Challenges in Deploying Machine Learning

What is rarely talked about at ML conferences

Andrei Paleyes, University of Cambridge

Industry Expert Insights, Cambridge Spark

August 2021

About myself

- A software engineer for a decade...
- ... including few years deploying ML in Amazon
- Now PhD student with Neil Lawrence, ML@CL group, University of Cambridge
    - People think we do ML, but we really mostly do software systems research

ML Adoption

Bright side

- ML adoption in businesses growth 25% year-to-year

Dark side

- Over 60% of companies report difficulties
- Lots of failures: 1 in 4 companies report 50% failure rate

"Global AI Survey...", McKinsey, 2019

"Artificial Intelligence Global Adoption Trends and Strategies", IDC, 2019

The questions

Where do the challenges arise? What stages of the deployment cause concerns?

Hence the paper

Challenges in Deploying Machine Learning: a Survey of Case Studies

- With Raoul-Gabriel Urma and Neil D. Lawrence
- Accepted to ML-Retrospectives, Surveys & Meta-Analyses @ NeurIPS 2020 Workshop
- Under review in one of ACM journals
- Available on arXiv

How to answer it?

1. Fix deployment workflow definition. We use Ashmore et al. 2019.
2. Review existing literature on deployments.
3. Identify practical challenges that were reported.
4. Map them to the ML deployment workflow steps.
5. Analyze and draw conclusions.

"Assuring the machine learning lifecycle: Desiderata, methods, and challenges", Ashmore et al., 2019

Literature

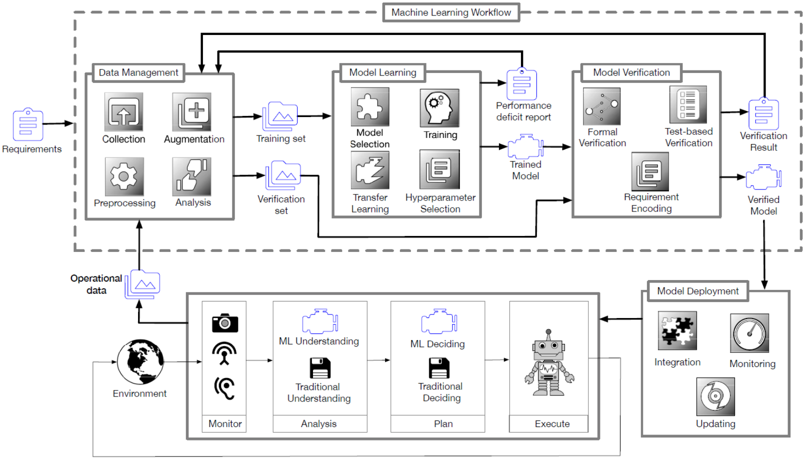
Types:

- Case studies
- Reviews of ML applications in a field
- Lessons learned
- Interview studies among practitioners
- Regulations

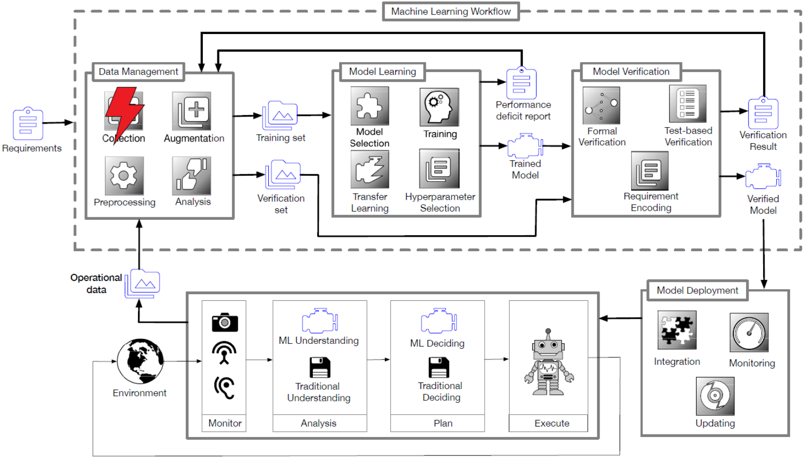
Conditions:

- Not older than 5 years
- All industries
- Don't ignore blog posts

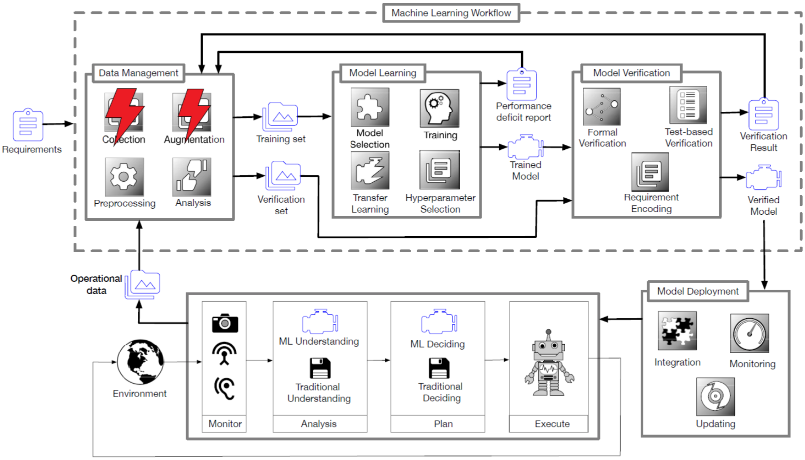
ML workflow



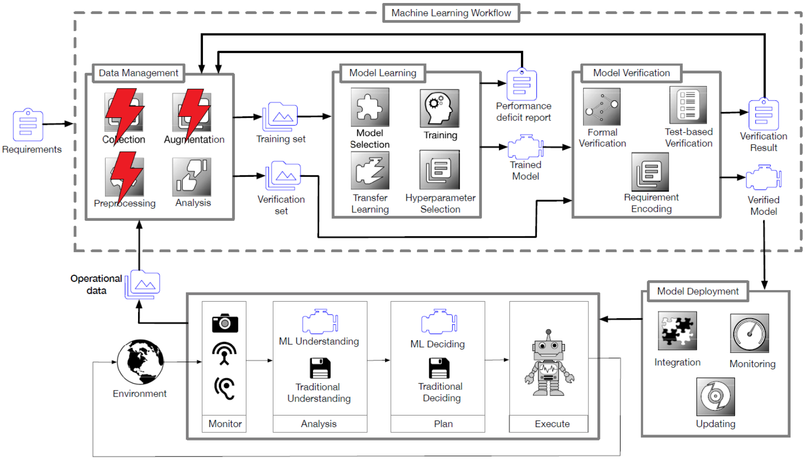
"Assuring the machine learning lifecycle: Desiderata, methods, and challenges", Ashmore et al., 2019

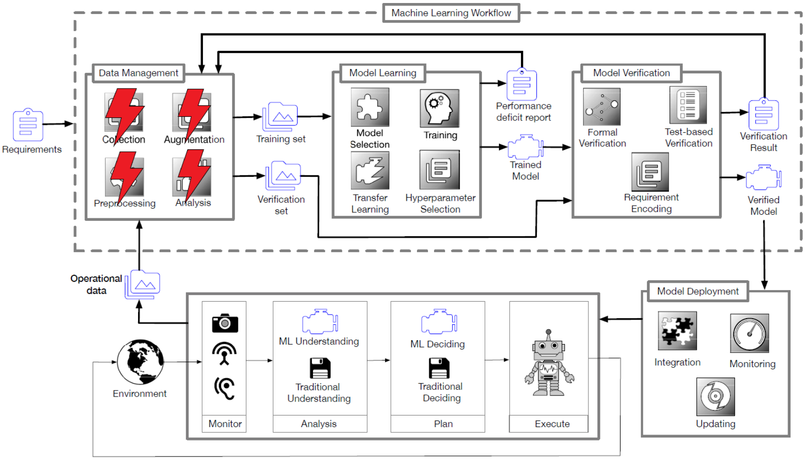


Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Operational data
Model Deployment
Integration
Monitoring
Updating
Environment
ML Understanding
Traditional Understanding
ML Deciding
Traditional Deciding
Monitor
Analysis
Plan
Execute

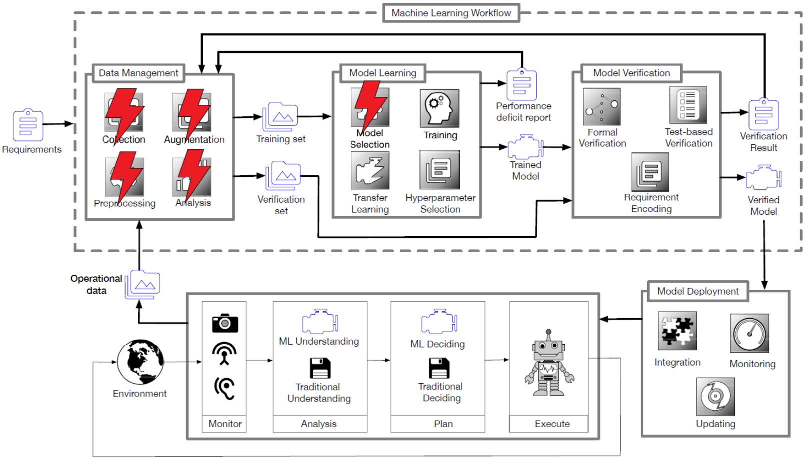


Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Model Deployment
Integration
Monitoring
Updating
Operational data
Environment
ML Understanding
Traditional Understanding
ML Deciding
Traditional Deciding
Monitor
Analysis
Plan
Execute

