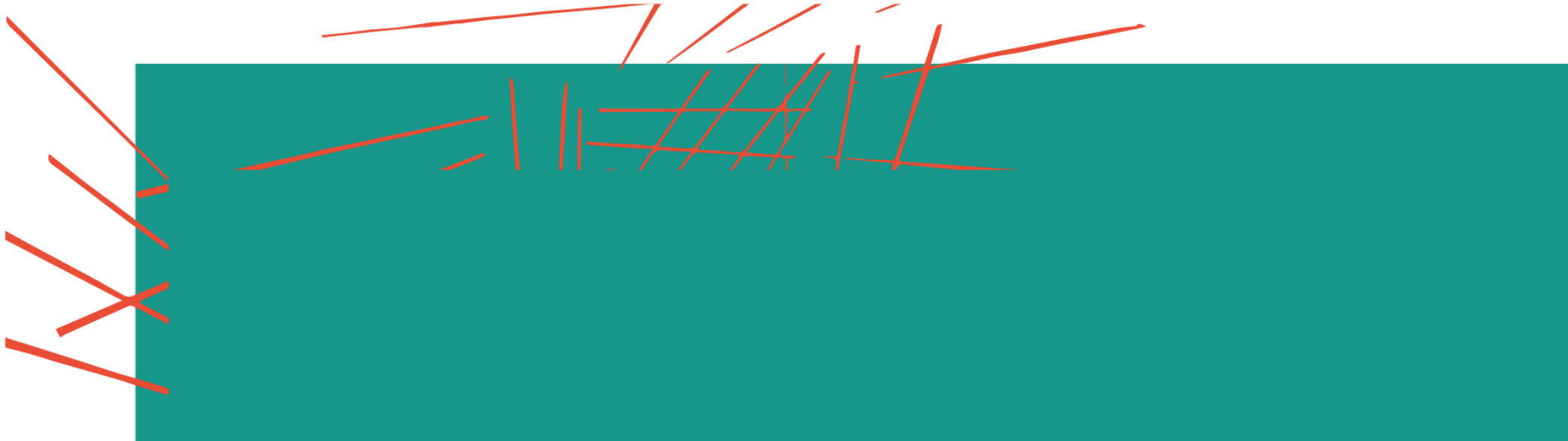
Engineering and Physical Sciences Research Council

EPSRC Update

Steph Williams, Senior Portfolio Manager, EPSRC Digital Security and Resilience theme
stephanie.williams@epsrc.ukri.org



Digital Security & Resilience

EPSRC has established a 'Digital Security & Resilience' (DS&R) theme to put a spotlight on digital technologies relevant to the security, defence, and resilience of the UK.

The research supported will aim **to create a more secure and resilient digital society, that is robust and prepared to withstand shocks and challenges in an increasingly interconnected digital world.**

Engineering and Physical Sciences Research Council

Digital Security & Resilience

Broadly, the Digital Security & Resilience theme's investments and activities fall within the following areas:

1. **Mitigating risk** - research to embed security and resilience within digital technologies.
2. **Creating opportunities** - research into the development of digital technologies and national capabilities that improve the security, defence and resilience of the UK, its organisations, systems, infrastructure and society.
3. **Shaping the future** - through collaboration and thought leadership, promote safe adoption of digital technologies and maximise impact on the UK economy and society.

UKRI Engineering and Physical Sciences Research Council

Digital Security and Resilience

Cybersecurity

- Cybersecurity research projects aligned to NCSC Research Institutes
- Digital Security by Design ISCF
- Securing Digital Technology at the Periphery SPF (SDTaP) and PETRAS
- Phase 3 of Centre for Secure Information Technologies IKC (CSIT)
- 4 Cybersecurity-related CDTs
- New EPSRC cybersecurity ecosystem investment

Digital Twinning

- EPSRC Hub for Applied Research in Digital Twinning for Decarbonising Transport
- UKRI Digital Twinning for Energy Grid Operation and Resilience
- UKRI Research Community Building and Thought Leadership in Digital Twinning
- EPSRC Core Research in Digital Twinning, with co-funding

Digital Resilience

- Defence-related investments
- RBOC+

Digital Security, Identity, Privacy & Trust

- Security, Privacy, Identity and Trust Engagement network plus (SPRITE+)
- Future of the Internet community engagement and research programme
- Digital Forensics part of the UKRI Forensics Sandpit

UKRI Engineering and Physical Sciences Research Council

Developing the cybersecurity research ecosystem

A community building activity to be awarded in 2024

A **key objective** of the UK Cyber Strategy: Strengthen the structures, partnerships and networks necessary to support a whole-of-society approach to cyber

By getting the right people working together in the right ways across the whole **public sector, industry and academia**, pulling together the whole cyber community.

