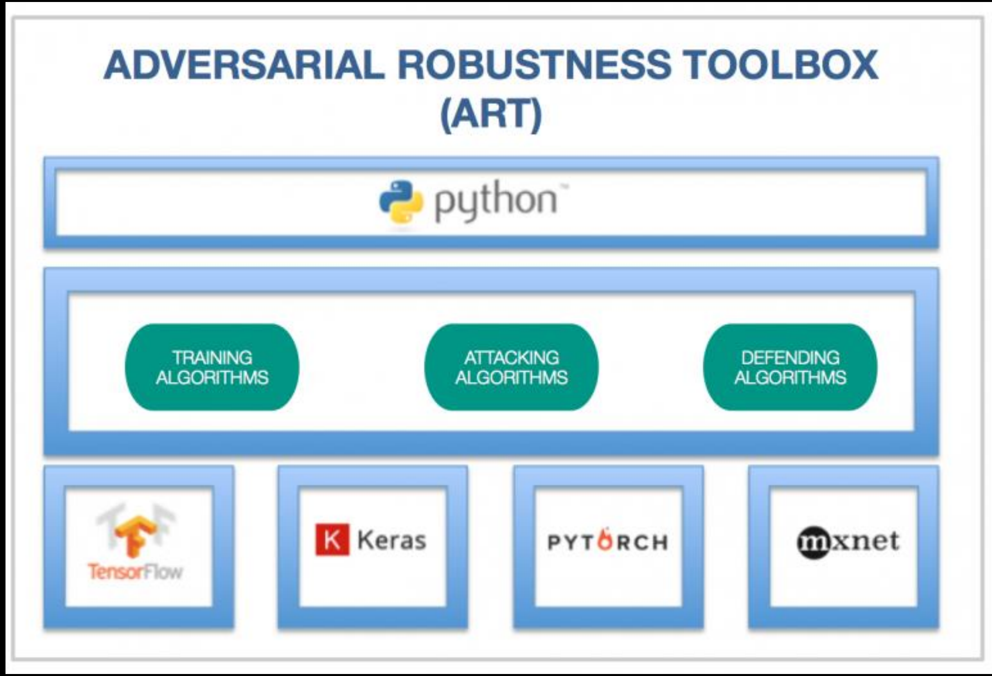
AI Fairness Enhanced Edition

Early access offering

- New state-of-the-art bias mitigation algorithms from IBM Research and the external community
- Support for natural language processing
- Bias mitigation and verification for individual fairness
- Support for transfer learning
- Fair generation
- Protected attribute extraction

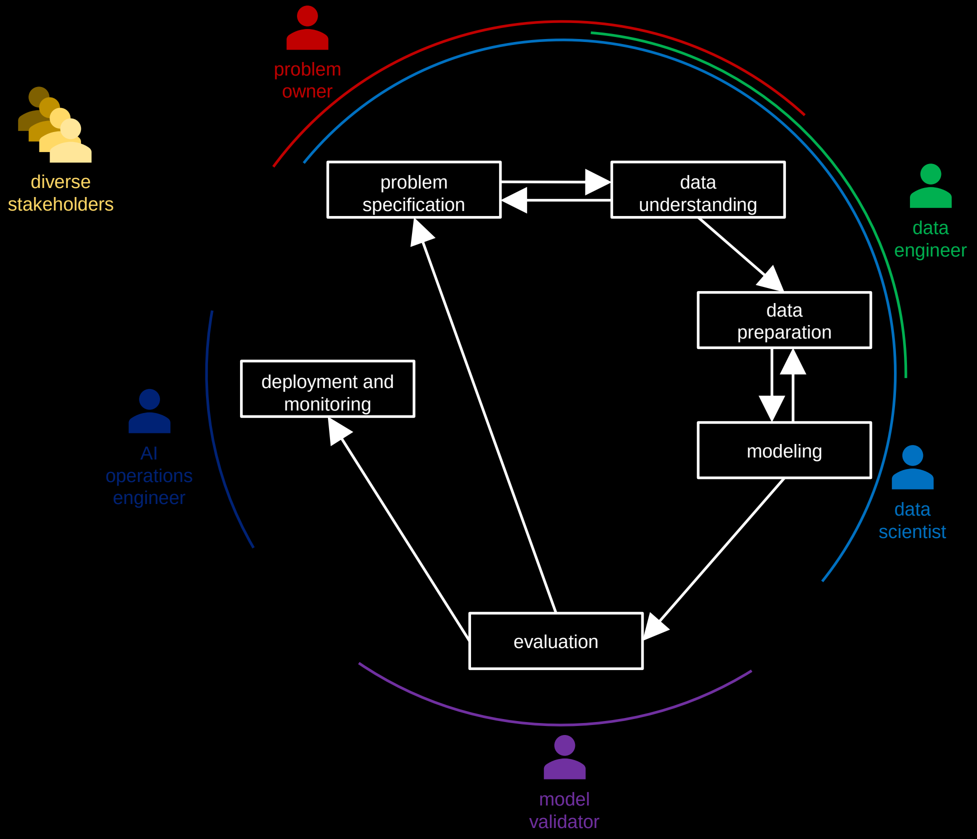
Adversarial robustness
https://art-demo.mybluemix.net



