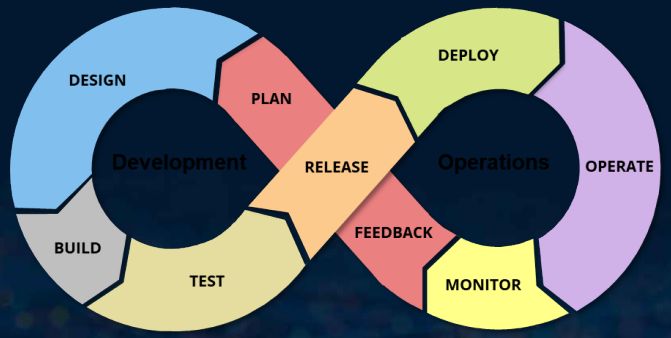
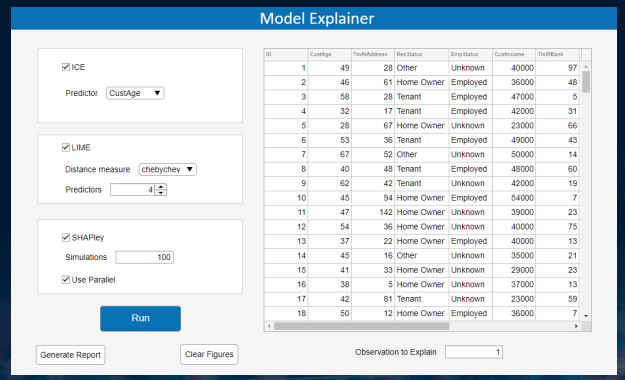
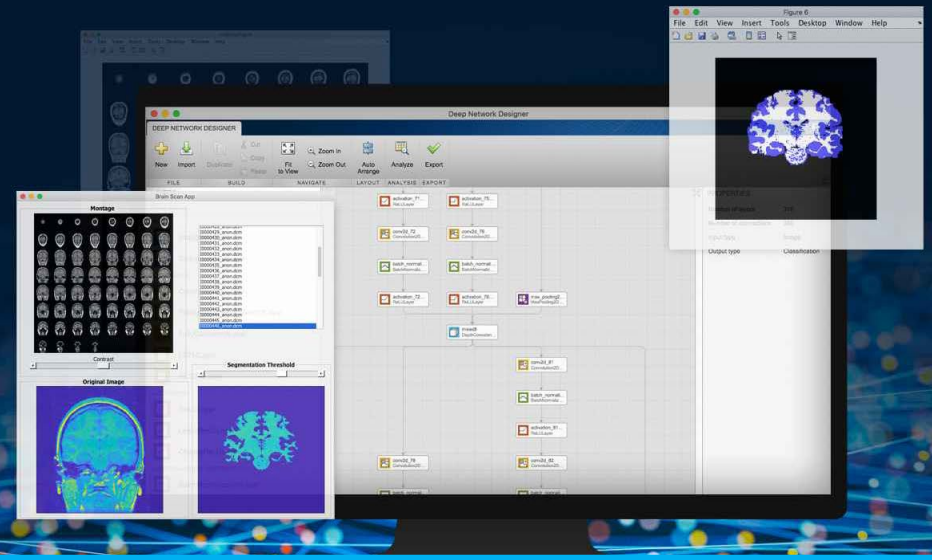






Building a responsible AI pipeline

MathWorks Computational Finance Conference

Stuart Kozola

September 29th, 2021

MathWorks®

Outline

Introduction to Responsible AI

Scaling model development and use in the era of AI

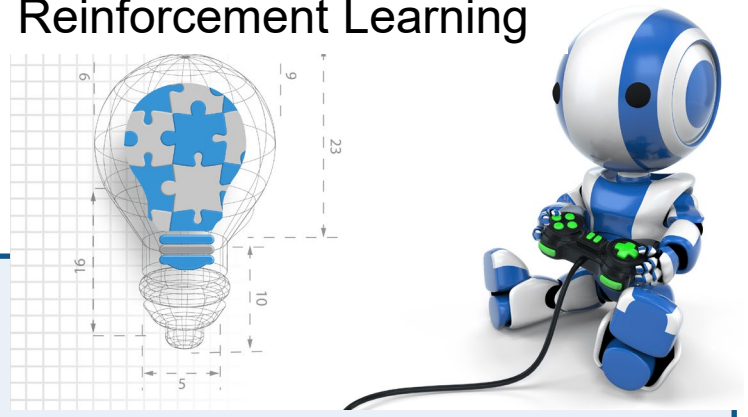
Best practices for building agile cross functional teams

Introduction to Responsible AI

Trusted AI

Reputational considerations and regulatory scrutiny are increasing the focus on “fairness” in the use of AI in Financial Services. This combines not just **explainability, but bias, ethics and transparency**, and is a focus area for **regulators, solution providers and consultancies**.

AI, Machine Learning and Deep Learning



Artificial Intelligence

- Reasoning
- Perception
- Knowledge Representation
- Machine Translation
- Computer Board Games
- Interactive Programs
- Expert Systems

Machine Learning

- Weather Forecasting
- Spam Detection
- Sentiment Analysis
- Recommender Systems
- Algorithmic Trading
- Fraud Detection
- Bioinformatics
- Medical Diagnosis
- Health Monitoring

Deep Learning

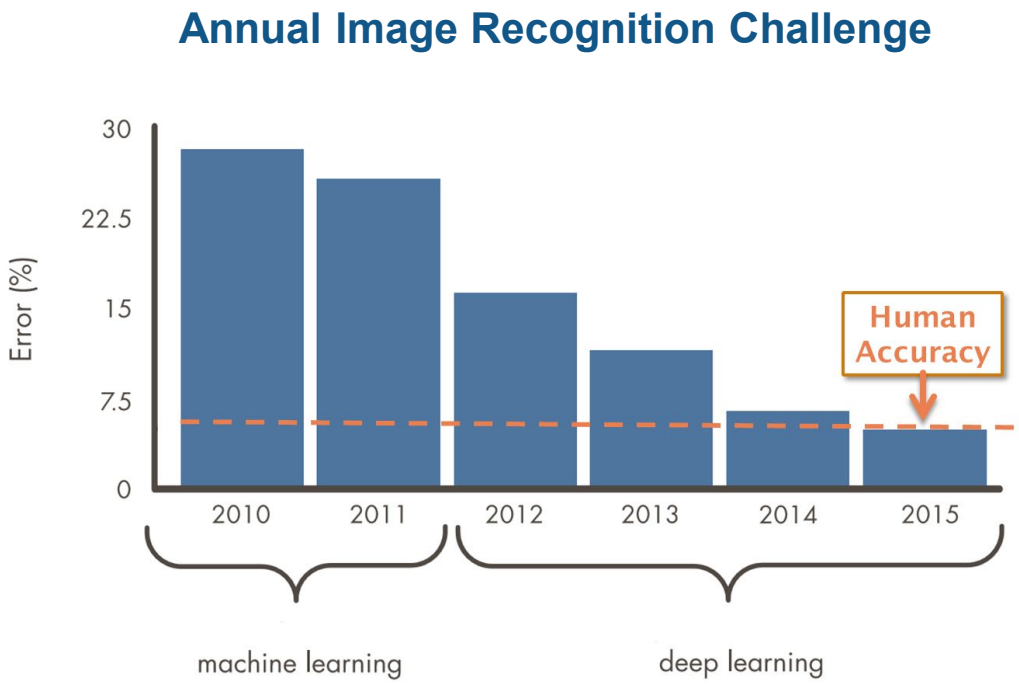
- Automated Driving
- Trading
- Object Recognition
- Robotics
- Speech Recognition

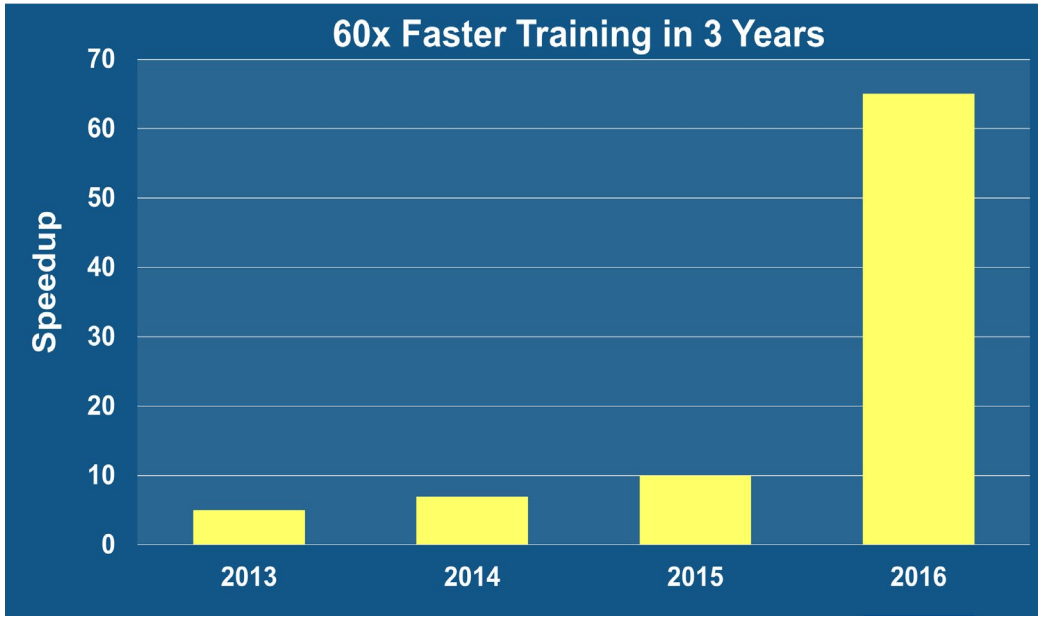
Application Breadth

1950s 1980s Today

Timeline

Enabling Technology: Deep Learning and GPUs





Source: ILSVRC Top-5 Error on ImageNet

Principles of fairness require explainability, bias, ethics, and transparency

2. Use of personal attributes as input factors for AIDA-driven decisions is **justified**.

4. AIDA-driven **decisions** are regularly **reviewed** so that models behave as designed and intended.

8. Firms using AIDA are **accountable** for both internally developed and externally sourced AIDA models.

13. Data subjects are provided, upon request, clear **explanations** on what data is used to make AIDA-driven decisions about the data subject and how the data affects the decision.

"AIDA" refers to artificial intelligence or data analytics, which are defined as technologies that assist or replace human decision-making.

Source: FEAT Principles, Monetary Authority of Singapore

RFIs and Evolution of Regulatory Guidance

PUBLISHED DOCUMENT

AGENCY:

Board of Governors of the Federal Reserve System, Bureau of Consumer Financial Protection, Federal Deposit Insurance Corporation, National Credit Union Administration, and Office of the Comptroller of the Currency (agencies).

ACTION:

Request for information and comment.

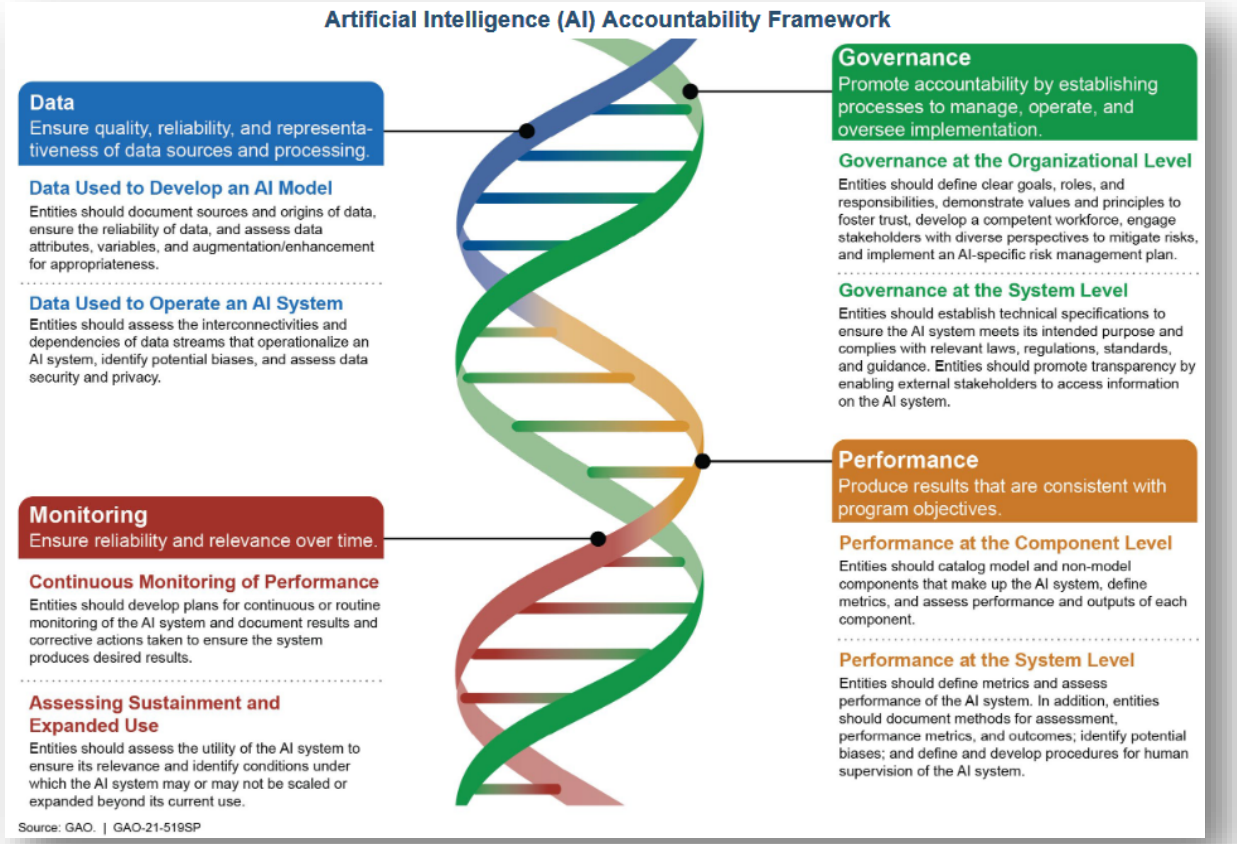
SUMMARY:

The agencies are gathering information and comments on financial institutions' use of artificial intelligence (AI), including machine learning (ML). The purpose of this request for information (RFI) is to understand respondents' views on the use of AI by financial institutions in their provision of services to customers and for other business or operational purposes; appropriate governance, risk management, and controls over AI; and any challenges in developing, adopting, and managing AI. The RFI also solicits respondents' views on the use of AI in financial services to assist in determining whether any clarifications from the agencies would be helpful for financial institutions' use of AI in a safe and sound manner and in compliance with applicable laws and regulations, including those related to consumer protection.

DATES:

Comments must be received by June 1, 2021.

