Assurance Argument Patterns and Processes for Machine Learning in Safety-Related Systems.

Chiara Picardi, Colin Paterson, Richard Hawkins, Radu Calinescu and Ibrahim Habli

UNIVERSITY of York

ASSURING AUTONOMY
INTERNATIONAL PROGRAMME

Motivation

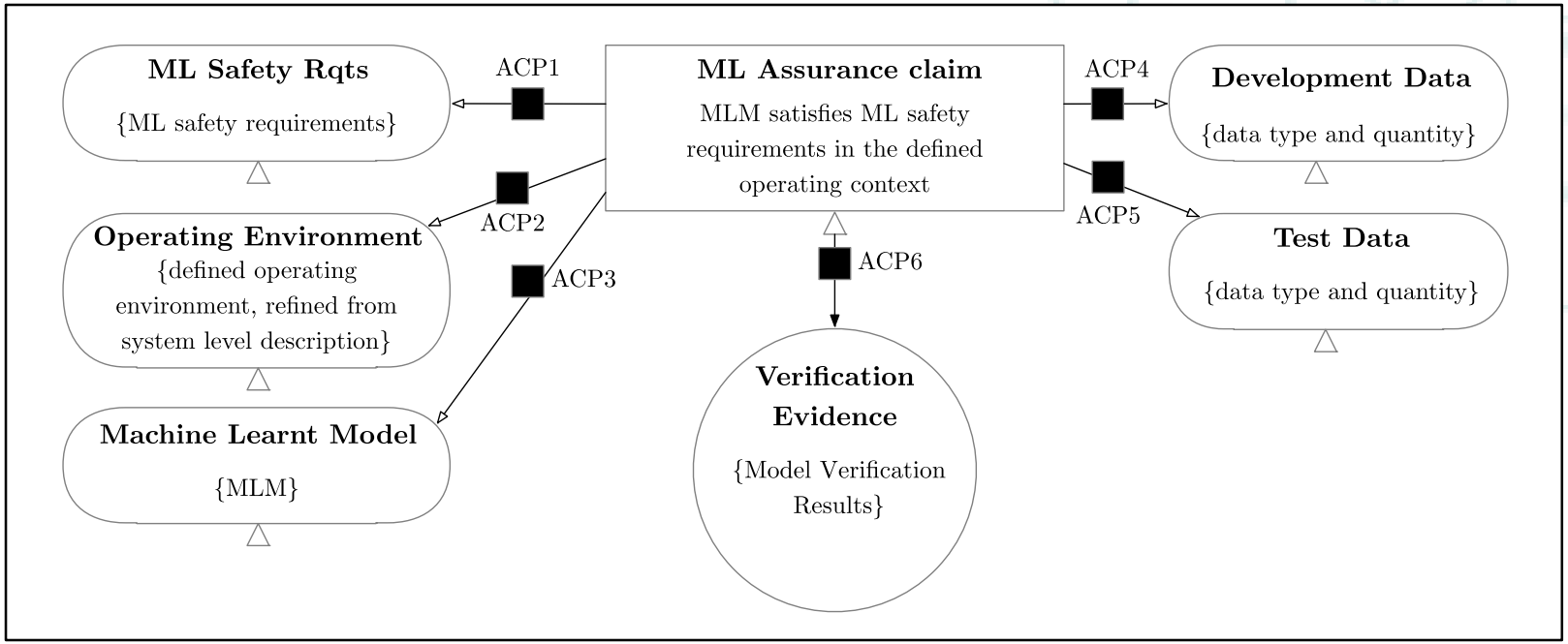
- ML is widely accepted for use in a large range of applications
- Claims are being made that Human performance levels can be met or exceeded