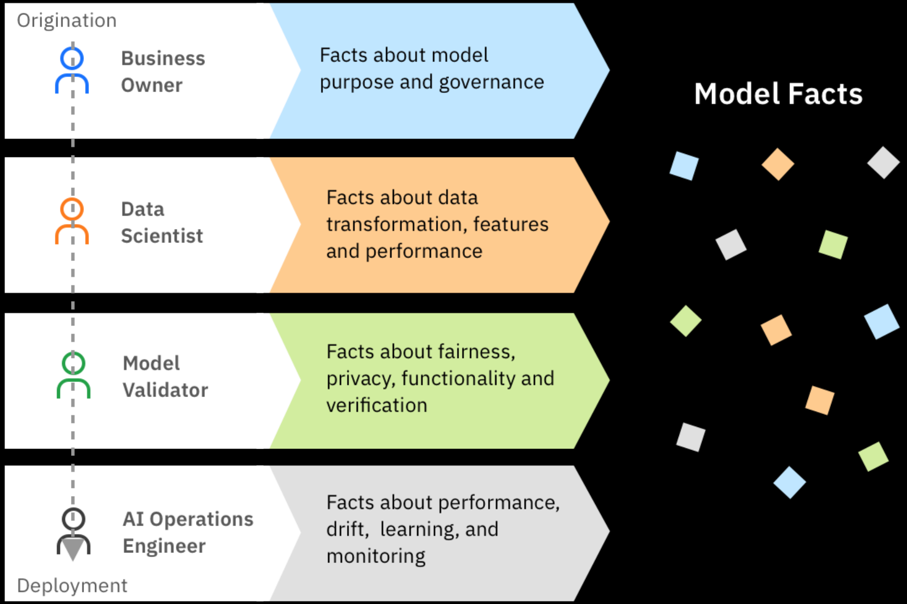


No shortcuts in evaluation, deployment and monitoring

Take advice from a panel of diverse voices

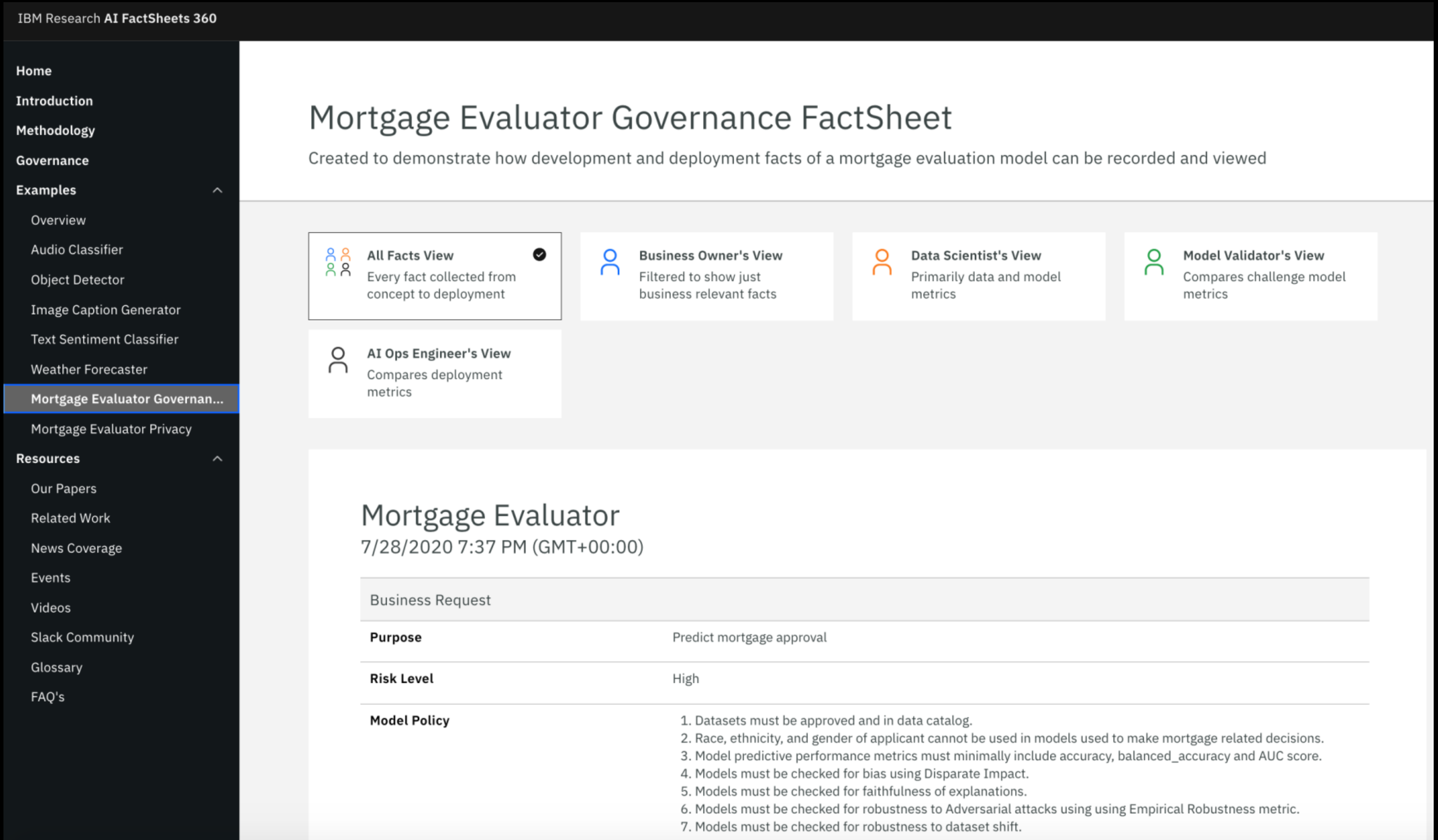


No shortcuts in governance



Predictive Performance	Data Scientist	Model Validator	AI Ops Engineer
Accuracy	0.95	0.94	0.92
Balanced Accuracy	0.63	0.63	0.61
AUC	0.79	0.78	0.77
F1	0.97	0.97	0.96
Fairness			
Disparate Impact	0.97	0.97	0.95
Statistical Parity Difference	-0.03	-0.03	-0.04
Adversarial Robustness			
Empirical Robustness	0.02	0.01	0.02
Explainabiity			
Faithfulness Mean	0.31	0.36	0.35

AI Factsheets 360 - https://aifs360.mybluemix.net/



Recap of the Toolkits

AI Fairness 360

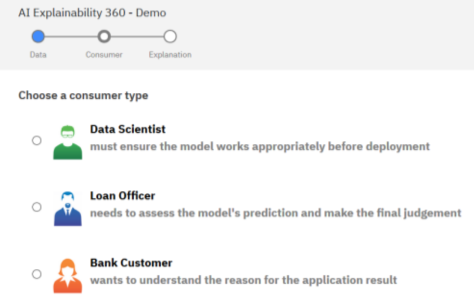
aif360.mybluemix.net

Most comprehensive **open source** toolkit for detecting & mitigating bias in ML models:

- 70+ fairness metrics
- 12 bias mitigators
- Interactive demo illustrating 5 bias metrics and 4 bias mitigators
- extensive industry tutorials and notebooks

AI Explainability 360

aix360.mybluemix.net



Most comprehensive **open source** toolkit for explaining ML models & data

- 8 explainability algorithms
- Interactive demo showing 3 algorithms in credit scoring application
- 13 tutorial notebooks: finance, healthcare, lifestyle, retention, etc.
- Extensive documentation and taxonomy of explainability algorithms

Adversarial Robustness 360

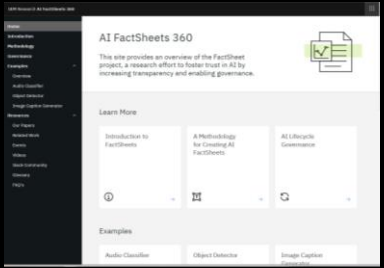
art-demo.mybluemix.net

Most comprehensive **open source** toolkit for defending AI from attacks

- Supports 10+ frameworks
- 19 composable and modular attacks (including adaptive white- and black-box)
- 10 defenses, including detection of adversarial samples and poisoning attacks
- Robustness metrics, certifications and verifications
- 30 notebooks covering attacks and defenses
- From dozens of publications

FactSheets 360

aifs360.mybluemix.net



Extensive website describing research effort to foster trust in AI by increasing transparency and enabling Governance

- 6 examples FactSheets
- 7-step methodology for creating useful FactSheets
- AI Governance
- Papers, videos, related work, FAQ, slack channel

