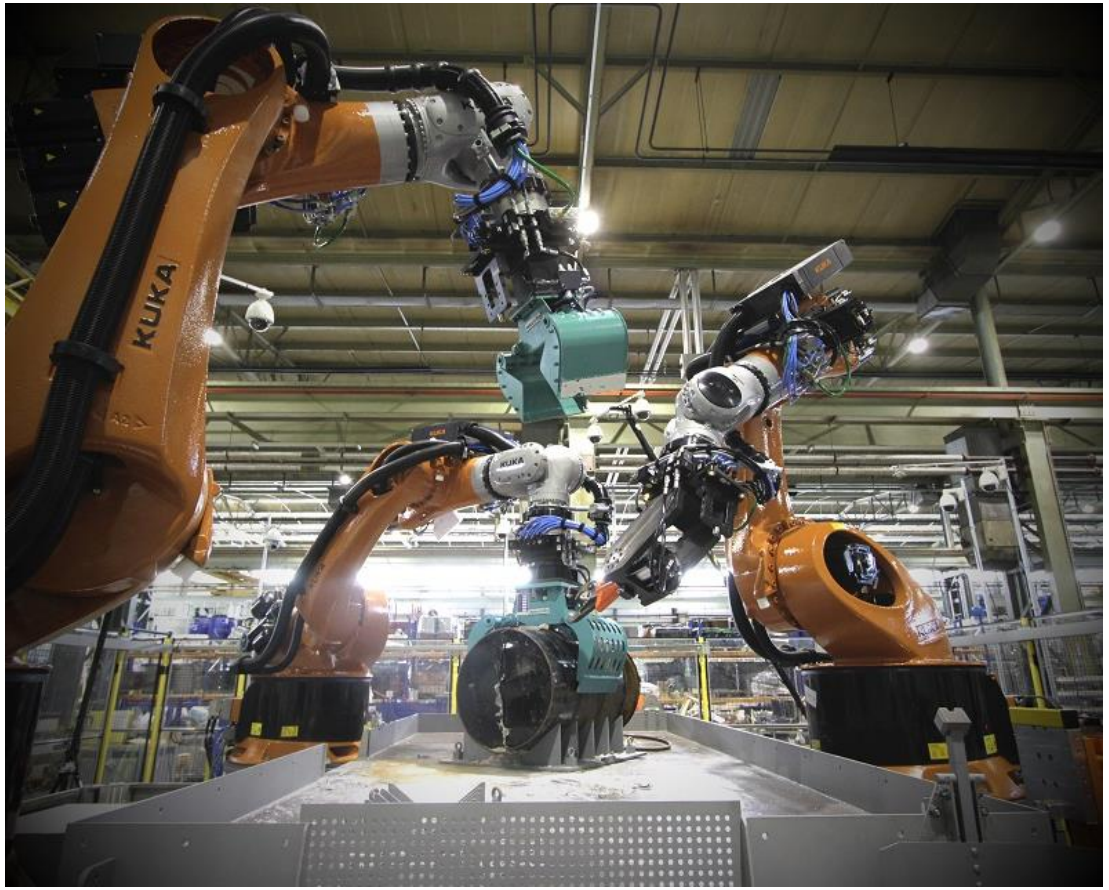
RAICo
ROBOTICS AND
AI COLLABORATION

Security and Safety for Autonomous Systems

Session Goals

RAICo
ROBOTICS AND
AI COLLABORATION



Cleaning up our nuclear past: faster, safer and sooner: Nuclear Decommissioning Authority.

1. **Understand the key challenges in securing autonomous and AI-driven systems, and in making them safe.**
2. **Determine the solutions to these problems** (in an ideal world).
3. **Elicit the key barriers preventing us from implementing these solutions.**
4. **Identifying any opportunities or accelerators which are underleveraged.**

Focus on:

- Autonomous systems and AI specifically.
- Any concrete challenges, barriers, and opportunities. Less focus on hypotheticals.

A Hot Topic

RAICo
ROBOTICS AND
AI COLLABORATION

- The UK hosted the AI Safety Summit in November 2023.
- It brought together 100 world leaders, executives from technology companies, and leading academics.
- A key outcome was the signing of a declaration by 28 countries to keep working together on AI safety and regulation.
- However, the outcomes were broad, focused on “frontier AI”, and concentrated on long-term existential risks.
- **What are the security challenges posed by/to AI and autonomous systems today?**

