34th Annual INCOSE international symposium

hybrid event

Dublin, Ireland
July 2 - 6, 2024

Friday 5 July 10:30am (note session starts at 9, speakers breakfast at 8am)
Presentation only #430

Safer Complex Systems

Overview

- Introduction and Motivation
- Introduction to Complex Systems Safety
- Highlights from Governance and Practice
- Case study on Human-AI Interaction

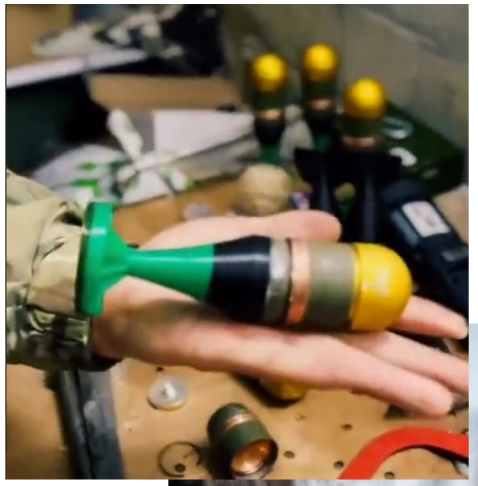
The presenters

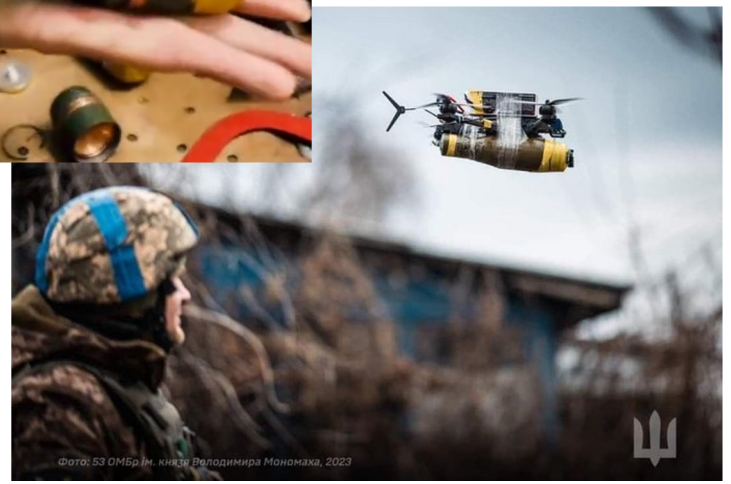
Duncan Kemp
duncan.kemp@incose.com

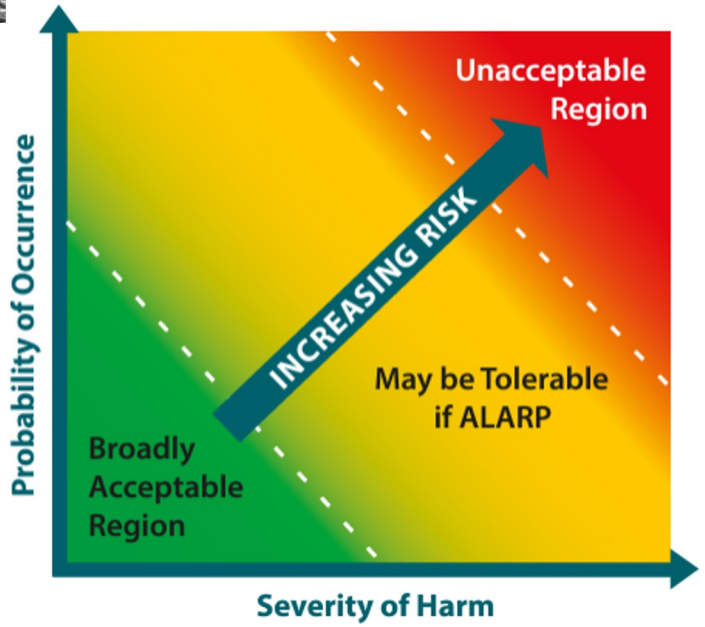
Meaghan O'Neil
System Design and Strategy
moneil@systemdesignstrategy.co.uk