Scaling and Automating AI

Dr. Lisa Amini, Director
IBM Research Cambridge

AI Agenda for the Enterprise

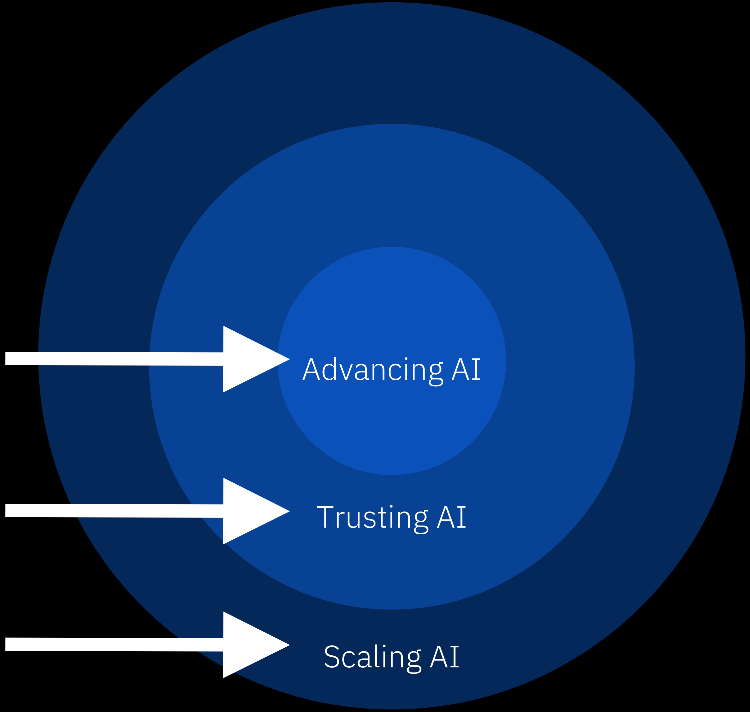
From narrow to broad AI

Advancing AI core algorithms for mastering language, learning and reasoning

Trusting AI through fairness, robustness, explainability, and transparency

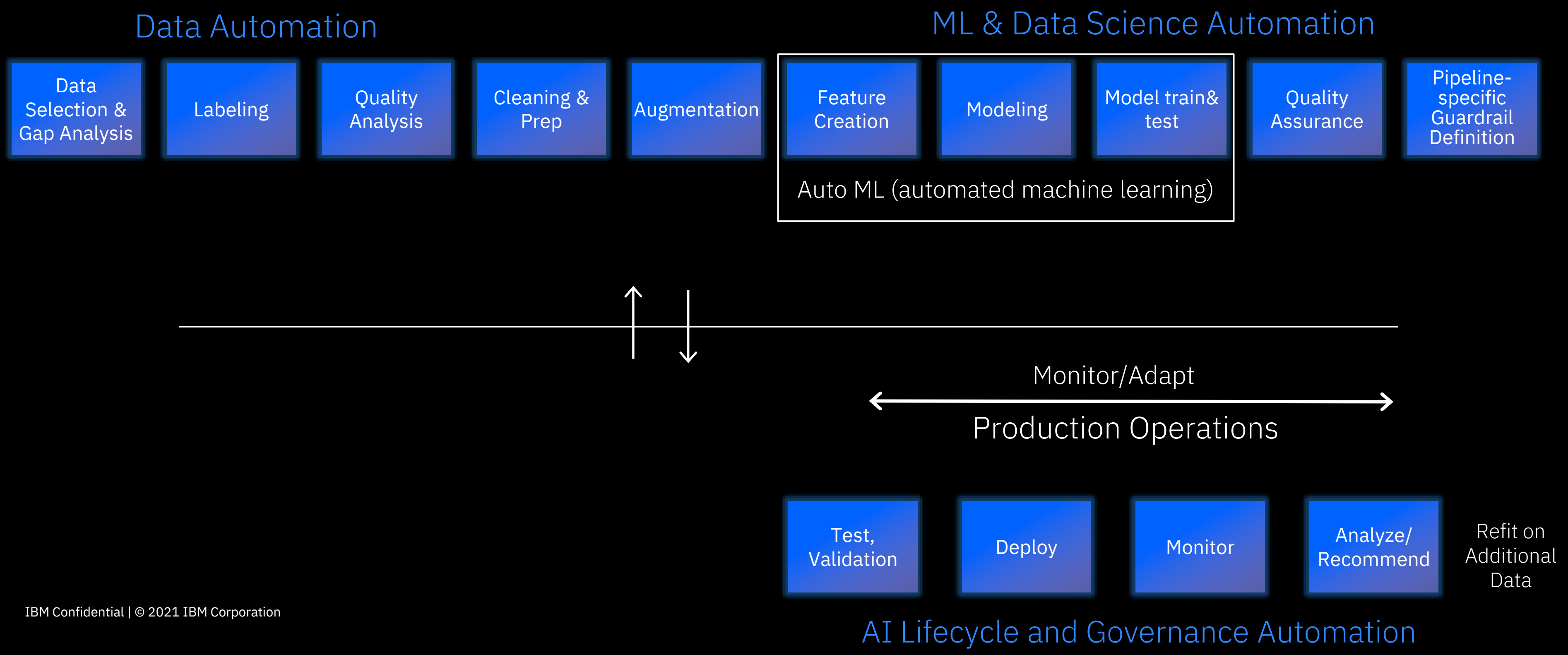
Operationalizing AI at scale

Scaling AI by managing, operating & automating its lifecycle

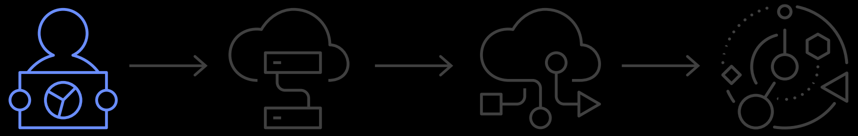


Scaling and automating AI

A holistic approach



Automation for data



Prepare Data

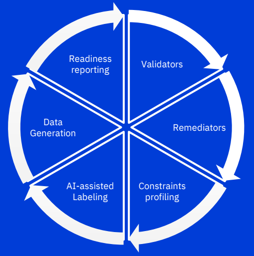
Create, Train Model

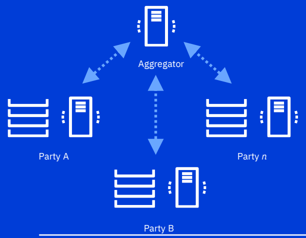
Validate & Deploy

Monitor & Adapt

Guide user in understanding **data readiness for ML** model building, and mitigating issues

Data Readiness Toolkit (DaRT)

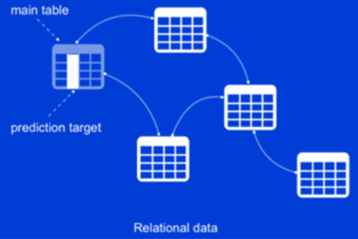


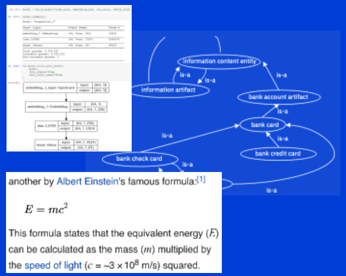


IBM Research's Framework for Federated Learning (FFL) and advanced Model Fusion algorithms enable training models across disparate systems, **keeping data in place**

Automate feature engineering for relational data with complex connections

OneButtonMachine





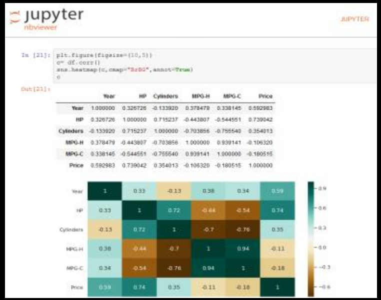
Knowledge Augmentation technology automatically **augments your data with domain knowledge from** *notebooks, documents, and other internal & external sources*

How expert data scientists drive up accuracy

External knowledge

Knowledge Graphs	*DBpedia, Wikidata, YAGO*
Open Data	*e.g. data.gov*
Web Tables	*HTML tables on the Web*
Blogs, newspapers	*e.g. NY Times*

Existing code developed by data scientists



Documents containing domain knowledge

Mathematical/ Engineering/ Business Domain Documents containing a mixture of natural language and mathematical formulae

$$\mathrm{BMI} = \frac{\mathrm{mass}_{\mathrm{kg}}}{\mathrm{height}_{\mathrm{m}}{}^{2}}$$

$$\mathrm{DTI} = \frac{\text{Total of Monthly Debt Payments}}{\text{Gross Monthly Income}}$$

another by Albert Einstein's famous formula:[1]

$$E = mc^2$$

This formula states that the equivalent energy (E) can be calculated as the mass (m) multiplied by the speed of light (c = ~3 x 10^8 m/s) squared.

