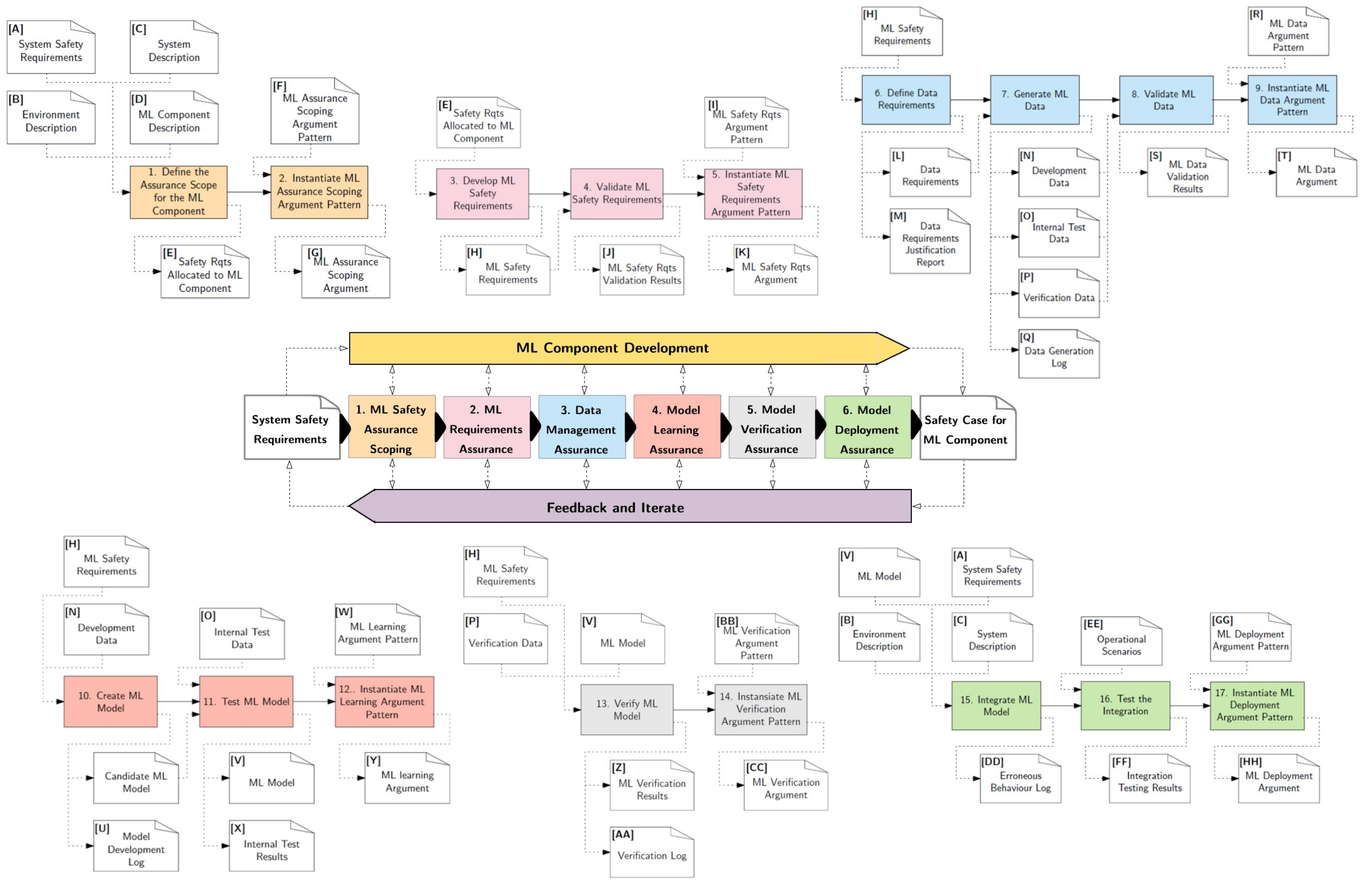
Assurance of Machine Learning for use in Autonomous Systems (AMLAS)