Chris J Ratcliffe/EPA, via The Guardian

Security and Safety

Wikimedia Commons, Public Domain

Photo by CEphoto, Uwe Aranas

Security

Protection against deliberate crime

Safety

Protection against accidents

Security versus Safety

RAICo
ROBOTICS AND
AI COLLABORATION

Safety

Emergency exits are required to leave a building in an emergency

Example from TÜV NORD GROUP

Security

Emergency exits are potential entry points for intruders

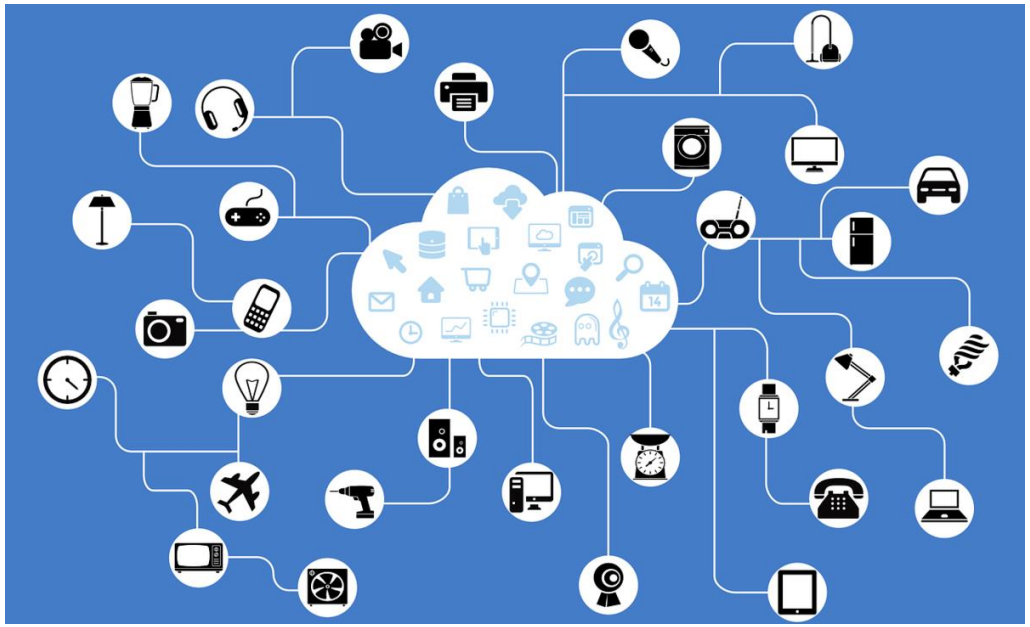
Functionality versus Safety

RAICo
ROBOTICS AND
AI COLLABORATION

Functionality

Autonomous systems must communicate with other systems to function

Wikimedia Commons, Public Domain

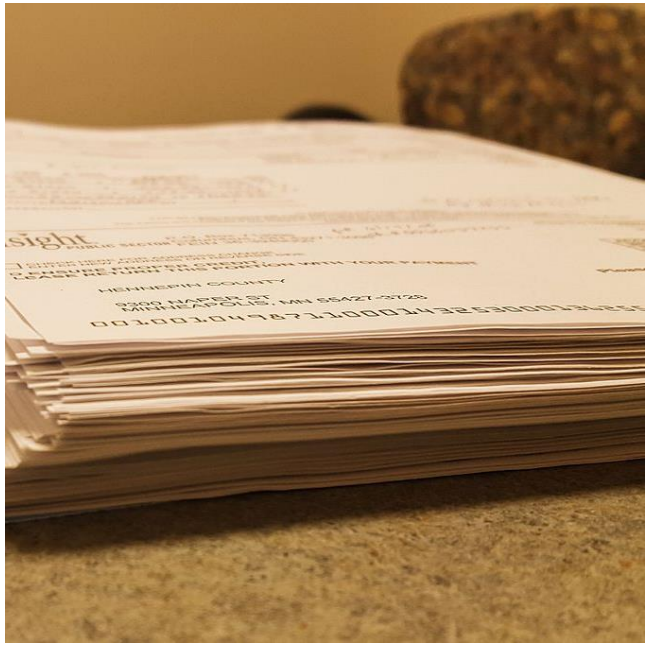


Security

Open ports are potential entry points for intruders

Domain – Not Just Cybersecurity

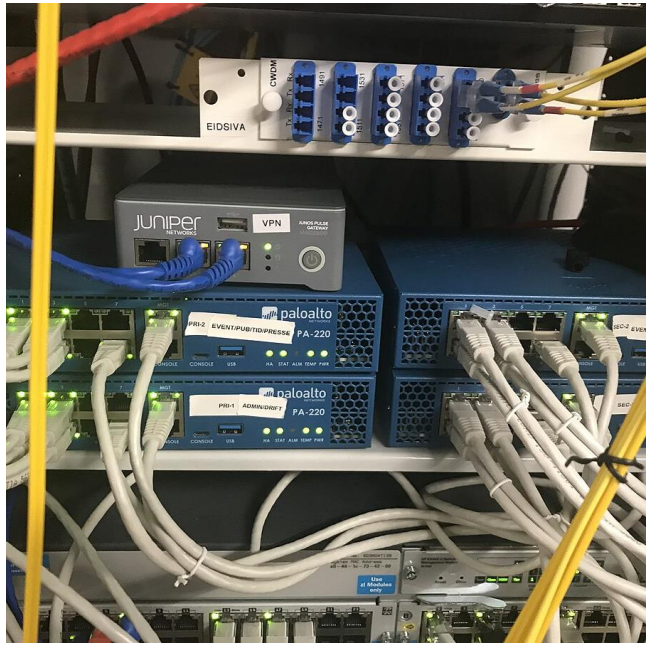
Tony Webster via Wikimedia Commons, CC BY-SA 2.0



Information Security

Protecting print or electronic sensitive information from unwanted access, use, disclosure, destruction, modification, or disruption.

DiFronzo via Wikimedia Commons, CC BY 2.0



Network Security

Physical or software measures to protect network infrastructure from unauthorised access, misuse, modification, or destruction, thus protecting critical systems.

Public Domain, CC0 (via Raw Pixel)

Physical Security

The protection of personnel, hardware, software, networks and data from physical actions and events that could cause serious loss or damage to an enterprise, agency or institution.