Automatically augmenting data with knowledge

NAME	AGE	INCOME	CREDIT
Hana Ja	34	$50,000	$20,000
Taro Tok	25	$40,000	$20,000

Input data to ML Problem

NAME	AGE	INCOME	CREDIT	Loan / Income
Hana Ja	34	$50,000	$20,000	0.4
Taro Tok	25	$40,000	$20,000	0.5

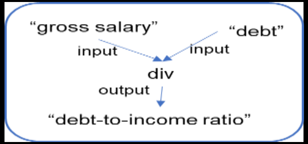
Augmented data

KAFE: Automated Feature Enhancement for Predictive Modeling using External Knowledge
Sainyam Galhotra, Udayan Khurana, Oktie Hassanzadeh, Kavitha Srinivas and Horst Samulowitz
KR2ML Workshop at NeurIPS, 2019

Automatically augmenting data with knowledge

NAME	AGE	INCOME	CREDIT
Hana Ja	34	$50,000	$20,000
Taro Tok	25	$40,000	$20,000

Input data to ML Problem



4 ***Feature Construction***

NAME	AGE	INCOME	CREDIT	Loan / Income
Hana Ja	34	$50,000	$20,000	0.4
Taro Tok	25	$40,000	$20,000	0.5

Augmented data

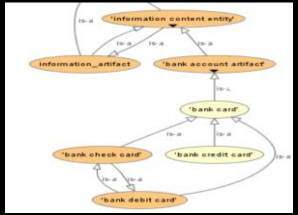
2 ***Feature-concept Mapping***

$$\text{DTI} = \frac{\text{Total of Monthly Debt Payments}}{\text{Gross Monthly Income}}$$

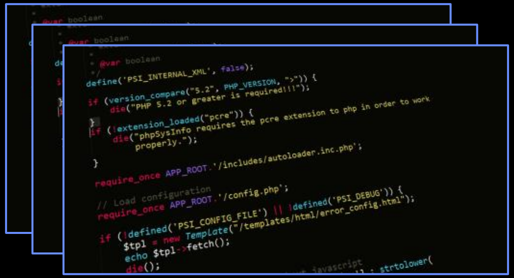
WIKIDATA

3 ***Expression Extraction***

Natural Language Documents (i.e. Wikipedia, internal documents)



1 **Knowledge Base**



---

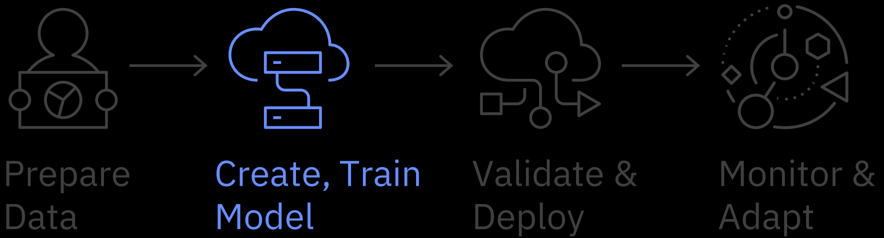
1 **Leverage Knowledge Bases** — Merging, unifying and smart indexing of existing knowledge bases

2 **Feature to concept mapping** — Map data and features to concepts (eg., numbers to '*income*')

3 **Expression extraction** — Extract formulas and expressions (such as '*income divided by loan amount*')

4 **Feature construction** — Compute novel features based on concept mapping and expressions

Automation for more complex modalities & tasks



AI automation for enterprise

Data scientist

Bank wants to determine an appropriate interest rate for a new loan

Task: Predict credit risk for an application

Business process & application owner

- Maximize business value to at least $12 million
- Limit false positives to less than 10%

AI operations

- Cloud budget < $1 million
- Timeline: 4 weeks

Compliance

- Minimize bias
- Provide explanations for approval/denial
- Race can never be used as a feature

Application developer

- Model size < 1GB
- Time per prediction < 30ms

IT operations

- Given high dimensions of data use support vector machines
- Apply previously generated models

How is it solved?

Alternating direction method of multipliers (ADMM) based algorithm solves this complex optimization problem by splitting into easier subproblems.

Black-box Constraints

- Deployment constraints
- Fairness
- Explainability & Interpretabiliity
- Problem & domain
- Physical & natural

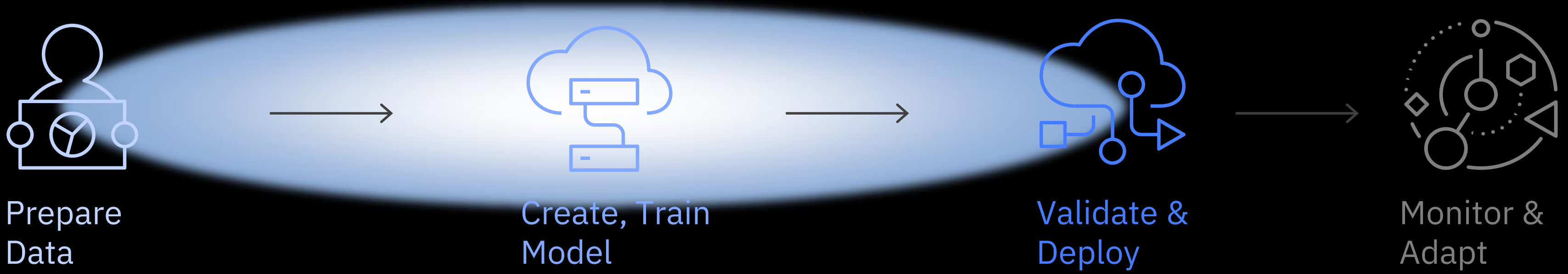
Method Selection

- Combinatorial multi-armed bandits
- Genetic algorithms
- Multi-fidelity approximation

Hyper-Parameter Optimization

- Random search (can be competitive)
- Sequential Model-Based Optimization
  - Gaussian Processes
- Trust-region DFO

Formulation enables automation well beyond ML model building



An ADMM Based Framework for AutoML Pipeline Configurations, AAAI, 2020

Consider Bias Mitigation Algorithms...

Pre-Processing Algorithms
Mitigates Bias in **Training Data**

Reweighing
Modifies the weights of different training examples

Disparate Impact Remover
Edits feature values to improve group fairness

Optimized Preprocessing
Modifies training data features and labels

Learning Fair Representations
Learns fair representations by obfuscating information about protected values

In-Processing Algorithms
Mitigates Bias in **Classifiers**

Meta Fair Classifier
Takes fairness metric as part of input & returns classifier optimized for the metric

Prejudice Remover
Adds a discrimination-aware regularization term to learning objective

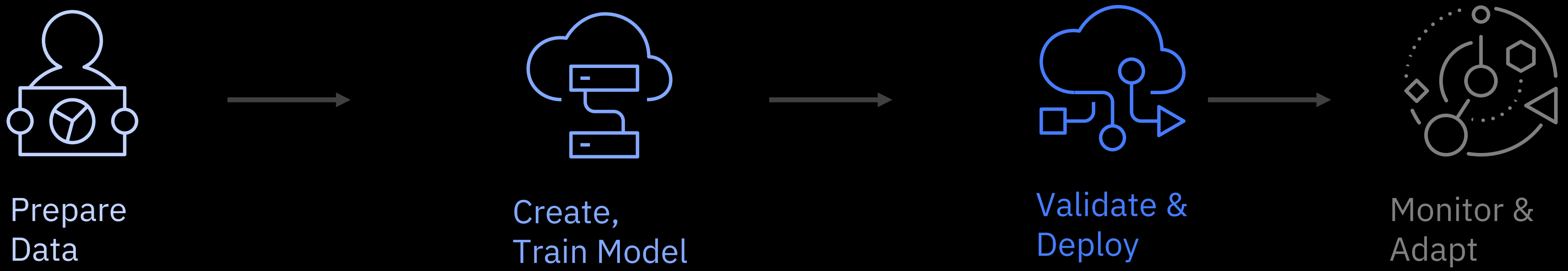
Adversarial Debiasing
Uses adversarial methods to maximize accuracy & reduce evidence of protected attributes in predictions

