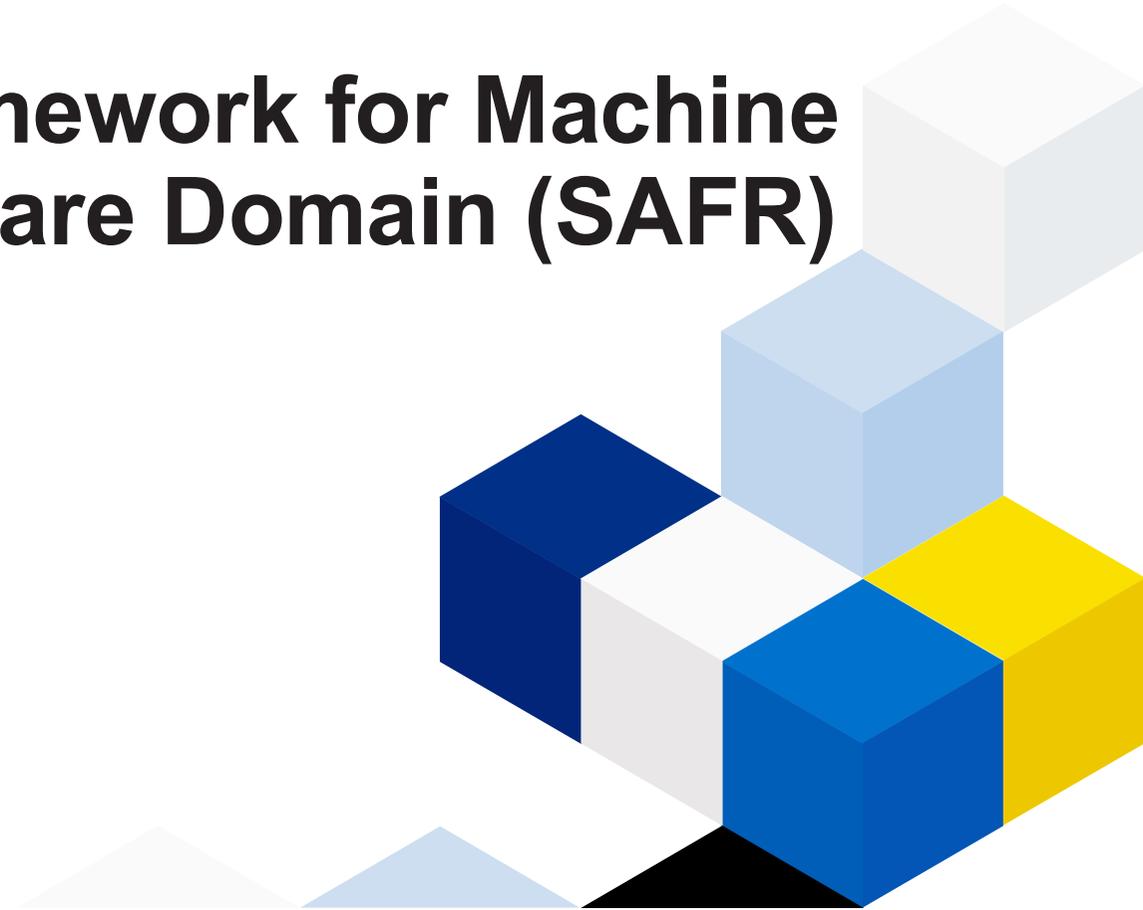


Safety Assurance Framework for Machine Learning in the Healthcare Domain (SAFR)

Presented by:
Shakir Laher
Shakir.laher1@nhs.net



Assessing Methods and Tools to Improve Reporting, Increase Transparency, and Reduce Failures in Machine Learning Applications in Health Care

Christian Garbin, MS • Oge Marques, PhD

From the College of Engineering & Computer Science, Florida Atlantic University, 777 Glades Rd, EE441, Boca Raton, FL 33431-0991. Received May 18, 2021; revision requested July 2; revision received December 17; accepted January 5, 2022. Address correspondence to O.M. (e-mail: omarques@fau.edu).