Through the initiative, EPSRC intends to support national priorities while achieving long-term stewardship of the research ecosystem by:

- **Providing and improving connectivity** between components of the UK cybersecurity research ecosystem
- **Promoting knowledge exchange** between academia, industry, and government
- **Informing our EPSRC's future investment strategy** and our dialogue with government
- **Providing an international perspective** to enable targeted collaboration with international partners, as well as benchmarking and horizon scanning to inform prioritisation of challenges
- **Promoting regional strengths**, for example via interaction with relevant regional groups such as the UK Cyber Clusters

UKRI Engineering and Physical Sciences Research Council

Summary: Investments into the AI Ecosystem

Skills & Training

£117M for UKRI AI CDTs R2
£20M for Turing Fellowships R2 & R3

AI Hubs

£80M for AI Hubs
AI for Science and Engineering
AI for Real Data
Foundational AI

National Institute for Data Science & AI

The Alan Turing Institute

Transition to Turing 2.0

AI for Research Challenges

£13M for AI for Health

£13M for AI for Net Zero

AI Innovation Programme

UKRI Innovate UK

£100m BridgeAI programme

Underpinning infrastructure

AI Compute

Enabling a Responsible AI Ecosystem

£31M for Responsible AI UK

RAiUK

+

Ada Lovelace Institute

UKRI Arts and Humanities Research Council

Bridging Responsible AI Divides

Plus additional cross-cutting programmes, and BAU activities.

Workshop on Artificial Intelligence and Security

Security for all in an AI enabled society*

Lucas Cordeiro[1] and Mustafa Mustafa[1,2]

[1]Department of Computer Science, The University of Manchester
[2]imec-COSIC, KU Leuven, Belgium

*Engineering and Physical Sciences Research Council (EPSRC)
The Digital Economy Theme (part of UKRI)
https://enncore.github.io/events/secaiws/

Objectives

Discuss recent achievements and future initiatives to build trustworthy AI systems by taking an applied and interdisciplinary approach

- Establish new **partnerships/collaborations** to
  - discuss the challenges at the intersection between AI and cyber-security: **security for AI and AI for security**
  - tackle our main obstacles to achieving **widespread adoption of trusted and secure AI systems**
  - create a **portfolio of adventurous flagship projects** to demonstrate the viability of different research approaches

Funded EPSRC Research Projects

The Digital Economy Theme, part of UKRI, committed **£7,025,040** (at 80% FEC) to support **four research projects involving eleven UK universities**, addressing challenges at the **intersection between artificial intelligence and cyber security**

- **Security for Artificial Intelligence**
  - Securing Int. Systems across their lifecycle
  - Understanding of attacks and defenses
  - Detection of Degradation of Behaviour
  - Data Supply Chains
  - Explainable AI

- **Artificial Intelligence for Security**
  - Analysing and Utilising Outputs
  - AI and Human Interaction
  - Novel Approaches
  - The AI Security Tool Lifecycle

Funded EPSRC Research Projects

- **CHAI: Cyber Hygiene in AI enabled domestic life**
  - Queen Mary University of London
  - University of Bristol
  - University of Greenwich
  - University of Reading

CHAI: Cyber Hygiene in AI enabled domestic life

- Cyber security has traditionally benefitted from the user's alertness and simple recommendations to minimise exposure to cyber risk. This is not true for AI-enabled life

CHAI explores how to:

Make **AI easier for the user to understand when things go wrong**

Developed testbed (smart heating application for home) with simple AI (schedule learning) and explainability/transparency capabilities

Bristol

Developed impact graph-based diagnostic approach for non-experts

Greenwich

Train the user to better protect themselves against malicious AI

Univ. of Reading

Determining what factors influence personal preparedness against AI threats

Queen Mary

Determining optimal cyber hygiene measures based on AI attack graphs

Developed and conducted training experiment with 10 households

UCL

Funded EPSRC Research Projects

- **CHAI: Cyber Hygiene in AI enabled domestic life**
  - Queen Mary University of London
  - University of Bristol
  - University of Greenwich
  - University of Reading
- **SAIS: Secure AI assistantS**
  - Kings College London
  - Imperial College London

SAIS: Secure AI assistantS

- AI assistants (e.g., Alexa, Siri…) are widely deployed and used (7M daily users in the UK)
- Despite this, concerns persist concerning security, transparency, and privacy
- Goal: propose methods to **specify**, formally **verify** and **monitor** the **security behaviour** of AI assistants
- Cross-disciplinary collaboration between KCL and ICL, with non-academic parners

Funded EPSRC Research Projects

- **CHAI: Cyber Hygiene in AI enabled domestic life**
  - Queen Mary University of London
  - University of Bristol
  - University of Greenwich
  - University of Reading
- **SAIS: Secure AI assistantS**
  - Kings College London
  - Imperial College London
- **AISEC: AI Secure and Explainable by Construction**
  - Heriot-Watt University
  - University of Strathclyde
  - University of Edinburgh

