

EVIDENCE-BASED AI SAFETY, ASSURANCE AND EVALUATION

Expertise and Partner Contribution for MSCA Doctoral Networks

LEAD CONTACT Dr Abdul Razzaq Assistant Professor, CSIS	ROLES OFFERED Beneficiary lead • WP leader Doctoral supervisor • Trainer	BEST-FIT NETWORKS Trustworthy AI • Agentic AI Autonomous and human-centred systems
--	---	---

01 PARTNER PROFILE AND CONSORTIUM VALUE

Dr Abdul Razzaq is an Assistant Professor in the Department of Computer Science and Information Systems at the University of Limerick. His work integrates software engineering with trustworthy and agentic AI, focusing on how claims about safety, reliability, explainability and governance can be supported by transparent, reproducible and verifiable evidence.

Why this expertise strengthens a Doctoral Network

Many AI-focused networks contain strong model-development projects but lack a common assurance layer connecting safety claims, empirical evidence, software behaviour, human oversight and real-world validation. UL offers this cross-network integration: shared methods, benchmarks, evidence structures and training that make individual doctoral projects scientifically coherent and collectively stronger.

02 CORE EXPERTISE

AI SAFETY & ASSURANCE Evidence-based safety claims, assurance cases, risk reasoning and lifecycle evaluation.	AGENTIC & TRUSTWORTHY AI Evaluation of reasoning, planning, tool use, autonomy, adaptation and unintended behaviour.	SOFTWARE ENGINEERING Architectures, traceability, empirical methods and dependable AI-enabled systems.
ROBUSTNESS & FAILURE Uncertainty, failure modes, drift, adversarial conditions and resilience assessment.	HUMAN OVERSIGHT Intervention, escalation, accountability, explainability and human-AI responsibility.	REPRODUCIBLE EVALUATION Benchmark design, cross-project protocols, provenance and evidence comparability.

03 CONTRIBUTION TO MSCA EVALUATION DIMENSIONS

DIMENSION	UL CONTRIBUTION
EXCELLENCE	Integrated evidence, evaluation and assurance methodology across doctoral projects; interdisciplinary research linking AI, software engineering and human oversight.
IMPACT	Training researchers to translate technical results into defensible safety and reliability claims; stronger industry, regulatory and standards relevance.
IMPLEMENTATION	A defined WP, shared protocols, deliverables, milestones, training modules, supervision and cross-project validation mechanisms.

03 A CROSS-NETWORK ASSURANCE PIPELINE

UL connects doctoral research outputs through a shared evidence pipeline so that technical results become defensible, auditable and reusable assurance knowledge.

1 CLAIMS Define what safety, reliability or control must be demonstrated.	2 EVIDENCE Link data, tests, software artefacts and operational observations.	3 EVALUATION Assess quality, coverage, robustness and uncertainty.	4 ASSURANCE Build transparent reasoning from evidence to justified conclusions.	5 LEARNING Update claims and evidence as systems and environments change.
---	---	--	---	---

CONCRETE CONTRIBUTION TO AN EXISTING CONSORTIUM

A ready-to-integrate work package, doctoral project and training offer

04 PROPOSED WORK PACKAGE

WP title: Evidence, Evaluation and Assurance for Trustworthy AI Systems

Objective: establish a common scientific and engineering framework that enables all doctoral projects to define, generate, connect and evaluate evidence supporting claims about AI safety, reliability, robustness, explainability and human control.

T1	SAFETY AND ASSURANCE REQUIREMENTS: DEFINE CLAIM STRUCTURES, EVIDENCE NEEDS AND DOMAIN-SPECIFIC ACCEPTANCE CRITERIA.
T2	Evidence and provenance: connect datasets, tests, model behaviour, software artefacts and human decisions to assurance claims.
T3	Evaluation and benchmarking: establish comparable cross-project protocols, metrics, baselines and reproducibility requirements.
T4	Robustness and failure analysis: assess uncertainty, drift, adversarial conditions, cascading failures and unintended behaviour.
T5	Human oversight and accountability: design intervention, escalation, explanation and responsibility-allocation mechanisms.
T6	Cross-project validation: synthesise evidence across doctoral projects and demonstrate assurance in realistic application settings.

05 PROPOSED DOCTORAL RESEARCH PROJECT

Doctoral project: Evidence-Based Assurance of Adaptive and Agentic AI Systems

The project will investigate how evidence about AI behaviour, robustness, uncertainty, human intervention and safety can be generated, linked, evaluated and maintained as systems plan, act, learn and change over time.

O1	O2	O3
Formalise evidence structures linking claims to tests, observations, software artefacts and operational context.	Develop methods for continuous assurance under adaptation, uncertainty, drift and changing environments.	Validate the approach through consortium use cases and compare assurance quality, coverage and decision usefulness.

06 TRAINING, SUPERVISION AND NETWORK SERVICES

NETWORK-WIDE TRAINING • Evidence-based AI safety and assurance • Empirical evaluation and benchmark design • Reproducibility, provenance and traceability • Software engineering for dependable AI • Research methodology, writing and communication	SUPERVISION & CONSORTIUM SERVICE • Recruitment and supervision of a UL doctoral researcher • Cross-institutional co-supervision and progress review • WP coordination, deliverables and milestone management • Hosting research visits, workshops and secondments • Industry-facing validation and knowledge exchange
---	--

07 TARGET CONSORTIUM THEMES

Trustworthy AI	AI safety	Responsible agentic AI	Human-AI collaboration
Autonomous systems	Cyber-physical systems	Clinical and health AI	Critical infrastructure

08 COOPERATION SOUGHT

Invitation to coordinators and consortium builders

University of Limerick is ready to join an MSCA Doctoral Network as a beneficiary and to lead the evidence, evaluation and assurance contribution. We welcome contact from coordinators seeking a partner that can connect technical AI research with software engineering, human oversight, reproducible evaluation and defensible safety claims. Contact Dr Abdul Razzaq through the MSCA Matchmaking Platform to discuss scientific fit, WP integration, doctoral-project links and consortium responsibilities.