IFIP WG 10.4 Workshop Winter'22

Karthik Pattabiraman, UBC
Marco Vieira, U. Coimbra
Mootaz Elnozahy, KAUST

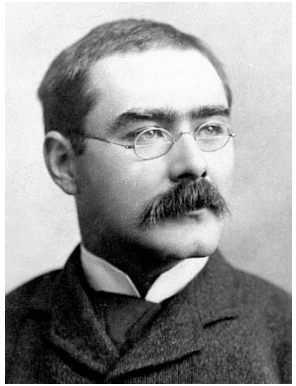
Workshop Theme

AI and Dependability: And the Twain Shall Meet

With due apologies to Rudyard Kipling (1865-1936)

"*East is East and West is West,*

And never the twain shall meet"



State of AI/ML today

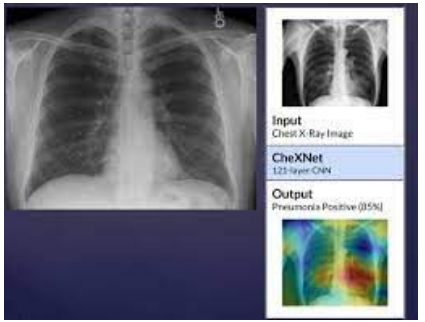
Most AI is ML, and in particular using CNNs (Convolutional Neural Networks)

Used in safety-critical applications - self-driving cars, home robots, healthcare, etc.

Very little guarantees on correctness, formal or otherwise

Easy to attack ML systems (evasion, poisoning attacks, privacy leaks etc.)

Issues around fairness, explainability, ethics ….



What can the dependability community offer AI ?

Dependability community has a long history of building safety-critical systems

- Avionics, healthcare, Banking etc.
- Lot of wisdom on what works, what doesn’t work etc. at scale

But, AI is different …. (non-exhaustive list)

- No clear notion of correctness, or even what’s an acceptable outcome
- Difficult to peek into the black box, test the system for corner cases etc.
- Not clear what’re the blind spots, failure modes and how to avoid them
- AI systems have much larger attack surface than conventional software

What can AI offer the dependability community ?

Provides new ways to analyze and understand failure data and extract patterns

Can help automate or speed up dependability evaluation (e.g., fault injection)

Software testing and verification can be considerably improved with AI

AI can also help with software reliability and security modeling

Attempt at a Definition of Dependable AI/ML

A dependable AI/ML system is one that satisfies the following constraints:

1. Satisfies the specification of the overall system (Reliability)
2. Does not perform harmful or unsafe actions (Safety)
3. Does not allow attackers to compromise it or cause malfunctioning (Security)
4. Allows users to understand how decisions were made (Transparency)
5. Does not suffer from bias due to flawed or incomplete training (Fairness)
6. Allows society to place their trust in the overall system (Legal, Ethical, etc.)

Any other thoughts ?

Workshop Agenda (All times are in CET)

Jan 20th (Day 1): Total time: 3 hours

Introductions (10 mins) - 2:00 PM

Session 1: Software Engineering and Testing (2:10 PM to 3:30 PM)

1. Michael Lyu, CUHK, *Software Dependability Modeling with A Data-Driven AI Paradigm*
2. Lionel Briand, Univ. of Ottawa, *Trustworthy Machine Learning-Enabled Systems*

10 minute break

Session 2: Self-Driving Cars and Safety (3:40 PM to 5:00 PM)

1. Paolo Rech, University of Trento, *Can we Rely on Self-Driving Cars? Evaluation and Mitigation of Neutron-Induced Errors in Convolutional Neural Networks for Autonomous Vehicles*
2. Daniel Schneider, Fraunhofer IESE, *Safety-critical systems with Machine Learning components: Challenges and Solutions*

Jan 21st (Day 2): Total time: 3 hours

Session 1: Safety and other considerations (2:00 PM to 3:20 PM)

1. Andre Lourenco, *CardioID, Integrating Physiological Monitoring in the Industry*
2. Timothy Tsai, Nvidia, *What Safety Challenges for Autonomous Systems Would Benefit from Research?*

10 minute break

Session 2: Summary and free form discussion (3:30 PM to 4:50 PM)