Hortense Gerardo
University of California, San Diego
Hgerardo@ucsd.edu

Loss, ALARP and Tolerability







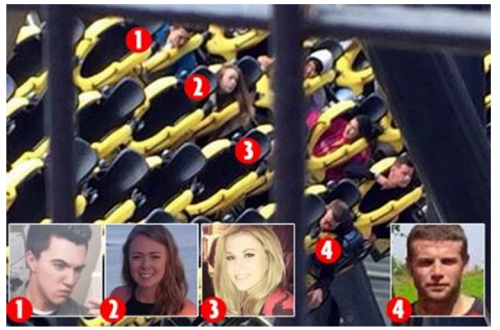
Source of hazards

Management System

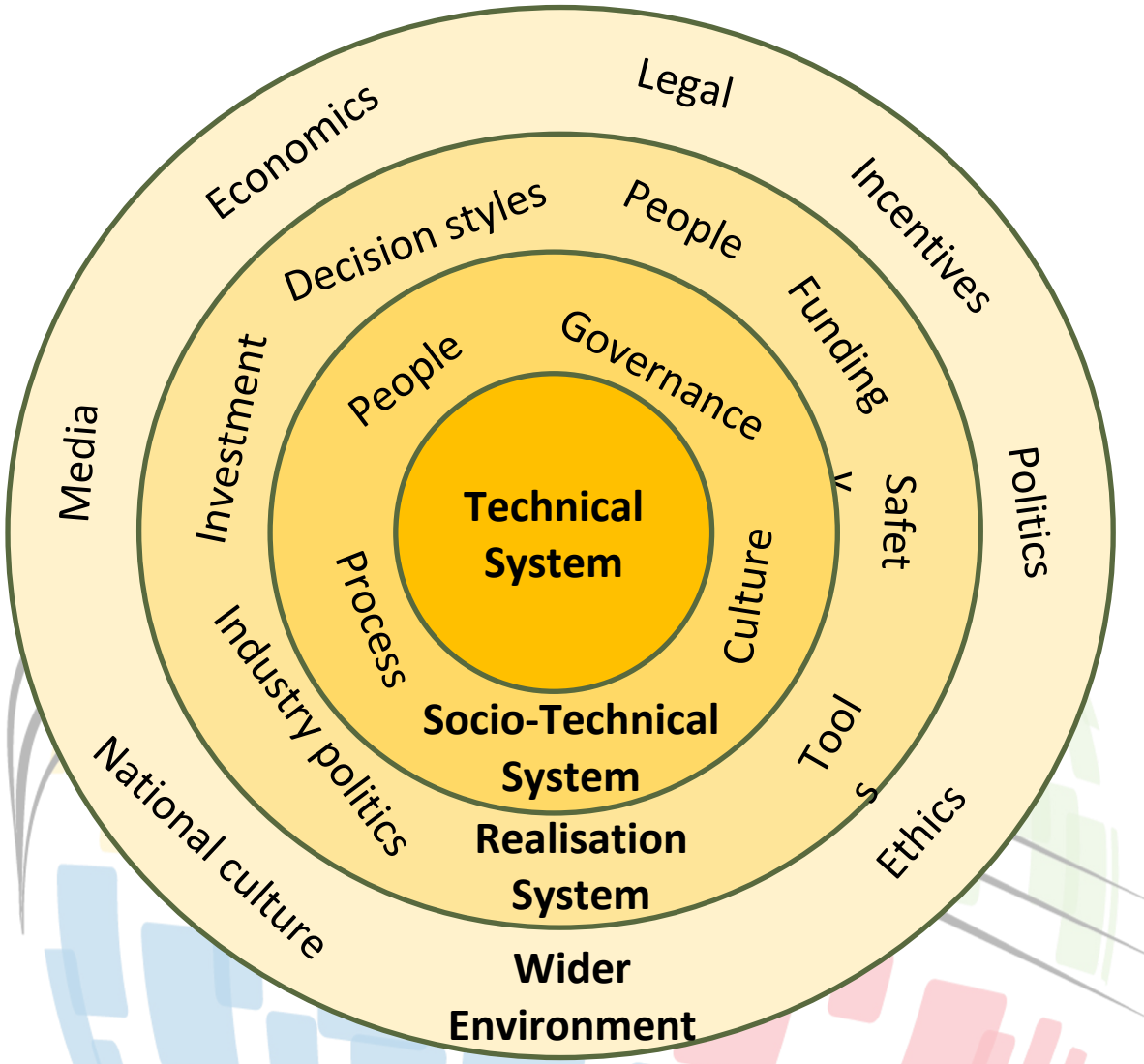
Monitoring and control

Physical System

Sub-system



The systems landscape



What are the Systems Safety challenges for systems containing artificial intelligence, quantum technology, additive manufacture, … ?

How do we enable safer socio-technical systems using psychology, system science, social science … ?

How can we do better Systems Safety Engineering using artificial intelligence, quantum technology, psychology, social science, leadership and management … ?

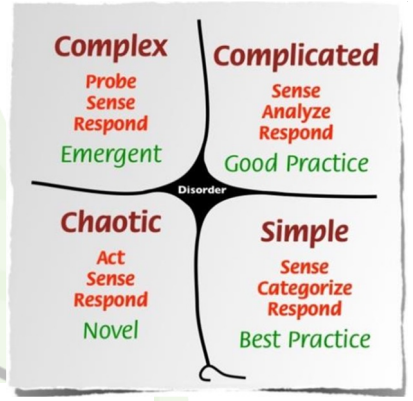
How do we influence the wider environment to help us realise safe and effective systems?