Authors declared no funding for this work.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2022; 4(2):e210127 • <https://doi.org/10.1148/ryai.210127> • Content code: **AI**

Artificial intelligence applications for health care have come a long way. Despite the remarkable progress, there are several examples of unfulfilled promises and outright failures. There is still a struggle to translate successful research into successful real-world applications. Machine learning (ML) products diverge from traditional software products in fundamental ways. Particularly, the main component of an ML solution is not a specific piece of code that is written for a specific purpose; rather, it is a generic piece of code, a model, customized by a training process driven by hyperparameters and a dataset. Datasets are usually large, and models are opaque. Therefore, datasets and models cannot be inspected in the same, direct way as traditional software products. Other methods are needed to detect failures in ML products. This report investigates recent advancements that promote auditing, supported by transparency, as a mechanism to detect potential failures in ML products for health care applications. It reviews practices that apply to the early stages of the ML lifecycle, when datasets and models are created; these stages are unique to ML products. Concretely, this report demonstrates how two recently proposed checklists, datasheets for datasets and model cards, can be adopted to increase the transparency of crucial stages of the ML lifecycle, using ChestX-ray8 and CheXNet as examples. The adoption of checklists to document the strengths, limitations, and applications of datasets and models in a structured format leads to increased transparency, allowing early detection of potential problems and opportunities for improvement.

Supplemental material is available for this article.

© RSNA, 2022

1. Define the problem to be solved.
2. Procure a dataset
3. Train the model
4. Test the product
5. Release the product
6. Monitor the behaviour of the product

	Traditional	Machine learning
Specific piece of code, written for a specific purpose	Write code	Procure dataset Large amounts of high-dimensional data
Human-readable	Inspect/review	Train model Generic piece of code, customized with a dataset and hyperparameters
White and black-box tests	Test code	Test model Black box tests

Current AI Guidance



A Buyer's Guide to AI in Health and Care

10 questions for making well-informed procurement decisions about products that use AI



Guidance:

Medical device stand-alone software including apps (including IVDMDs)



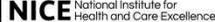
Artificial Intelligence: How to get it right

Putting policy into practice for safe data-driven innovation in health and care



BS EN ISO 14971:2012

BSI Standards Publication



Evidence standards framework for digital health technologies

As digital health technologies develop at an increasing pace, we've worked with partners to develop standards that ensure new technologies are clinically effective and offer economic value.

Partners

- Public Health England
- MedCiv



Medical devices — Application of risk management to medical devices (ISO 14971:2007, Corrected version 2007-10-01)

...making excellence a habit™



A guide to good practice for digital and data-driven health technologies

Updated 19 January 2021

Contents

- How to operate ethically
- Have a clear value proposition
- Usability and accessibility
- Technical assurance
- Clinical safety
- Data protection
- Data transparency
- Cybersecurity
- Regulation
- Interoperability and open standards
- Generate evidence that the product achieves clinical, social, economic or behavioural benefits
- Define the commercial strategy

Introduction

Across the country and around the globe, digital innovators are helping us deliver our commitment to the digital transformation of health and social care, to bring benefits to patients, the workforce and the system as a whole. NHS England's Long Term Plan sets the direction towards widespread digitally-enabled care. The Secretary of State's Technology Vision goes on to articulate a clear ambition for the generation of more digital services designed around user need and adhering to key principles of privacy, security, interoperability and inclusion.

It is our duty as NHS England and central government to capitalise on these opportunities responsibly. The healthcare system is a unique space where a variety of regulatory ecosystems overlap. Due to the privileged nature of dealing with people's health and their protected data, the system is covered by various pieces of legislation as well as professional and ethical standards. Innovators in this field may come from sectors that are not necessarily familiar with medical ethics and research regulation, and may utilise data sets or processing methods that sit outside existing NHS safeguards.

At the same time, the wider NHS – patients, professionals, commissioners, purchasers – need a means of obtaining assurance and confidence across these domains so they



Digital Technology Assessment Criteria (DTAC)

For health and social care

The Digital Technology Assessment Criteria for health and social care (DTAC) gives staff, patients and citizens confidence that the digital health tools they use meet our clinical safety, data protection, technical security, interoperability and usability and accessibility standards.

