Promoting responsible innovation in AI for peace and security

Session 1

AI and international peace & security:

Risk definition and identification

Session 1 – Learning goals

- **Recognise** the value of assessing the **societal impacts** of technology.
- **Identify** the **impacts** of AI's development, deployment and use on international peace and security.
- **Understand** the **risk** posed by (mis)uses of civilian AI to international peace and security.

1. Societal impact of technology

Why does studying the societal impact of technology matter?

Risk identification, analysis, and mitigation

Responsible innovation → aims at the acceptability, sustainability and societal desirability of the innovation process and its marketable products

1. Societal impact of technology

Useful distinction regarding the **consequences of technology**

Intended effects vs. Unintended effects

Intended use vs. Misuse

Malfunction vs. Well-functioning

Desired effects vs. undesired effects

Expected effects vs. unexpected effects

Specific effects vs. Systemic effects

Specific **innovations can create or exacerbate problems**, even **without hostile intent** (e.g., generative AI → deepfakes and disinformation)

*This matters for how we understand risk

2.1. What is international **peace and security?**

How are these pictures related to **peace and security**?

U.S. President Joe Biden waves as he walks with Chinese President Xi Jinping at Filoli estate on the sidelines of the Asia-Pacific Economic Cooperation (APEC) summit, in Woodside, California, U.S., November 15, 2023. REUTERS/Kevin Lamarque *Acquire Licensing Rights*

Former Google employees speak about Google's Project Nimbus at a pro-Palestine protest at UC Berkeley, Calif., on April 25, 2024. Photo: Tayfun Coskun/Anadolu via Getty Images

ISRAELI WEAPONS FIRMS REQUIRED TO BUY CLOUD SERVICES FROM GOOGLE AND AMAZON

Google downplays its military work with Israel, but "Project Nimbus" documents tie the American tech giants to Israel's deadly military capabilities.

Sam Biddle
May 1 2024, 3:58 p.m.

2.1. What is international **peace and security?**

- "**International**" goes **beyond states**
  - Traditionally, P&S were a state-centric concepts. Focused on territorial integrity and international use of force.
  - Today, there is a **human-centric** approach to **peace and security**
    - Not only about states, war, and military threats
    - Focus on people's needs and rights as a basis for sustainable peace

2.2 How can AI impact intl. peace and security?

2.2 How can **AI undermine** international **peace and security (P&S)**?

Group discussion:

- In your groups, find examples of how AI could undermine P&S
- Remember P&S go beyond states, human-centric approach,…

Groups of 4-5

Short presentation of examples

15’ Group discussion

Open discussion

2. Impact of AI on intl. peace and security

How can AI undermine international peace and security?

Pathway 1: Military adoption

Intelligence, Surveillance, and Reconnaissance (ISR);Command, Control, and Communication; Force delivery; Logistics and Maintenance…

Dual Use

Pathway 2: Civilian adoption

Transportation; Health; Logistics; Business; Entertainment and Knowledge

3. Risks posed to intl. peace and security

3.1 **Definition** of risk and type of risks

Who?

States, organisations, individuals…

What?

Death, injury, economic loss, rule of law…

Risk

Likelihood x Impact

Type of harm

Physical, digital, and political

Order of effect

First, second, third…

Direct, ripple effects…

3. Risks posed to intl. peace and security

3.1 Definition of risk and **type of risks**

Accidental Risk

Unintended negative consequences from intended use

Misuse Risk

Harmful unintended use

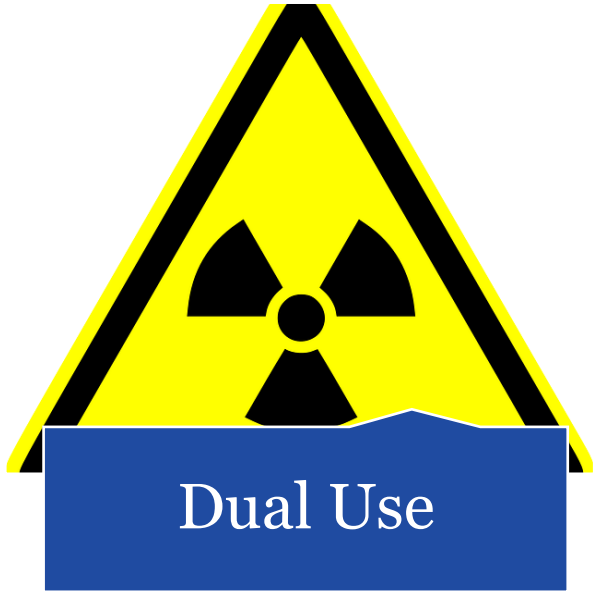
Structural Risk

Impact broader environment that could be disruptive or harmful

3. Risks posed to intl. peace and security

Zooming on the **risk of misuse**:

- Are there **misuse potentials** in your area of research? What makes it easy or difficult to **misuse**?
- Are these risk stemming from **development** decisions or from **diffusion** decisions?