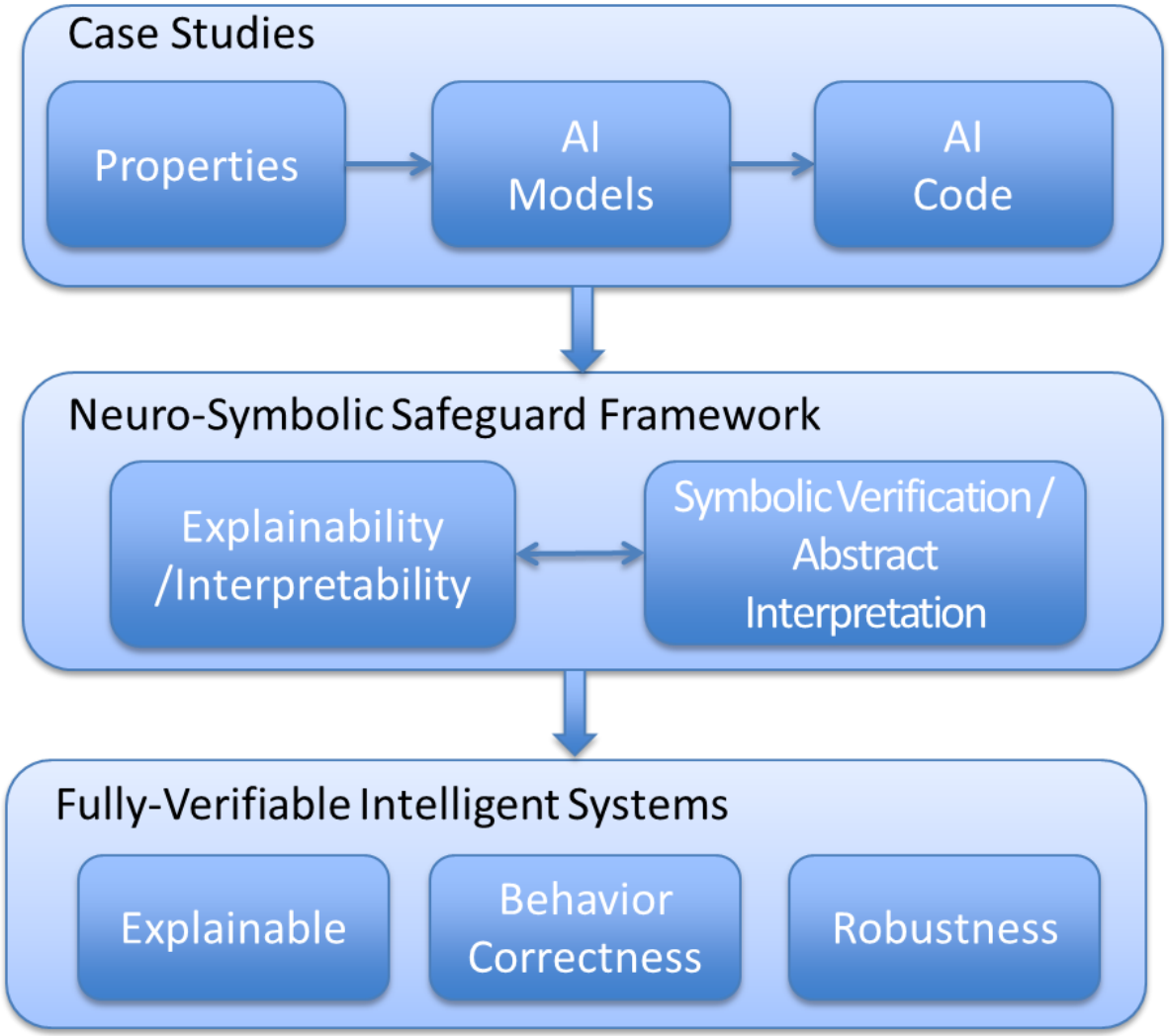
AISEC: AI Secure and Explainable by Construction

- Investigate **neural network verification** to ensure the safety and security of **complex intelligent systems**
- Design a tool Vehicle (poster session) that:
  - provides a **user-friendly language** for reasoning about properties of neural networks
  - leverages the power of existing **advanced neural network verifiers**
  - integrates **property-driven training** into neural network verification
  - enables **integration of neural network verification** into the verification of complex systems with **neural network components**
- Investigate applications: **NLP** (including LLMs) (talk today), **autonomous systems**, **neural networks in security applications** (Luca Arnaboldi)

Funded EPSRC Research Projects

- **CHAI: Cyber Hygiene in AI enabled domestic life**
  - Queen Mary University of London
  - University of Bristol
  - University of Greenwich
  - University of Reading
- **SAIS: Secure AI assistantS**
  - Kings College London
  - Imperial College London
- **AISEC: AI Secure and Explainable by Construction**
  - Heriot-Watt University
  - University of Strathclyde
  - University of Edinburgh
- **EnnCore: End-to-End Conceptual Guarding of Neural Architectures**
  - University of Manchester
  - University of Liverpool

EnnCore: End-to-End Conceptual Guarding of Neural Architectures



- Creation of the evaluation benchmarks
- Use case deployment & usability study

- Develop neural interpretability methods
- Reason over security properties in AI model/code

- Evaluation of security properties in real case studies from health and energy
- Validation of the results