It is the new national baseline criteria for digital health technologies into the NHS and social care. It is designed to be used by suppliers to build technology and healthcare organisations to build and to buy technologies that meet our minimum baseline standards.

The DTAC is available as [a document](#) (ODT, 132KB) that you can download and print.

Please download the assessment criteria at the point of use, to make sure you're using the most up to date version.



And more...

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

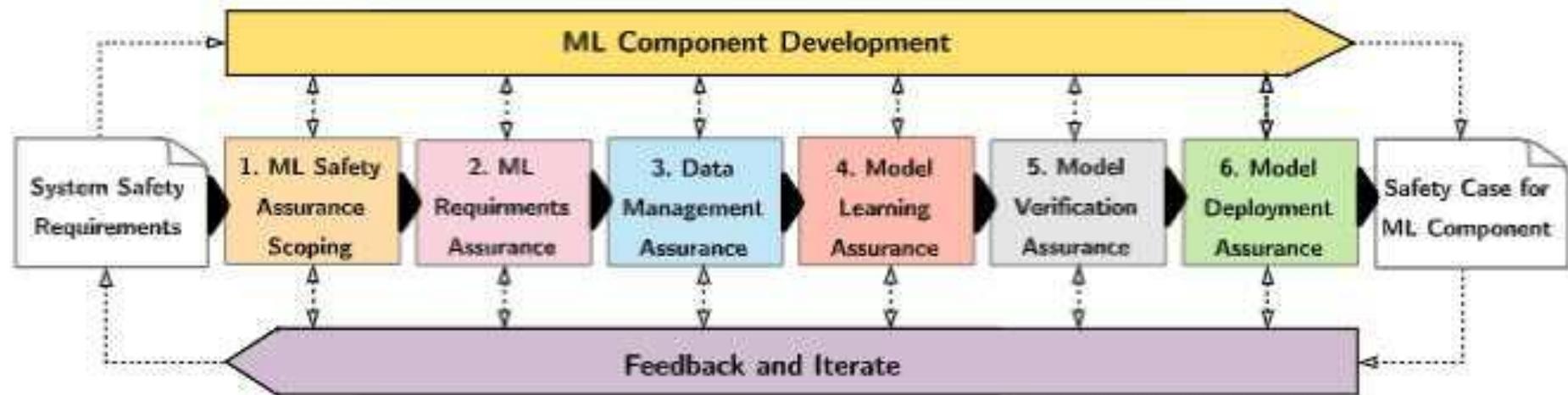
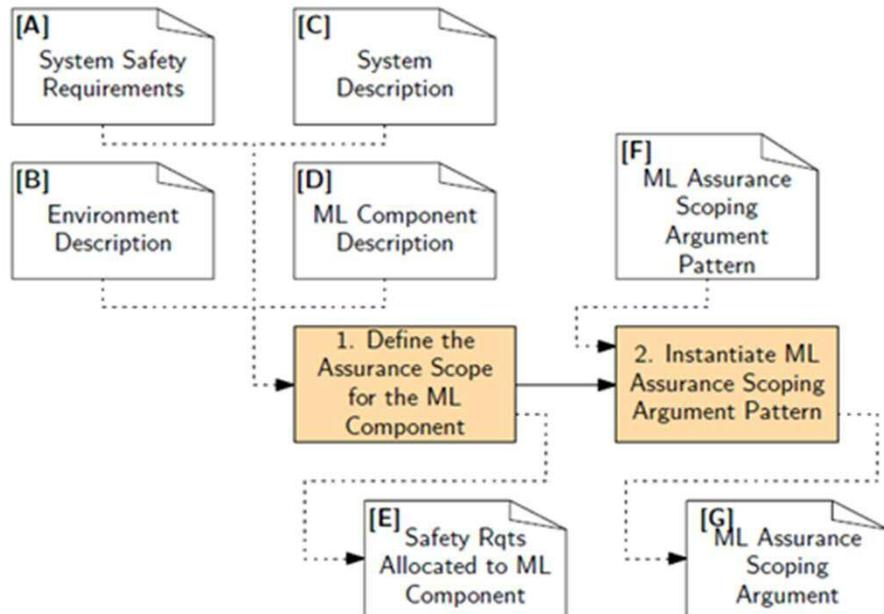


