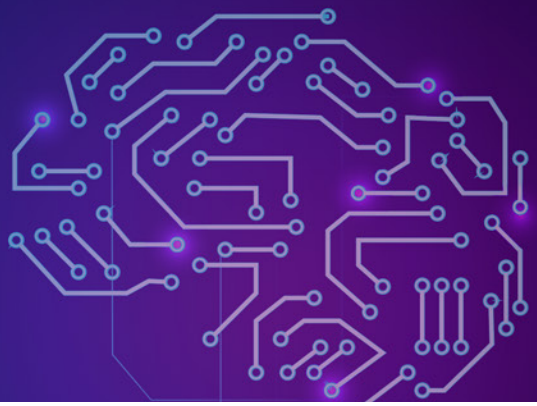
the 2022 innovations dialogue.

AI DISRUPTION, PEACE & SECURITY



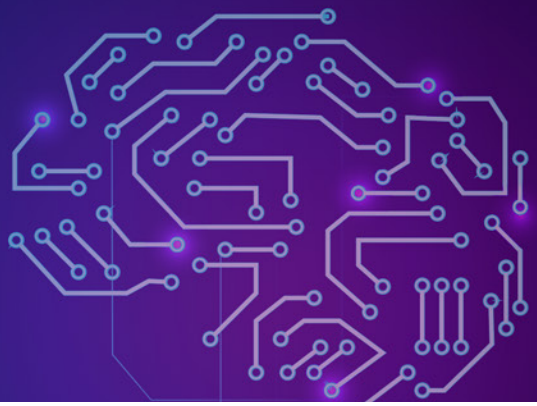
HIGHLIGHTS

AI and its state of play

- At present there is no widely accepted definition for AI as **AI is a broad discipline that is defined in different ways for different purposes**.
- AI can simply be thought of **as a system in which algorithms use data to make decisions** (or perform tasks) on our behalf or help humans make decisions (or perform tasks).
- **AI systems are what we make them**, and they will be what we want them to be. Essentially, they are human artifacts or human constructs designed by humans and trained largely on socially generated data.
- As the field of AI is evolving rapidly, it is becoming apparent that **AI presents unprecedented opportunities** to augment human capabilities, particularly in problem-solving and decision-making. However, at the same time, **significant ethical, legal, safety and security concerns** remain and are coming to the fore as AI systems are increasingly adopted across sectors, including the defence sector.
- **These concerns range across issues** related, but not limited, to transparency, reliability, predictability, understandability, accountability, bias and discrimination, and technical robustness.
- **A key issue that causes AI systems to make errors or fail is that AI can be biased.** There are three main dimensions of bias – *pre-existing bias*, which concerns bias in data, *technical bias*, which is introduced by the operation of the technical system itself and may amplify pre-existing bias, and *emergent bias*, which arises in the context of use of a system.
- **Deciding how much autonomy should be given to an AI system to perform which tasks in which contexts is crucial** because errors or failures in performing safety- and security-critical tasks can have adverse consequences for individuals, organizations and societies.
- **At present, AI systems are becoming good at performing narrow and specific, well-defined tasks** that are often repeatable and have clear criteria for success, on the basis of which developers and users can judge whether the system has achieved its purpose.
- **AI systems nonetheless remain brittle when it comes to their performance in dynamic and cluttered environments**, where these systems encounter uncertain conditions.
- While it is hard to predict the exact trajectory of AI advancements, what is becoming evident is that the **future will witness new forms and structures of collaboration and coordination between humans and AI systems**.

the 2022 innovations dialogue.