Chiara Picardi, Richard Hawkins and Ibrahim Habli
Assuring Autonomy International Programme, University of York

AMLAS process

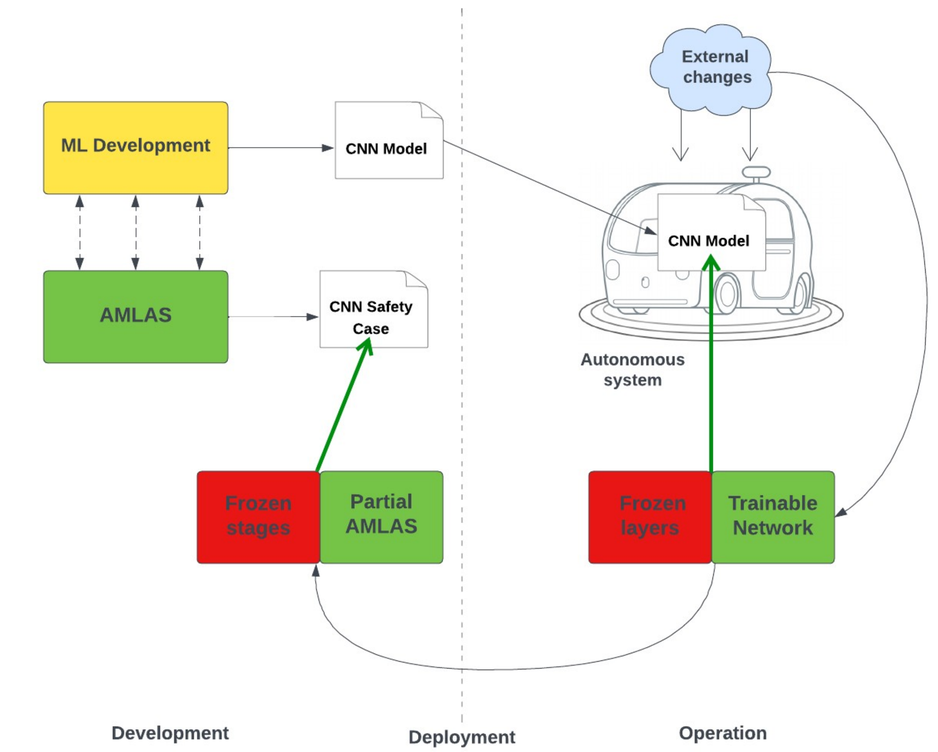


We have developed a methodology for the Assurance of Machine Learning for use in Autonomous Systems (AMLAS). AMLAS provides the first detailed guidance for creating a safety case for an ML component. It comprises a set of safety case patterns and a process for (1) systematically integrating safety assurance into the development of ML components and (2) for generating the evidence base for explicitly justifying the acceptable safety of these components when integrated into autonomous system applications. AMLAS covers the entire ML lifecycle and describes assurance activities that are integrated with ML development.
The diagram shows the overall process in the middle along with the details of the activities and artifacts of each stage around the outside. The artifacts generated by following AMLAS are used to create a safety case for the ML component.

You can download AMLAS at: https://www.york.ac.uk/assuring-autonomy/guidance/amlas/

SCAN ME

Transfer Assurance for ML



During operation, changes to the autonomous system or its operating environment may lead to a divergence between the input data to the ML component and the training data. This may mean that an assured ML component may need to be retrained to reflect these changes. On the RAILS project (Responsible AI for Long-term trustworthy autonomous Systems) we are investigating how assurance in the ML component can be preserved through-life in the presence of changes. Transfer learning enables partial re-training of Convolutional Neural Network (CNN) models where some layers remain fixed and only the last layers are fine-tuned. Similarly, we hope to be able to define an approach for the partial re-use of assurance from the AMLAS process. In the figure, red is used to highlight reused elements while green is used for new elements.

SCAN ME