Source: Federal Register, Board of Governors of the Federal Reserve System



U.S. Government Accountability Office

Questions to answer during AI Validation

Does the AI model or system produce consistent and reliable results?

Has the AI model or system been developed with safety as a key component?

Is the AI model or system immune from spoofing or other common attacks? Does it provide privacy protection?

Can you explain the workings of an AI model or system in human understandable terms?

Does the AI model or system make decisions or recommendations that are fair and free of bias?

Can you observe and trace cause and effect in AI model or system and explain the rationale of the decision?

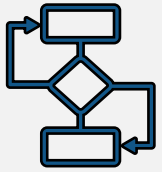
Common approaches used to identify answers

Explainable

- Model Re-specification/simplification
- Global Explanation
- Local Explanation
- Visual Explanation

Secure

- Federated Learning
- Differential Privacy
- Evasion, Poisoning, Extraction, Interference

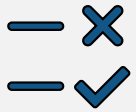


Interpretable

- Model Re-specification/simplification
- Quantitative Validation
- Global vs. Local

Robustness

- Adversarial Attack w/Coverage Metrics
- Reference/Benchmark Model
- Residual Deviation/Explanation



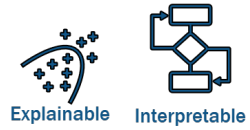
Fairness

- Disparate Error Analysis
- Adversarial Debiasing
- Feature Decomposition
- Reasoning
- Example Base

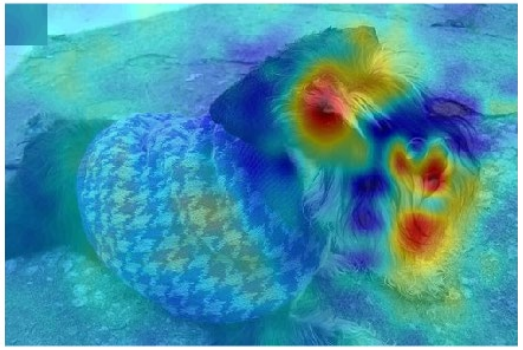
Safety

- Redundancy (with voting)
- Failure Mode Effect Analysis
- Monitoring
- Audit trail
- Best Practices

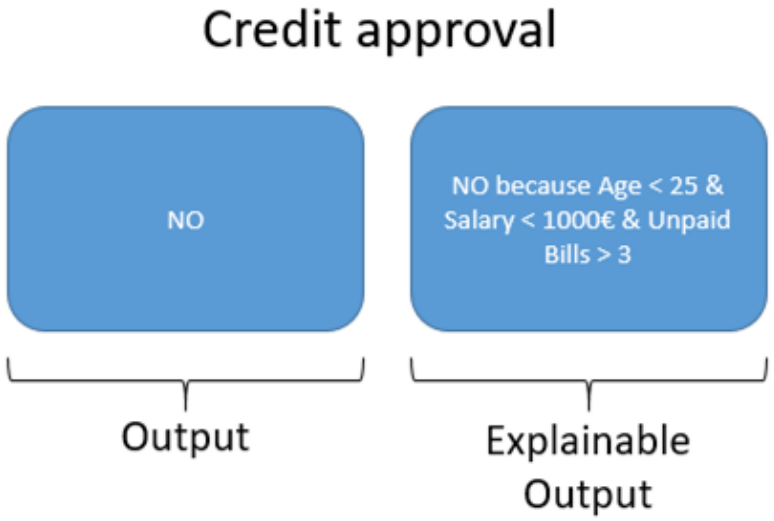
Explainability vs Interpretability



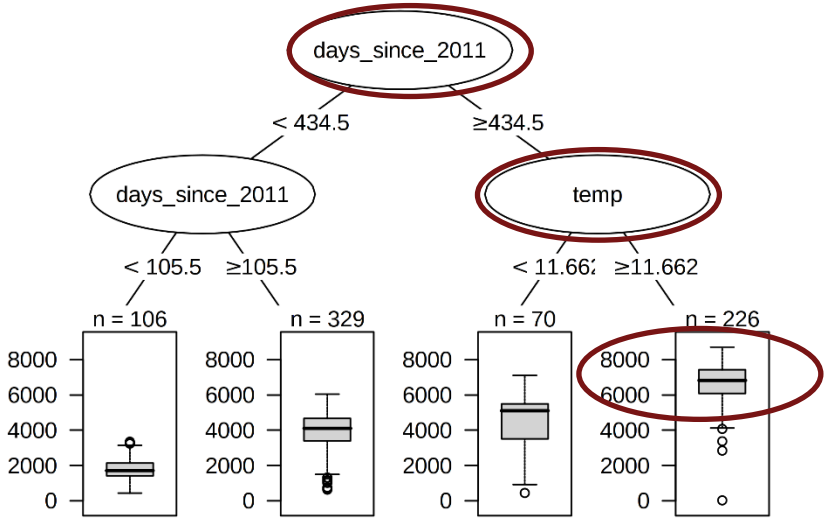
Explainability = “why is this happening?”



e.g., this is a picture of a “Schnauzer” because of the eyebrows and moustache.



Intrepretability = discern the mechanics without necessarily knowing why



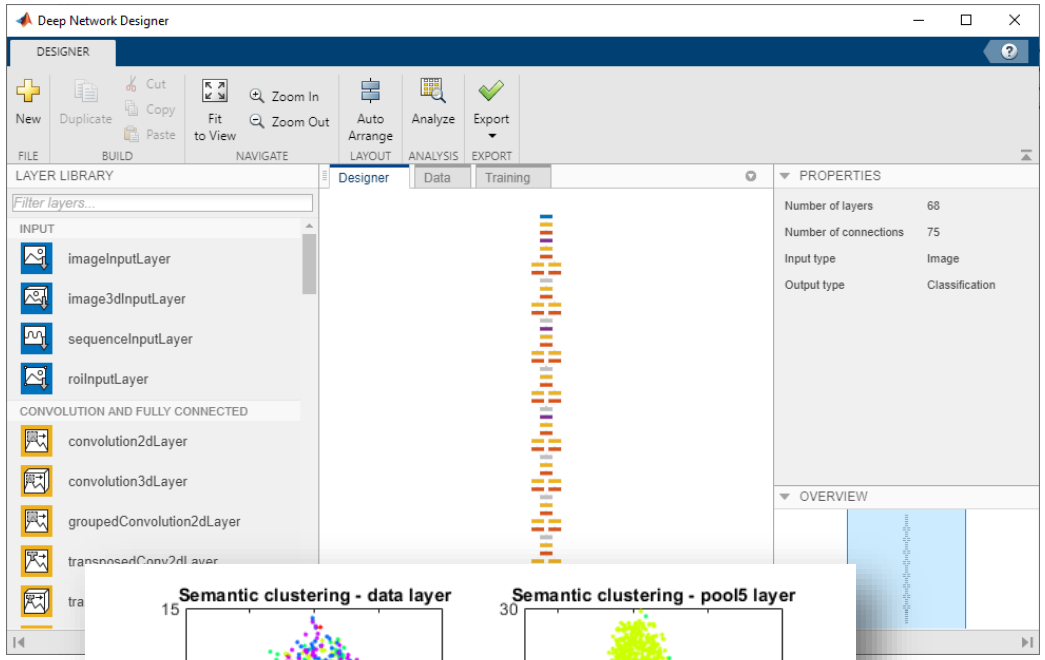
e.g., if it’s more that 434 days since 2011 and the temperature is more that 11.6 degrees, then 7000 bicycles will be rented

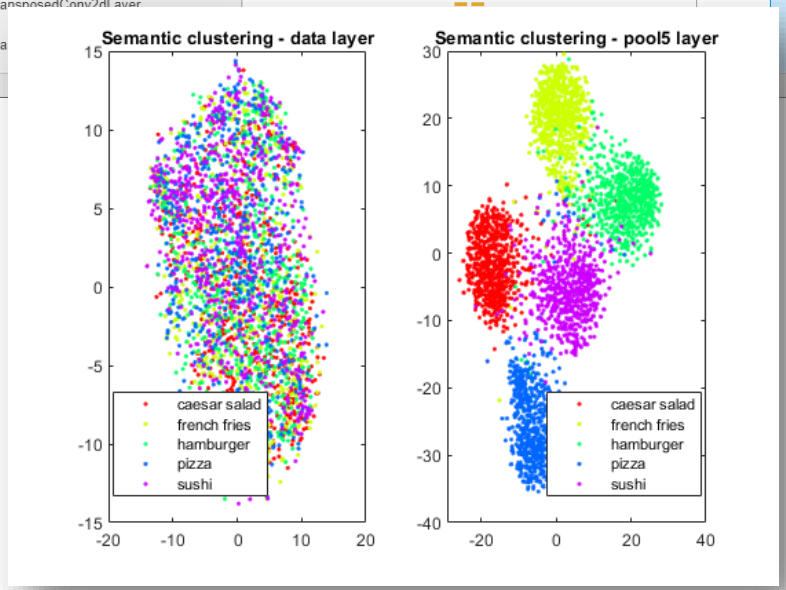
https://christophm.github.io/interpretable-ml-book/tree.html

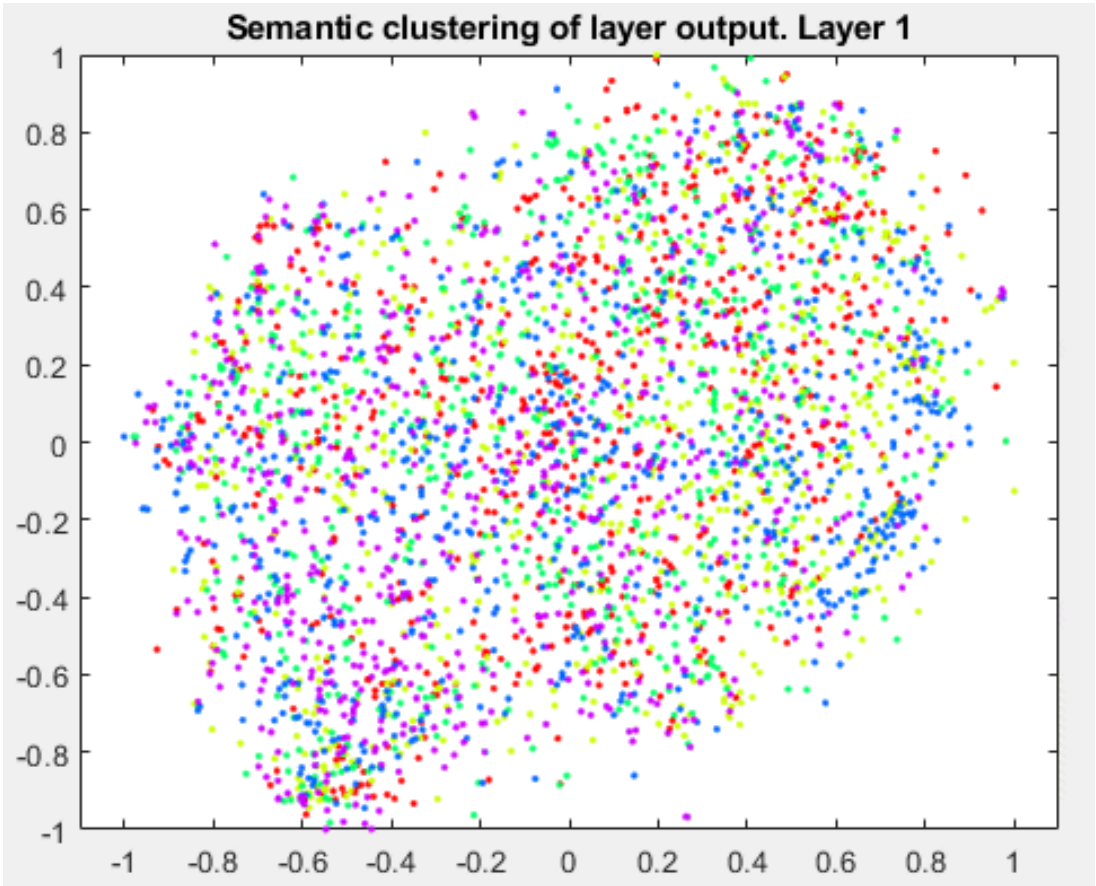
Visual Interpretation of Features Across Layers

Explainable

Interpretable







https://blogs.mathworks.com/deep-learning/2019/01/18/neural-network-feature-visualization/

Sources of Bias

Bias: Unequal treatment of different groups by an ML model

Fairness in Responsible AI: Detecting and mitigating bias against unprivileged groups in ML modeling

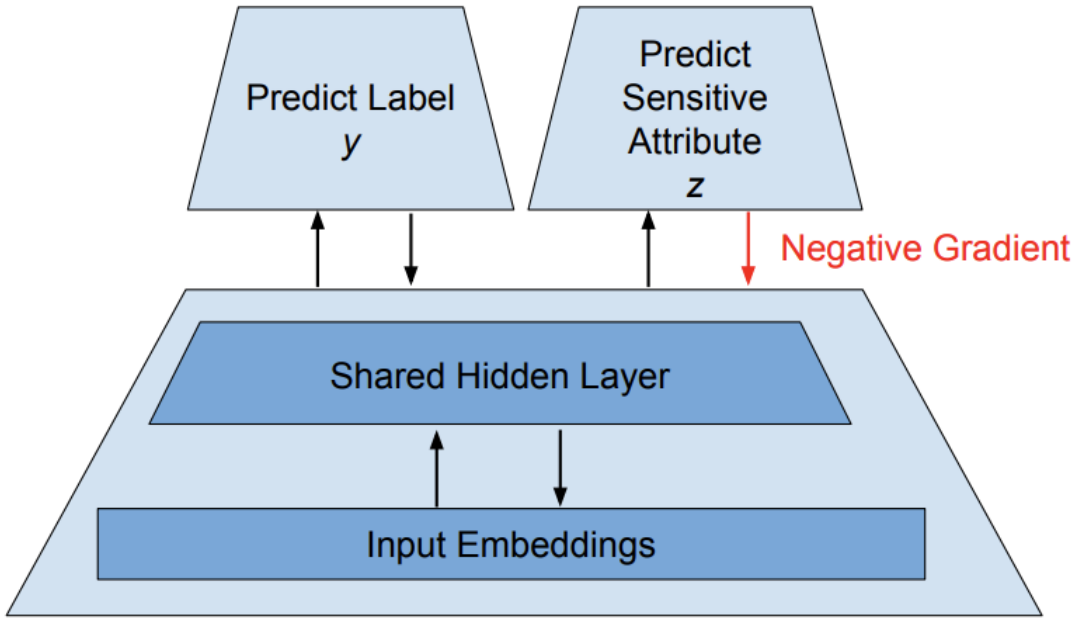
Not enough data, Bias through Selection

Legacy bias in the data, Bias through Behavior

Model issues

In-processing Bias Mitigation
Adversarial Debiasing

- Objective is to maximize the model's ability to predict Y, while minimizing the adversary's ability to predict Z
- One of the more effective methods of bias mitigation
- Can be optimized for any bias metric