Complexity and safety

"We don't really know how to manage safety in complex systems"

"This is complex, what you need to do is …"

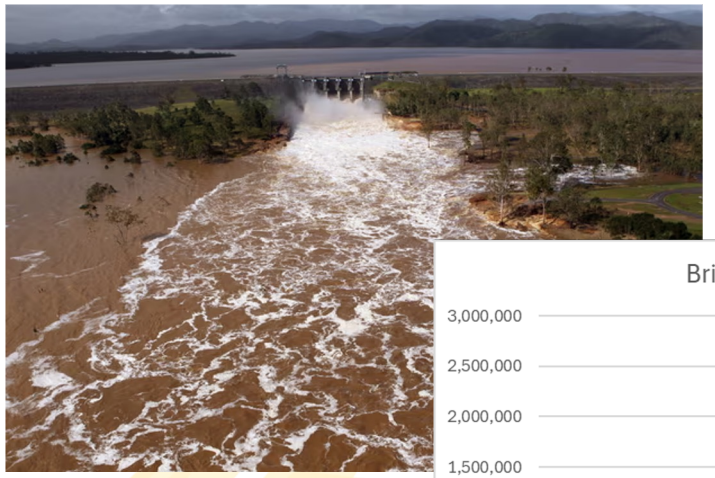
"You need to commission me, the expert, to fix this. We just need better analysis, more modelling and simulation. Let's build a digital twin!"

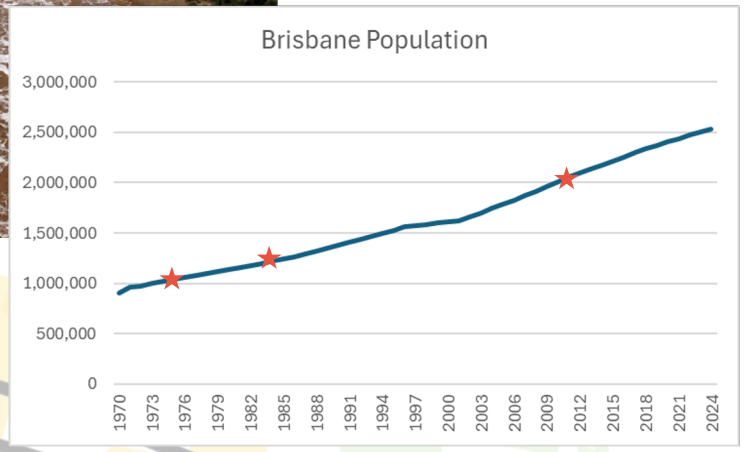


"The system is too complex to manage. Just leave it. If it does fall over, then I will take charge and tell you what to do"

"You are over-complicating things. Just tell me the rules and regulations, and I will follow them. Business needs certainty."