Not Just Sensitive Information

Company worker in Hong Kong pays out £20m in deepfake video call scam

Police investigate after employee says she was tricked into sending money to fraudsters posing as senior officers at her firm

Police said the woman made 15 transactions to banks accounts totalling HK$200m. Photograph: Blend Images/Alamy

Hong Kong police have launched an investigation after an employee at an unnamed company claimed she was duped into paying HK$200m (£20m) of her firm's money to fraudsters in a deepfake video conference call.

The Guardian

- On the 4th February 2024, a working in Hong Kong was tricked into transferring $25m to fraudsters.
- The attackers used recordings of company employees (including the chief financial officer) to create deepfakes of the employees, and staged a Teams call.
- **Your "non-sensitive" data can now be used against you**, for deepfakes, phishing attacks, adversarial AI attacks, training data poisoning etc.

Threat Actors

Ebrahim via Wikimedia Commons, CC BY-SA 4.0

Highly Capable State Threat Actors

B Klug via Flickr, CC BY-NC 2.0

Organised Cyber-Crime Groups

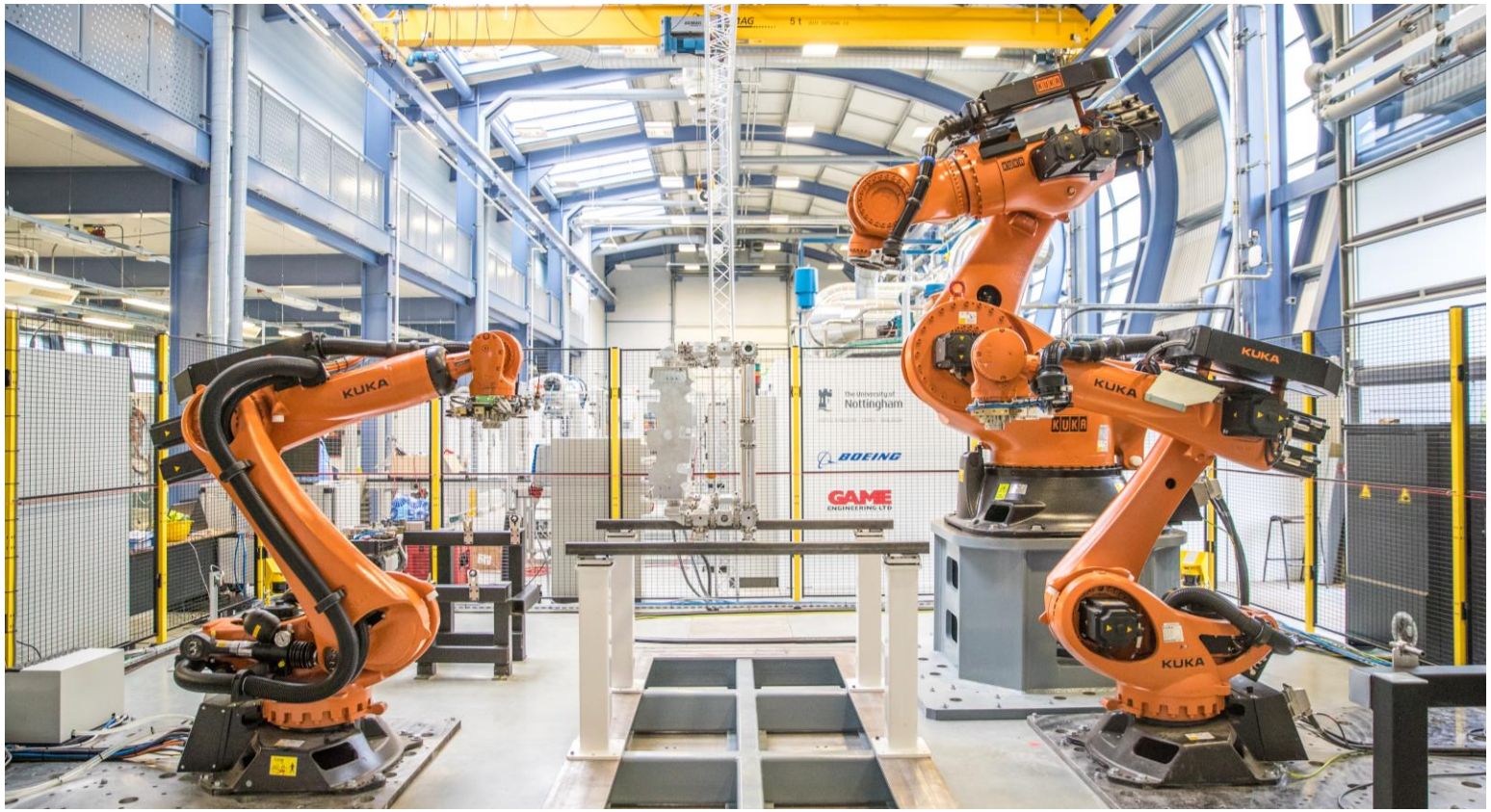
Marco Virch via Flickr, CC BY 2.0

Opportunistic Cyber-Criminals, Hacktivists, Disgruntled (ex)Employees

Definitions adapted from The National Cyber Security Centre – The Near-Term Impact of AI on the Cyber Threat