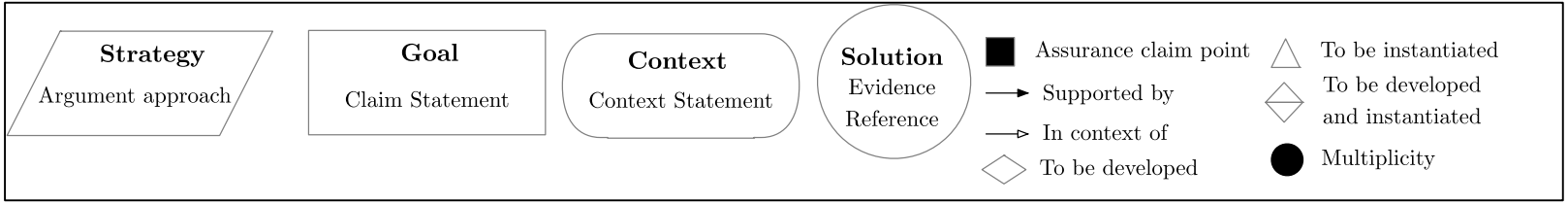
Motivation

- ML is widely accepted for use in a large range of applications
- Claims are being made that Human performance levels can be met or exceeded

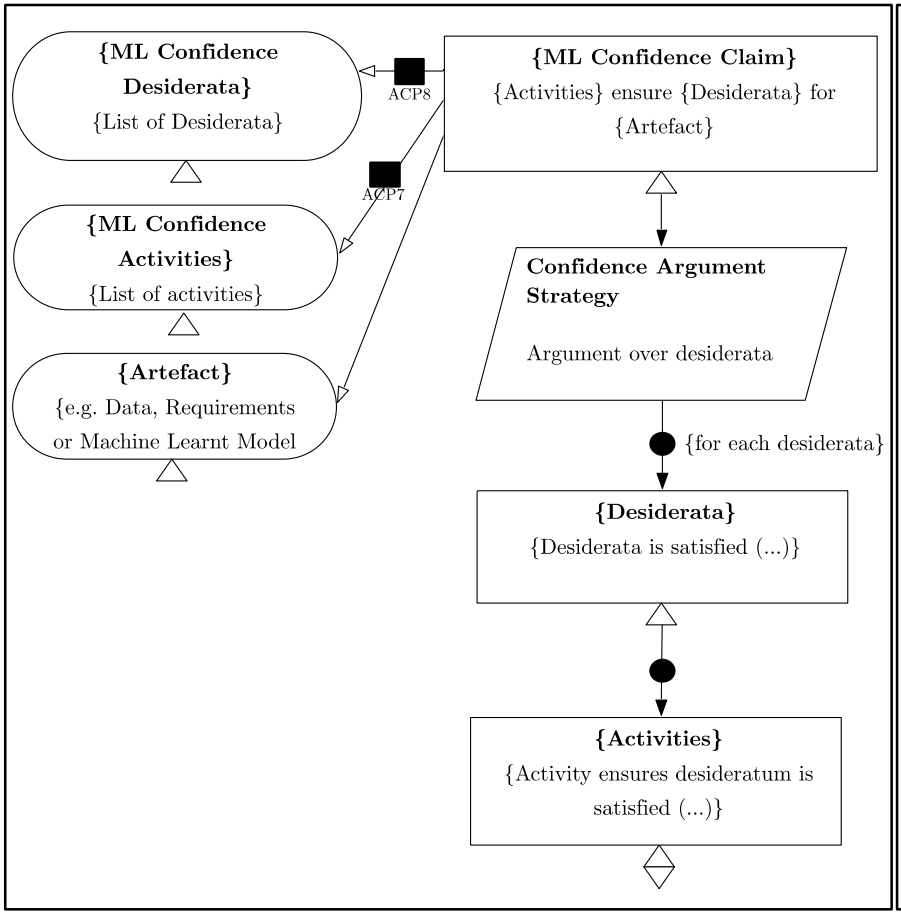
- Evidence for these claims is limited
- Existing ML development processes rarely consider the need for assurance