Key safety challenges of complex systems





Different people see the system differently

Unpredictable / opaque

Fuzzy / open boundaries

Significant interdependencies

Nonlinear / dynamic feedback

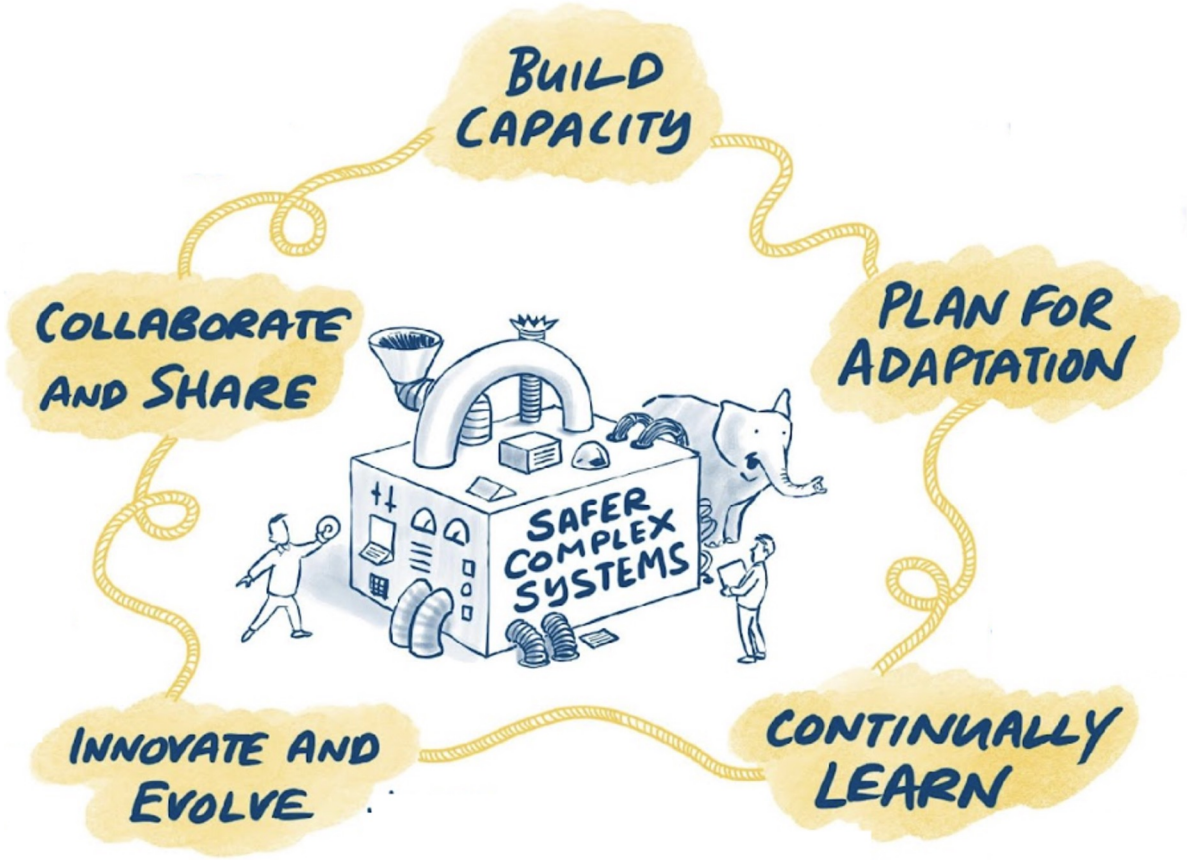
Novel / unplanned emergence

Emergent architecture

Managing safety of complex systems

Action	Why	How
Continue application of existing safety techniques where applicable	• Don't "throw the baby out with the bathwater" • The current approaches work when used in the right space and the right way	• Don't give in to pressure to over-simplify or remove regulations because 'it's all too complex, so lets simplify things'
Increased use of existing safety techniques where they will add value	• Even complex systems aren't wholly complex all the time • Some people think their system is complex – but it isn't!	• Continue to define procedures and rules for simple systems • Continue to develop safety management systems built around a safety argument for complicated ones.
Develop the safety approach for the complex space	• Management of complex situations is different to management of complicated, simple or chaotic ones	• Develop and use new modelling and analysis approaches to make the (subjectively) complex complicated • Develop new approach that can meet the challenges of complex system safety for the genuinely complex systems
Develop an approach to help people recognise they are in the complex space	• Understanding the situation you are in is harder that it appears • People prefer the approaches they are familiar with	• Develop tools to help understand the genuinely complex from the highly complicated • Develop tools to understand when situation is changing • Educate teachers, researchers, practitioners and key stakeholders that different situations require different approaches

Highlights from Work on Governance



Highlight: Formal/Informal Governance

Value and relationship of formal and informal governance

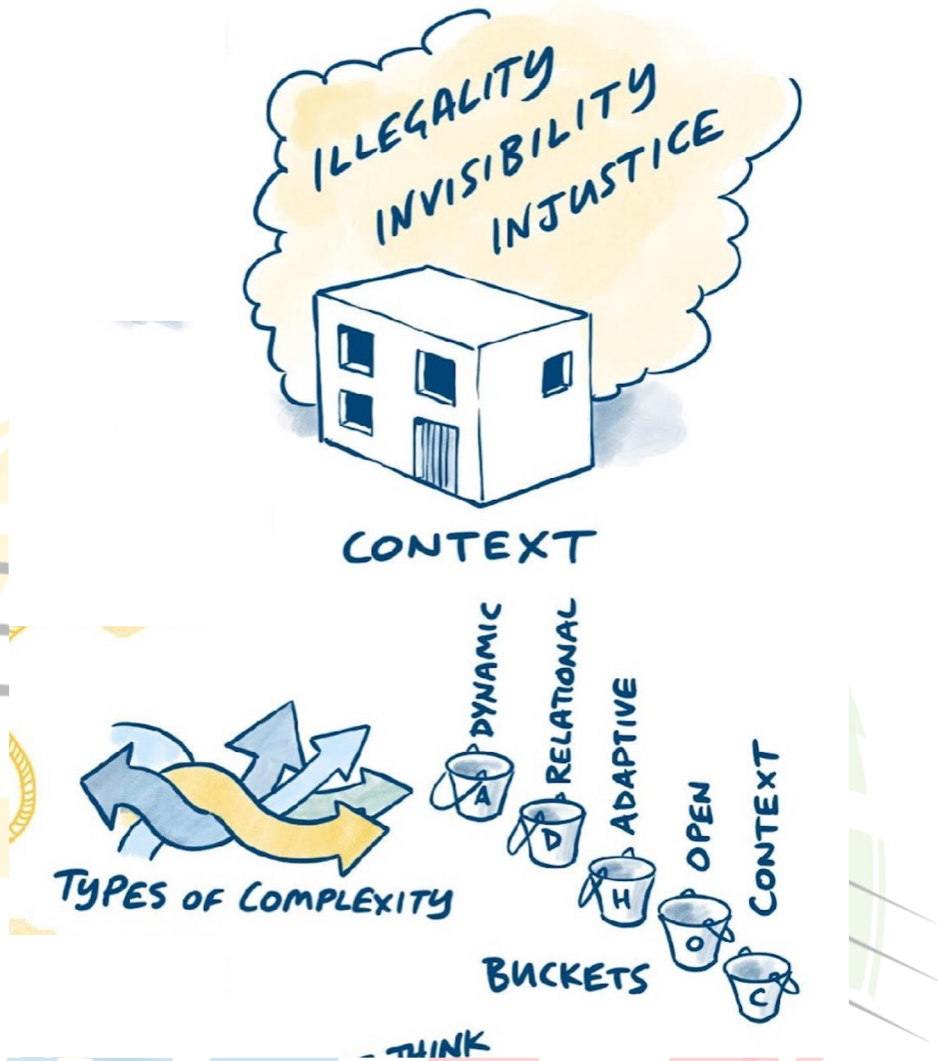
Highlight: Planned Adaptation

How do you make a rule (or a law) for something you don't understand?