Adversarial Examples

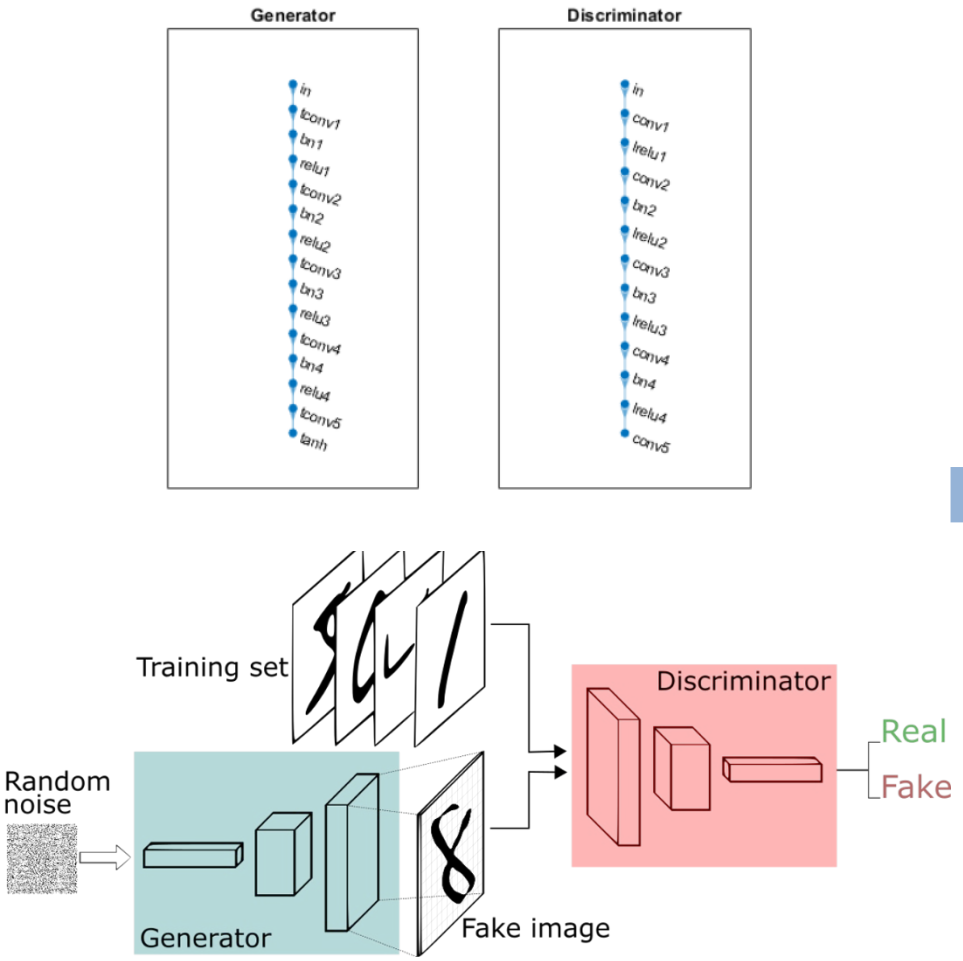
Robustness

Even rotating an image can cause it to be misclassified

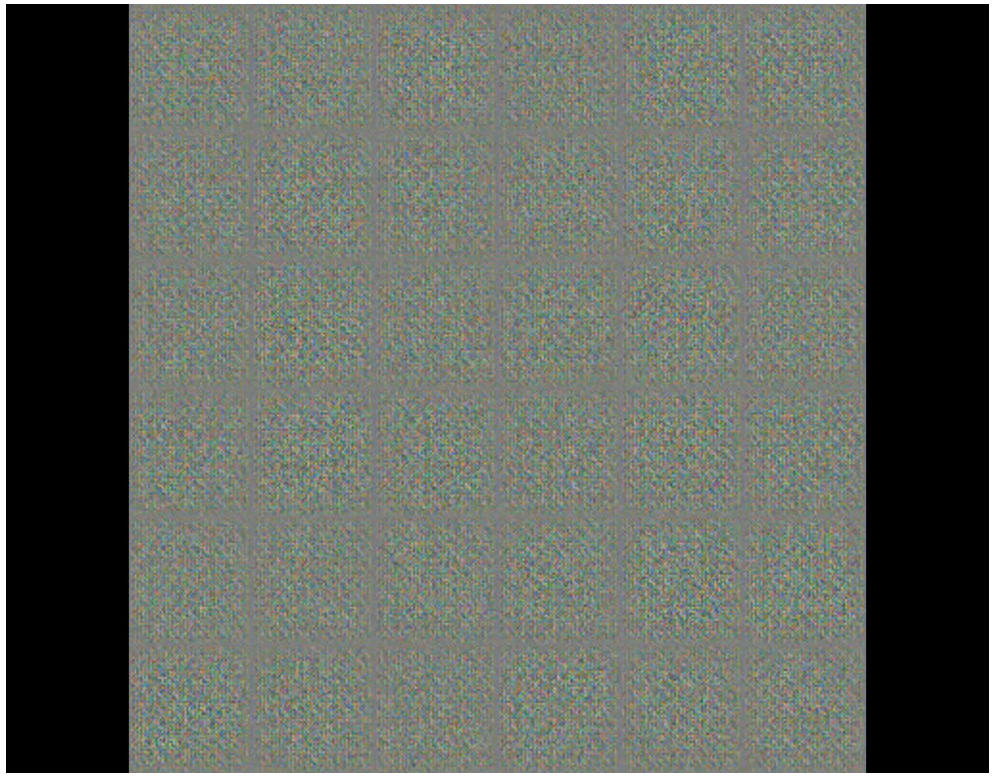
Original Image
Class: street sign

Adversarial Image (Rotation = -90 degrees)
Class: laptop

GAN Network that Creates Synthetic Data

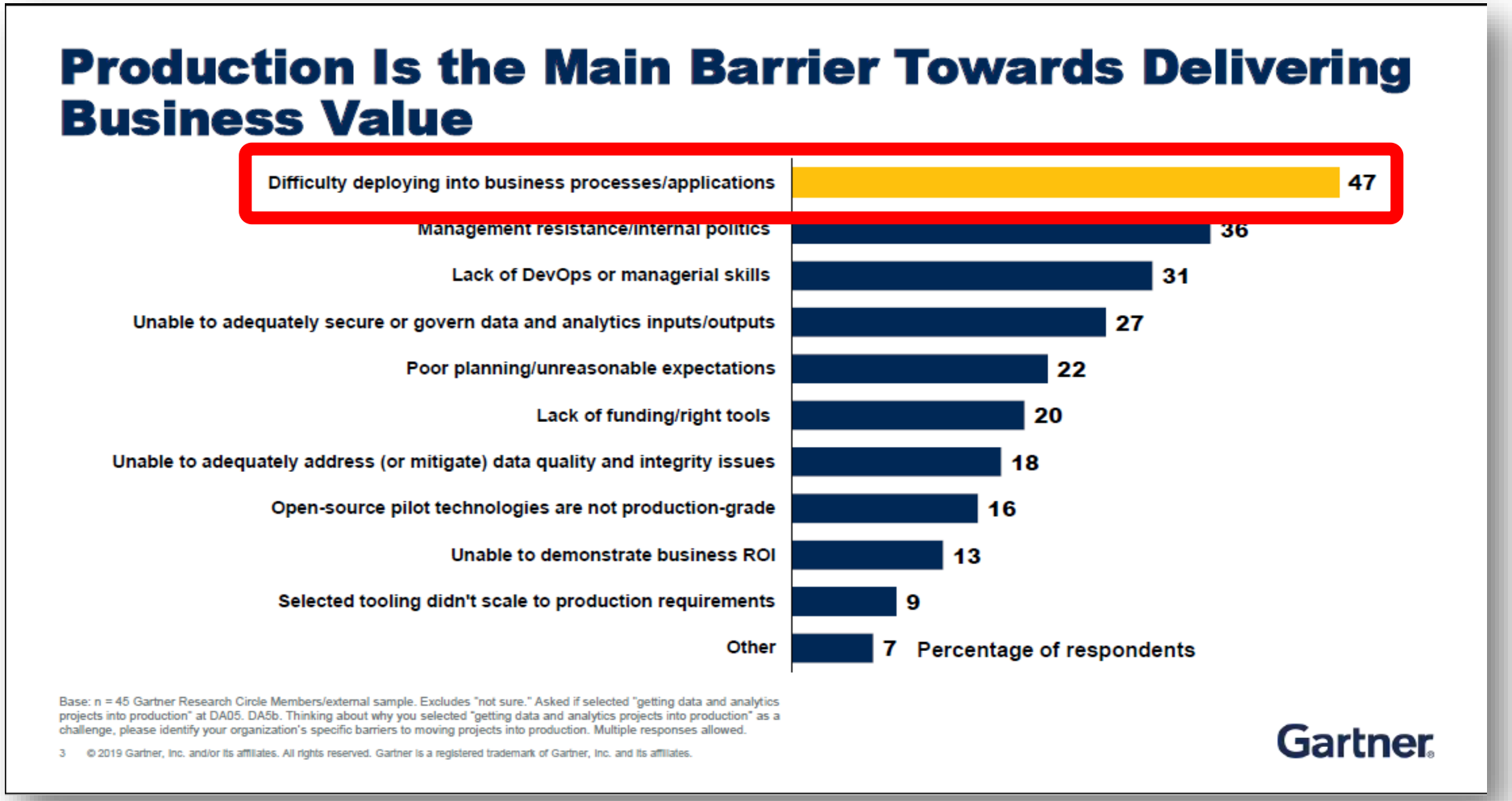


Network



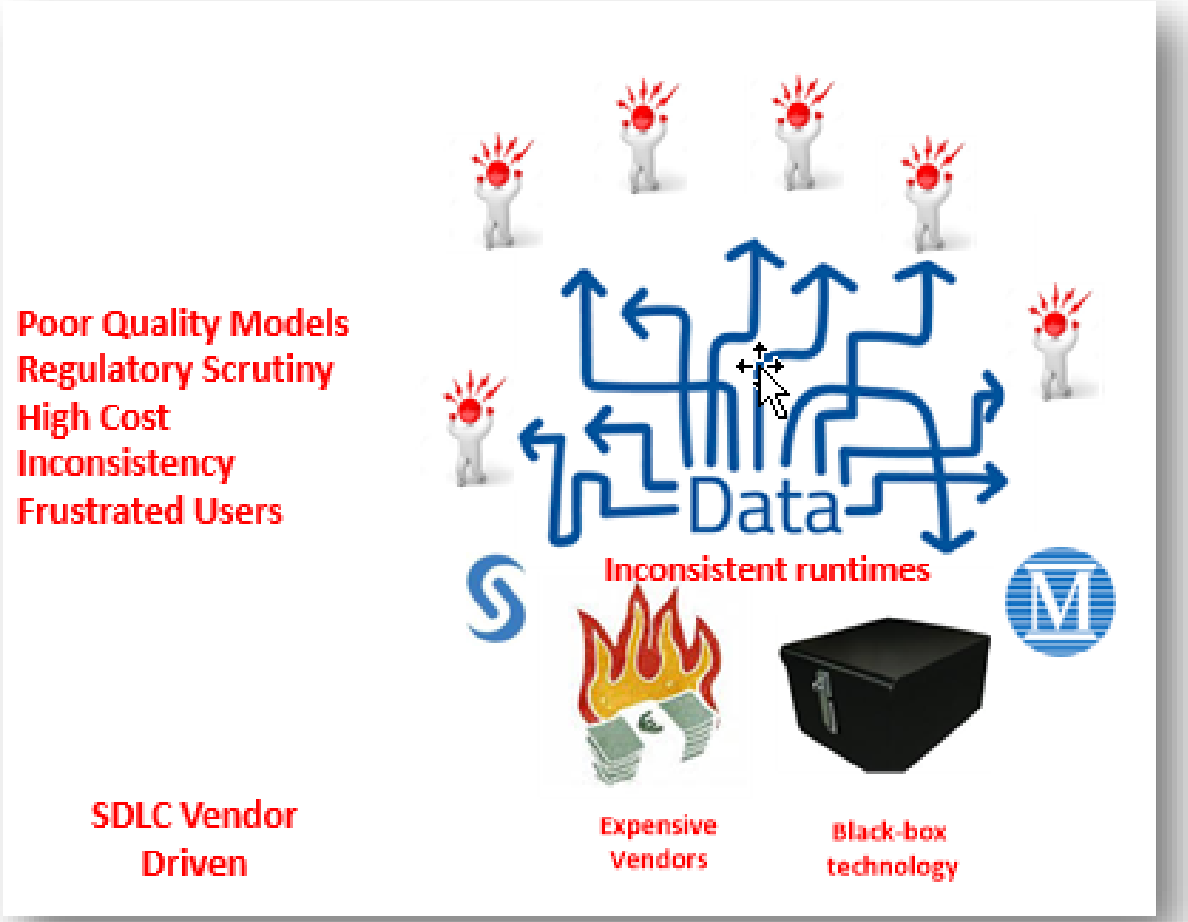
Scaling model development and use in the era of AI

Operationalizing Analytics, Modeling and Simulation is hard



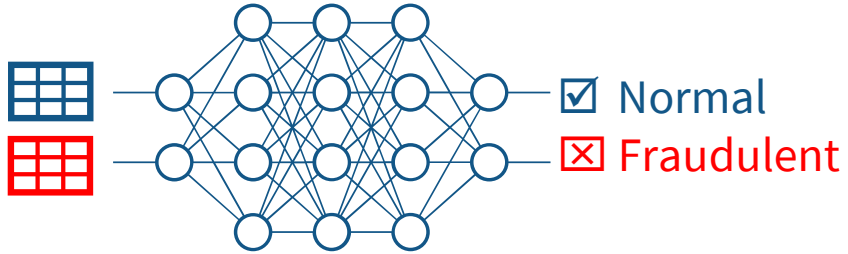
[1]Source: https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2019-and-beyond/