Post-Processing Algorithms
Mitigates Bias in **Predictions**

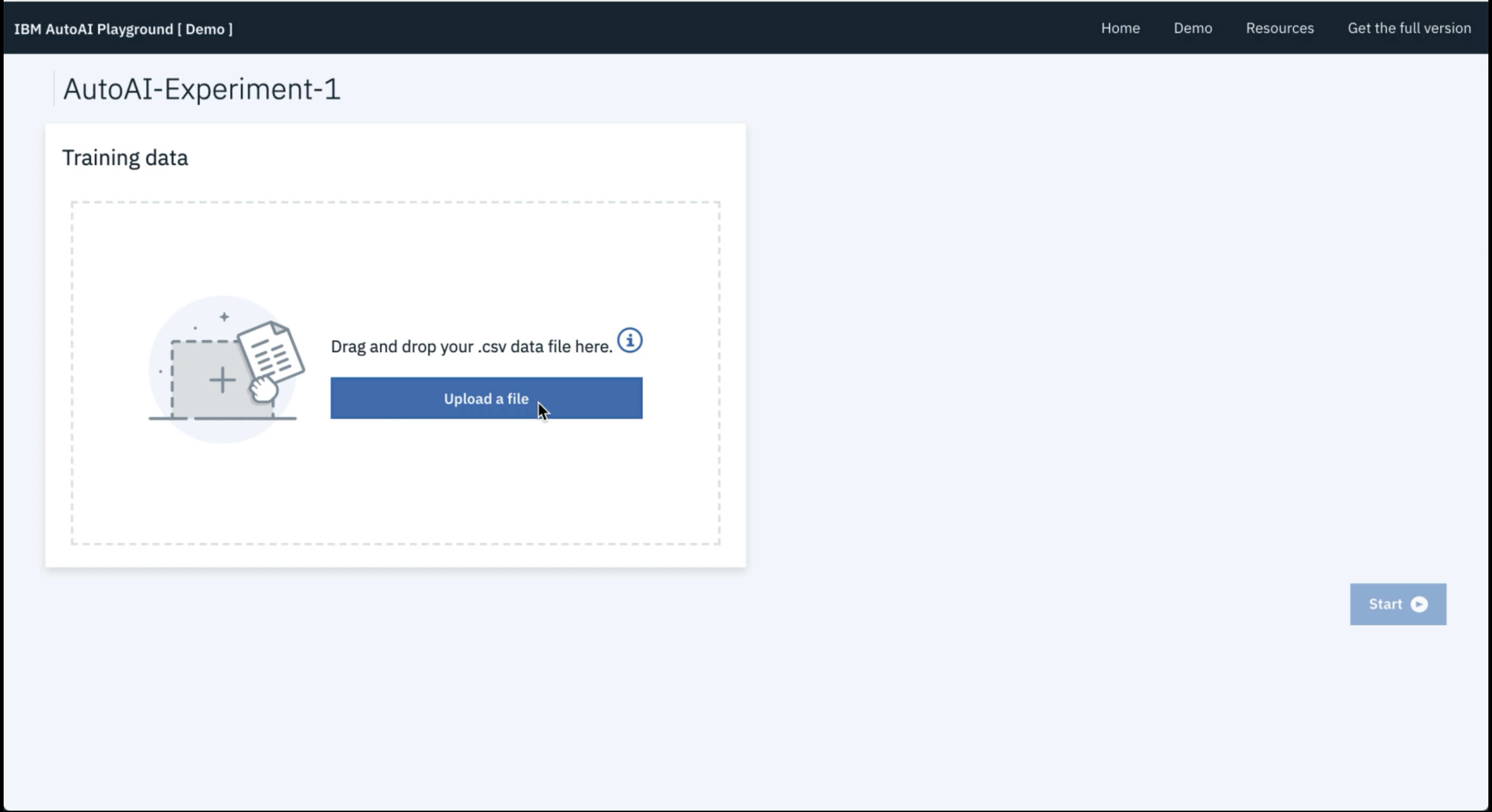
Rejection Option Classification
Changes predictions from a classifier to make them fairer

Calibrated Equalized Odds
Optimizes over calibrated classifier score outputs that lead to fair output labels

Equalized Odds
Modifies the predicted label using an optimization scheme to make predictions fairer

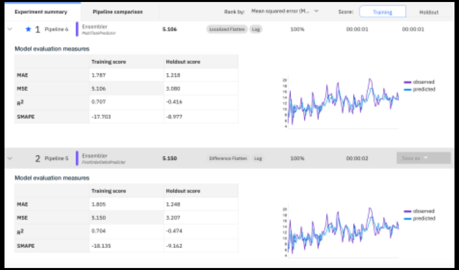


https://aif360.mybluemix.net



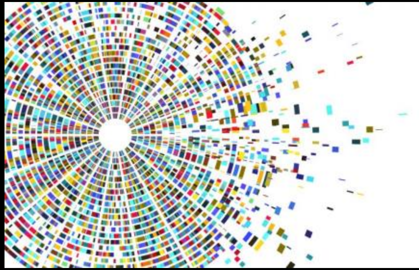
How do Data Science Workers Collaborate?: Roles, Workflows, and Tools, CSCW, 2020.
How Data Scientists Work Together with Domain Experts in Scientific Collaborations: To Find the Right Answer or To Ask the Right Question?, GROUP, 2020.

Time series forecast, scaling, and automation



Auto generate forecasting pipelines with algorithms:

- Auto feature engineering and hyperparameter optimization tailored for time series
- Support multi-series and multi-step forecasting
- Configurable back testing setup

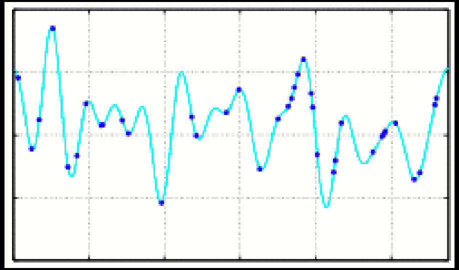


Support large scale time series data:

Typical data sets from many use cases may contain:

- 1000s of time series
- 100k+ data points per series

Handling this scale requires special architectures and techniques.



Handle irregularly sampled data:

- Data isn’t always well behaved (uniformly sampled, with no missing values).
- Conventional approaches for resampling or imputation may not yield good results.
- Leverage state of the art deep learning models to address these challenges.

Enable augmentation with external variables:

Signals from weather and other data sources can help the prediction process for the target time series.