EnnCore

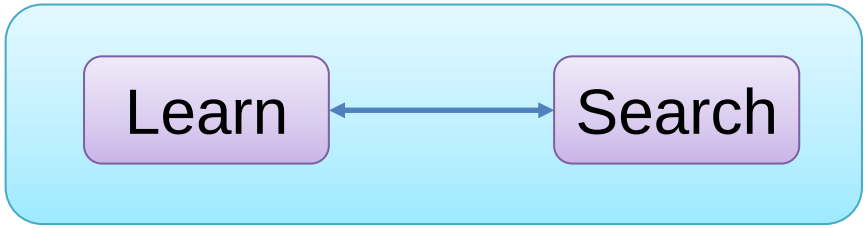
Program

Timing	Talk
12:30 - 13:00	Arrival, sandwiches, and coffee (Mercury room)
13:00 - 13:30	Welcome and introduction by Lucas Cordeiro, Stephanie Williams (EPSRC UKRI), Mustafa Mustafa (LT1.4)
13:30 - 14:30	Evaluating Privacy in Machine Learning (by Andrew Paverd) (LT1.4)
14:30 - 14:45	Coffee break (Mercury room)
14:45 - 15:15	A Tale of Two Oracles: Defining and Verifying when AI Systems are Safe (by Edoardo Manino) (LT1.4)
15:40 - 15:55	Coffee break (Mercury room)
15:15 - 15:45	One Picture Paints a Thousand Words: Using Abstract Interpretation for NLP Verification (by Marco Casadio) (LT1.4)
15:45 - 16:00	Coffee break (Mercury room)
16:00 - 16:30	Efficiently Training Neural Networks for Verifiability (by Alessandro De Palma) (LT1.4)
16:30 - 17:00	Cyber Hygiene in AI-enabled domestic life (by George Loukas) (LT1.4)
17:00 - 18:00	Drinks reception (Mercury room)
18:00 - 18:30	Free time
18:45 - 21:00	Dinner at Bem Brasil Deansgate (44 King St W, Manchester M3 2GQ)

The Bitter Lesson by Rich Sutton March 13, 2019

“The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation”

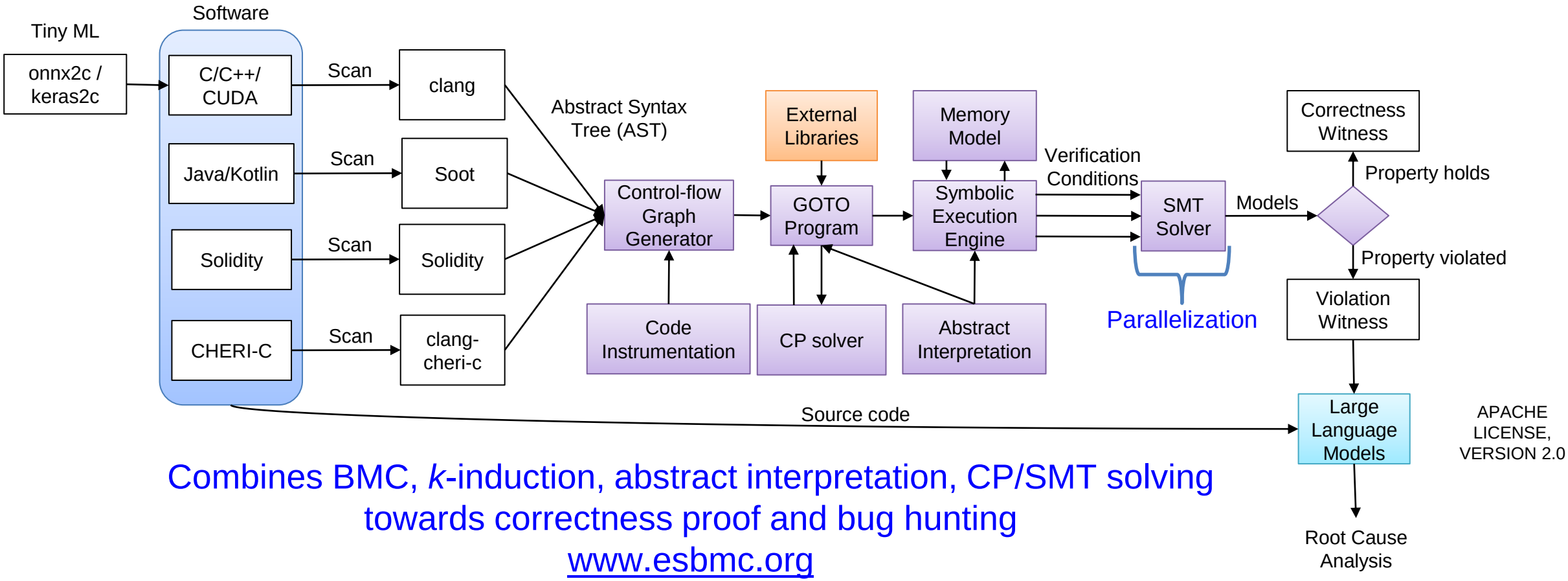
Parallelization



“The two methods that seem to scale arbitrarily in this way are *search* and *learning*”