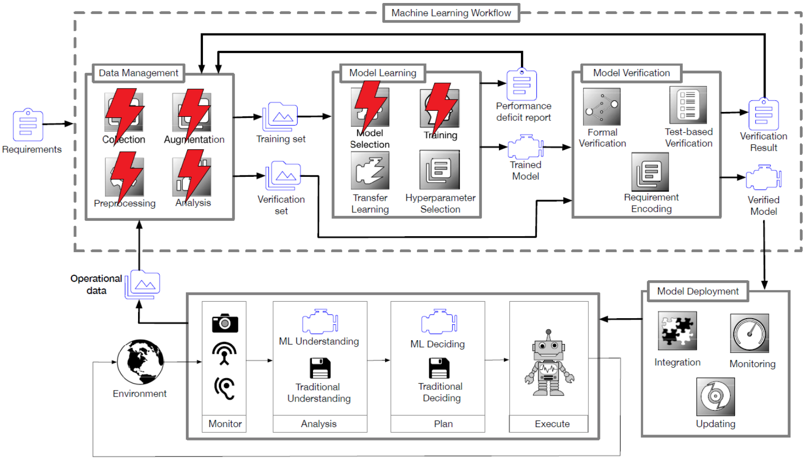

Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Operational data
Environment
Monitor
ML Understanding
Traditional Understanding
Analysis
ML Deciding
Traditional Deciding
Plan
Execute
Model Deployment
Integration
Monitoring
Updating



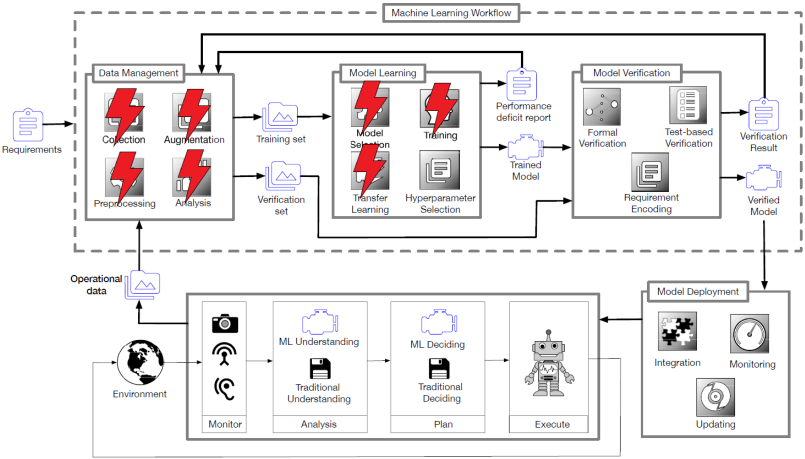
Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Operational data
Model Deployment
Integration
Monitoring
Updating
Environment
ML Understanding
Traditional Understanding
ML Deciding
Traditional Deciding
Monitor
Analysis
Plan
Execute



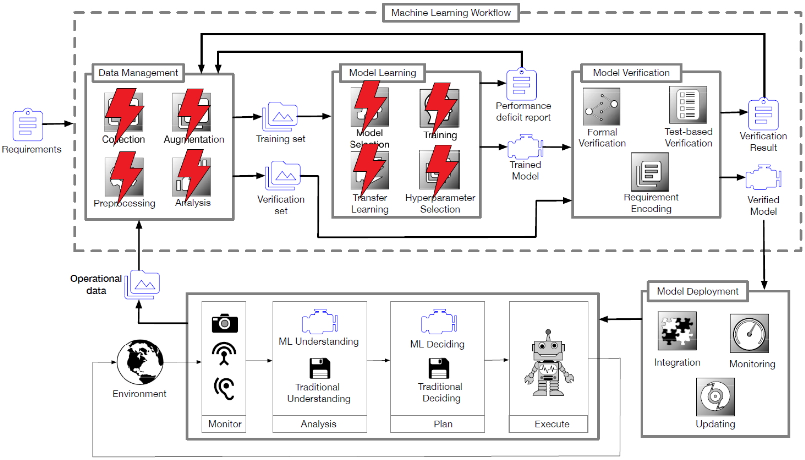
Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Model Deployment
Integration
Monitoring
Updating
Operational data
Environment
Monitor
ML Understanding
Traditional Understanding
Analysis
ML Deciding
Traditional Deciding
Plan
Execute