Highlight: Planned Adaptation

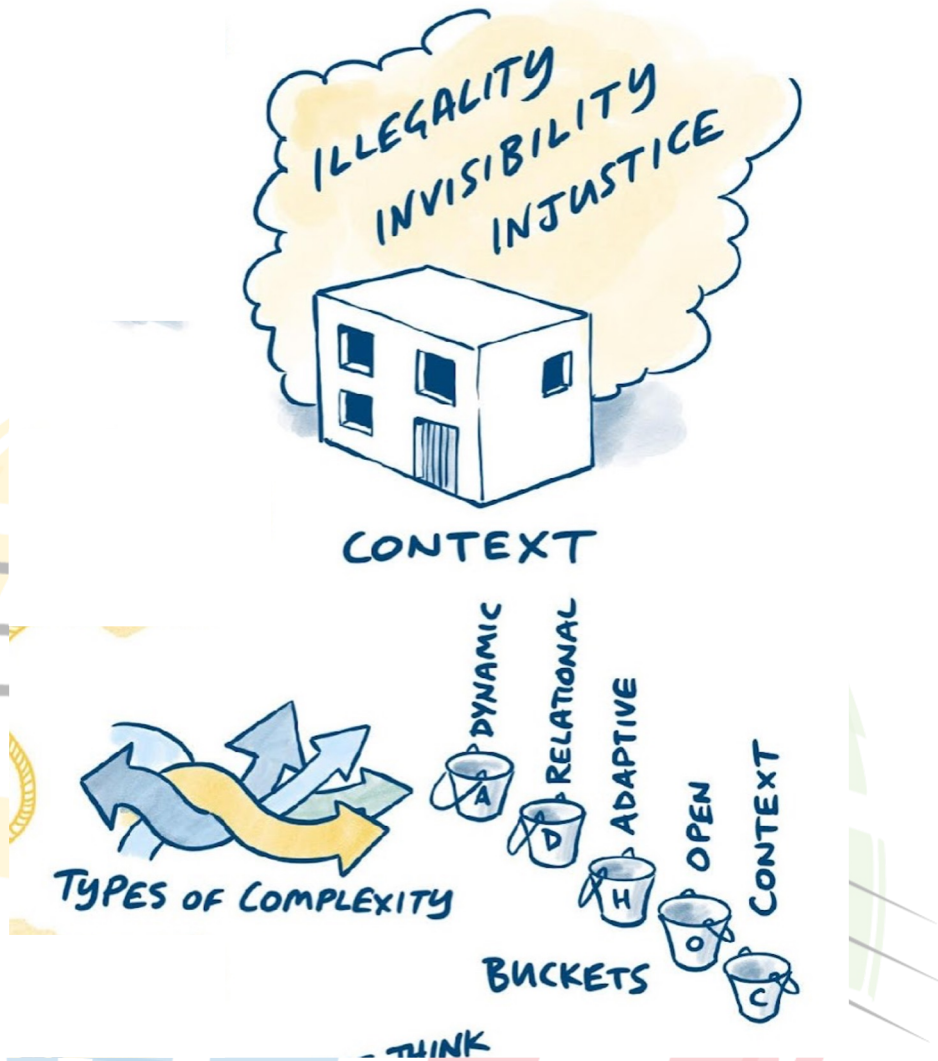
Developing continuous learning capability

Highlight: Context



Context matters

Highlight: Context



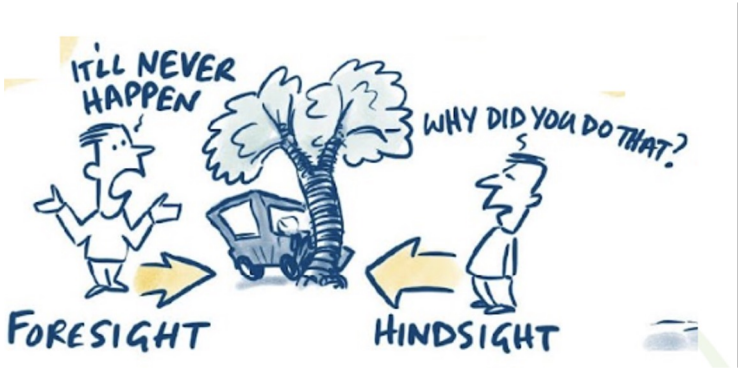
One size fits "none"

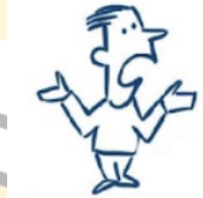

GOVERNANCE
DIVERSE VOICES
INFORMAL
BOUNDARY
FORMAL
SYSTEMS OF EXCLUSION
PRINCIPLES
WHY?
RULE MAKING
FAIRS
WE'RE STUFFED!
WHAT FORECAST?
IT'S STILL WORTHWHILE
COMPLEXITY
NAVIGATE
PLAN-LEARN-ADAPT
NOT CONNECTED
UNRELIABLE
GRID
BUILD CAPACITY
COLLABORATE AND SHARE
INNOVATE AND EVOLVE
PLAN FOR ADAPTATION
CONTINUALLY LEARN
SAFER COMPLEX SYSTEMS
IT'S BETTER THAN NOTHING!
SAFETY
ILLEGALITY
INVISIBILITY
INJUSTICE
CONTEXT
TYPES OF COMPLEXITY
DYNAMIC
RELATIONAL
ADAPTIVE
OPEN
CONTEXT
BUCKETS
IT'LL NEVER HAPPEN
WHY DID YOU DO THAT?
FORESIGHT
HINDSIGHT
OH NO! NOT AGAIN!
JUST THINK WHAT WE LEARNED!
I'M NOT TO BLAME!