But the current environment is … complicated



Source: HSBC MATLAB Expo Presentation

Complex Models



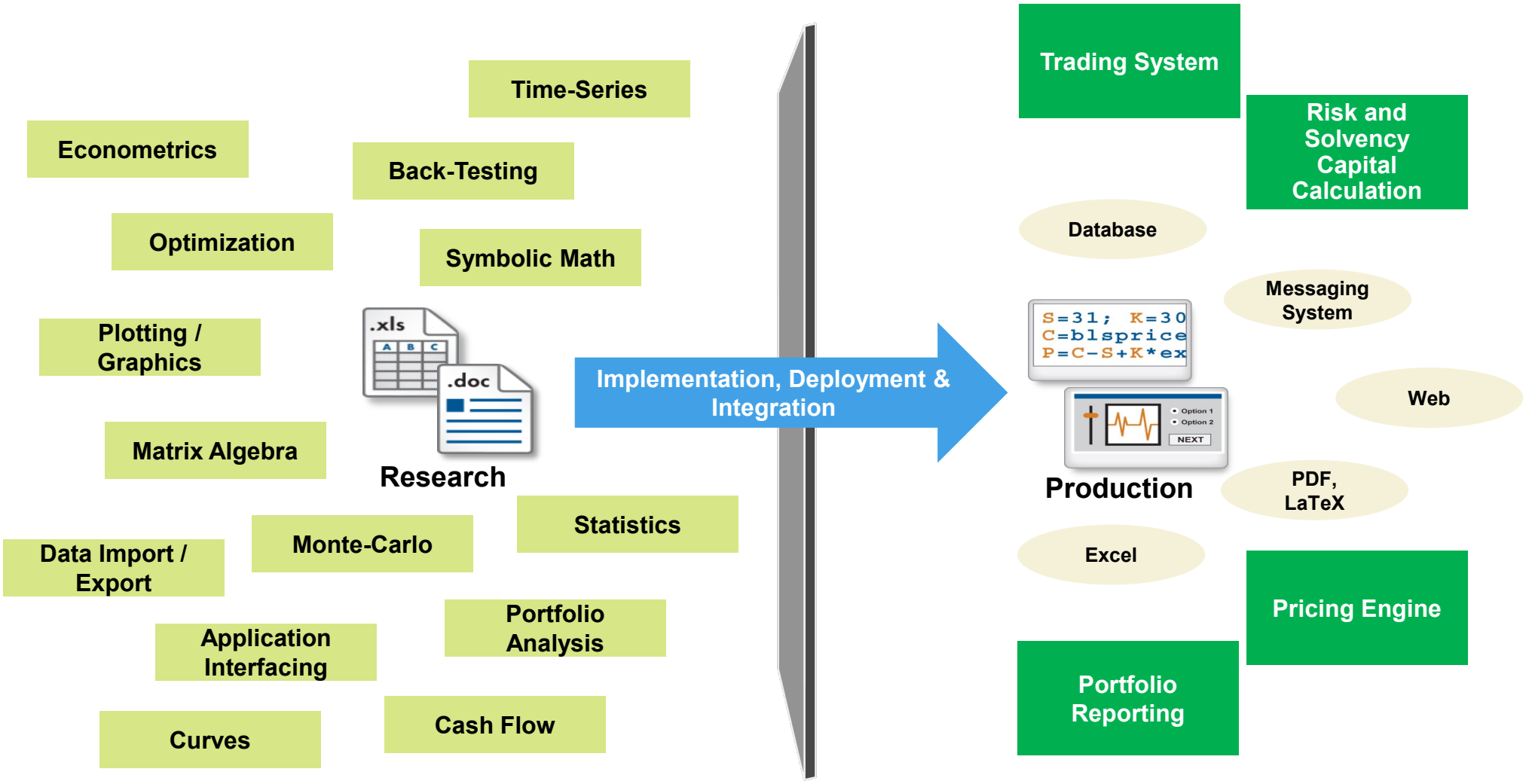
Disruption

Covid-19

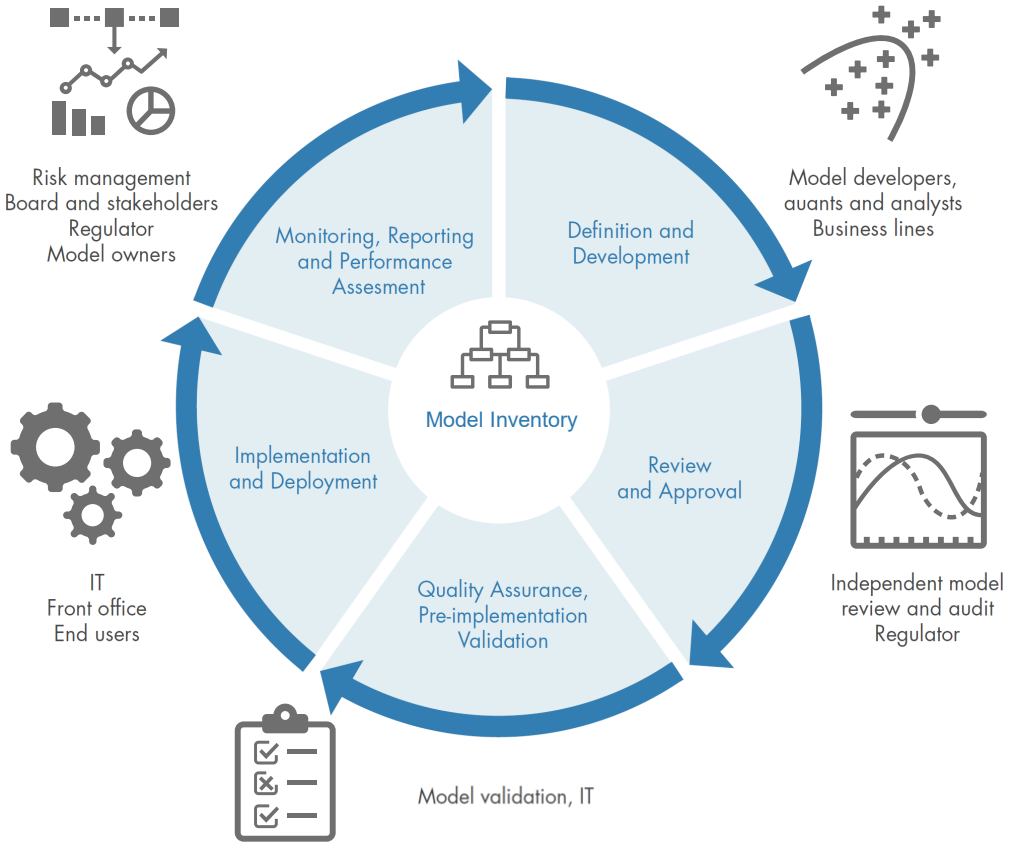
E S G

Climate change

Goal it to go from silos and walled gardens …



… to continuous model delivery



Model Inventory & Repository — MIR

- Centralized access to models, dependencies, meta-data, lineage, audit trail, risk scoring, and model risk reporting

Model Monitoring Dashboard — MMD

- Configure performance thresholds and alerts for breaches and generate reports
- Summarize model execution results using a customizable web dashboard
- Analyze the model usage to determine candidate models for retirement

Model Development Environment — MDE

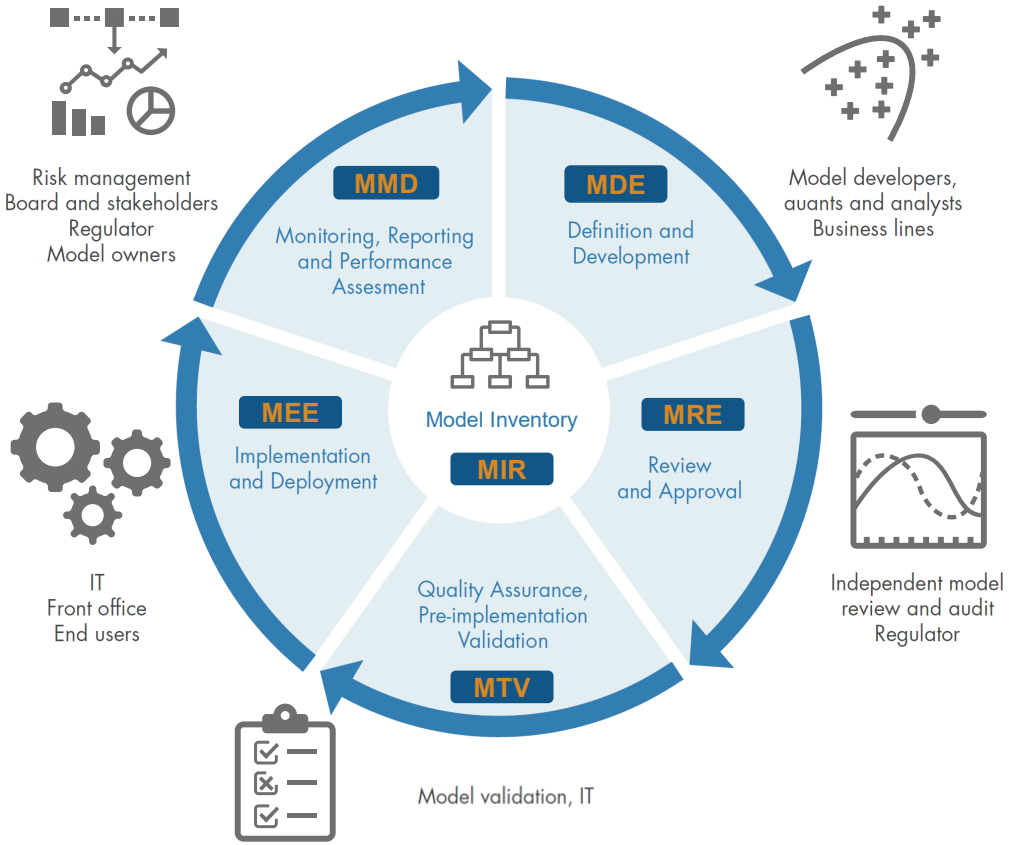
- Explore, develop, back-test, and document models and methodologies
- Improve transparency and reproducibility of model development process
- Create reusable model templates
- Auto-generate model documentation

Model Execution Environment — MEE

- Deploy models in production environment without recoding
- Integrate with existing technology infrastructures
- Host production models and scale to end users in a secure controlled environment “on-prem” or “cloud”

Model Review Environment — MRE

- Perform independent model reviews
- Perform interactive what-if and sensitivity analysis on model parameters
- Comment and flag various aspects for response and resolution



Model Test & Validation — MTV

- Automatically run unit tests and generate test reports
- Perform preproduction testing and validation for approved models
- Compare tests of preproduction model with a production model

Secure

Cloud

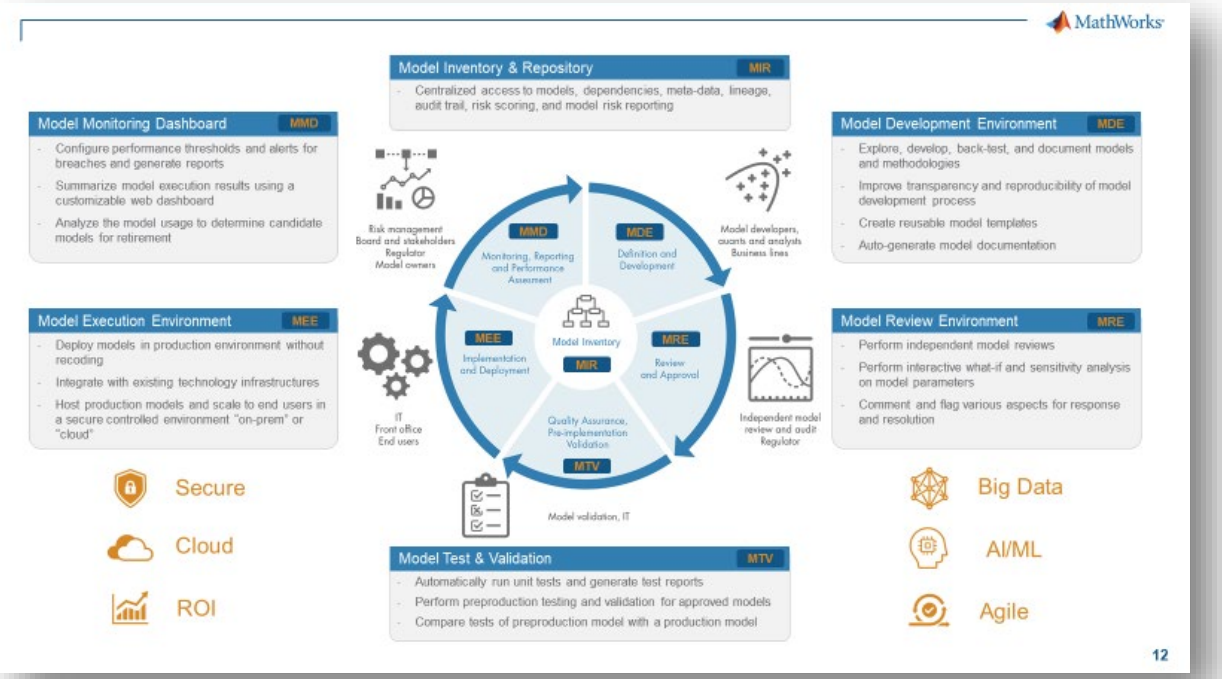
ROI

Big Data

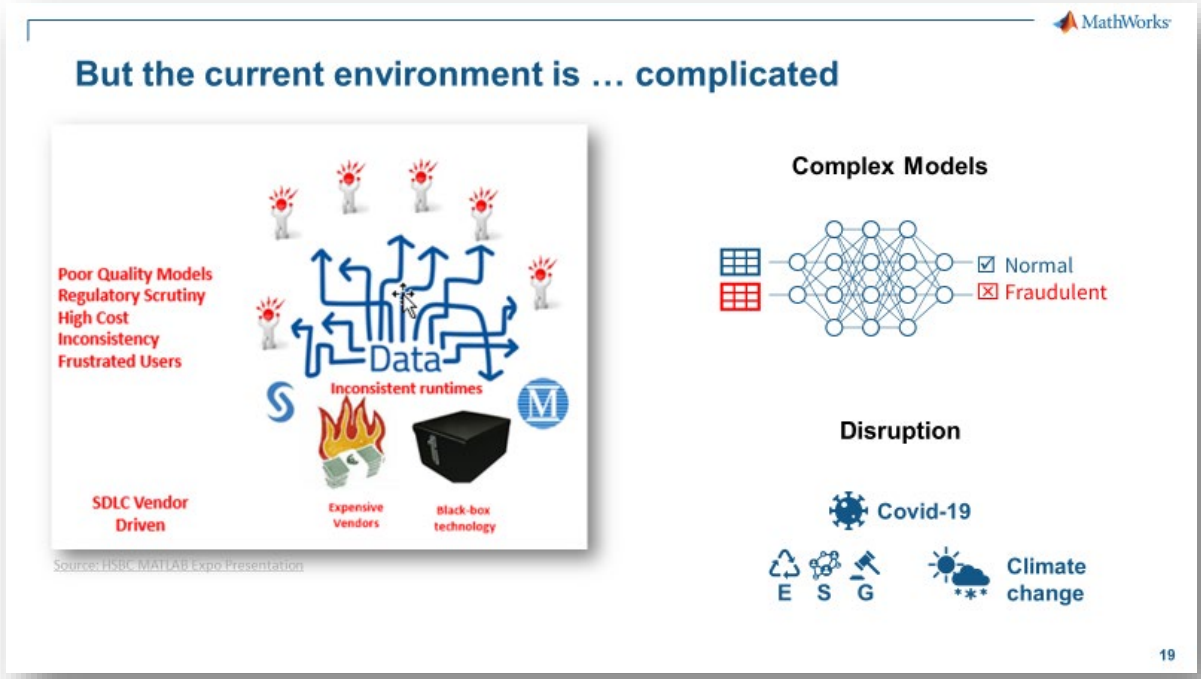
AI/ML

Agile

How do we get to here …



… from there?