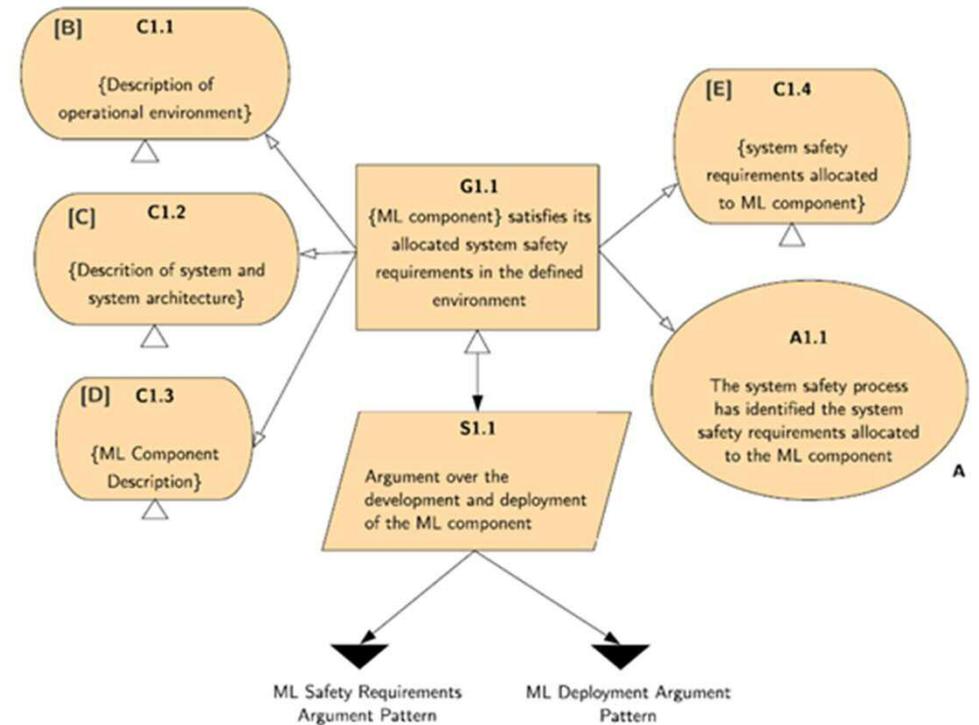
Figure 1: Overview of the AMLAS Process

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

Process



Argument Pattern



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

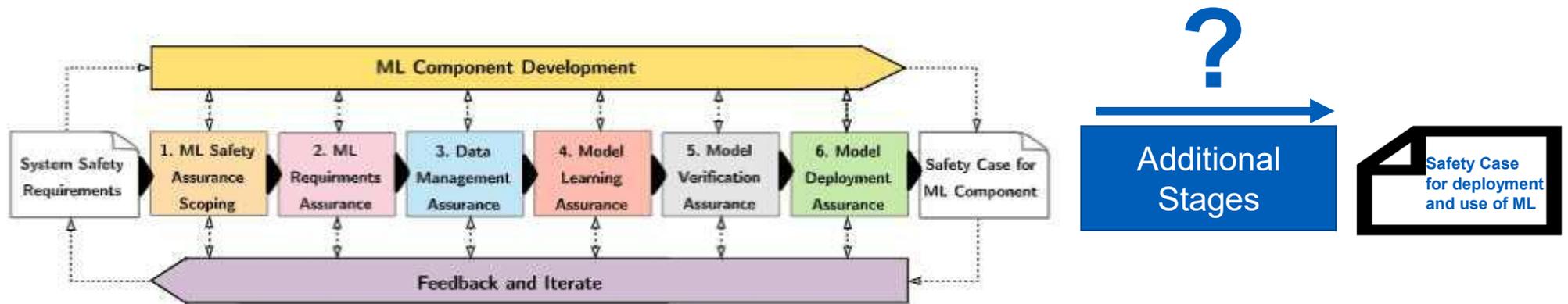
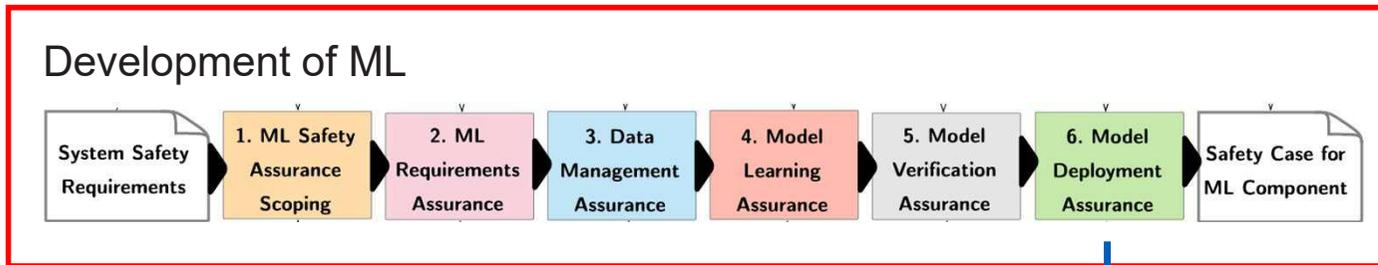