Domain – AI and Autonomous Systems

RAICo
ROBOTICS AND AI COLLABORATION



Future Automated Aerospace Assembly Demonstrator, University of Nottingham

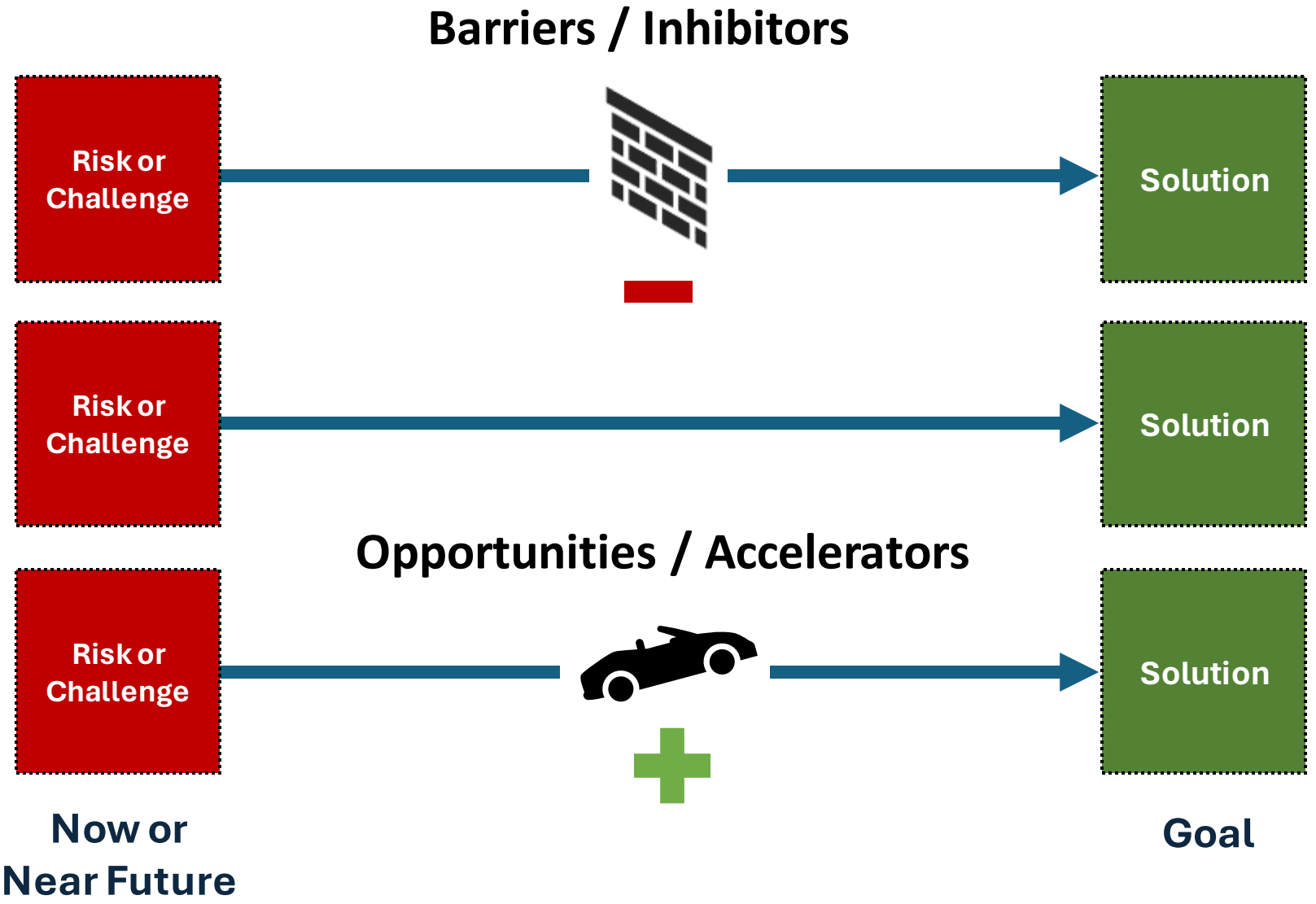
Discussions of (cyber) security can very easily get very broad.

Remember our focus here is security issues for **AI and autonomous systems.**

We also want to focus on your real, concrete experience. Not just hypotheticals.

Session Aims

RAICo
ROBOTICS AND
AI COLLABORATION



Session Activity 0 (10 minutes) Icebreaker

RAICo ROBOTICS AND AI COLLABORATION

What Does Security Impact?

- What is “security” looking to protect?
- What security is required in your industry?

- *Get to know your group!*
- *Be specific with answers.*

NASA, Public Domain

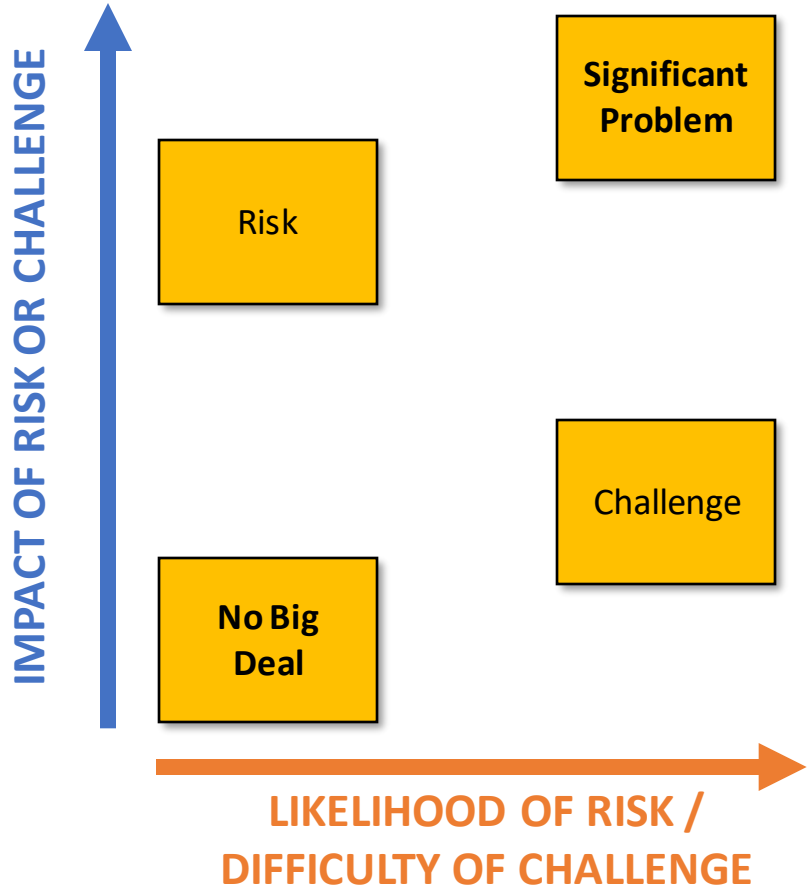
Session Activity 1 (25 minutes) Security Risks and Challenges