Challenges in Practice

that's complex!

that's complicated!





that proposed intervention sounds too obvious and simple…. it can't be right!

what is the system boundary?

SAFER COMPLEX SYSTEMS:
Case study on Human-AI Interaction

International Council on Systems Engineering International Symposium (INCOSE IS)
Dublin, Ireland July 5, 2024

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Anthropology, Performance,
and Technology Program

Hortense Gerardo, Ph.D
the Anthropology, Performance, and Technology Program
Jacobs School of Engineering
University of California, San Diego

THE APT PROGRAM

BUILDING THE EXPERIENTIAL LANGUAGE OF ENGINEERING

ANTHROPOLOGY

a language to provide engineering students a multicutural perspective of what the field can achieve. (teaching).

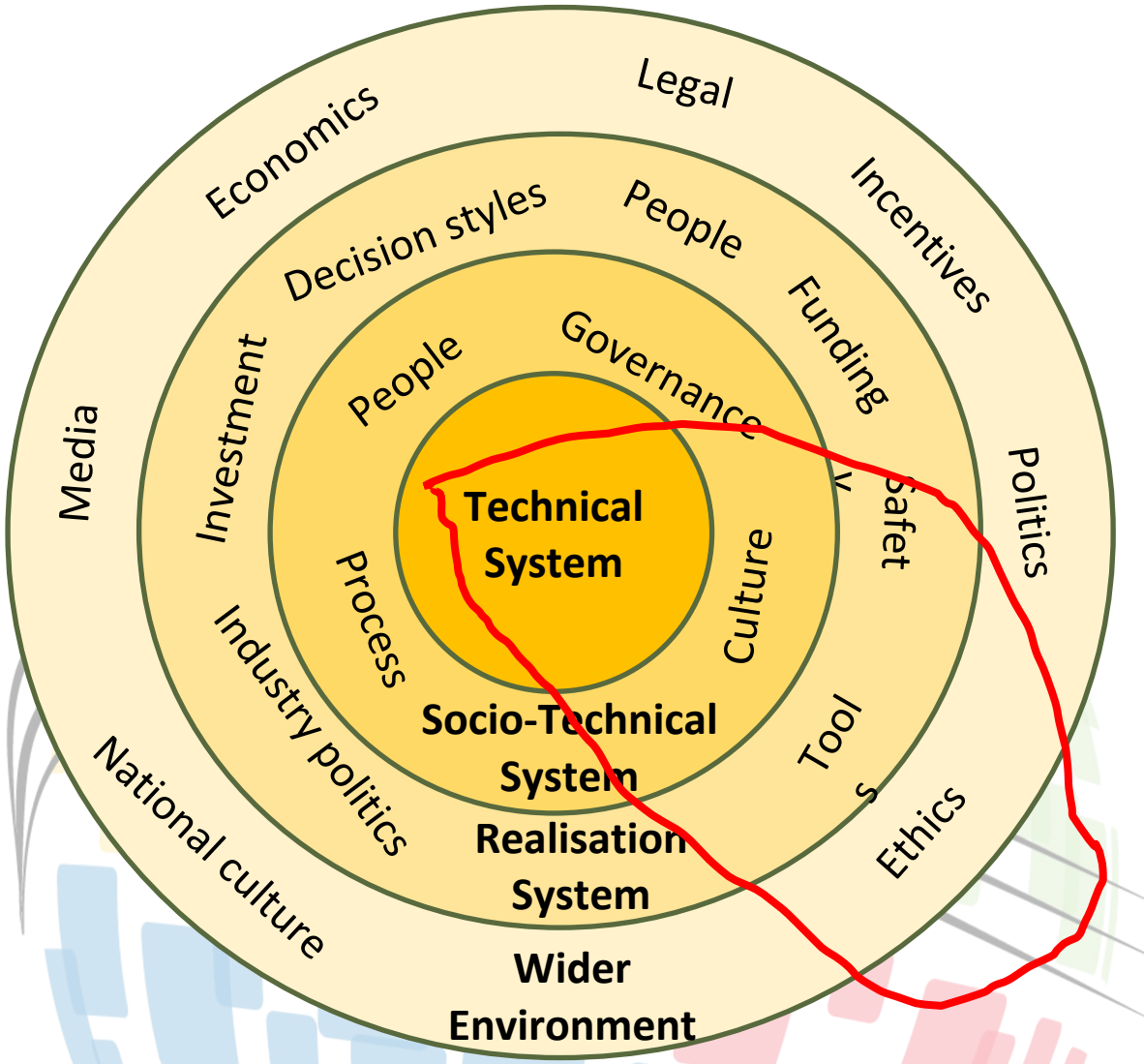
PERFORMANCE

a language for engineers to describe the experiential sense of a system, product or service (storytelling).