Session 2

AI governance:

How to mitigate AI-related risks

Session 2 – Learning goals

- **Map out** the international **AI governance landscape**
- **Identify** how AI governance initiatives address **peace and security risks**
- **Understand** tech developer's role in **mitigating the risks** posed by (mis)uses of AI

1. How is the AI governance landscape structured?

What is international governance?

Policies, norms, stakeholders, and initiatives through which transnational challenges are addressed

Challenges that **require collective responses**

1. How is the AI governance landscape structured?

How is AI governed?

By whom? How?

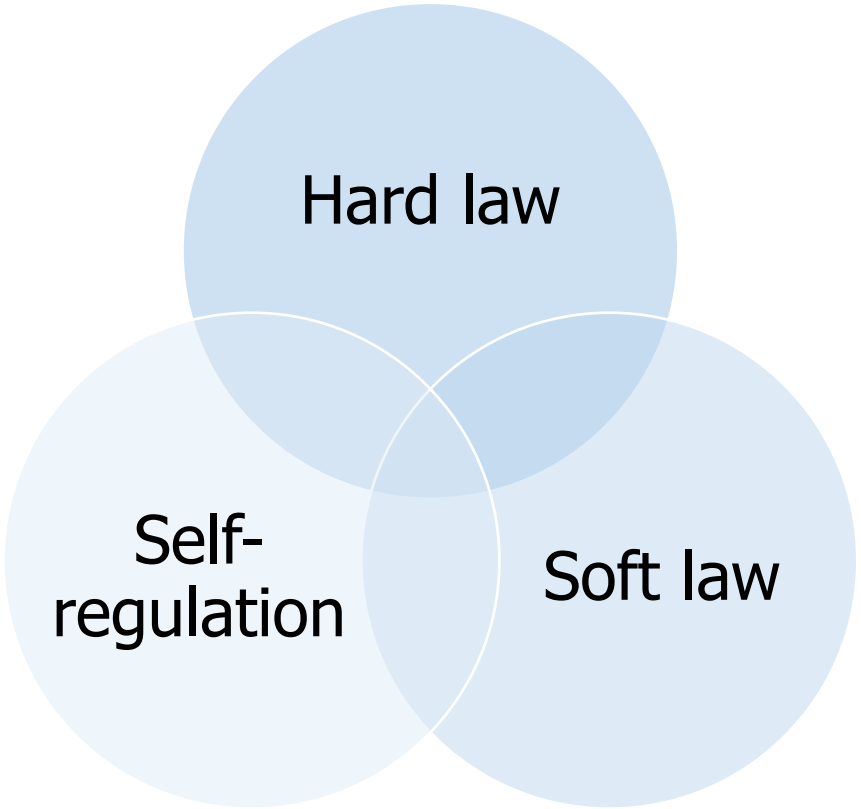
1. How is the AI governance landscape structured?

Is AI Governance a Wild West?

Not really, let’s unpack your answers…

1. How is the AI governance landscape structured?

Intl. Org. | States | Private sector | Academia | Civil Society



Enabling

- **Facilitating tech cooperation**

Limiting

- **Limiting the development of certain tech**

1. How is the AI governance landscape structured?

United Nations

- UN Bodies and Agencies
- Member states

States

- States' interests
- National regulations

Private sector

- Professional organisations
- Standard setting organisations

Academia

- Research Institutes
- Universities

Civil Society

- Advocacy
- Campaigners
- Individuals

1. How is the AI governance landscape structured?

OpenAI logo displayed on a mobile phone screen in front of computer screen on Sept. 5, 2023 in Ankara, Turkey. Photo: Dilek Nente/Anadolu Agency via Getty Images

OPENAI QUIETLY DELETES BAN ON USING CHATGPT FOR "MILITARY AND WARFARE"

The Pentagon has its eye on the leading AI company, which this week softened its ban on military use.