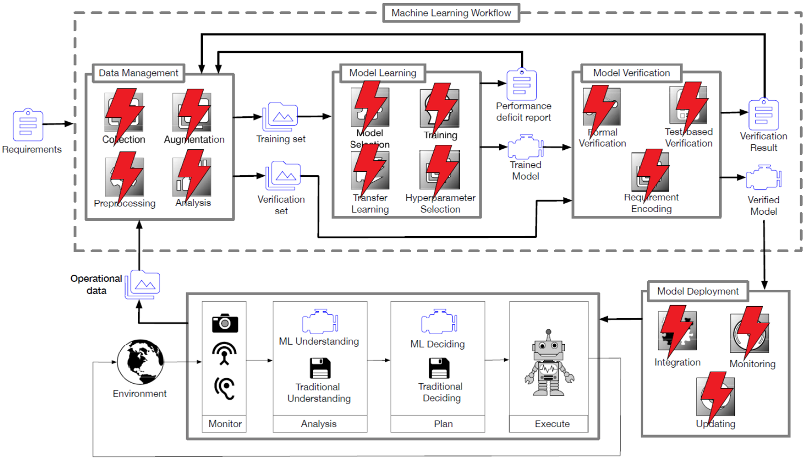
Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Operational data
Model Deployment
Integration
Monitoring
Updating
Environment
ML Understanding
Traditional Understanding
ML Deciding
Traditional Deciding
Monitor
Analysis
Plan
Execute



Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Operational data
Environment
ML Understanding
Traditional Understanding
ML Deciding
Traditional Deciding
Monitor
Analysis
Plan
Execute
Model Deployment
Integration
Monitoring
Updating



Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Model Deployment
Integration
Monitoring
Updating
Operational data
Environment
Monitor
ML Understanding
Traditional Understanding
Analysis
ML Deciding
Traditional Deciding
Plan
Execute


Machine Learning Workflow
Requirements
Data Management
Collection
Augmentation
Preprocessing
Analysis
Training set
Verification set
Model Learning
Model Selection
Training
Transfer Learning
Hyperparameter Selection
Performance deficit report
Trained Model
Model Verification
Formal Verification
Test-based Verification
Requirement Encoding
Verification Result
Verified Model
Operational data
Environment
ML Understanding
Traditional Understanding
ML Deciding
Traditional Deciding
Monitor
Analysis
Plan
Execute
Model Deployment
Integration
Monitoring
Updating

Don't forget cross-cutting aspects!

- Ethics
- End users' trust
- Security

Data management

Data management	Data collection	Data discovery
	Data preprocessing	Data dispersion Data cleaning
	Data augmentation	Labeling of large volumes of data Access to experts Lack of high-variance data
	Data analysis	Data profiling

Model learning

Model learning	Model selection	Model complexity Resource-constrained environments Interpretability of the model
	Training	Computational cost Environmental impact
	Hyper-parameter selection	Resource-heavy techniques Hardware-aware optimization

Model Verification

Model verification	Requirement encoding	Performance metrics Business driven metrics
	Formal verification	Regulatory frameworks
	Test-based verification	Simulation-based testing

Model Deployment

Model deployment	Integration	Operational support Reuse of code and models Software engineering anti-patterns Mixed team dynamics
	Monitoring	Feedback loops Outlier detection Custom design tooling
	Updating	Concept drift Continuous delivery

Cross-cutting aspects

Cross-cutting aspects	Ethics	Country-level regulations Focus on technical solution only Aggravation of biases Authorship Decision making
	End users' trust	Involvement of end users User experience Explainability score
	Security	Data poisoning Model stealing Model inversion

Conclusions

There is no single “bottleneck” stage. ML deployment projects face serious challenges every step of the way, from data collection to model monitoring.

It is worth ML community’s time and focus to think about these challenges.

Reports are scarce. Lots of knowledge goes unpublished. Please share your practical experience more!

What can be done? - Tools

- Cloud platforms. Examples: AWS SageMaker, AzureML, TensorFlow TFX, MLflow
- Quality assurance. Example: CheckList for NLP
- Weak labeling. Examples: Snorkel, Snuba, cleanlab

Pros: specific tool for specific problem
Cons: dependencies management, maintenance

What can be done? - Holistic approaches

- *Data Oriented Architectures*, Neil Lawrence, 2019
- *Technology Readiness Levels for AI & ML (TLR4ML)*, Alex Lavin et al, 2020
- *Data meshes*, Zhamak Dehghani, 2019
- Cookbooks, for example *Rules of machine learning: Best practices for ML engineering*, Martin Zinkevich, 2017

Pros: ML first mindset
Cons: big investment

Summary

ML deployment is hard
Every part of the workflow presents its own challenges
Some aspects affect the whole process
There are tools and approaches that can help

https://paleyes.info
https://mlatcl.github.io
Get in touch: ap2169@cam.ac.uk