Current challenges and desired state

	Current State	Pain	Desired State	Capability
 Velocity	Fragmented data sources Fragmented tools Inconsistent runtimes Vendor driven timelines Manual disjoint processes	1 **Unable to integrate and automate the Model Lifecyle**	Unified data access Integrated toolchain Choice of runtimes Extensible / customizable	OPEN & INTEROPERABLE
 ROI	Redundant data / systems Expensive vendors Labor intensive processes Compliance / auditing Flexible compute resources	2 **High cost reduces investments in the business**	Connected data systems Choice of technology solutions Plug-n-play, low configuration Streamlined documentation & reporting Leverage on-prem and cloud	SCALABLE
 Quality	Silos: data + tools + teams Black-Box tools Lack of documentation Escaped defects in models Inappropriate model use	3 **Limited visibility + traceability of model quality across the Lifecyle**	Collaborative agile teams Open and automatable Always up-to-date reporting/docs High quality models Full model lineage	MAXIMIZE AUTOMATION
 Risk	Regulatory scrutiny Limited model reproducibility Lost opportunities Loss exposure	4 **Lack of timely access to risk metrics for decision making**	Set the bar Full model reproducibility Experiment and innovate more Minimized loss potential	FAST FEEDBACK

Minimum Requirements for Success

Open & Interoperable

- ✓ Integrates with structured/unstructured, historical and live data sources and big data systems
- ✓ Integrates models across languages, desktop tools, and runtimes
- ✓ Extensible: Proprietary, commercial, and community add-ons
- ✓ Enables use by novices and expert users

Scalable

- ✓ Integrated documentation and reporting
- ✓ Horizontal and vertical compute scaling
- ✓ Integrates with 3rd party platforms and applications
- ✓ Desktop to on-prem compute to public/private/hybrid cloud architectures

Maximize Automation

- ✓ Always up-to-date models and documents
- ✓ Automated testing and validation
- ✓ Reproducible model experimentation
- ✓ Automate cross-team workflows

Fast Feedback

- ✓ Real-time customizable dashboards and reports
- ✓ Full model traceability across Lifecyle
- ✓ Unlock experimentation and idea exploration by all users
- ✓ Monitoring and alerting across the Lifecyle

Minimum Requirements for Success

Open & Interoperable

- ✓ Integrates with structured/unstructured, historical and live data sources and big data systems
- ✓ Integrates models across languages, desktop tools, and runtimes
- ✓ Extensible: Proprietary, commercial, and community add-ons
- ✓ **Enables use by novices and expert users**

Scalable

- ✓ **Integrated documentation and reporting**
- ✓ Horizontal and vertical compute scaling
- ✓ Integrates with 3rd party platforms and applications
- ✓ Desktop to on-prem compute to public/private/hybrid cloud architectures

Maximize Automation

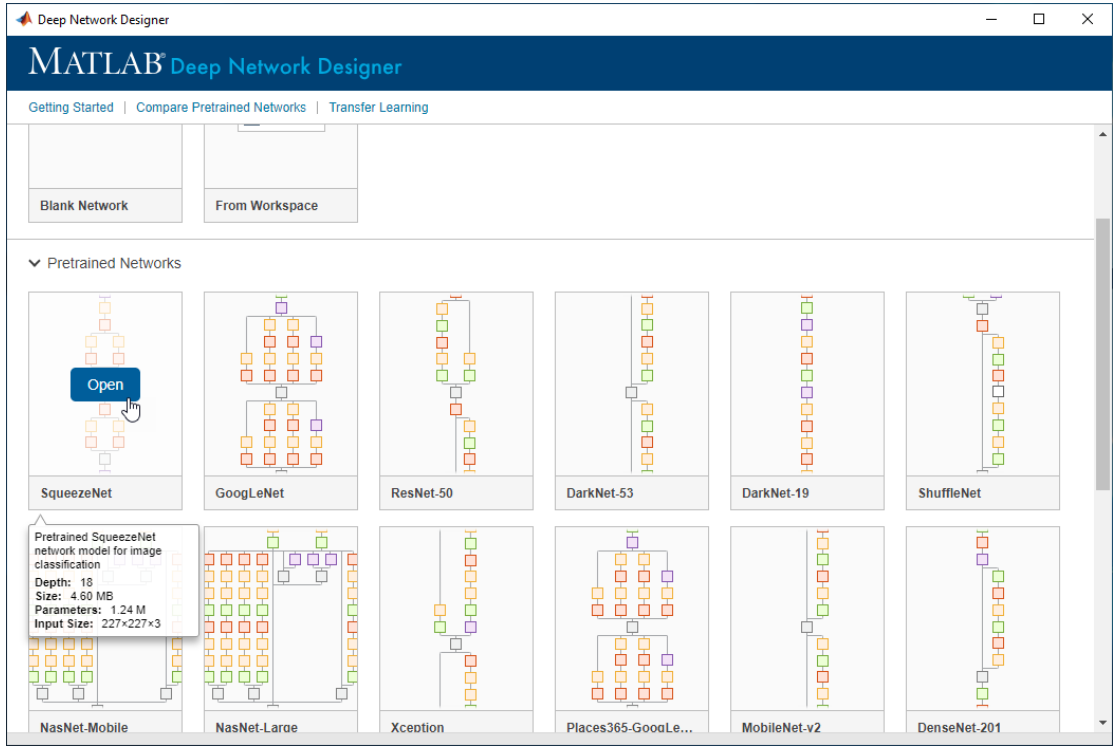
- ✓ **Always up-to-date models and documents**
- ✓ Automated testing and validation
- ✓ Reproducible model experimentation
- ✓ Automate cross-team workflows

Fast Feedback

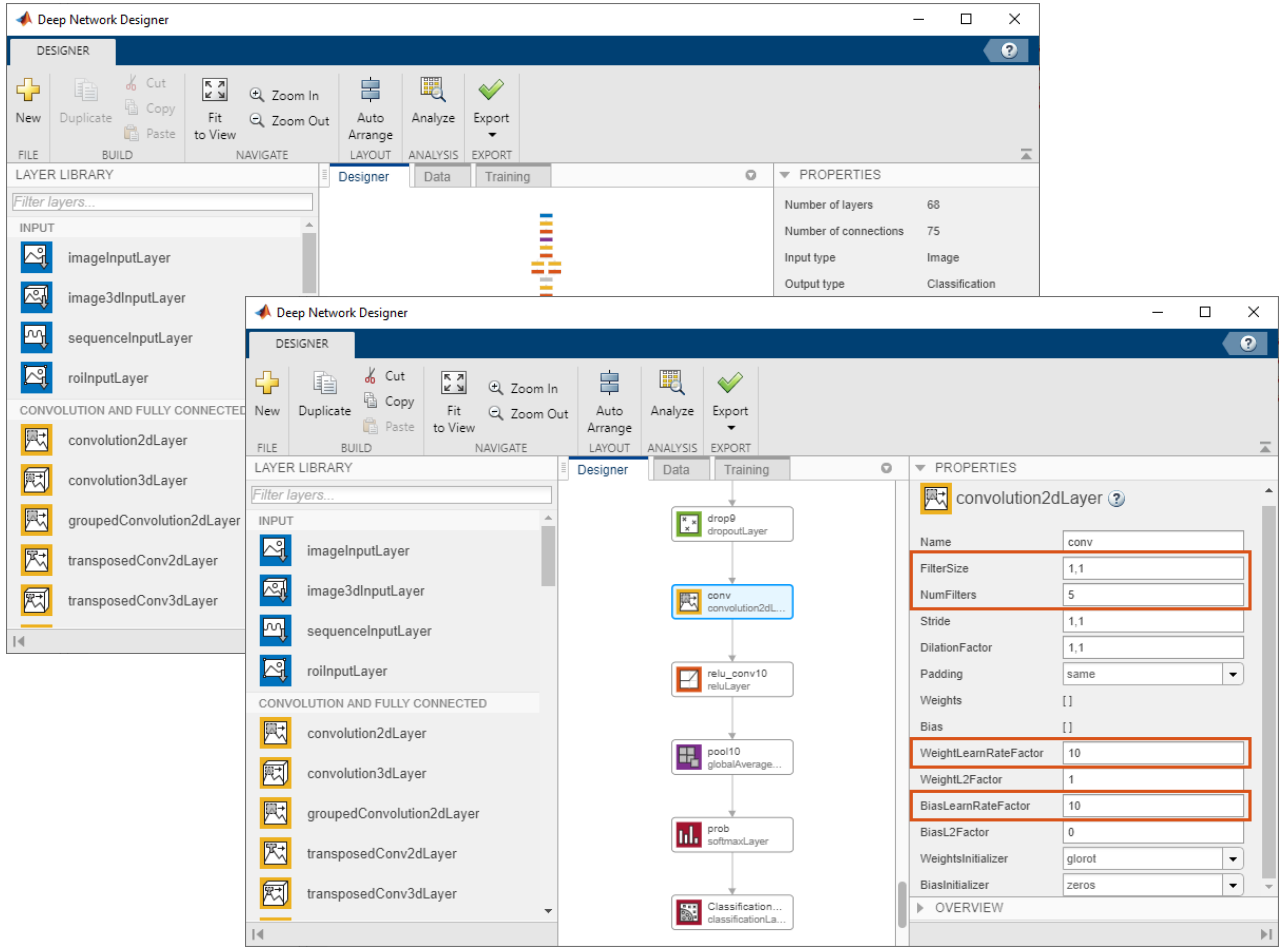
- ✓ Real-time customizable dashboards and reports
- ✓ Full model traceability across Lifecyle
- ✓ **Unlock experimentation and idea exploration by all users**
- ✓ Monitoring and alerting across the Lifecyle