Sam Biddle
January 12 2024, 2:07 p.m.

AI

Microsoft bans US police departments from using enterprise AI tool for facial recognition

Kyle Wiggers
1:57 PM PDT • May 2, 2024

Comment

Image Credits: Fabrice Coffrini / AFP / Getty Images

Hard law

Self-regulation

Soft law

EU AI Act

Proposal for a

Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

2021/0106 (COD)

European Commission

G7 2023 HIROSHIMA SUMMIT

Hiroshima Process International

Guiding Principles for Organizations

Developing Advanced AI System

1. How is the AI governance landscape structured?

Enabling

- Consumer friendly competition, setting **technical standards**, facilitating **technological cooperation**
- US EO 141110, EU AI Act…
- NIST Standards, ISO/IEC 420001:2023,

Limiting

- Poses **limits** to the development and use of certain technologies, both in the **civilian** and **military** sector
- Some elements of the EU AI Act, Strategic Export Controls…

1. What issues does AI governance face?

Conceptual

- Definitional challenge, intangibility
- Broad vs. Narrow regulation
- Dual/general purpose of AI

Institutional

- Speed of action, institutional gridlock
- Centre of action
- Diffusion of knowledge and resources

(Geo)Political

- Great powers competition
- Innovation vs. regulation
- Digital divide

2. How are AI governance initiatives addressing P&S risks?

Impact of AI on peace and security

Pathway 1: Military adoption

Pathway 2: Civilian adoption

2. How are AI governance initiatives addressing P&S risks?

Pathway 1: Military adoption

United Nations A/C.1/79/L.43

General Assembly

Distr.: Limited
16 October 2024

Original: English

United Nations A/C.1/79/L.77

General Assembly

Distr.: Limited
18 October 2024

Original: English

Seventy-ninth session
First Committee
Agenda item 98
General and complete disarmament

Austria, Belgium, Brazil, Canada, Chile, Czechia, Denmark, [illegible] Greece, Kenya, Luxembourg, Netherlands (Kingdom of the), Portugal, Republic of Korea, Singapore, Switzerland, Türkiye Great Britain and Northern Ireland and United States of Am[illegible]

Artificial intelligence in the military domain and [illegible] for international peace and security

[illegible]l autonomous

[illegible]sta Rica, Guatemala, Ireland, Kiribati, [illegible], New Zealand, Philippines, Sierra Leone, Trinidad and Tobago: draft resolution

[illegible]pons systems

CAMPAIGN TO STOP KILLER ROBOTS

U.S. DEPARTMENT of STATE

POLICY ISSUES COUNTRIES & AREAS BUREAUS & OFFICES ABOUT

Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy

OTHER RELEASE

BUREAU OF ARMS CONTROL, VERIFICATION AND COMPLIANCE

FEBRUARY 16, 2023

An increasing number of States are developing military AI capabilities, which may include using AI to enable autonomous systems.[1] Military use of AI can and should be ethical, responsible, and enhance international security. Use of AI in armed conflict must be in accord with applicable international humanitarian law, including its fundamental principles. Military use of AI capabilities needs to be accountable, including through such use during military operations within a responsible human chain of command and control. A principled approach to the military use of AI should include careful consideration of risks and benefits, and it should also minimize unintended bias and accidents. States should take appropriate measures to ensure the responsible development, deployment, and use of their military AI capabilities, including those enabling autonomous systems. These measures should be applied across the life cycle of military AI capabilities.

The following statements reflect best practices that the endorsing States believe should be implemented in the development, deployment, and use of military AI capabilities, including those enabling autonomous systems:

A. States should take effective steps, such as legal reviews, to ensure that their military AI capabilities will only be used consistent with their respective obligations under international law, in particular international humanitarian law.

B. States should maintain human control and involvement for all actions critical to informing and executing sovereign decisions concerning nuclear weapons employment.

C. States should ensure that senior officials oversee the development and deployment of all military AI capabilities with high-consequence applications, including, but not limited to, weapon systems.

D. States should adopt, publish, and implement principles for the responsible design, development, deployment, and use of AI capabilities by their military organizations.

E. States should ensure that relevant personnel exercise appropriate care, including appropriate levels of human judgment, in the development, deployment, and use of military AI capabilities, including weapon systems incorporating such capabilities.

F. States should ensure that deliberate steps are taken to minimize unintended bias in military AI capabilities.

G. States should ensure that military AI capabilities are developed with auditable methodologies, data sources, design procedures, and documentation.

