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

- **ML Safety Rqts**: ML safety requirements
- **Operating Environment**: Defined operating environment, refined from system level description
- **Machine Learnt Model**: MLM
- **ML Assurance claim**: MLM satisfies ML safety requirements in the defined operating context
- **Verification Evidence**: Model Verification Results
- **Development Data**: Data type and quantity
- **Test Data**: Data type and quantity

**Strategy**
- Argument approach

**Goal**
- Claim Statement

**Context**
- Context Statement

**Solution**
- Evidence Reference
  - Assurance claim point
  - Supported by
    - In context of
    - To be developed
  - To be instantiated
  - To be developed and instantiated
  - Multiplicity
Confidence Argument Pattern for MLM

**ML Confidence Desiderata**
- List of Desiderata

**ML Confidence Activities**
- List of activities

**Artefact**
- e.g. Data, Requirements or Machine Learnt Model

---

**ML Confidence Claim**
- Activities ensure Desiderata for Artefact

**Confidence Argument Strategy**
- Argument over desiderata
  - {for each desiderata}
  - Desiderata is satisfied (...)

**Activities**
- Activity ensures desideratum is satisfied (...)

---

**Associated activities**

<table>
<thead>
<tr>
<th>Method</th>
<th>Model Selection</th>
<th>Hyperparam Selection</th>
<th>Transfer Learning</th>
<th>Performant Robust Reusable Interpretable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use appropriate performance measures [52, 167]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Statistical tests [112, 118]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ensemble Learning [145]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Optimise hyperparameters [71, 172]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Batch Normalization [73]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Prefer simpler models [3, 144]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Augment training data</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Regularization methods [58]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Use early stopping</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Use models that intrinsically support reuse [2]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Transfer Learning [173]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Use model zoos [58]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Post-hoc interpretability methods [3, 93, 105]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

---

**Supported desiderata**

- Performance deficit report

---

**Model Learning**
- Model Selection
- Training
- Hyperparameter Selection
- Trained Model

**Training set**
- Model Selection
- Training

**Verification set**
- Transfer Learning
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.