RAICo
ROBOTICS AND AI COLLABORATION

Key Security Risks and Challenges for Autonomous Systems

- How could autonomous systems be attacked or disrupted?
- What challenges exist to securing autonomous systems?

- *Prioritise ideas by severity and likelihood.*



Session Activity (10 minutes) Break

Refresh for the second half of activities.

- *The most important activity of the day!*

James Joel via Flickr, CC BY-ND 2.0

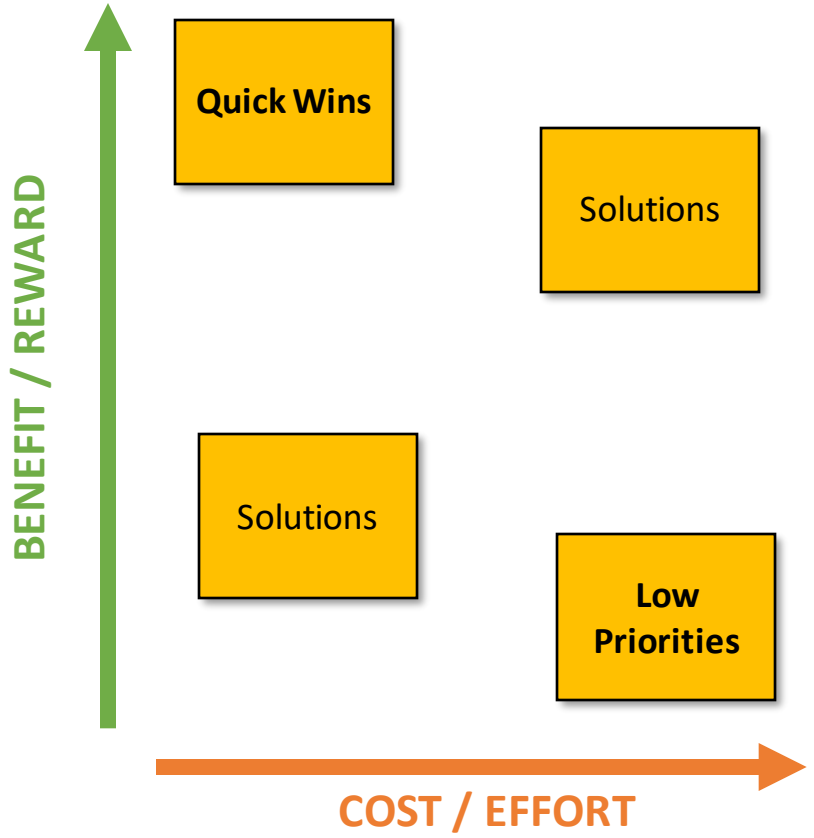
Session Activity 2 (20 minutes) Solutions to Risks and Challenges

RAICo ROBOTICS AND AI COLLABORATION

In a perfect world, how would you solve your high priority risks and challenges?

- Focus on the bigger challenges you identified.
- Solutions could span more than one problem.

- *Prioritise ideas by benefit and cost.*



Session Activity 3 (30 minutes) What's Stopping Us?