ESBMC: Software Verification Platform

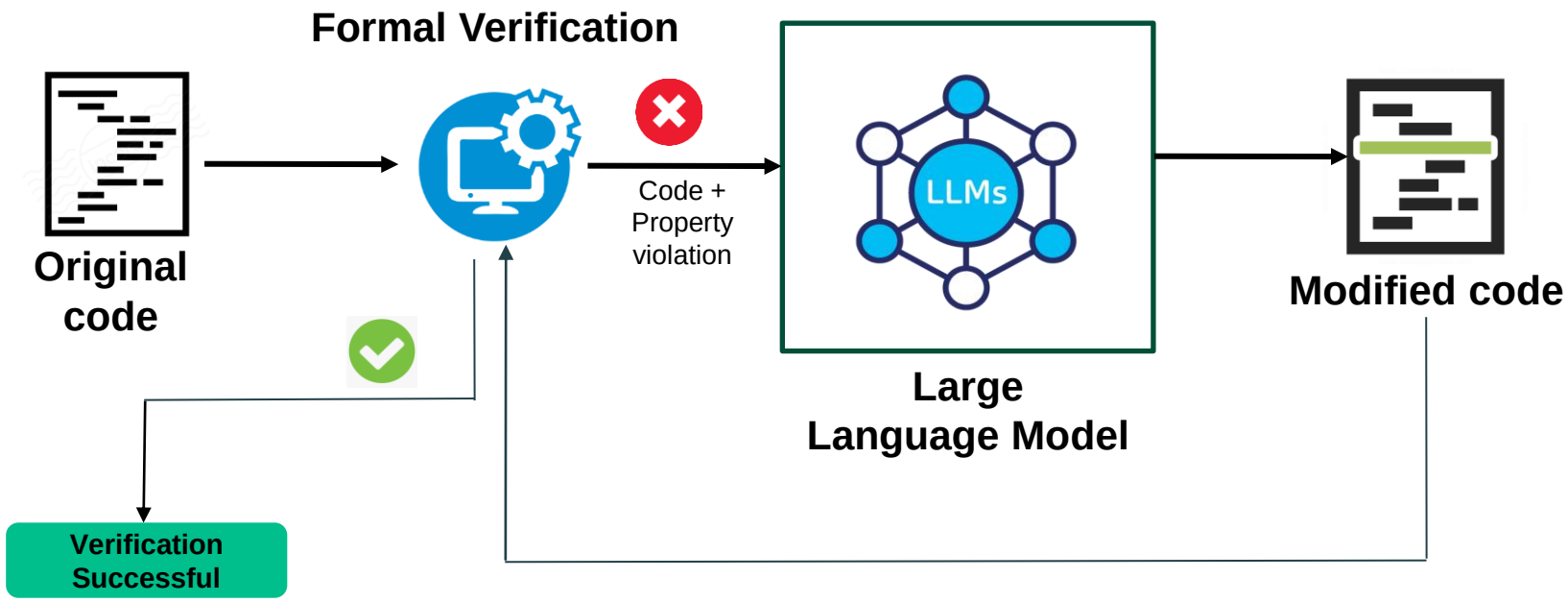
Logic-based automated reasoning for checking the **safety** and **security** of **single- and multi-threaded programs**



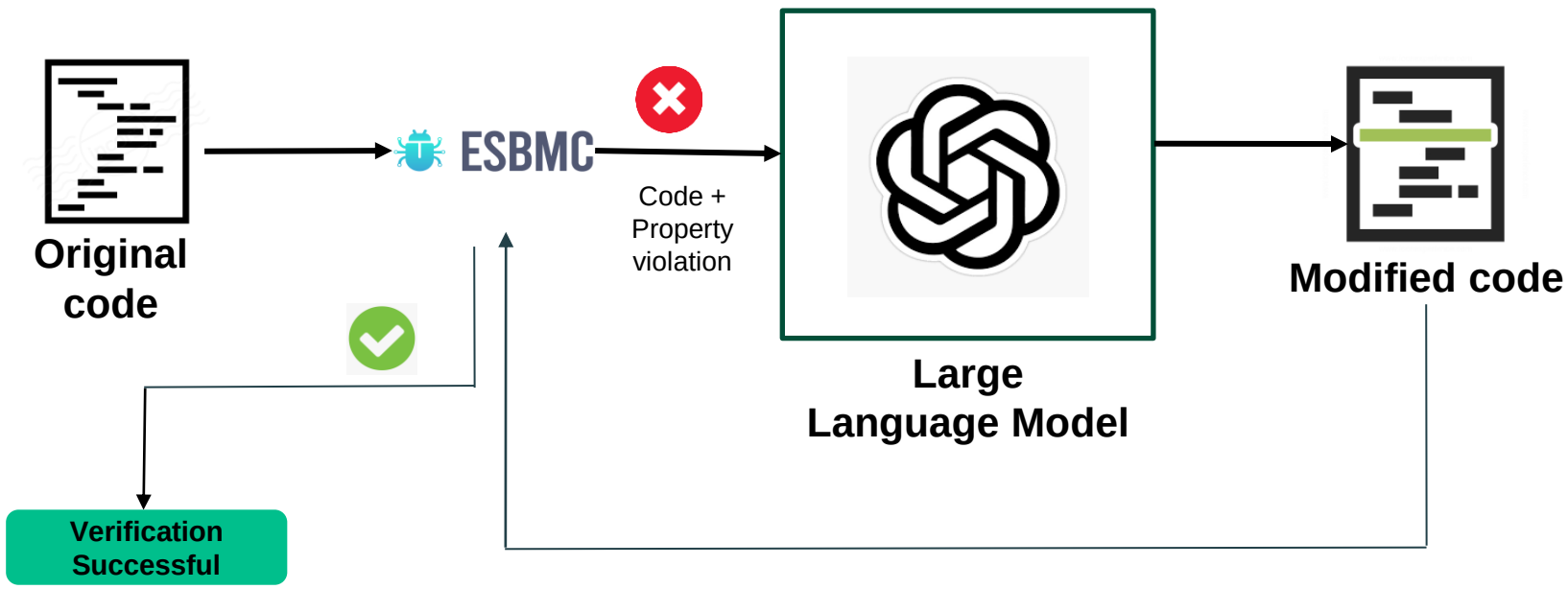
Combines BMC, *k*-induction, abstract interpretation, CP/SMT solving towards correctness proof and bug hunting

www.esbmc.org

Towards Self-Healing Software: LLM + Formal Verification



Towards Self-Healing Software: LLM + Formal Verification



Experimental Evaluation

Set-up

Code Generation

- **Processor:** AMD Ryzen Threadripper PRO 3995WX
- **Cores:** 16
- **RAM:** 256 GB