Assurance Argument Pattern for MLM

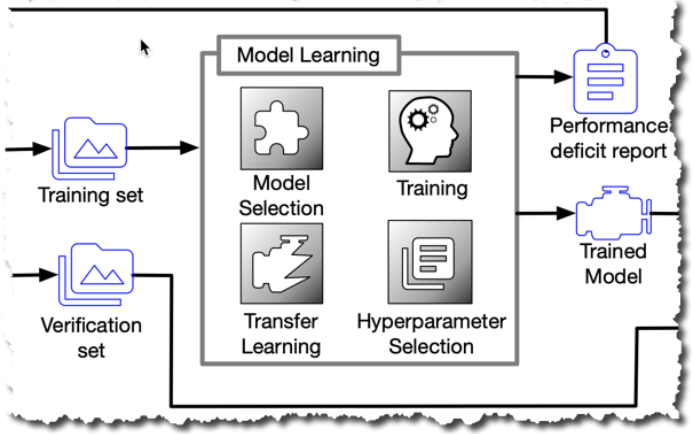


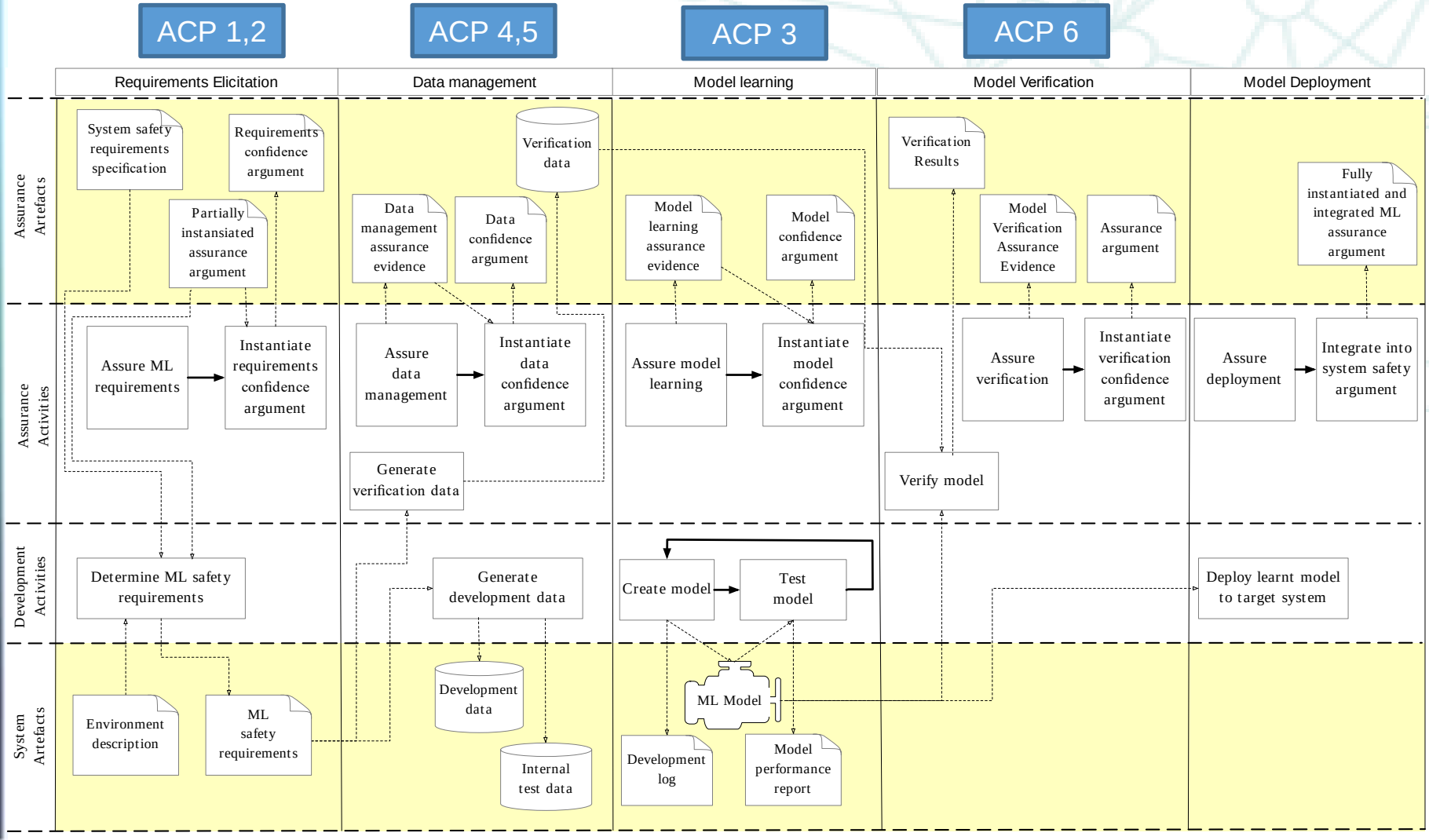


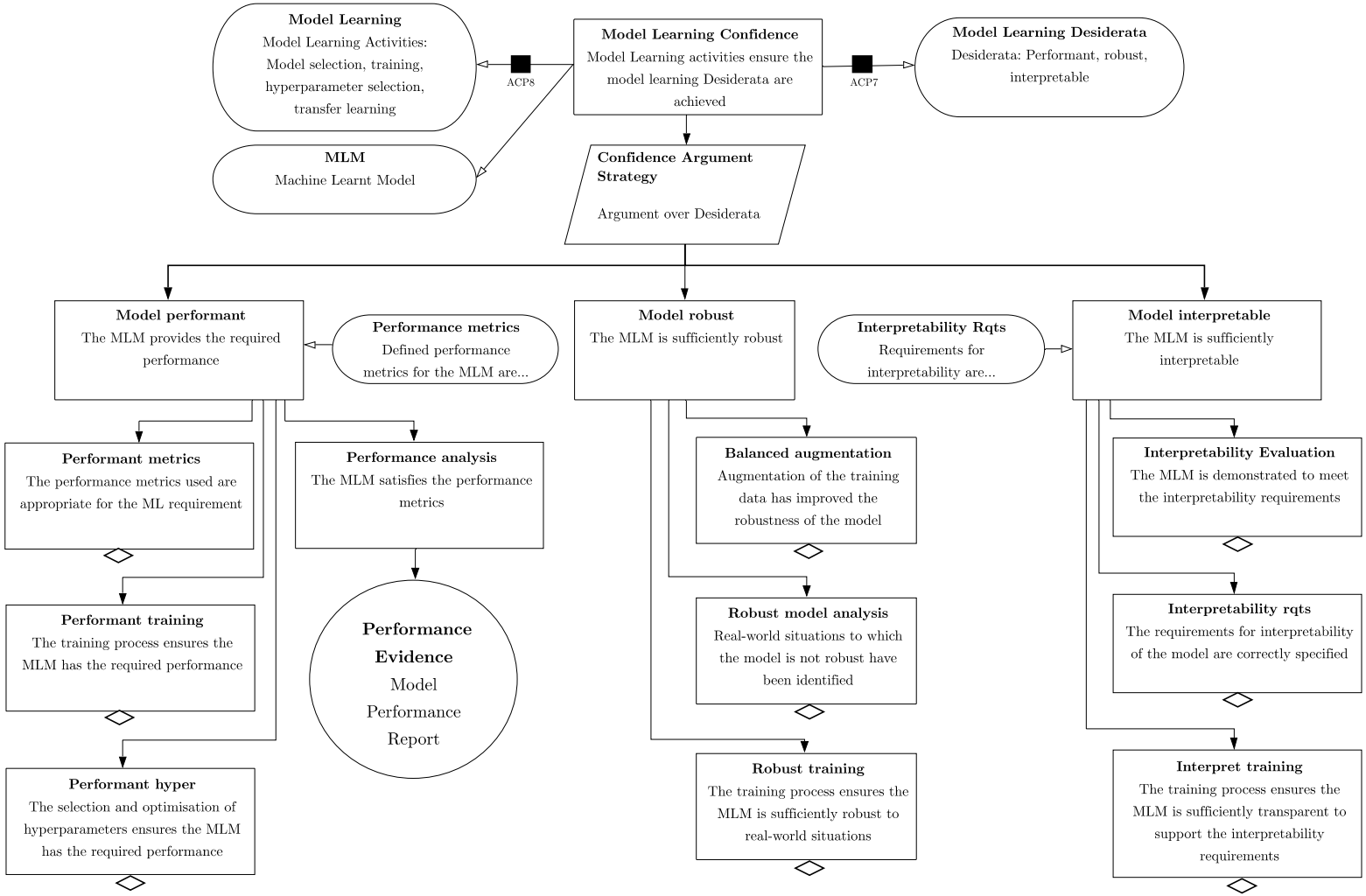
Confidence Argument Pattern for MLM



Method	Associated activities†				Supported desiderata‡			
	Model Selection	Training	Hyperparam. Selection	Transfer Learning	Performant	Robust	Reusable	Interpretable
Use appropriate performance measures [52, 167]	✓	✔			★	★		
Statistical tests [112, 118]	✓	✔			★			
Ensemble Learning [145]	✔	✔		✔	★	★		
Optimise hyperparameters [71, 178]		✔	✔		★	★		
Batch Normalization [73]		✔	✔		★	★		
Prefer simpler models [3, 143]	✔	✓			☆	★		☆
Augment training data		✔			★	★		
Regularization methods [58]		✔	✔			★		
Use early stopping		✔	✔			★		
Use models that intrinsically support reuse [2]	✔			✔			★	☆
Transfer Learning [173]	✔	✓		✔			★	☆
Use model zoos [58]	✔	✓		✔			★	
Post-hoc interpretability methods [3, 93, 105]		✔						★