Enable use by novices and expert users

OPEN & INTEROPERABLE

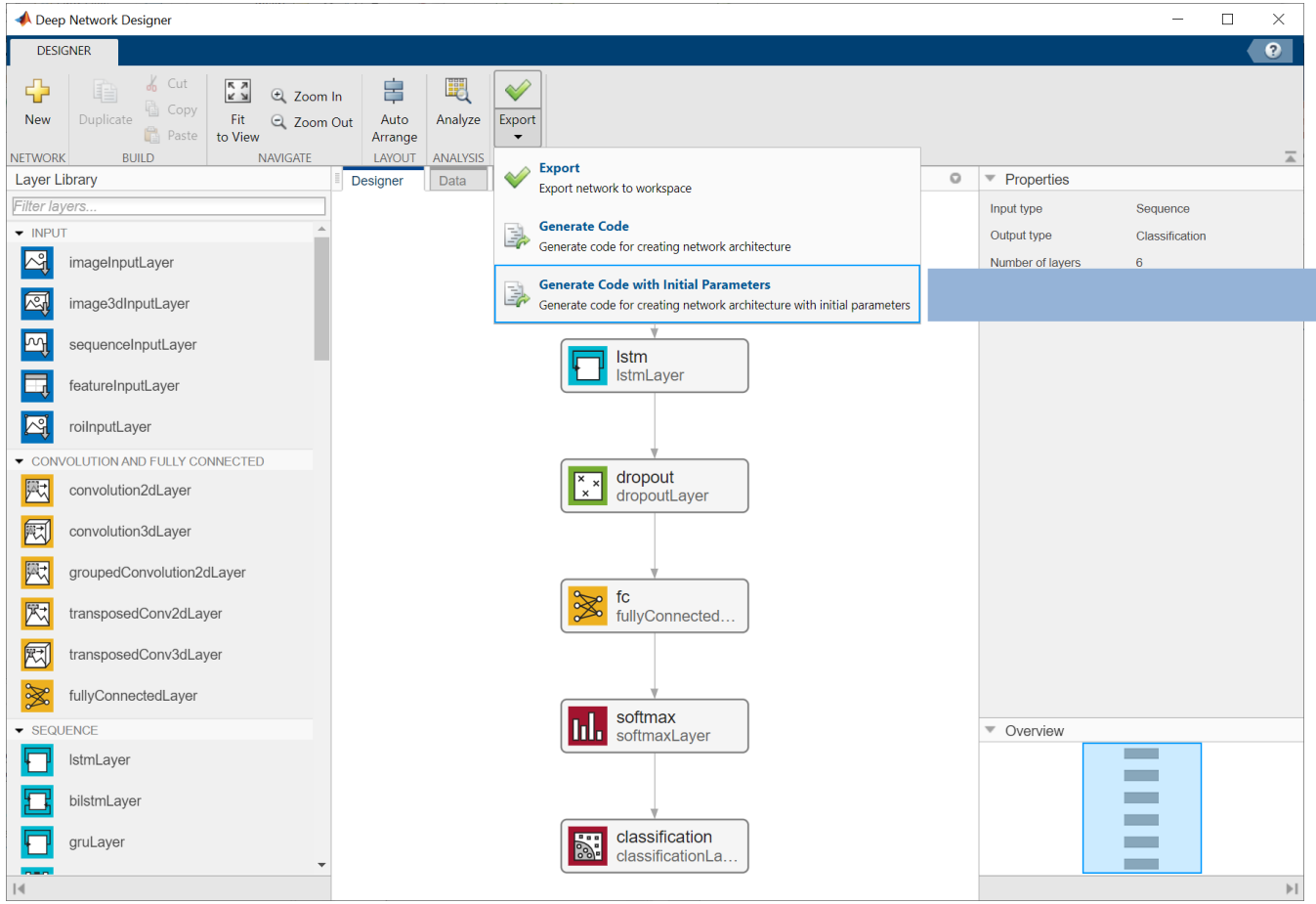


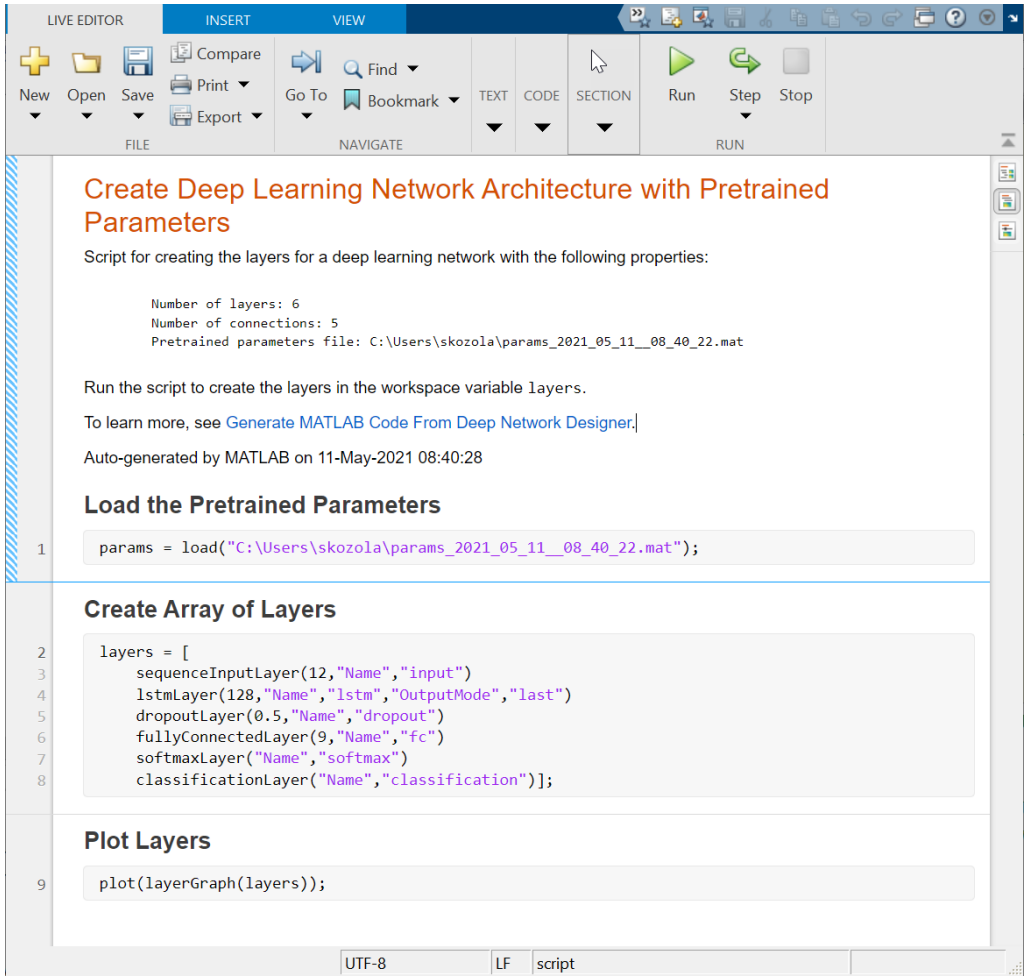
Library of pre-trained models (enables transfer learning)



Design of new or customize existing networks

Code Generation for Automation/Expert tuning





Integrated documentation and reporting

SCALABLE

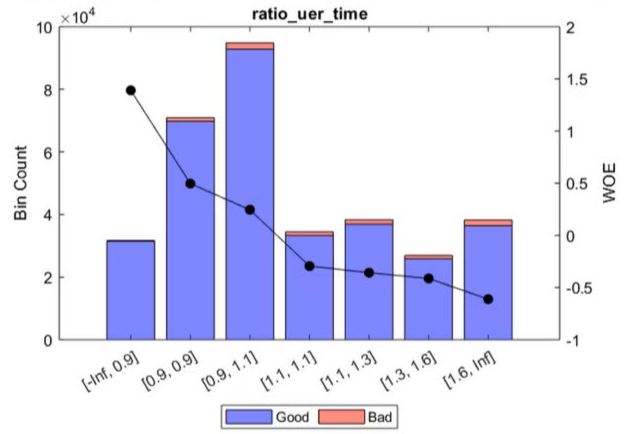
MortgageAnalysis.mlx

Visualize the Binning Results:

View the results of the binning. Here, 'Bad' and 'Good' represent those customers who have or have not defaulted on their mortgage respectively. Choose a predictor variable from the drop down menu to view the WOE binning results. To prevent plot creation, set the plotFlag input to `exploreWOE` (third input) to `false`.

```
28 predictor = 'ratio_uer_time';
29 figure
30 bInfo = displayWOE(binnedScorecard, predictor)
```

```
[INFO WOEBinningInfo 2018-10-17 07:01:44.840 ppeeling] Binning information for ratio_uer_time
```



bInfo = 8×7 table

	ratio_ue...	Good	Bad	Odds	WOE	InfoValue	Percent...
1	"[-Inf, 0.9]"	31503	210	150.01	1.3903	0.1007	9.4561
2	"[0.9, 0.9]"	69789	1137	61.38	0.49663	0.041535	21.149
3	"[0.9, 1.1]"	92851	1944	47.763	0.2458	0.015227	28.266
4	"[1.1, 1.1]"	33287	1196	27.832	-0.29427	0.010261	10.282
5	"[1.1, 1.3]"	36898	1413	26.113	-0.35801	0.01741	11.423
6	"[1.3, 1.6]"	25876	1047	24.714	-0.41306	0.016736	8.0278
7	"[1.6, Inf]"	36423	1797	20.269	-0.61137	0.057469	11.396

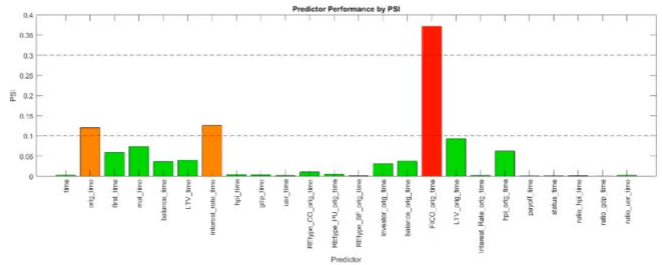
MortgageAnalysis.mlx

Population Stability Index Thresholds

```
62 Thresholds.psiLowTh =  0.1 ;
63 Thresholds.psiHighTh =  0.3 ;
```

Visualize the PSI thresholds:

```
64 plotPSIPredictorSelection(Metrics, Thresholds, predictorNames)
```

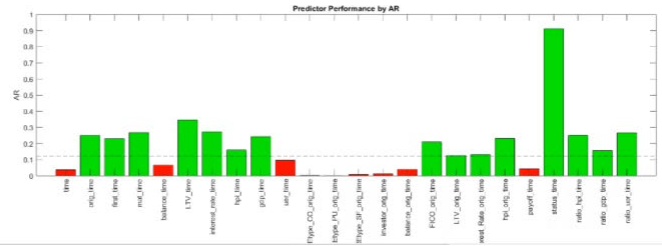


Accuracy Ratio Thresholds

```
65 Thresholds.arTh =  0.12 ;
```

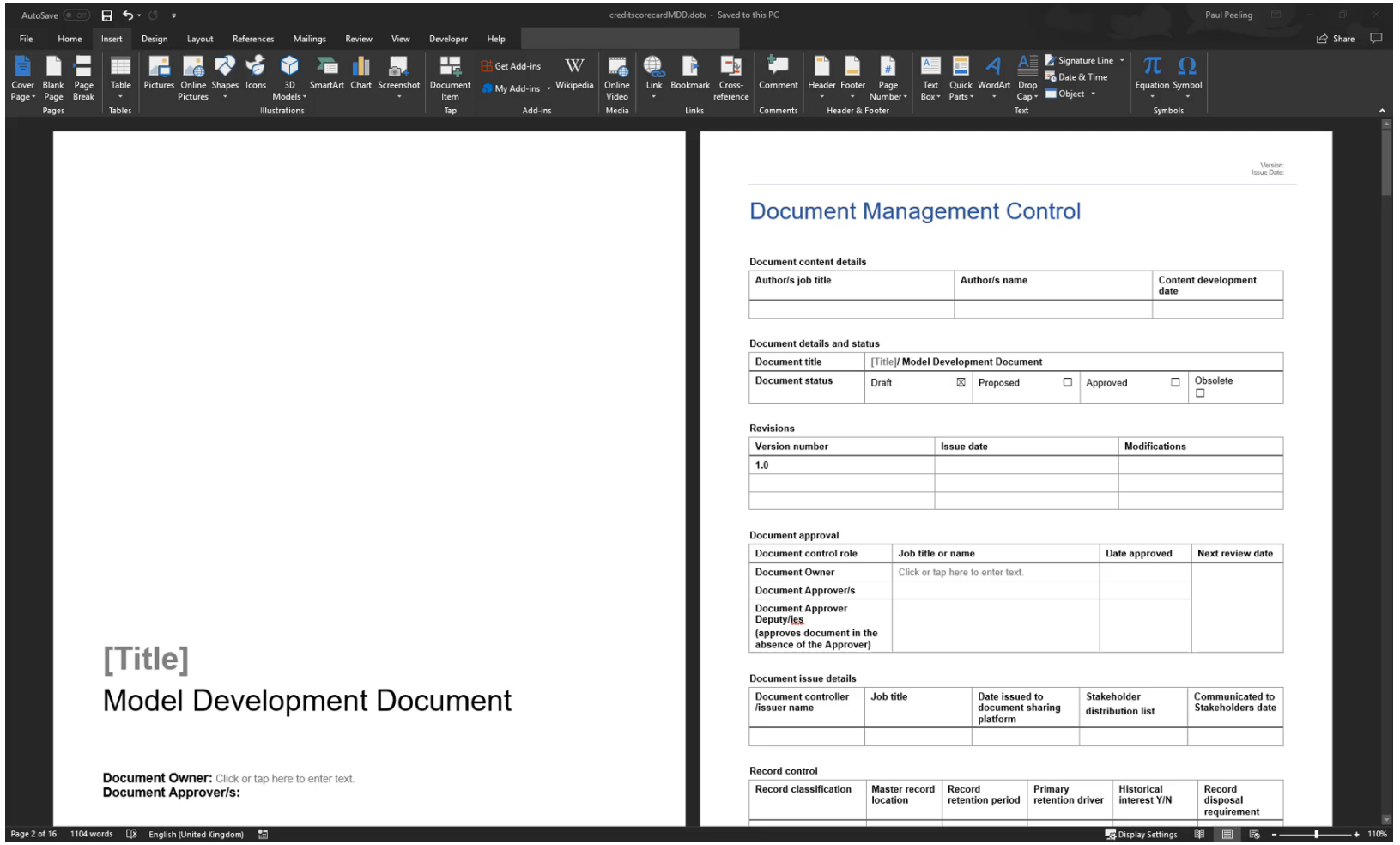
Visualize the AR thresholds:

```
66 plotARPredictorSelection(Metrics, Thresholds, predictorNames)
```



Always up-to-date models and documentation

MAXIMIZE AUTOMATION



[Title]

Model Development Document

Document Owner: Click or tap here to enter text.
Document Approver/s:

Version:
Issue Date:

Document Management Control

Document content details

Author/s job title	Author/s name	Content development date

Document details and status

Document title	[Title]/ Model Development Document			
Document status	Draft ☒	Proposed ☐	Approved ☐	Obsolete ☐

Revisions

Version number	Issue date	Modifications
1.0		

Document approval

Document control role	Job title or name	Date approved	Next review date
Document Owner	Click or tap here to enter text.		
Document Approver/s			
Document Approver Deputy/ies (approves document in the absence of the Approver)			

Document issue details

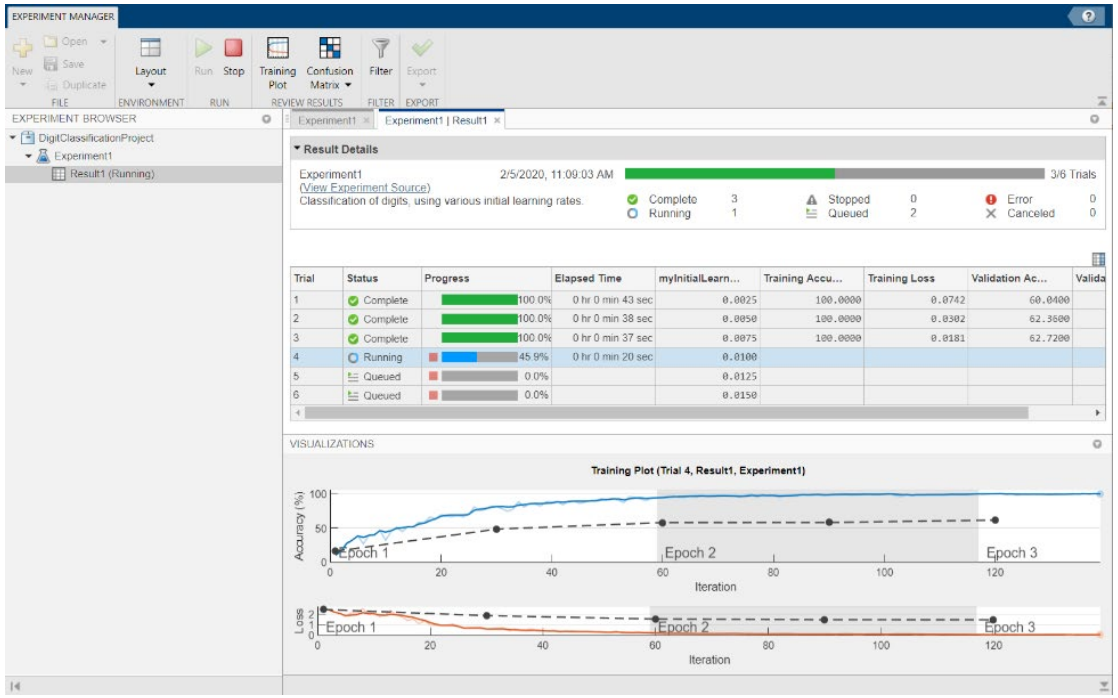
Document controller /issuer name	Job title	Date issued to document sharing platform	Stakeholder distribution list	Communicated to Stakeholders date