2. How are AI governance initiatives addressing P&S risks?

Pathway 2: Civilian adoption

2. How are AI governance initiatives addressing P&S risks?

Pathway 2: Civilian adoption

What's common across these initiatives?

- Stress the **role of developers** engaging in responsible innovation:
  - Through the conduct of **technology impact assessment**;
  - Risk **identification, evaluation and mitigation**

3. What's the role of the AI community?

- Is there a **role** for the **AI community**? What could this role be?

Expertise

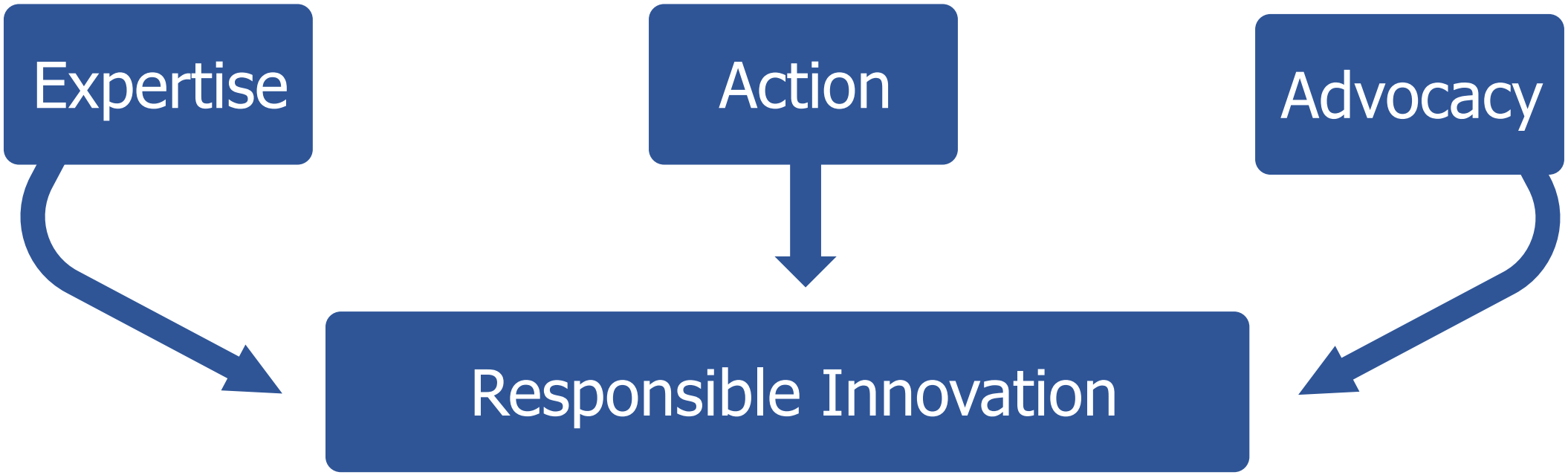
Action

Advocacy

- **Contribute** to **overcome the challenges** mentioned before: conceptual, institutional, and (geo)political…

3. What's the role of the AI community?

3.1. **How can** the AI community contribute to AI governance?