TECHNOLOGY

a language informed by social consciousness to generate innovative systems (product design and deployment).

The systems landscape



What are the Systems Safety challenges for systems containing artificial intelligence, quantum technology, additive manufacture, … ?

How do we enable safer socio-technical systems using psychology, system science, social science … ?

How can we do better Systems Safety Engineering using artificial intelligence, quantum technology, psychology, social science, leadership and management … ?

How do we influence the wider environment to help us realise safe and effective systems?

AI and LONELINESS

In May 2023, U.S. Surgeon General Vivek Murthy, M.D., M.B.A., called loneliness **a public health epidemic.**

We’re existentially alone on the planet. I can’t know what you’re thinking and feeling and you can’t know what I’m thinking and feeling. And the very best works construct a bridge across that abyss of human loneliness.

David Foster Wallace

Writer

In an age when many people describe their lives as lonely, there may be value in having AI companionship as a form of reciprocal social interaction that is stimulating and personalized.

Tony Prescott

Professor, Cognitive Robotics

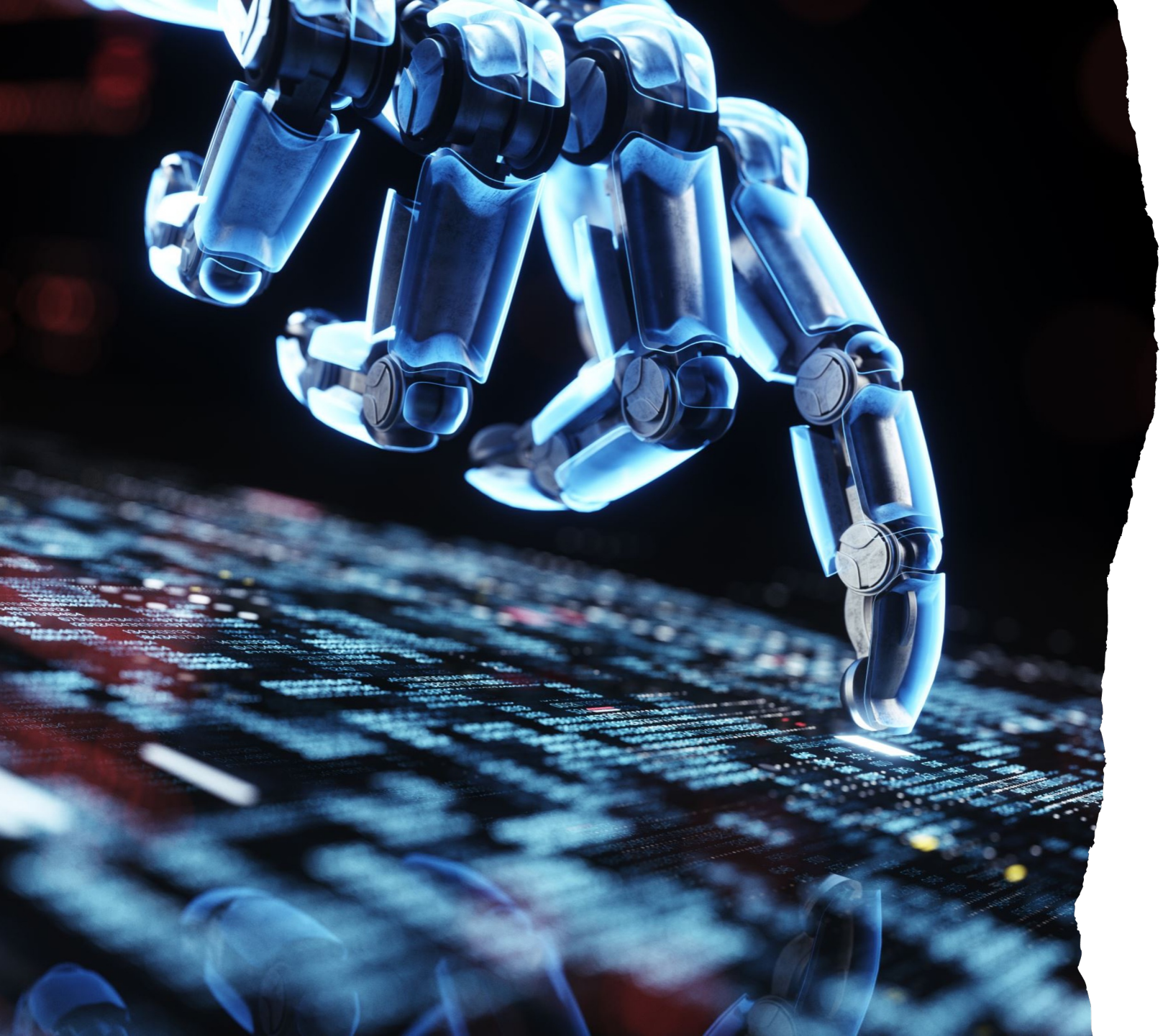
University of Sheffield

The Psychology of Artificial Intelligence

Four Types of AI

experience to aid in decision-making processes. e.g. Self-driving cars