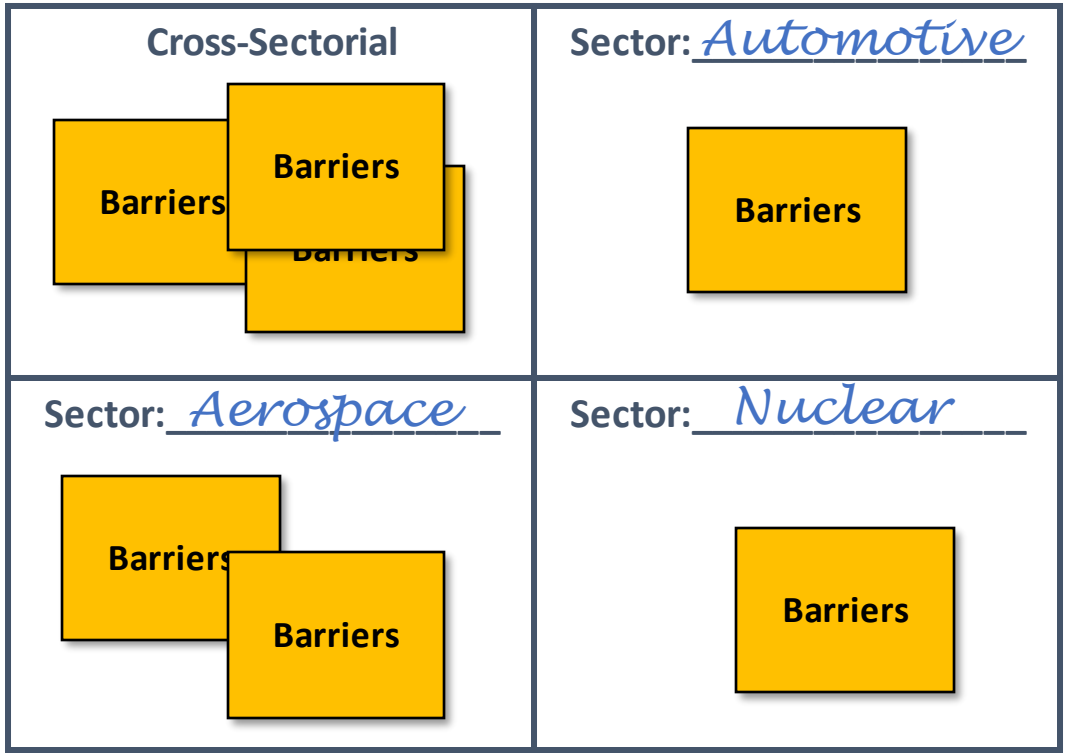
RAICo ROBOTICS AND AI COLLABORATION

Why can't we implement these solutions now?

And

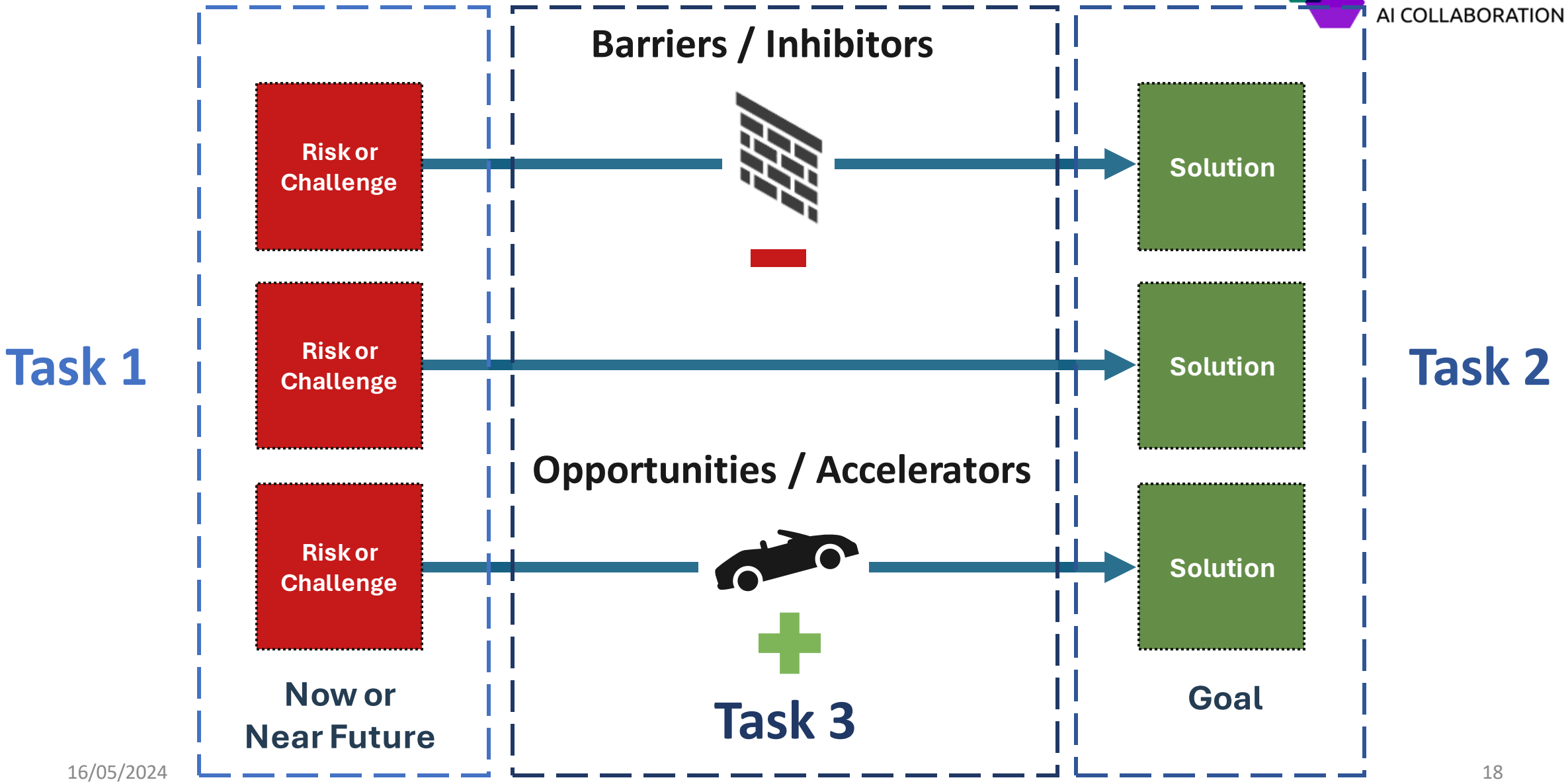
What opportunities exist which need leveraging better?