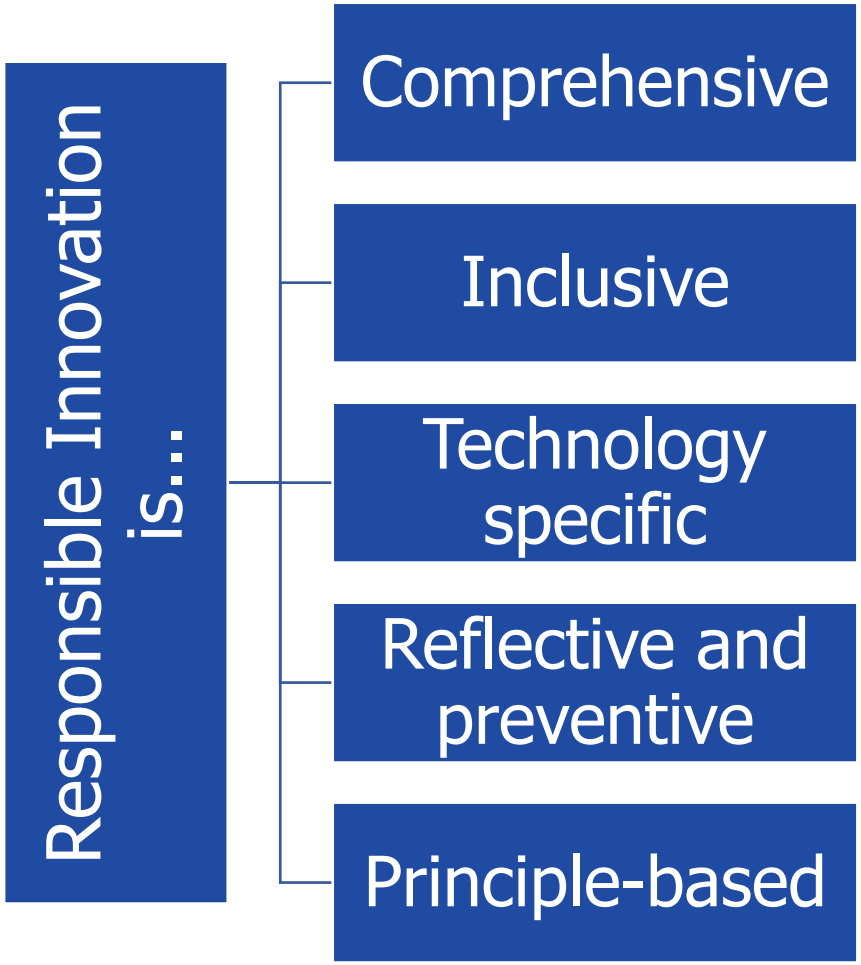


3. What's the role of the AI community?

What is Responsible Innovation?

- Anticipatory approach to research and innovation
- Provides guidance amid uncertainties about emerging technology
- Enhances acceptability, sustainability and social desirability of innovation processes

3.1. Why is Responsible Innovation useful?



- **Identifies** issues throughout tech's lifecycle
- **Involves** different stakeholders
- **Focuses** on specific applications
- **Seeks** to respond before problems occur
- **Facilitates** by-passing geopolitical deadlock

3.3. Good practice in responsible innovation

When?

- It's a process
- Through the research, development, deployment process

Who?

- Tech designers
- Decision makers within the development process

With whom?

- The process should include external stakeholders
- e.g., user's perspectives, ethicists, lawmakers...

Responsible AI for Peace and Security

Session 3

Responsible innovation in practice:

Risk assessment framework

Session 3 – Learning goals

- **Recognise** specific **risks** associated with the development, deployment, and use of AI technologies or applications.
- **Engage** in **risk mitigation** through responsible innovation practice.
- **Reflect** on the **implementation** of the risk assessment framework.

2.1. Introduction to the risk assessment approach

How to **engage** in **Responsible Innovation**?

Step 1: Mapping the risk

- What could go wrong?

Step 2: Assess the tolerability of the risk

- What can be done about it?

Step 3: Identify means of intervention

- What can I do as researcher/engineer?
- When do I need to transfer the issue to others?

2.1. Introduction to the risk assessment approach

Step 1: Mapping the risk

- What could go wrong?
  - **Define and characterise** the spectrum of **peace and security risks** associated with the development, diffusion, and (mis)use of a technology
  - Different methodologies: Red-teaming, fault trees/event trees, Delphi technique, Fishbone method…

2.1. Introduction to the risk assessment approach

Step 2: Assess the tolerability of the risk

- What can be done about it?

- **Accept**
  - Unlikely in case of P&S risk

- **Avoid**
  - Stop the development

Most likely

- **Mitigate**
  - Technical fixes, diffusion fixes, other…

- **Transfer**
  - Horizontal, vertical…

2.1. Introduction to the risk assessment approach

Step 3: Identify means of intervention

- What can I do as researcher/engineer?
- When, how, and to whom do I need to transfer the issue?
    - Practical steps **in** and **beyond** the R&D process

Mitigation

- Technical fixes: …
- Diffusion fixes: …
- Other fixes: …

Transfer

- Horizontal: peers, community…
- Vertical: government, institution, lawmakers…

3. Wrap-up: overall we've covered…

- The **impact of civilian AI on peace and security**, looking at how research and innovation in AI can generate risks for peace and security.

- What are the **current efforts to mitigate the risks** that AI poses to international peace and why and how, as an AI developer, you can play a role.

- Finally, **responsible innovation** as an approach **to mitigate peace and security risks** posed by the research, development and deployment of AI.

Promoting responsible innovation in AI for peace and security

Scan the QR code to access additional resources!