Record control

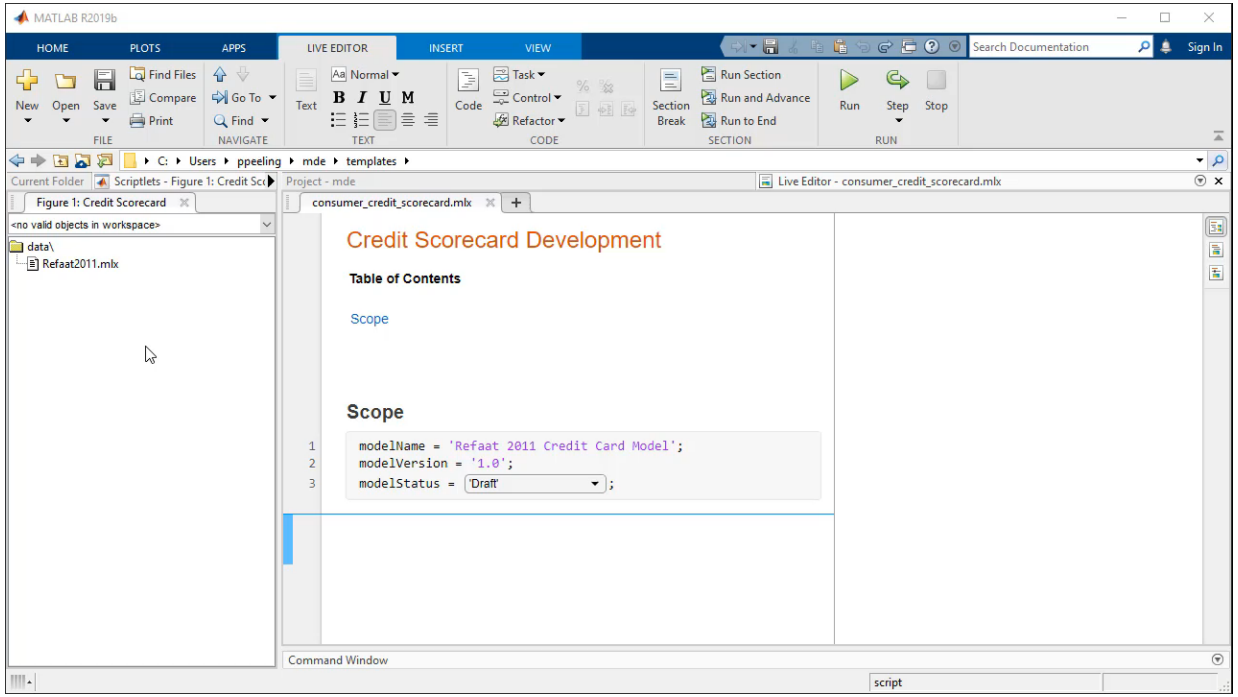
Record classification	Master record location	Record retention period	Primary retention driver	Historical interest Y/N	Record disposal requirement

Unlock experimentation and idea exploration for all users

FAST FEEDBACK



Experiment Manager



Templated model development

Model Inventory & Repository — MIR

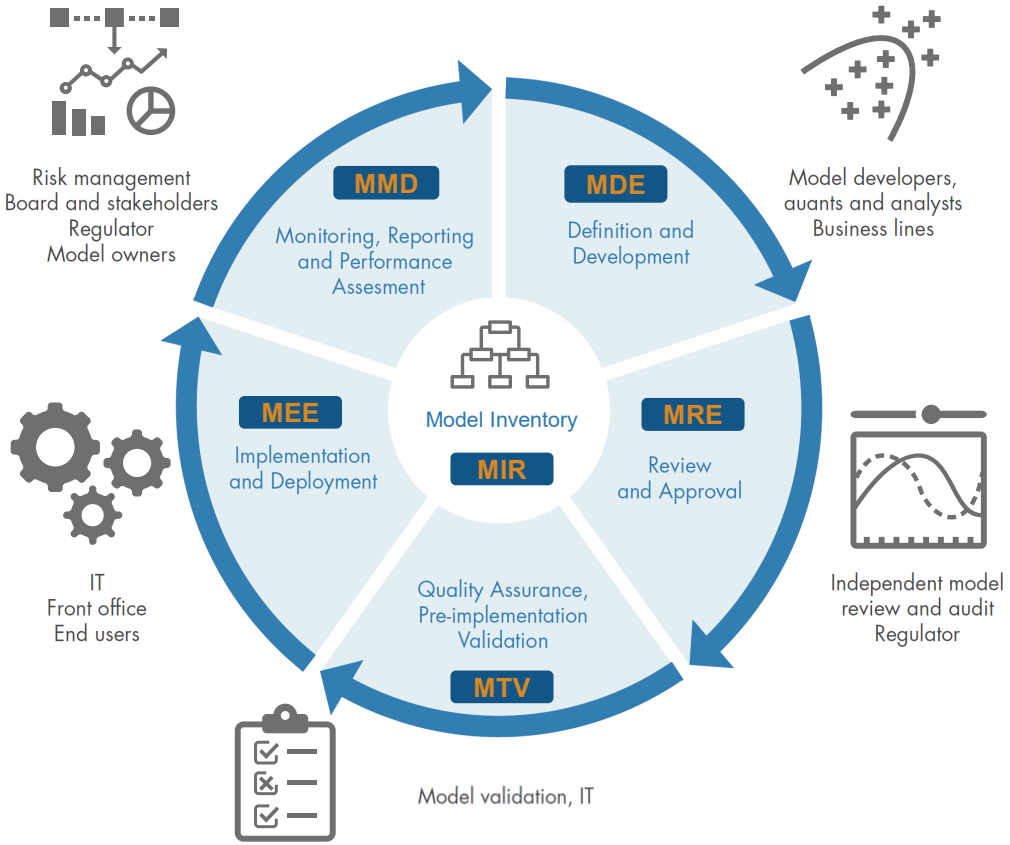
- Centralized access to models, dependencies, meta-data, lineage, audit trail, risk scoring, and model risk reporting

Model Monitoring Dashboard — MMD

- Configure performance thresholds and alerts for breaches and generate reports
- Summarize model execution results using a customizable web dashboard
- Analyze the model usage to determine candidate models for retirement

Model Development Environment — MDE

- Explore, develop, back-test, and document models and methodologies
- Improve transparency and reproducibility of model development process
- Create reusable model templates
- Auto-generate model documentation



Model Execution Environment — MEE

- Deploy models in production environment without recoding
- Integrate with existing technology infrastructures
- Host production models and scale to end users in a secure controlled environment “on-prem” or “cloud”

Model Review Environment — MRE

- Perform independent model reviews
- Perform interactive what-if and sensitivity analysis on model parameters
- Comment and flag various aspects for response and resolution

Secure

Cloud

ROI

Big Data

AI/ML

Agile

Model Test & Validation — MTV

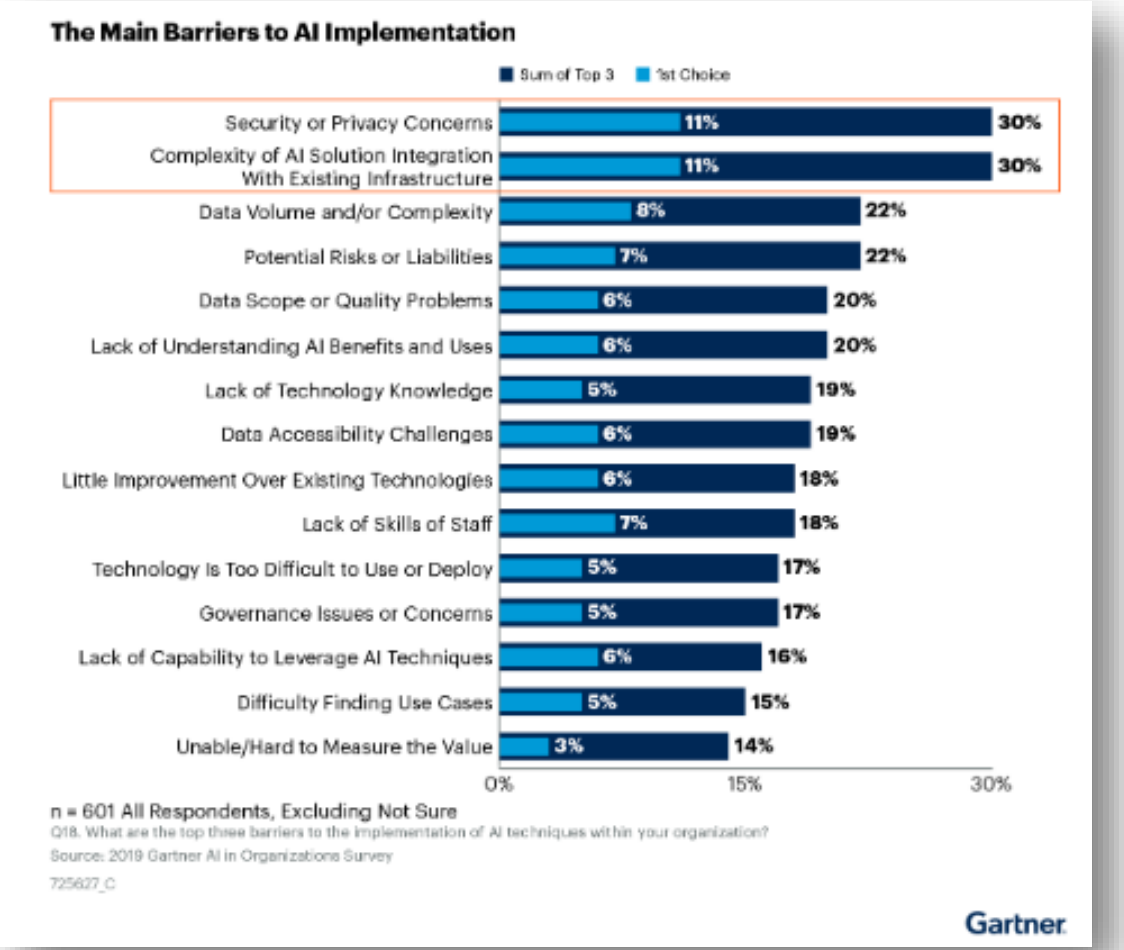
- Automatically run unit tests and generate test reports
- Perform preproduction testing and validation for approved models
- Compare tests of preproduction model with a production model

Best practices for building agile cross functional teams

How successful is your team at operationalizing AI?

"Approximately half of all AI models never make it into production due to lack of ModelOps"

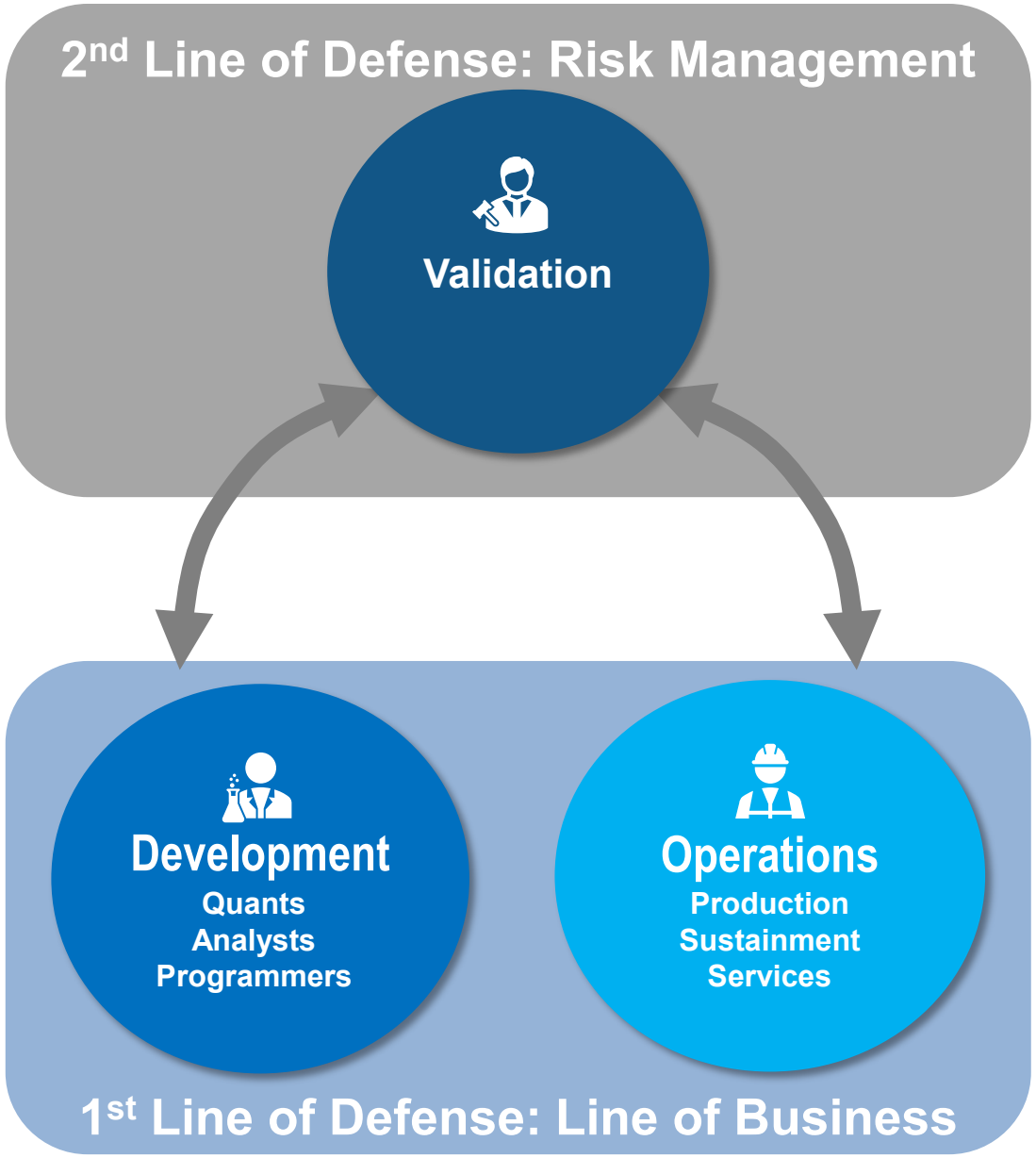
Gartner



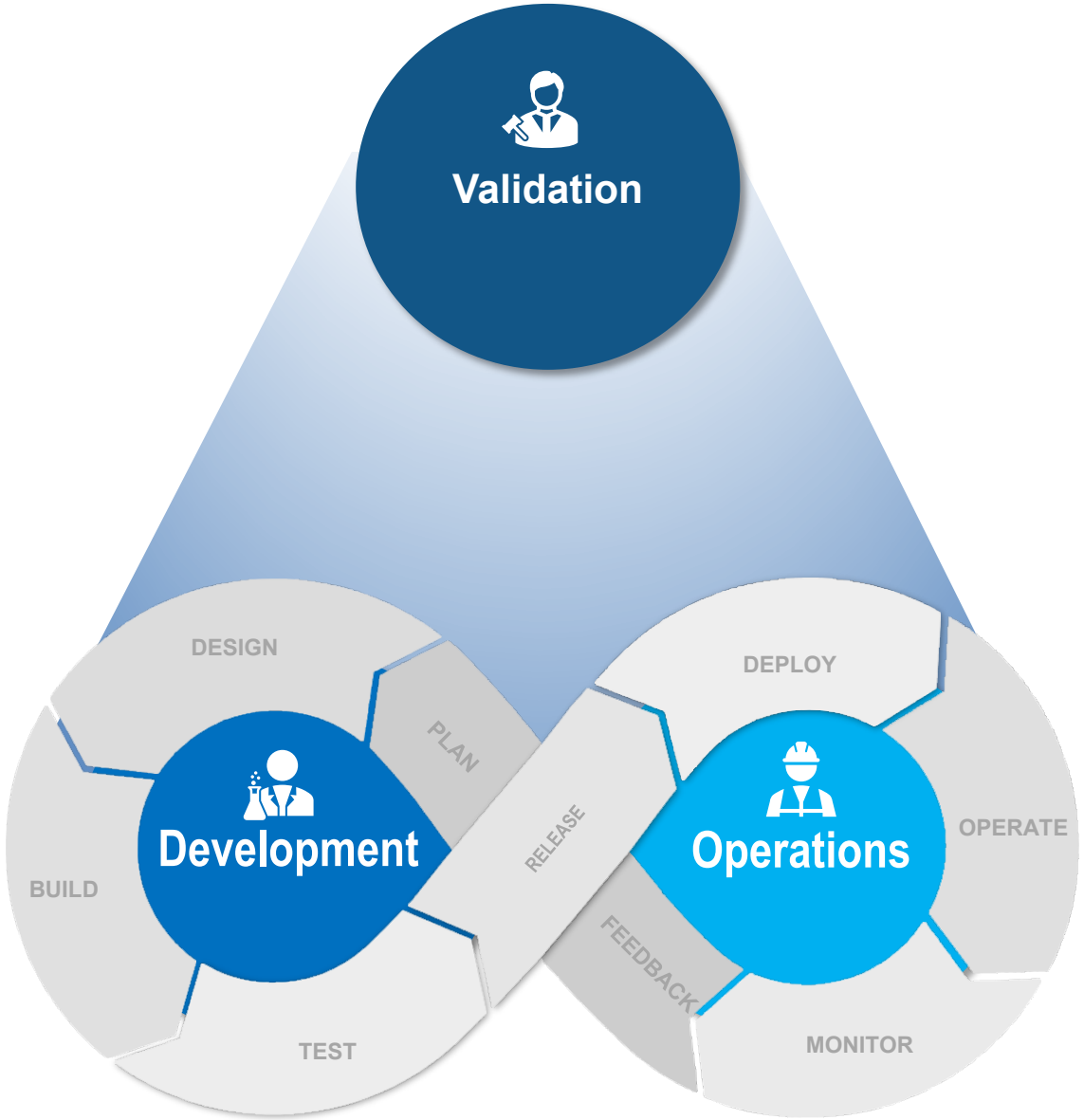
Source: Gartner "Innovation Insight for ModelOps"; Farhan Choudhary; Gartner ID G00729419; August 6, 2020.