Figure 1: Overview of the AMLAS Process

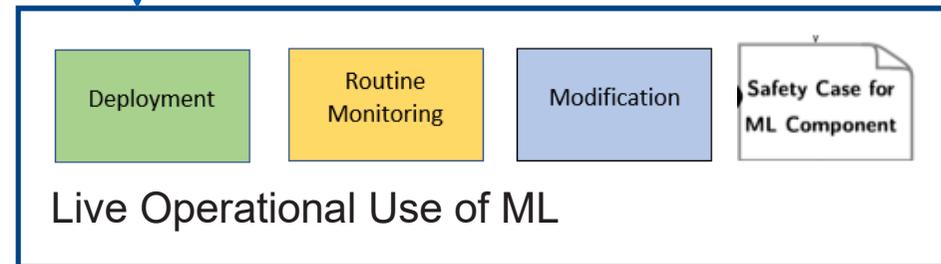
- Summary of findings:
- Development of healthcare specific guidance for AMLAS
- Develop additional stages for deploying healthcare organisations (Adopters)

Manufacturer



GAP

Adopter



Risk-Confidence-Conformance Approach

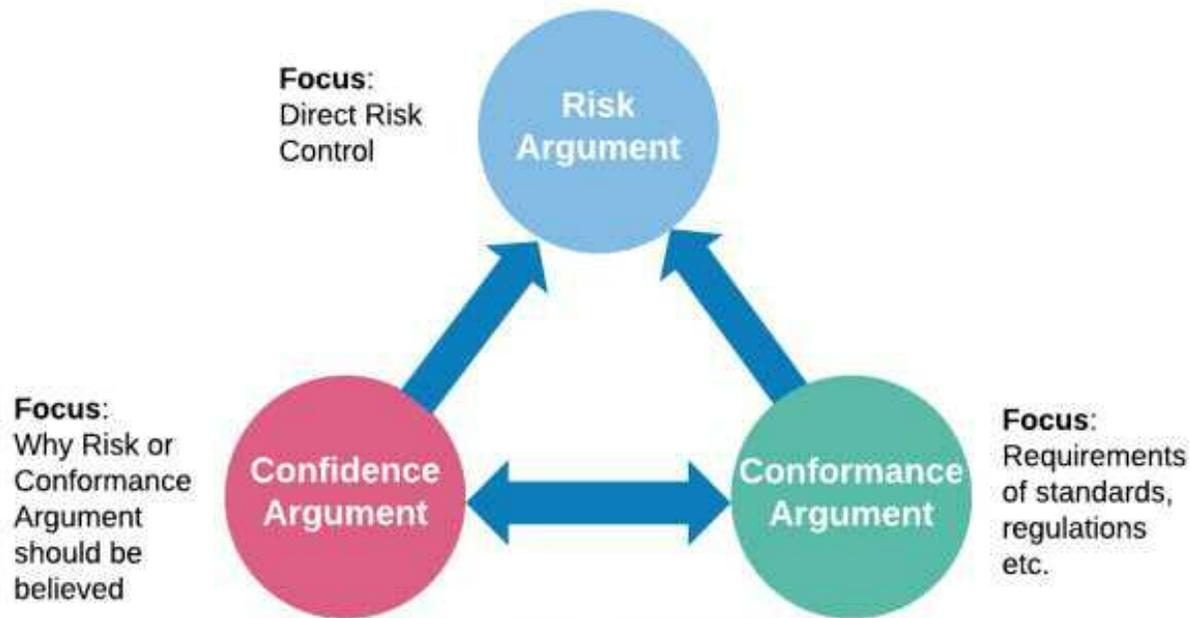
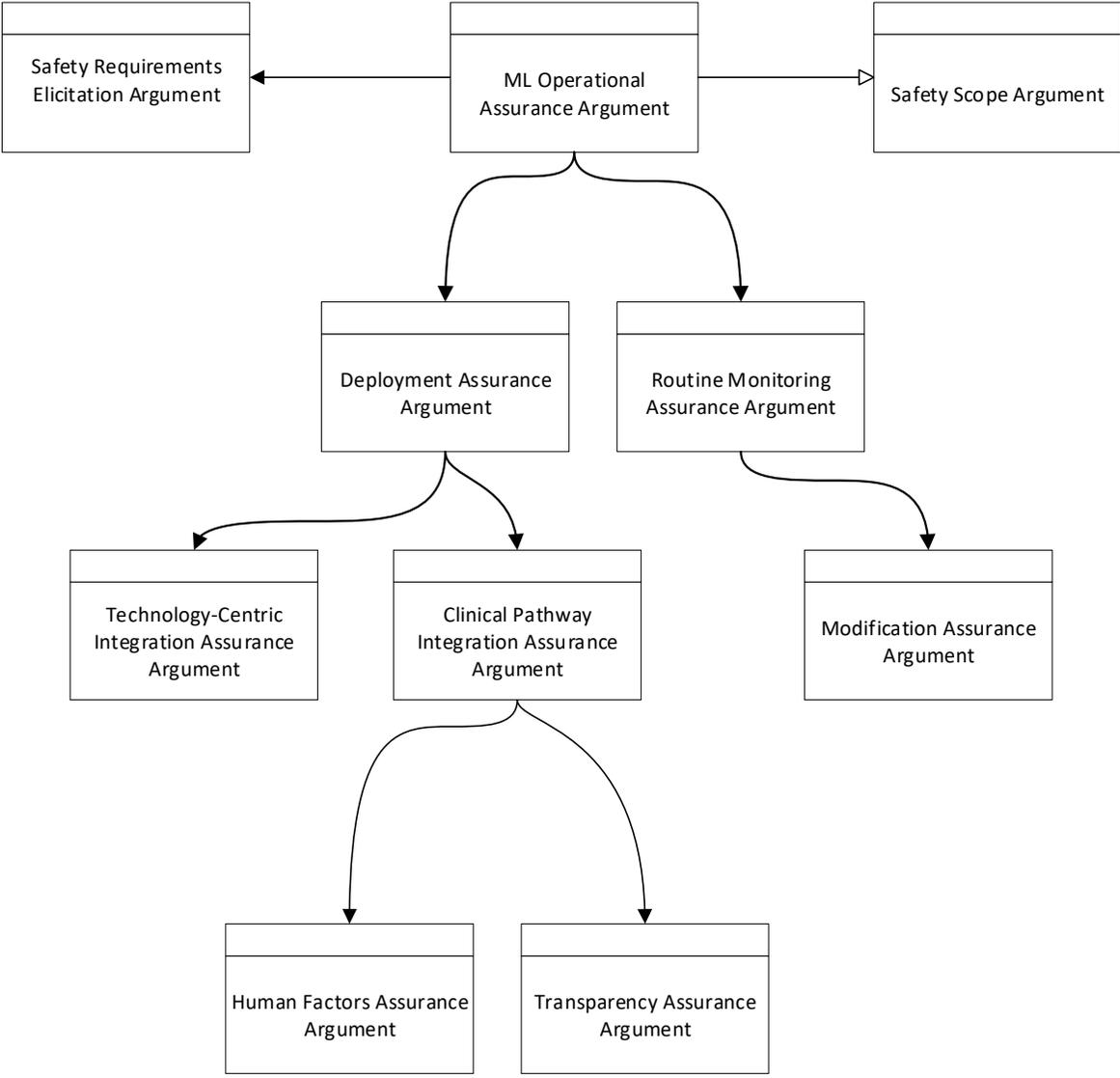