- *Are there any sector specific barriers or opportunities that could cross boundaries?*



Session Aims

RAICo
ROBOTICS AND
AI COLLABORATION



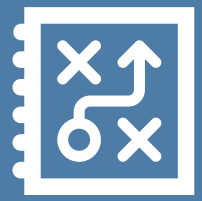
Any Questions?

RAICo
ROBOTICS AND
AI COLLABORATION

RAICo
ROBOTICS AND AI COLLABORATION

0: Icebreaker

What Does Security Impact?



1: Problems

Key Security Risks and Challenges for Autonomous Systems

2: Solutions

In a perfect world, how would you solve your high priority risks and challenges?

3: Barriers / Opportunities

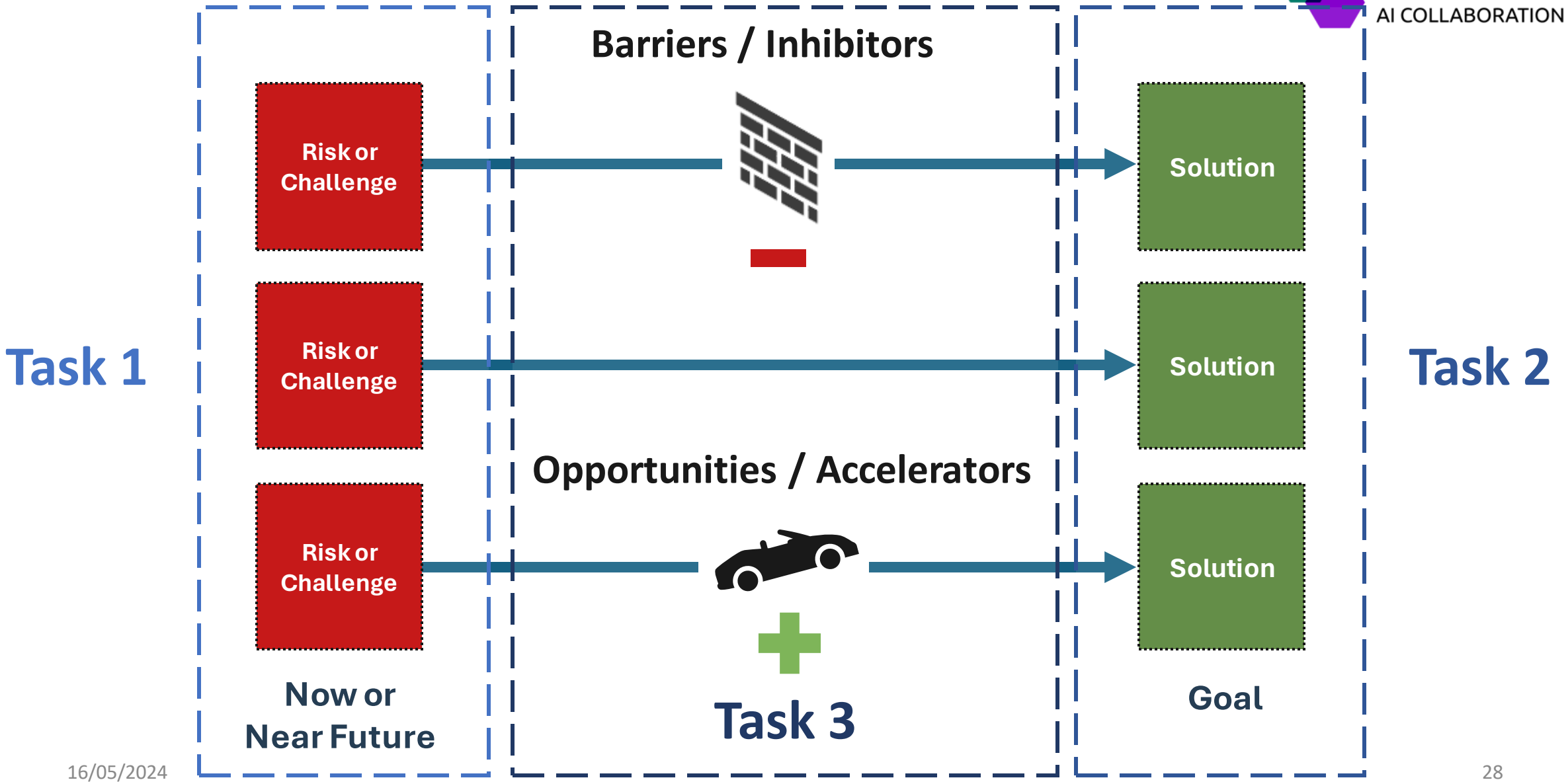
Why can't we implement these solutions now?

and

What opportunities exist which need leveraging better?

