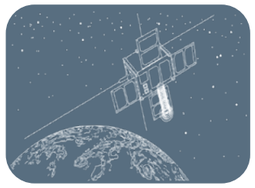
AMLAS

Develop compelling safety cases for machine learning

What is AMLAS?

AMLAS is the only methodology for developing compelling safety cases for machine learning (ML) within autonomous systems. It's a downloadable framework being used in industries from healthcare to transport, embedded in safety processes, referenced in new standards, and recommended by safety engineers.

AMLAS supported the development of a dedicated assurance model to support Clinical Safety Officers evaluate AI technologies.



AMLAS helped demonstrate the safety of Craft Prospect's Autonomous System to the European Space Agency.

AMLAS provided Luffy AI with a framework to integrate safety assurance of neural networks and build a compelling argument for its safety case.

Who is AMLAS for?

Our AMLAS guidance is aimed at three primary audiences:

Safety Engineers

AMLAS enables safety engineers to determine the Machine Learning-specific safety considerations, evaluate the impact of the ML component within the autonomous technology and create a safety case for the ML component.

Machine Learning Developers

AMLAS assists Machine Learning developers in the development of ML components that satisfy safety requirements and generate the artefacts required to support the safety case.

Decision Makers

AMLAS offers assurance and confidence around the deployment of AI, ML and autonomous systems to senior management teams, board members and regulators through the creation of robust safety cases.

The expertise behind AMLAS

"Despite the interest from government and industry, there is currently no established process for certifying and assuring an AI component. AMLAS fills this gap by providing us with a framework to integrate safety assurance into our development process and build a compelling argument for our safety case."

Matthew Carr, Luffy AI

AMLAS has been developed by the Centre for Assuring Autonomy, a £10m partnership between Lloyd's Register Foundation and the University of York. Our team of globally recognised experts has been advancing the safety of complex systems through pioneering research for over 30 years.

We worked with an international community of developers, regulators, and researchers to develop and release AMLAS as freely accessible guidance, with proven real-world impact and benefit. Our AMLAS guidance is based on:

1. Sound research – peer-reviewed and published in leading academic venues.
2. Empirical evaluation – evaluated in real-world, credible contexts.
3. Accessibility – practical guidance disseminated online and through CPD training.

Through AMLAS we, at the Centre for Assuring Autonomy, are leading the way in ensuring society can safely benefit from AI and autonomous systems.

How is AMLAS used?

AMLAS provides practical, easy-to-use guidance to assure the safe deployment of an ML component in a particular system context.

Our guidance:

- Is a structured process which will generate a compelling and detailed safety case.
- Provides assurance that an autonomous system with an ML component will perform safely and as expected.
- Is a practical framework which can be used alongside existing ML development processes.

AMLAS guidance creates a compelling safety case which positively impacts market readiness, regulatory compliance and public buy-in.

Since its launch in 2022, AMLAS has been increasingly referenced and adopted in multiple domains and organisations.

For example:

- AMLAS is referenced in NHS Healthcare Guidance.
- AMLAS is currently being used in the defence sector.
- AMLAS has been used to demonstrate the assurance of a robotic arm in a nuclear decommissioning setting.

How to use AMLAS

The assurance activities of AMLAS run in parallel to, and should be integrated with, the ML development activities. AMLAS consists of six stages which are:

Stage 1: ML safety assurance scoping

Assuring the safety of ML components is only possible with consideration of the systems and environment into which the component is to be deployed. Indeed, without considering the system and environmental context, it is not possible to make any claim about the safety of an ML component. The aim of this stage is to define the scope under which we are able to demonstrate the safety of the ML.

Stage 2: ML requirements assurance

In this stage, ML safety requirements are defined and justified. The ML safety requirements must include requirements for the performance and robustness of the ML model. ML performance considers quantitative performance metrics, such as classification accuracy, whereas ML robustness considers the model's ability to perform well when the real-world inputs encountered differ from those in the training data.

Stage 3: Data management assurance

This stage provides assurance for the data used to develop the ML. This involves firstly developing data requirements which encode the ML safety requirements as features against which the data sets are assessed. ML data requirements include consideration of the relevance, completeness, accuracy and balance of the data. Data sets for development, internal testing, and verification of the ML model are generated and validated in accordance with the data requirements.

Stage 4: Model learning assurance

This develops the ML model using the development data obtained in the previous stage. The internal test data is used to assess the extent to which the ML model is able to meet the ML safety requirements. A Model Development Log documents and justifies all key decisions made during the learning process and how those choices impact the required performance or robustness of the model.

Stage 5: Model verification assurance

This stage seeks to demonstrate that the model will meet the ML safety requirements when exposed to inputs not present during the development of the model. It is important that the verification activities are sufficiently independent of the development activities. This includes ensuring that the information concerning verification data is hidden from the developers to ensure the models generated are robust to the whole class of failures and not just specific examples present in the verification data.

Stage 6: Model deployment assurance

This stage considers the safe integration of the ML component into the target system. The aim is to demonstrate that the allocated system safety requirements are still satisfied during operation of the system in the real-world environment. This involves integration of the ML model, ensuring sufficient and effective monitoring is in place, and system-level testing against a defined set of operating scenarios.

Centre *for* Assuring Autonomy

Work with us

There are several ways the Centre for Assuring Autonomy can support your organisation in developing safety assurance cases.

Training

Our bespoke training and education sessions enable industry, regulators, and policy makers to develop the expertise necessary to ensure that autonomous systems are brought safely to market and into operation.

Research

We can help you solve emerging and critical research questions around safe AI through our multidisciplinary team approach.

Partnerships and collaborations

Our team of experts works with organisations to help develop safety cases, address particular challenges, or develop compliance procedures.

Consultancy

Our bespoke consultancy service enables you to build a package of support which meets organisational needs and access the right experts for your AI and AS challenges.

Our safety assurance guidance and frameworks

AMLAS is part of a suite of frameworks and guidance freely available which organisations can use to develop safety cases for autonomous systems and AI, including Frontier AI.

Read our other factsheets in this series

- BIG Argument
- SACE
- PRAISE

University of York

LR Foundation

+44 (0)1904 325345
assuring-autonomy@york.ac.uk

cfaa.bsky.social
assuring-autonomy