Expanding the scope of Auto AI to decision optimization

data-driven generation example: supply chain inventory

Historical Decisions

Historical Business Impact

Date	Demand
01/02/2020	500
02/02/2020	300
03/02/2020	200
04/02/2020	600
05/02/2020	700
06/02/2020	100
07/02/2020	200
08/02/2020	100
09/02/2020	300
10/02/2020	400
11/02/2020	500
12/02/2020	100
13/02/2020	50

Demand forecasting from "Standard AutoAI"

Date	Demand	Inventory	Restocking	Cost
01/02/2020	500	1000		500
02/02/2020	300	500		300
03/02/2020	200	200	300	200
04/02/2020	600	300	1000	2100
05/02/2020	700	700	300	700
06/02/2020	100	300		100
07/02/2020	200	200	100	200
08/02/2020	100	100		100
09/02/2020	300	0	700	1800
10/02/2020	400	400	300	400
11/02/2020	500	300		1500
12/02/2020	100	-200	500	1600
13/02/2020	50	200		50

Additional information to model decision making *from historical data*

Reinforcement Learning

$$\max \sum_{t=0}^{\infty} \gamma^t E[R_{t,(X_t,A_t)}]$$

$$P(X_{t+1} = y / X_t = x, A_t = a)$$

Automatically generate model based on history (e.g. using regression, state/action reward)

- Many patterns amenable to data-driven automation: Set point optimization, Continuous Flow Maximization, Inverse Optimization, RL for sequential decision optimization.
- Enables data-driven learning of risk-measures, scenarios, or uncertainty sets

Training Stronger Baselines for Learning to Optimize, NeurIPS 2020

IBM AutoAI-SDK Auto-Generated Notebook v1.12.2

Note: Notebook code generated using AutoAI will execute successfully. If code is modified or reordered, there is no guarantee it will successfully execute. This pipeline is optimized for the original dataset. The pipeline may fail or produce sub-optimium results if used with different data. For different data, please consider returning to AutoAI Experiments to generate a new pipeline. Please read our documentation for more information:
Cloud Platform

Before modifying the pipeline or trying to re-fit the pipeline, consider:
The notebook converts dataframes to numpy arrays before fitting the pipeline
(a current restriction of the preprocessor pipeline). The known_values_list is passed by reference and populated with categorical values during fit of the preprocessing pipeline. Delete its members before re-fitting.

Notebook content

This notebook contains steps and code to demonstrate AutoAI pipeline. This notebook introduces commands for getting data, pipeline model, model inspection and testing.

Some familiarity with Python is helpful. This notebook uses Python 3.

Notebook goals

- inspection of trained pipeline via graphical vizualization and source code preview.
- pipeline evaluation.
- pipeline deployment and webservice scoring

Contents

This notebook contains the following parts:

1. Setup
   a. AutoAI experiment metadata
2. Pipeline inspection
   a. Get historical optimizer instance
   b. Get pipeline model
   c. Preview pipeline model as python code
   d. Visualize pipeline model
   e. Read training and holdout data
   f. Test pipeline model locally
3. Pipeline refinery
   a. Pipeline definition source code
   b. Lale library
4. Deploy and score
   a. Insert WML credentials
   b. Create deployment
   c. Score webservice
   d. Delete deployment
5. Authors

Setup

Before you use the sample code in this notebook, you must perform the following setup tasks:

- `watson-machine-learning-client` uninstallation of the old client
- `watson-machine-learning-client-V4` installation
- `autoai-libs` installation/upgrade
- `lightgbm` or `xgboost` installation/downgrade if they are needed

```
In [ ]: !pip uninstall watson-machine-learning-client -y
```

```
In [ ]: !pip install -U watson-machine-learning-client-V4
```

```
In [ ]: !pip install -U autoai-libs
        !pip install -U lightgbm==2.2.3
```

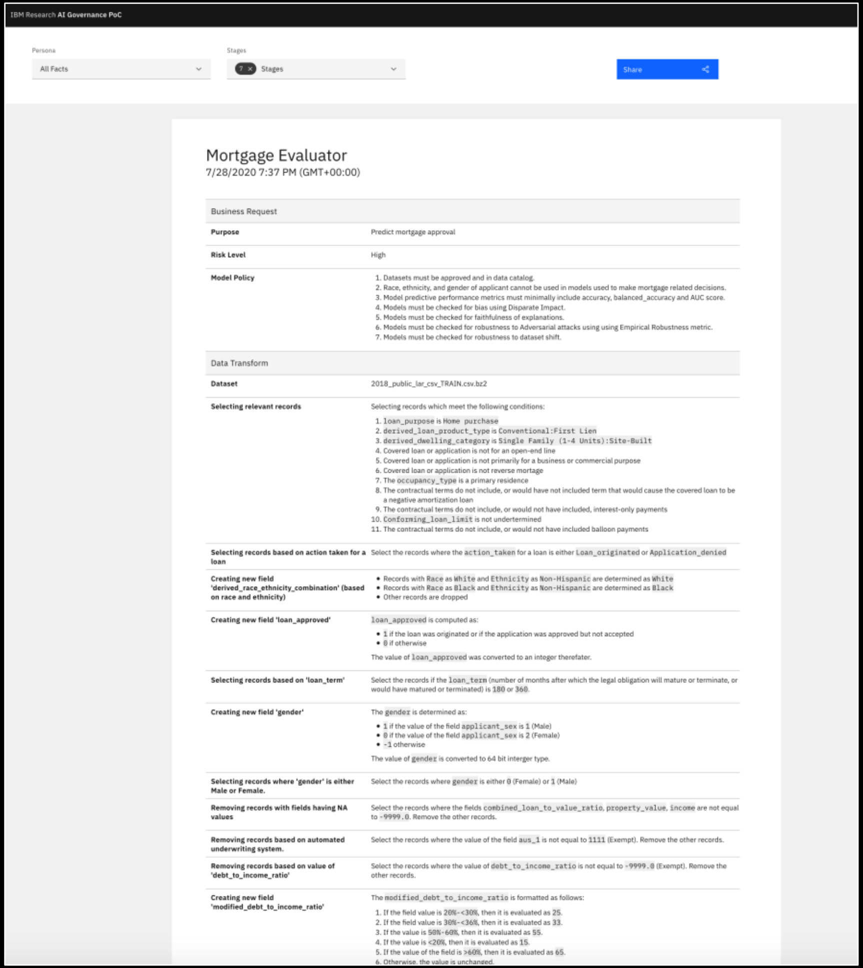
AutoAI experiment metadata

This cell contains input parameters provided to run the AutoAI experiment in Watson Studio and COS credentials required to retrieve AutoAI pipeline.

```
In [1]: from watson_machine_learning_client.helpers import DataConnection, S3Connection, S3Location

        experiment_metadata = dict(
           prediction_type='classification',
           prediction_column='grade',
           test_size=0.1,
           scoring='accuracy',
```

Watson AutoAI: Can I get the model? Medium.com



IBM Research AI Governance PoC

Persona: All Facts
Stages: 7 × Stages
Share

Mortgage Evaluator

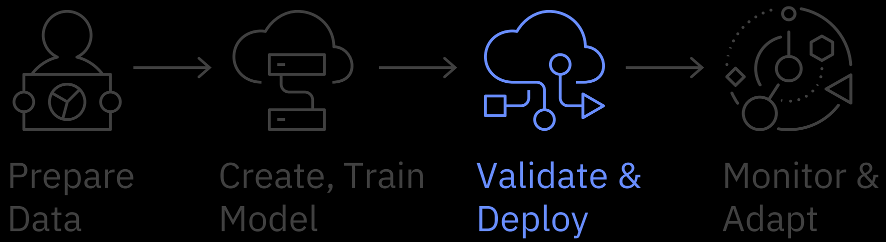
7/28/2020 7:37 PM (GMT+00:00)

Business Request	
Purpose	Predict mortgage approval
Risk Level	High
Model Policy	1. Datasets must be approved and in data catalog. 2. Race, ethnicity, and gender of applicant cannot be used in models used to make mortgage related decisions. 3. Model predictive performance metrics must minimally include accuracy, balanced_accuracy and AUC score. 4. Models must be checked for bias using Disparate Impact. 5. Models must be checked for faithfulness of explanations. 6. Models must be checked for robustness to Adversarial attacks using using Empirical Robustness metric. 7. Models must be checked for robustness to dataset shift.