Code Repair

- **Model:** MacBook Pro (2017)
- **RAM:** 16 GB RAM of LPDDR3 RAM (2133 MHz)
- **Processor:** 2.5 GHz Intel Core i7-7660U

Benchmarks

Generate 1000 programs with GPT-3.5 turbo with the following prompt

Code generation prompt

Generate a minimum of 10 and a maximum of 50 lines of C code. Use at least two functions. Use strings, arrays, bit manipulations, and string manipulations inside the code. Be creative! Always include every necessary header. Only give me the code without any explanation. No comment in the code.

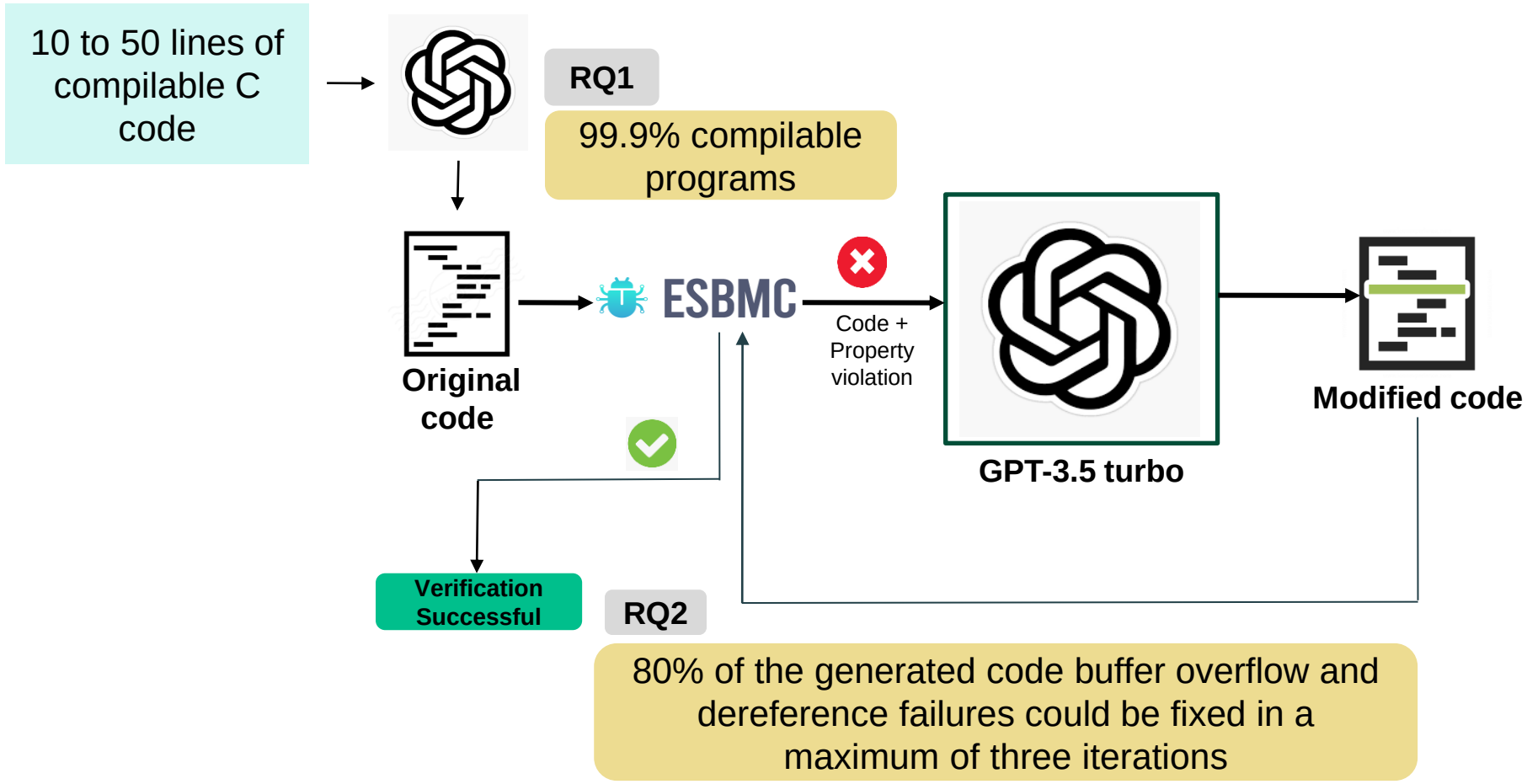
Objectives

To answer the following research questions.