AI DISRUPTION, PEACE & SECURITY



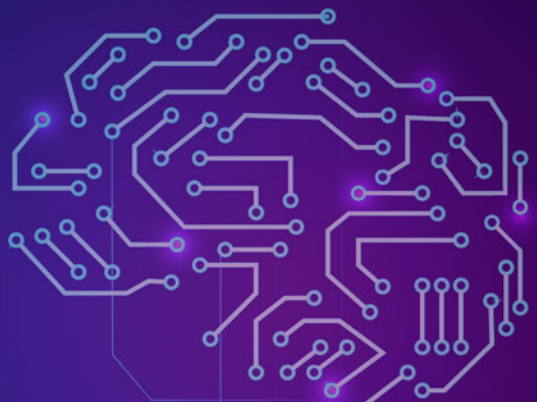
HIGHLIGHTS

The disruptive impact of AI on international peace and security

- The steady integration of AI technologies into an increasing number of military applications could **transform the conduct of military operations by enhancing military capabilities** in terms of efficiency, speed, precision, survivability and coordination.
- **In the military domain, at present the uses of AI technologies are rudimentary and not at scale**, but they are perhaps groundbreaking in the sense that they have not been attempted before. The most prominent uses of AI technologies today are in intelligence, surveillance and reconnaissance (ISR) operations, strengthening cyber defences, conducting as well as mitigating AI-enhanced influence operations, and enhancing combat simulations for military training and planning.
- While AI technologies could offer benefits in military operations, they also create unique risks. For example, **increased speed in military decision-making could result in miscalculation and inadvertent escalation**. Current AI technologies are also brittle and prone to making errors and being fooled by adversarial spoofing or hacking. Due to this, they may ultimately be less accurate and precise than human operators in complex battlefields.
- As governments around the world are increasingly seeking to harness AI technologies across sectors, including in the military domain, they are developing national AI strategies and even defence-specific AI strategies. To ensure that such high-level strategy documents can have the desired operational impact, governments must remain cognizant of three key considerations – **AI is all about trade-offs, AI innovation involves uncertainty, and not every nail needs an AI-enabled hammer**.
- Given that the civilian AI industries work for or with militaries to build AI systems, it is imperative that governments and their militaries, regardless of the AI governance approach they adopt, take measures **to ensure that civilian and military governance frameworks align with respect to military applications of AI technologies**.
- As an enabling technology, the integration of **AI technology across domains of warfare – from cyber and biological to nuclear, and especially in convergence with other powerful dual-use technologies** – can have benefits for international peace and security as well as pose novel risks.
- **AI advancements can be harnessed to build and sustain peace**. In recent years, United Nations agencies have explored and even deployed AI-enabled applications for conflict prevention and peacebuilding around the globe, including to facilitating dialogue among different communities and better understanding the different needs and concerns within a local context. However, there remain practical implementation challenges to deploying AI solutions for conflict prevention and peacebuilding at scale.

the 2022 innovations dialogue.

AI DISRUPTION, PEACE & SECURITY



HIGHLIGHTS

Towards Responsible AI

- Given the ethical, legal, safety and security concerns that AI technologies present, **governments, intergovernmental organizations, private sector entities and members of civil society are developing normative instruments such as principles and standards to guide the AI system lifecycle**. These aim to ensure that AI systems are researched, designed, developed, deployed and used in a responsible manner in accordance with legal requirements and ethical values. This approach to AI governance is broadly known as Responsible AI.
- **RAI can be understood as a principles-based, socio-technical approach to the research, design, development, deployment, use, maintenance and governance of AI systems** across sectors that is conscious of and considers the effects (both positive and negative) that such systems may have on individuals, communities and society at large.
- **This RAI approach** helps to prevent unethical or irresponsible applications of AI technologies and consequently to **build trust in AI systems. The trust in turn is an enabler for the rapid adoption and deployment of AI systems**.
- Through ethical principles or guidelines and a combination of tools such as (but not limited to) testing standards, risk-assessment frameworks, conformity-assessment schemes, accountability checks and employment guidance, the **RAI approach can proactively ensure that decisions made through the AI system lifecycle result in intended outcomes**.
- Since Responsible AI is a lifecycle approach towards managing risks and preventing possible harms while facilitating responsible use, **stakeholders involved and concerned with every stage of the AI system lifecycle, from research to use, have a shared role to play**.
- **RAI efforts usually begin with the adoption of broad AI principles or guidelines** that encompass technical, legal and ethical requirements that AI systems should meet in order to be responsible and trustworthy. Committing to principles is, however, not sufficient to achieve responsible and trustworthy AI.
- **Broad principles need to translate into practice.** Thus, beyond the commitment to principles, governments and organizations that create or use AI, at their own levels, should develop detailed practical guidance for AI actors involved in the AI system lifecycle and put in place tools, processes, and governance structures and mechanisms for the operationalization of AI principles.
- **Scaling up RAI practices** and realizing the adoption of responsible and trusted AI systems ultimately **requires cultivating and sustaining a culture of Responsible AI** in which RAI-related considerations and values are instilled in the organizational culture and viewed as an integral and enabling part of AI development, rather than barriers to it, at both the system-wide and individual levels.