Data Transform	
Dataset	2018_public_lar_csv_TRAIN.csv.bz2
Selecting relevant records	Selecting records which meet the following conditions: 1. `loan_purpose` is `Home purchase` 2. `derived_loan_product_type` is `Conventional:First Lien` 3. `derived_dwelling_category` is `Single Family (1-4 Units):Site-Built` 4. Covered loan or application is not for an open-end line 5. Covered loan or application is not primarily for a business or commercial purpose 6. Covered loan or application is not reverse mortage 7. The `occupancy_type` is a primary residence 8. The contractual terms do not include, or would have not included term that would cause the covered loan to be a negative amortization loan 9. The contractual terms do not include, or would not have included, interest-only payments 10. `Conforming_loan_limit` is not undertermined 11. The contractual terms do not include, or would not have included balloon payments
Selecting records based on action taken for a loan	Select the records where the `action_taken` for a loan is either `Loan_originated` or `Application_denied`
Creating new field 'derived_race_ethnicity_combination' (based on race and ethnicity)	• Records with `Race` as `White` and `Ethnicity` as `Non-Hispanic` are determined as `White` • Records with `Race` as `Black` and `Ethnicity` as `Non-Hispanic` are determined as `Black` • Other records are dropped
Creating new field 'loan_approved'	`loan_approved` is computed as: • `1` if the loan was originated or if the application was approved but not accepted • `0` if otherwise The value of `loan_approved` was converted to an integer therefater.
Selecting records based on 'loan_term'	Select the records if the `loan_term` (number of months after which the legal obligation will mature or terminate, or would have matured or terminated) is `180` or `360`.
Creating new field 'gender'	The `gender` is determined as: • `1` if the value of the field `applicant_sex` is `1` (Male) • `0` if the value of the field `applicant_sex` is `2` (Female) • `-1` otherwise The value of `gender` is converted to 64 bit interger type.
Selecting records where 'gender' is either Male or Female.	Select the records where `gender` is either `0` (Female) or `1` (Male)
Removing records with fields having NA values	Select the records where the fields `combined_loan_to_value_ratio`, `property_value`, `income` are not equal to `-9999.0`. Remove the other records.
Removing records based on automated underwriting system.	Select the records where the value of the field `aus_1` is not equal to `1111` (Exempt). Remove the other records.
Removing records based on value of 'debt_to_income_ratio'	Select the records where the value of `debt_to_income_ratio` is not equal to `-9999.0` (Exempt). Remove the other records.
Creating new field 'modified_debt_to_income_ratio'	The `modified_debt_to_income_ratio` is formatted as follows: 1. If the field value is `20%-<30%`, then it is evaluated as `25`. 2. If the field value is `30%-<36%`, then it is evaluated as `33`. 3. If the value is `50%-60%`, then it is evaluated as `55`. 4. If the value is `<20%`, then it is evaluated as `15`. 5. If the value of the field is `>60%`, then it is evaluated as `65`. 6. Otherwise, the value is unchanged.

https://pages.github.ibm.com/factsheets/ai-gov-demo/
https://www.youtube.com/watch?v=-4d3kEVsu-s&t=2s

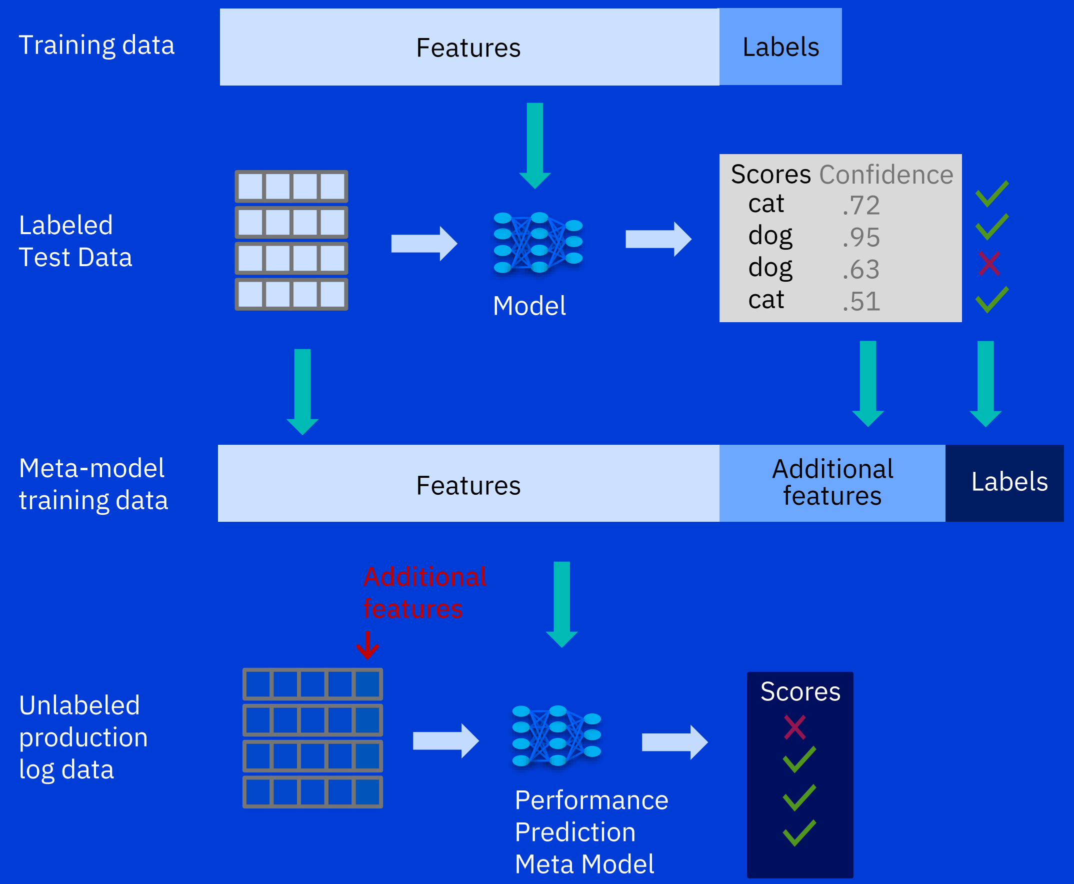
AI lifecycle automation



Validation

Meta-learning for performance prediction

1. Train the model
2. Score the test set with the model and evaluate
3. Construct new training data for the meta-model from the model’s scores on the test data
4. Train the meta-model
5. Score production data with the meta-model to obtain accuracy predictions



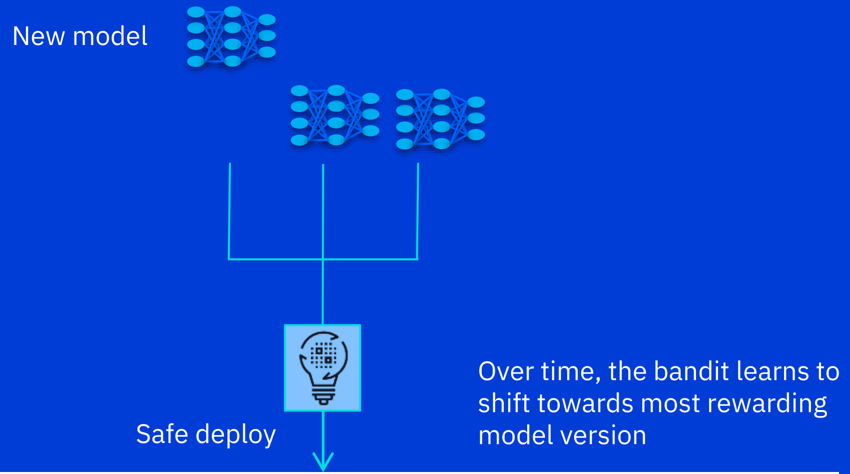
Deployment

Problem

I need to introduce a new version of a model into my application with minimal risk