Figure 2:5-1 - Types of assurance case argument

Proposed Assurance Argument Overview



Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study

Paisan Ruamviboonsuk, Richa Tiwari, Rory Sayres, Variya Nganthavee, Kornwipa Hemarat, Apinpat Kongprayoon, Rajiv Raman, Brian Levinstein, Yun Liu, Mike Schaekermann, Roy Lee, Sunny Virmani, Kasumi Widner, John Chambers, Fred Hersch, Lily Peng, Dale R Webster

Summary

Background Diabetic retinopathy is a leading cause of preventable blindness, especially in low-income and middle-income countries (LMICs). Deep-learning systems have the potential to enhance diabetic retinopathy screenings in these settings, yet prospective studies assessing their usability and performance are scarce.

Methods We did a prospective interventional cohort study to evaluate the real-world performance and feasibility of deploying a deep-learning system into the health-care system of Thailand. Patients with diabetes and listed on the national diabetes registry, aged 18 years or older, able to have their fundus photograph taken for at least one eye, and due for screening as per the Thai Ministry of Public Health guidelines were eligible for inclusion. Eligible patients were screened with the deep-learning system at nine primary care sites under Thailand's national diabetic retinopathy screening programme. Patients with a previous diagnosis of diabetic macular oedema, severe non-proliferative diabetic retinopathy, or proliferative diabetic retinopathy; previous laser treatment of the retina or retinal surgery; other non-diabetic retinopathy eye disease requiring referral to an ophthalmologist; or inability to have fundus photograph taken of both eyes for any reason were excluded. Deep-learning system-based interpretations of patient fundus images and referral recommendations were provided in real time. As a safety mechanism, regional retina specialists over-read each image. Performance of the deep-learning system (accuracy, sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]) were measured against an adjudicated reference standard, provided by fellowship-trained retina specialists. This study is registered with the Thai national clinical trials registry, TCRT20190902002.

Google – Diabetic Retinopathy

- Real-world images lower quality than training set
- Technical hardware limitations
- Workflow variance

MHRA Work Packages

WP 4 Post market

1. Implement a robust post market surveillance system that produces a strong and clear safety signal, allowing for quicker and thorough capture of adverse incidents for SaMD
2. Utilise real world evidence to provide further intended, maintains performance, and continues to safety
3. Articulate clear change management requirements

WP 9 Project AI RIG (AI Rigour)

1. Utilise existing and broadly accepted frameworks to ensure AI as a medical device (AIaMD) placed on the market provides robust assurance that it is safe and effective
2. Develop supplementary frameworks to ensure AIaMD placed on the market provides robust assurance with respect to safety and effectiveness, with a special emphasis on ensuring that AIaMD is fit for purpose for all populations in which it is intended to be used
3. Develop technical methods to test AIaMD to ensure the device is safe and effective

WP 10 Project Glass Box (AI Interpretability)

1. Articulate how opacity in AIaMD can translate into safety or effectiveness issues
2. Develop frameworks regarding interpretability of AIaMD to ensure that AI models are sufficiently transparent to be robust and testable or are otherwise properly validated
3. Develop frameworks regarding interpretability of AIaMD to ensure that interpretability's relationship to usability is made plain and emphasised in relation to safety and effectiveness

CQC – Healthcare Services Regulator

Using machine learning in diagnostic services

A report with recommendations from CQC's regulatory sandbox

- Audit Services not ML
- Recognised need to update their approach

Questions?