Source: Gartner: "Use **Gartner's 3-Stage MLOps Framework** to Successfully Operationalize Machine Learning Projects", Shubhangi Vashisth, Gartner ID: G00725627; July 2, 2020

Challenges arise as groups need to collaborate in new ways



Agile DevOps practices can improve the collaboration



Data

Organization

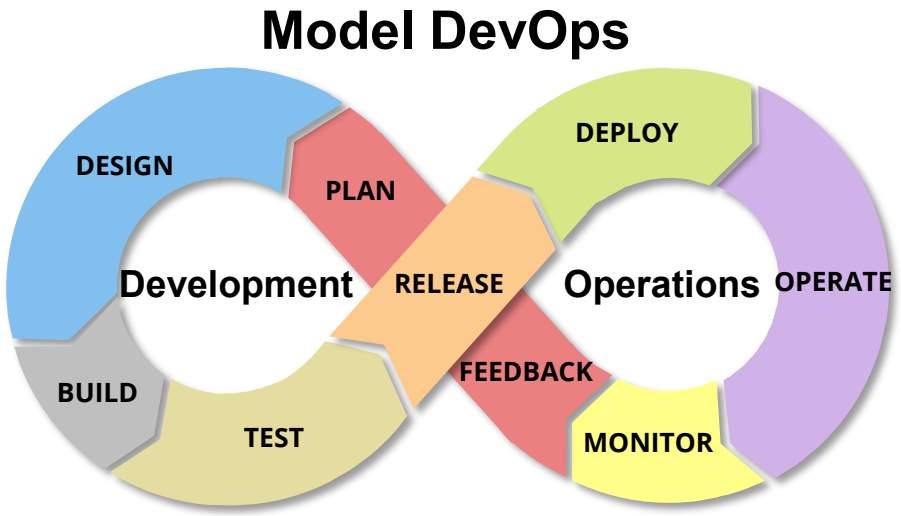
Technologies

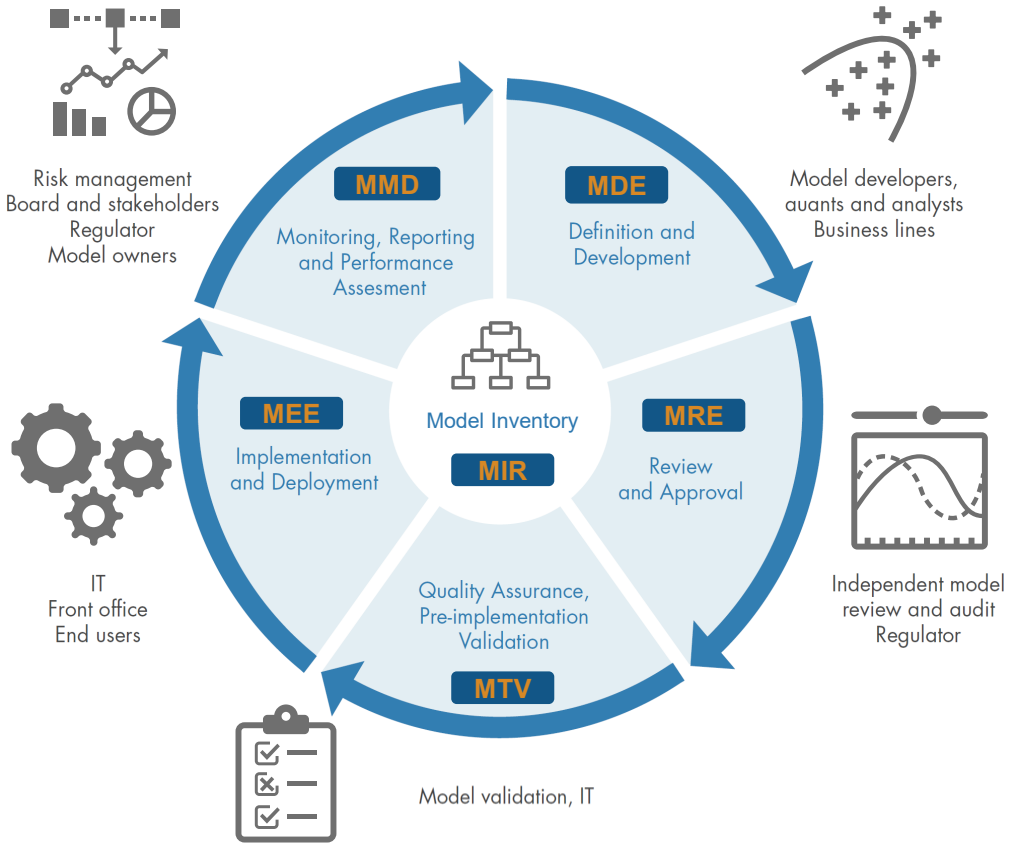
Processes

Culture

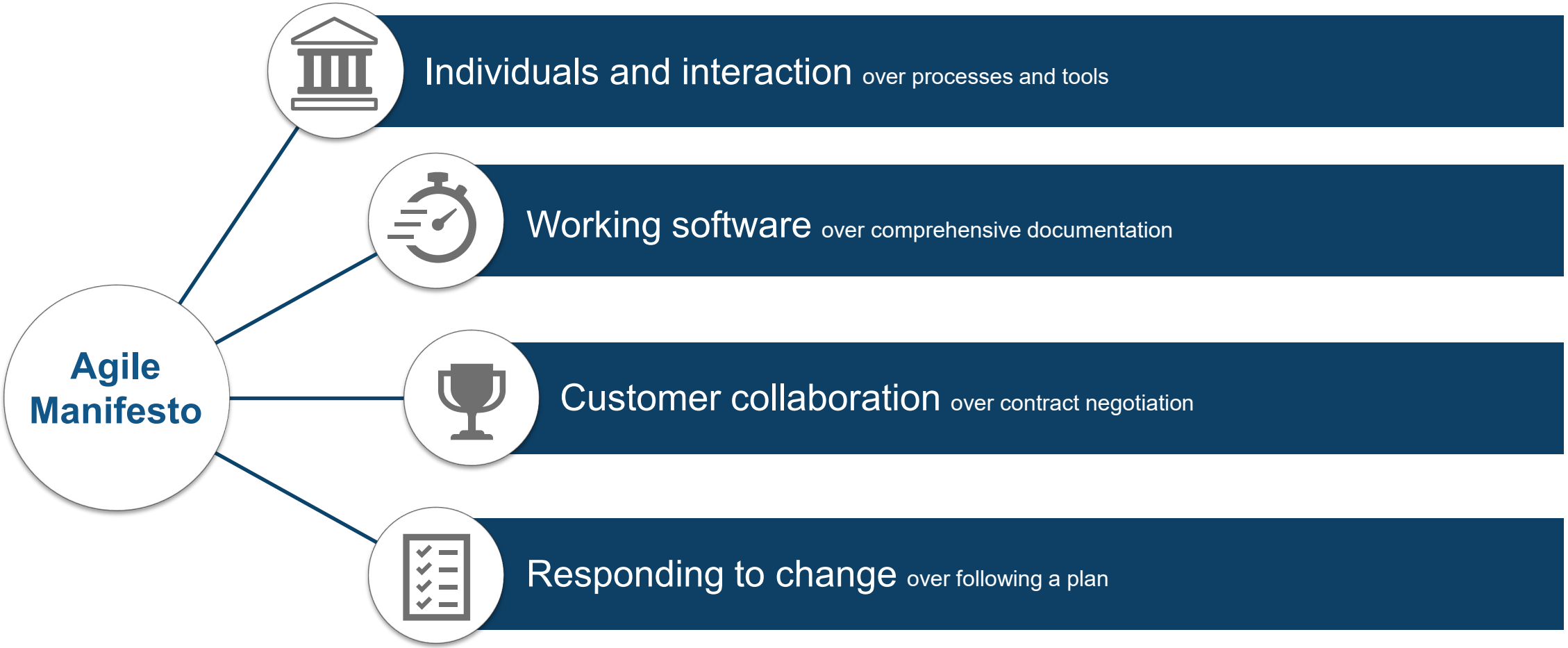
“The best architectures, requirements, and designs emerge from self-organizing teams”

Source: AgileManifesto.org, Principles of Agile.





Agile is about empowering small teams



Source: AgileManifesto.org

Best Practices for Agile Risk Management

Individuals and Interaction

1. Small teams of 6-10 (two pizza teams)
2. Integrated cross-functional team make-up
3. Daily “status/blocker” standups
4. Trust the team to get job done

Working “Models”

1. Favor simplicity over complexity
2. Models as code
3. Documentation as code
4. Infrastructure as code

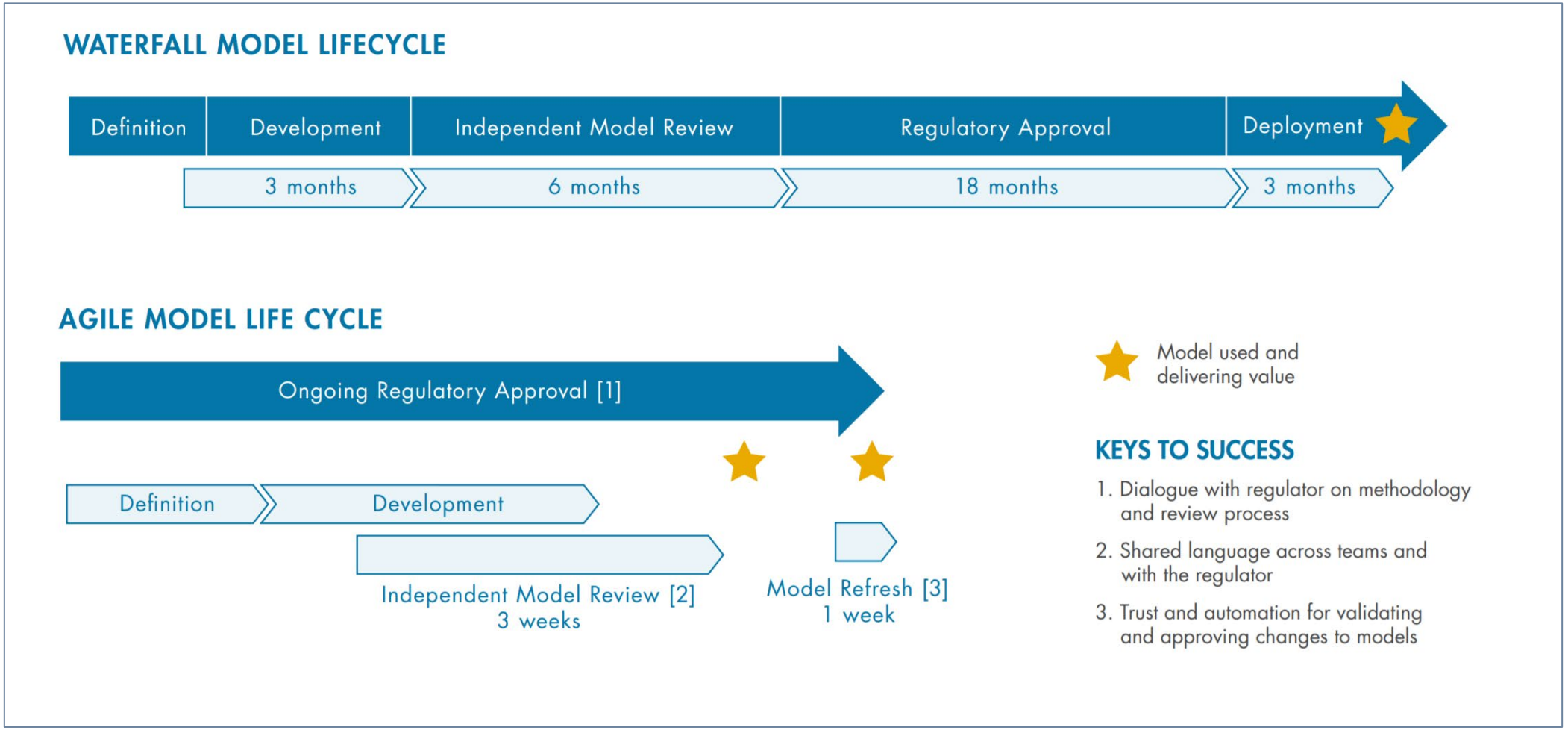
Customer Collaboration

1. Face-to-face demos/meeting
2. Deliver concepts early and often

Responding to Change

1. Change is expected, welcome it
2. Automate: Practice continuous
   - model validation
   - documentation
   - integration
   - model delivery

Reducing model time-to-market and refresh



"Move modelling from a bespoke artisan activity dependent on individuals, to an industrialised process for systematising institutional knowledge."

Learn more about MathWorks, our products, and our services at mathworks.com and on social media:

MathWorks®