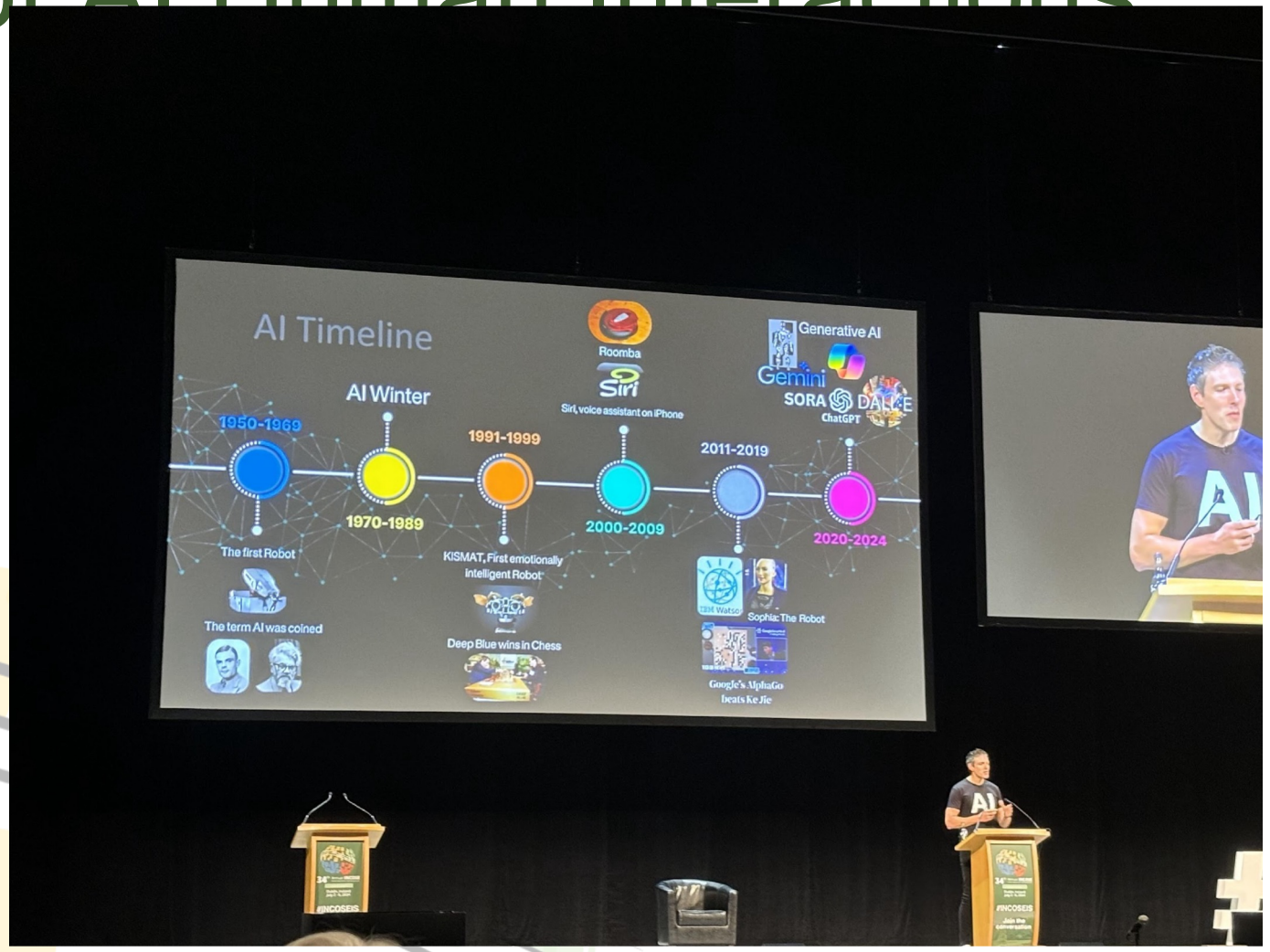
3. **Theory of Mind AI**, the current state of advanced AI, are programmed with decision-making abilities that mimic



HOW DO YOU INTERACT WITH COGNITIVE ARTIFICIAL INTELLIGENCE?

Join at menti.com | use code 5494 7153

PROS of AI Human Interactions



CONS of AI Human Interactions

- Tendency to hallucinate
- Perpetuating implicit biases demonstrable in AI training (see Deep Hysteria)
- Unethical harvesting techniques and the need for updated copyright laws
- Increased rate of depression and suicide rates among teens related to Social Media

Example of Effects of Unconscious Bias

In the case of emotion detection algorithms, bias and subjectivity are at the core of the concept itself.

– Amy Alexander, *Deep Hysteria (2023)*

Questions

- What are the ethics behind testing the effects of AI on a general population?

- What protections are there for whistleblowers?
  - Board of Directors at Open AI and the firing of Sam Altman
  - Re David Snowden – Distributed Intelligence…

Conclusions

- Increased urgency because the technology is evolving quickly.
- There is an inherent difference between approving something in the pharmaceutical world, for example versus in AI, there is no taking it back once the genie is out of the bottle.
- The varying views PRO and CON are not new historically.

THANK YOU!

HORTENSE GERARDO – hgerardo@ucsd.edu

www.hortensegerardo.com