RQ1: (Code generation) Are the state-of-the-art GPT models capable of producing compilable, semantically correct programs?

RQ2: (Code repair) Can external feedback improve the bug detection and patching ability of the GPT models?

Evaluation: Towards Self-Healing Software: LLM + Formal Verification



Technical Report, Dataset, and Soure Code

FORMAI DATASET: A LARGE COLLECTION OF AI-GENERATED C PROGRAMS AND THEIR VULNERABILITY CLASSIFICATIONS

formAI DATASET

Citation Author(s): Norbert Tihanyi (Technology Innovation Institute)
Tamas Bisztray (University of Oslo)
Ridhi Jain (Technology Innovation Institute)
Mohamed Amine Ferrag (Technology Innovation Institute)
Lucas C. Cordeiro (University of Manchester)
Vasileios Mavroeidis (University of Oslo)

165 Views

Categories: Artificial Intelligence, Security

Keywords: artificial intelligence, Software Vulnerability Dataset

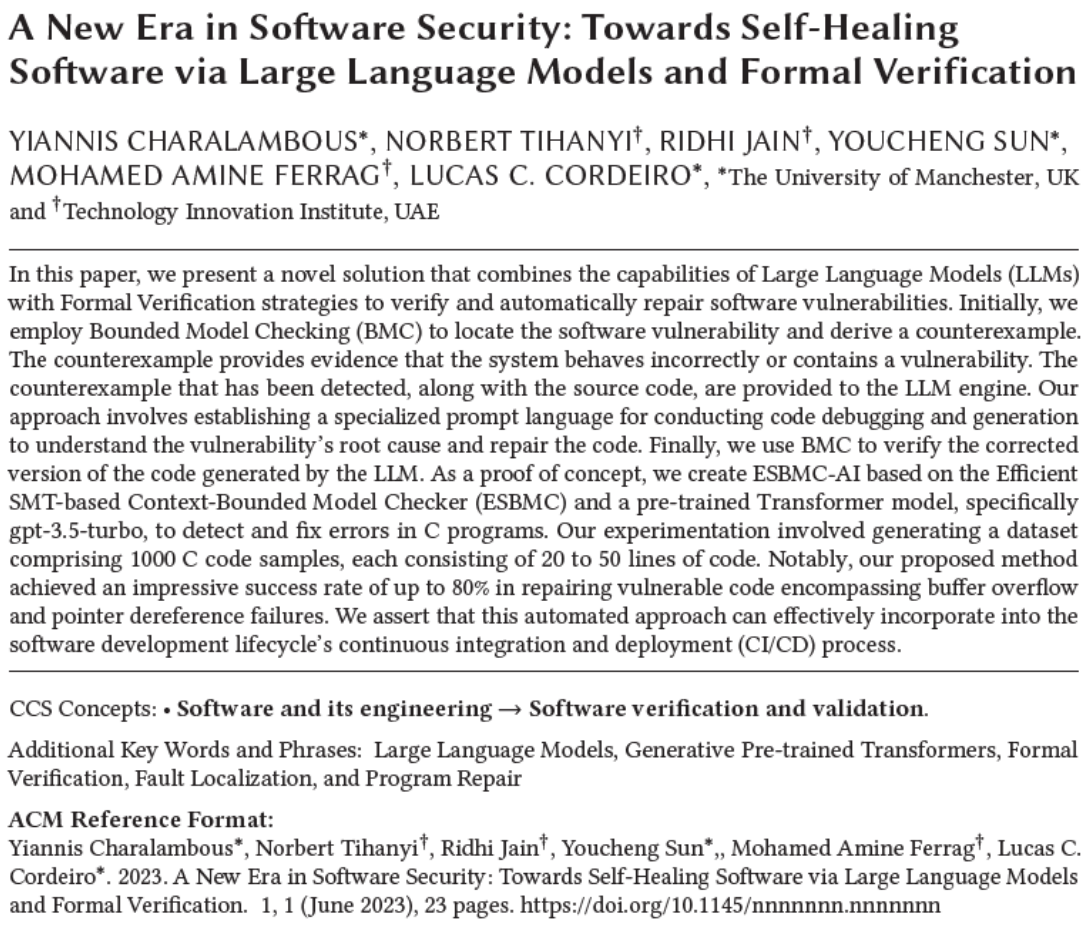
Submitted by: Norbert Tihanyi
Last updated: Mon, 06/19/2023 - 15:07
DOI: 10.21227/vp9n-wv96
Data Format: *.csv (zip);

Network Management, Table Games, Wi-Fi Signal Strength Analyzer, QR code reader, Image Steganography, Pixel Art Generator, Scientific Calculator Implementation, and Encryption, string manipulation, etc.