Session Aims

RAICo
ROBOTICS AND
AI COLLABORATION



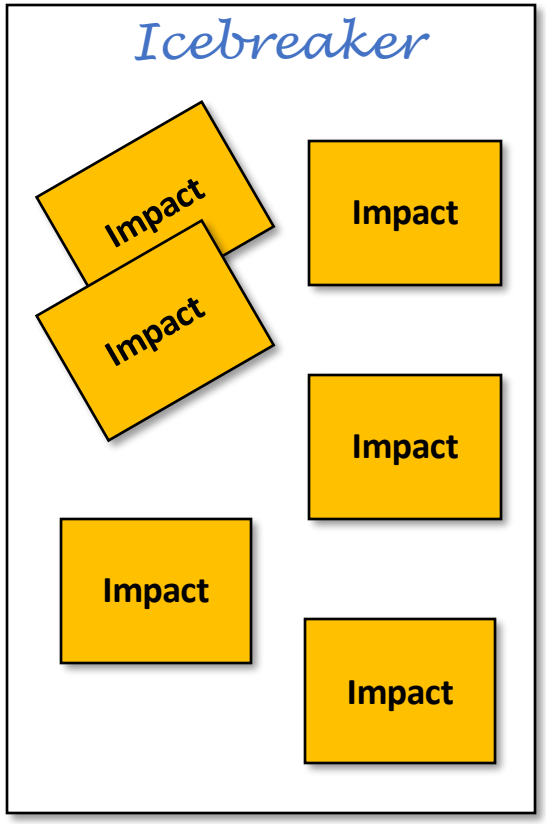
Session Activity 0 (10 minutes) Icebreaker

RAICo
ROBOTICS AND AI COLLABORATION

What Does Security Impact?

- What is “security” looking to protect?
- What security is required in your industry?

- *Get to know your group!*
- *Be specific with answers.*



Try and group post-its by theme!

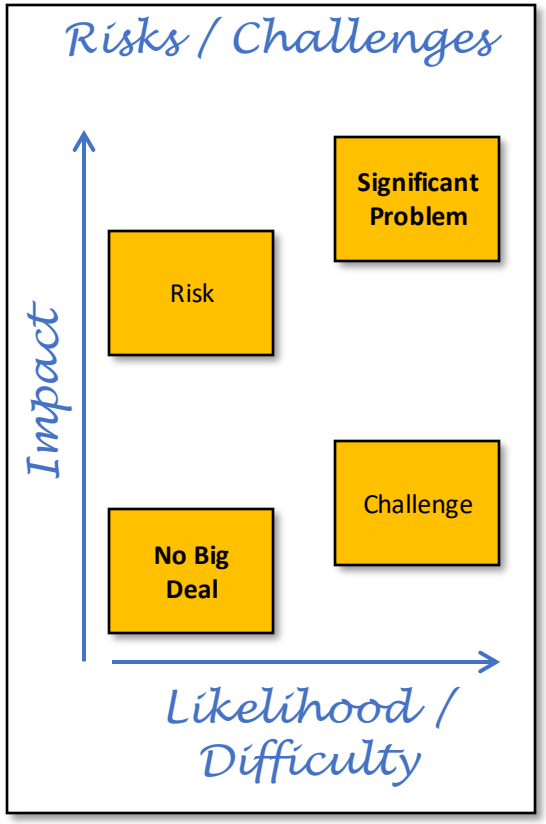
Session Activity 1 (25 minutes)
Security Risks and Challenges

RAICo
ROBOTICS AND AI COLLABORATION

Key Security Risks and Challenges for Autonomous Systems

- How could autonomous systems be attacked, disrupted, or behave in an unsafe way?
- What challenges exist to securing autonomous systems?

- *Prioritise ideas by severity and likelihood.*



What's a big problem that needs attention, and what's just a distraction?

Session Activity (10 minutes) Break

Refresh for the second half of activities.

- *The most important activity of the day!*

James Joel via Flickr, CC BY-ND 2.0

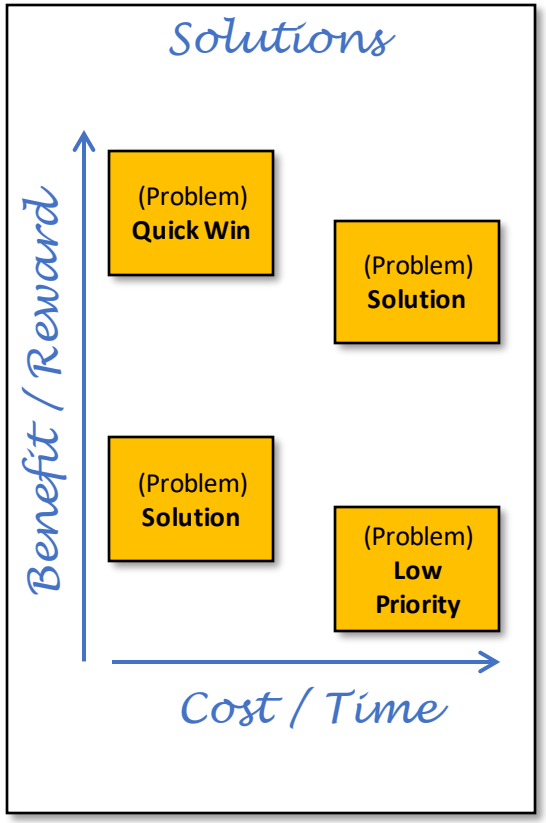
Session Activity 2 (20 minutes)
Solutions to Risks and Challenges

RAICo
ROBOTICS AND AI COLLABORATION

In a perfect world, how would you solve your high priority risks and challenges?

- Focus on the bigger challenges you identified.
- Solutions could span more than one problem.

- *Prioritise ideas by benefit and cost.*



Don't forget to mention what problem(s) the solutions are solving!

Session Activity 3 (30 minutes) What's Stopping Us?

RAICo ROBOTICS AND AI COLLABORATION

Why can't we implement these solutions now?

And

What opportunities exist which need leveraging better?

- *Are there any sector specific barriers or opportunities that could cross boundaries?*