Our Approach

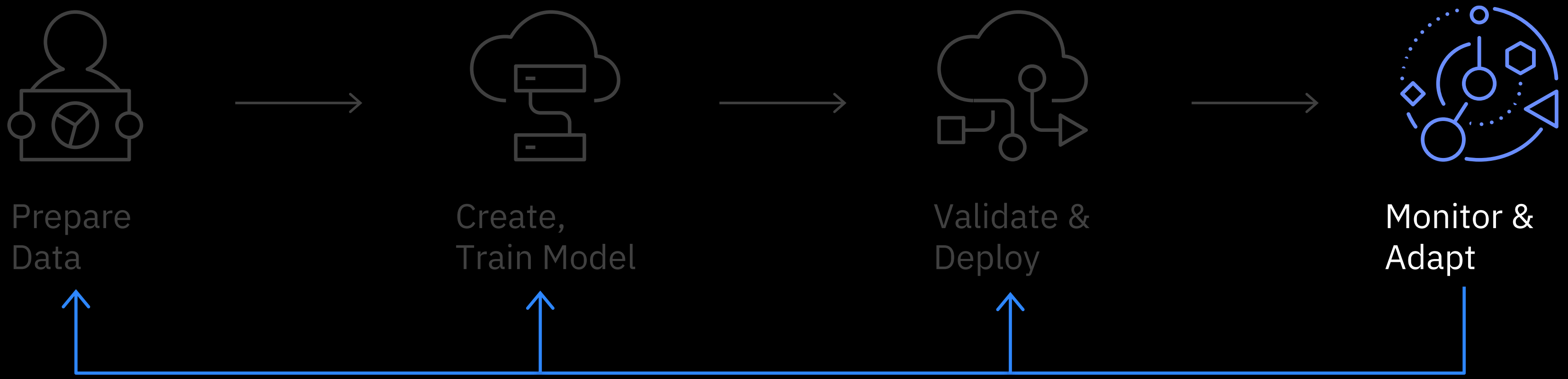
Staged deployment with Contextual Bandits that transparently deploys new versions safely without interrupting the application



Bandit explores available model versions to learn

Learning Exploration for Contextual Bandit, KDD (workshop), 2019

AI Lifecycle Automation



Composable capabilities across the AI lifecycle

Orchestrated through reusable pipelines

Data Selection, Quality, Management, Labeling

- DaRT Data Readiness Toolkit
- Active Learning Toolkit
- Conflict Analysis Toolkit
- New Class Detection Toolkit
- AI assisted Data Labeling Toolkit

Trust

- AIF 360 Toolkit
- AIX 360 Toolkit
- AI Robustness Toolkit

Insight, Analysis & Improvement

- Performance Prediction Toolkit
- Business KPI Insight Toolkit
- Improvement Recommender Toolkit
- Reward Based Learning Toolkit

Unsupervised Learning

Key areas of innovation

AI automation

Simplifying and accelerating the *entire,* end-to-end AI Lifecycle

from months to days to minutes

Continuously improving your AI models and applications

with transparency

Expanding our scope of Automation

to domain knowledge and decision optimization

References

AutoAI with Data Quality and Business Constraints

- AutoAI Playground: https://autoai.mybluemix.net/
- Lale Git repo: https://github.com/IBM/lale/
- Lale: Consistent Automated Machine Learning.
  Guillaume Baudart, et. al., KDD Workshop AutoML@KDD, 2020.
  https://arxiv.org/abs/2007.01977
- Mining Documentation to Extract Hyperparameter Schemas.
  Guillaume Baudart, et. al. ICML Workshop AutoML@ICML, 2020.
  https://arxiv.org/abs/2006.16984
- Solving Constrained CASH Problems with ADMM.
  Parikshit Ram, et al. ICML Workshop AutoML@ICML, 2020.
- Data Quality for AI: https://w3.ibm.com/w3publisher/ibm-research-ai-business-development/ai-portfolio/data-quality-for-ai
- An ADMM Based Framework for AutoML Pipeline Configuration.
  Liu, Sijia, et al. AAAI. 2020.
- Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI. Wang, Dakuo, et al. CSCW (2019)
- AutoAI: Automating the End-to-End AI Lifecycle with Humans-in-the-Loop.
  Dakuo Wang, et al. ICIUI, 2020.
- AutoAIViz: opening the blackbox of automated artificial intelligence with conditional parallel coordinates. Daniel Weidele, et al. ICIUI, 2020.
- Selecting Near-Optimal Learners via Incremental Data Allocation
  Ashish Sabharwal, Horst Samulowitz and Gerald Tesauro. AAAI 2016.

Scaling AI through Federated Learning and Fast Inference

- IBM federated learning Git repo: https://github.com/IBM/federated-learning-lib
- IBM federated learning web page: https://ibmfl.mybluemix.net
- IBM federated learning White paper: https://arxiv.org/abs/2007.10987
- Snap ML project web page: https://www.zurich.ibm.com/snapml/
- Snap ML documentation: https://ibmsoe.github.io/snap-ml-doc/
- Snap ML NeurIPS 2020 paper: https://arxiv.org/abs/2006.09745

Knowledge Augmentation for Supervised Learning:

- KAFE: Automated Feature Enhancement for Predictive Modeling using External Knowledge
  Sainyam Galhotra, Udayan Khurana, Oktie Hassanzadeh, Kavitha Srinivas and Horst Samulowitz
  KR2ML Workshop at NeurIPS, 2019
- Automated Feature Enhancement for Predictive Modeling using External Knowledge
  Sainyam Galhotra, Udayan Khurana, Oktie Hassanzadeh, Kavitha Srinivas, Horst Samulowitz, Miao Qi
  IEEE ICDM (demo track), 2019
- Semantic Search over Structured Data, Sainyam Galhotra and Udayan Khurana [Nominated for Best Paper CIKM2020]
  https://dl.acm.org/doi/10.1145/3340531.3417426

Auto AI Lifecycle for Time Series:

- FLOps: On Learning Important Time Series Features for Real-Valued Prediction.
  Dhaval Patel, et. al., IEEE Big Data 2020
- Smart-ML: A System for Machine Learning Model Exploration using Pipeline Graph.
  Dhaval Patel, et. al.,. IEEE Big Data 2020
- ThunderML: A Toolkit for Enabling AI/ML Models on Cloud for Industry 4.0S Shrivastava, D Patel, WM Gifford, S Siegel, J Kalagnanam, International Conference on Web Services, 163-180
- Providing Cooperative Data Analytics for Real Applications Using Machine Learning, A Iyengar, et. al. IEEE ICDCS, 2019
- Model Agnostic Contrastive Explanations for Structured Data.
  Dhurandhar et al.
- Boolean decision rules via column generation.
  Dash, Sanjeeb, Oktay Gunluk, and Dennis Wei. NeurIPS 2018.

IBM AutoAI Product Resources

- IBM Developer: https://developer.ibm.com/series/explore-autoai/
- Coursera course: https://www.coursera.org/learn/ibm-rapid-prototyping-watson-studio-autoai
- AutoAI demo: https://www.ibm.com/demos/collection/IBM-Watson-Studio-AutoAI/