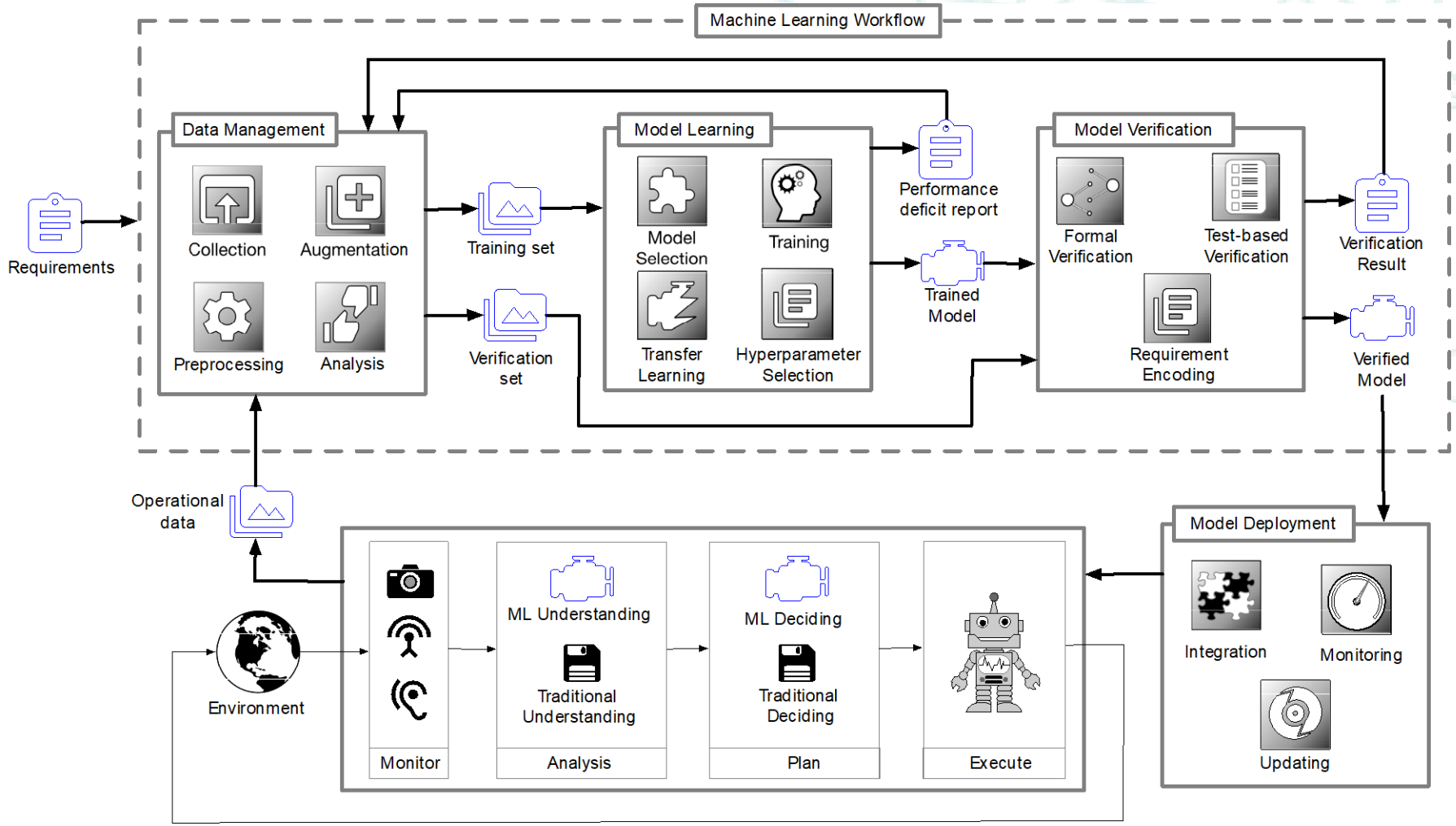



ACP 1,2
ACP 4,5
ACP 3
ACP 6
Requirements Elicitation
Data management
Model learning
Model Verification
Model Deployment
Assurance Artefacts
Assurance Activities
Development Activities
System Artefacts
System safety requirements specification
Requirements confidence argument
Partially instansiated assurance argument
Verification data
Data management assurance evidence
Data confidence argument
Model learning assurance evidence
Model confidence argument
Verification Results
Model Verification Assurance Evidence
Assurance argument
Fully instantiated and integrated ML assurance argument
Assure ML requirements
Instantiate requirements confidence argument
Assure data management
Instantiate data confidence argument
Generate verification data
Assure model learning
Instantiate model confidence argument
Assure verification
Instantiate verification confidence argument
Verify model
Assure deployment
Integrate into system safety argument
Determine ML safety requirements
Generate development data
Create model
Test model
Deploy learnt model to target system
Environment description
ML safety requirements
Development data
Internal test data
ML Model
Development log
Model performance report


Model Learning
Model Learning Activities: Model selection, training, hyperparameter selection, transfer learning
ACP8
Model Learning Confidence
Model Learning activities ensure the model learning Desiderata are achieved
ACP7
Model Learning Desiderata
Desiderata: Performant, robust, interpretable
MLM
Machine Learnt Model
Confidence Argument Strategy
Argument over Desiderata
Model performant
The MLM provides the required performance
Performance metrics