Dataset: https://ieee-dataport.org/documents/formai-dataset-large-collection-ai-generated-c-programs-and-their-vulnerability

Technical report: https://arxiv.org/abs/2305.14752

Source code: https://github.com/Yiannis128/esbmc-ai



A New Era in Software Security: Towards Self-Healing Software via Large Language Models and Formal Verification

YIANNIS CHARALAMBOUS*, NORBERT TIHANYI†, RIDHI JAIN†, YOUCHENG SUN*, MOHAMED AMINE FERRAG†, LUCAS C. CORDEIRO*, *The University of Manchester, UK and †Technology Innovation Institute, UAE

In this paper, we present a novel solution that combines the capabilities of Large Language Models (LLMs) with Formal Verification strategies to verify and automatically repair software vulnerabilities. Initially, we employ Bounded Model Checking (BMC) to locate the software vulnerability and derive a counterexample. The counterexample provides evidence that the system behaves incorrectly or contains a vulnerability. The counterexample that has been detected, along with the source code, are provided to the LLM engine. Our approach involves establishing a specialized prompt language for conducting code debugging and generation to understand the vulnerability's root cause and repair the code. Finally, we use BMC to verify the corrected version of the code generated by the LLM. As a proof of concept, we create ESBMC-AI based on the Efficient SMT-based Context-Bounded Model Checker (ESBMC) and a pre-trained Transformer model, specifically gpt-3.5-turbo, to detect and fix errors in C programs. Our experimentation involved generating a dataset comprising 1000 C code samples, each consisting of 20 to 50 lines of code. Notably, our proposed method achieved an impressive success rate of up to 80% in repairing vulnerable code encompassing buffer overflow and pointer dereference failures. We assert that this automated approach can effectively incorporate into the software development lifecycle's continuous integration and deployment (CI/CD) process.

CCS Concepts: • **Software and its engineering → Software verification and validation.**

Additional Key Words and Phrases: Large Language Models, Generative Pre-trained Transformers, Formal Verification, Fault Localization, and Program Repair

ACM Reference Format:
Yiannis Charalambous*, Norbert Tihanyi†, Ridhi Jain†, Youcheng Sun*,, Mohamed Amine Ferrag†, Lucas C. Cordeiro*. 2023. A New Era in Software Security: Towards Self-Healing Software via Large Language Models and Formal Verification. 1, 1 (June 2023), 23 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Systems and Software Security Research Group

L. Cordeiro

R. Banach

M. Mustafa

N. Zhang

B. Magri

Y. Sun

D. Dresner

A. Creswell

Centre for Digital Trust and Society

- A focal point for research across the University of Manchester that explores aspects of trust and security in our digital world
- USP -> Social science led cyber/digital security research agenda
- People
  - Director: Prof Nicholas Lord
  - Academic lead for Cyber Security: Prof Daniel Dresner
  - Lead for Early Career Researchers: Dr David Buil-Gil
- The centre constitutes of six clusters (more on the next slide)

MANCHESTER 1824
The University of Manchester

Research clusters

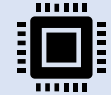
Digital Tech and Crime

David Buil-Gil
(Criminology)

Privacy and Trust

Mark Elliot
(Social Statistics)

Trusted Digital Systems



Mustafa Mustafa
(Computer Science) – S3, FM, MLO, APT

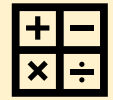
Workplace and Organisational Security

Richard Allmendinger
(AMBS)

Democracy and Trust

Peter Knight
(English and American Studies)

Advanced Mathematics



Theo Papamarkou
(Mathematics)

Projects

- External funding since 2018:
  - Wide range of projects, often >1 cluster
  - From seed corn to multi-million pound funding
  - Led by researchers across all career stages
  - Some with other DF themes

Project	Funder	Value (£K)
Soteria: Digital Security by Design	UKRI	£5800
i-Minds Digital Intervention for young people who have experienced online abuse	NIHR	£850
SPRITE+ (Security, Privacy, Identity, Trust NetworkPlus) (Combined Phase 1 2019-2023; Phase 2, 2023 to 2027)	UKRI	£4650
NW Partnership for Security and Trust	GCHQ	£760
National Centre for Research Methods (2020-24)	ESRC	£4250
VR and online child sexual exploitation	GCHQ	£32
Linguistic analysis of communications in online child sexual exploitation	UKRI	£30
ProvAnon (data anonymisation and provenance)	ATI	£71*
Manchester Digital Innovation and Security Hub	MCC	£200
CyberFoundry (project led from MMU)	ERDF	£1220*
PrivIoT: Understanding and mitigating privacy risks of IoT homes with demand-side management	PETRAS	£30*
Explainable AI for Digital Forensics Testing	UKRI	£120
Heilbronn Institute for Mathematical Research – North Note: funding confidential	GCHQ	(see note)
Developing Critical Mass in Cybersecurity	UKRI	£65
EnnCore (End-to-End Conceptual Guarding of Neural Architectures)	UKRI	£1720
Infodemic: Combatting COVID-19 Conspiracy Theories	UKRI	£278
The Intended and Unintended Consequences of Data-Driven Campaigning	NORFACE	£250
Digital Campaigning, Elections and Democracy	ERC	£2126
ELEGANT (Secure and Seamless Edge-to-cloud Analytics)	Horizon 2020	£508*
Digital Information Literacy: A Programme for Schools	ASPECT	£50
ScorCH (Secure Code for Capability Hardware)	ISCF via UKRI	£800K*
UKRI Impact Acceleration Account 373 EPSRC with SES Secure Ltd	UKRI	£86
Google ASPIRE Fund 2021 Program: Formal Verification Driven Fuzzing of Trusty OS	Google ASPIRE	£103 (US$118)
Manchester Turing Innovation Hub	Innovate UK	£4334
Total		£28333