Defined performance metrics for the MLM are...
Model robust
The MLM is sufficiently robust
Interpretability Rqts
Requirements for interpretability are...
Model interpretable
The MLM is sufficiently interpretable
Performant metrics
The performance metrics used are appropriate for the ML requirement
Performance analysis
The MLM satisfies the performance metrics
Balanced augmentation
Augmentation of the training data has improved the robustness of the model
Interpretability Evaluation
The MLM is demonstrated to meet the interpretability requirements
Performant training
The training process ensures the MLM has the required performance
Performance Evidence
Model Performance Report
Robust model analysis
Real-world situations to which the model is not robust have been identified
Interpretability rqts
The requirements for interpretability of the model are correctly specified
Performant hyper
The selection and optimisation of hyperparameters ensures the MLM has the required performance
Robust training
The training process ensures the MLM is sufficiently robust to real-world situations
Interpret training
The training process ensures the MLM is sufficiently transparent to support the interpretability requirements

Conclusions

1. ML promises to achieve human like capabilities in a range of safety critical domains
2. Assurance is necessary for the safe adoption of such technology
3. We propose a process to generate a compelling and credible assurance case
4. Our process allows us to generate assurances cases incrementally
5. Our process builds on existing practice
6. Our next steps involve validation using real world applications.


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