

Dark side of ICT

AI in warfare and cyber-diplomacy

The dark side of CyberWorld



Gian Piero Siroli, Physics & Astronomy Dept. Univ. of Bologna & CERN

IT seminar, CERN - November 2024



ICT is intrinsically a dual-use technology



Increasing integration of conventional warfare domains (land, sea, air, space) via battlefield digitization



~~ICT is intrinsically a dual-use technology~~



Inc
do



Operate OODA loop at a faster tempo than adversaries

.....

Less and less time to reflect/think

...above a certain speed of ops humans are overwhelmed...



Warfare digitization
&
Weaponization of cyber environment



Artificial Intelligence – Machine Learning

...and much more...



➤ **Big Data: acquisition, storage, analysis, transfer, visualization, querying, privacy. US DoD clouds: Air Force Cloud One, Flank Speed Transition, Cloud.mil**

➤ **AI being(!) integrated across full spectrum of military operations:**

- **Increases reliability/efficiency of cyber infrastructure defense in real time, especially against AI-enabled cyber threats on critical networks: real time, pattern recognition, anomaly detection**
- **Force multiplier in digital spaces, addressing challenges of scale & sophistication: threat monitoring, navigation, decision support, offensive/defensive cyber operations**
- **Human decision-making support & Lethal Autonomous Weapon Systems (LAWS) control (VERY advanced prototype phase)**

“AI Beyond Weapons: application and impact of AI in the military domain” (UNIDIR 2023)

➤ **Human *in, on, out* of the loop? (remote ctrl, semi-autonomous, autonomous). → “Meaningful Human Control” (MHC) to maintain oversight/accountability in compliance with ethical, moral, legal constrains. Humans must make critical decisions concerning life & death, not machines**

➤ **Cyber-intelligence? Current algorithms not capable of human level reasoning. Presently employed to process & manage (sensor) data, monitor systems integrity, support vocal commands, navigate**

➤ **Support C4ISTAR system: Command, Control, Communications, Computing, Information, Intelligence, Surveillance, Targeting Acquisition & Reconnaissance**

Artificial Intelligence – Machine Learning

...and much more...



We already have weapons that can use AI to search, select and engage targets in specific situations.

We are currently in an evolutionary phase of unmanned integration.

- Human *in, on, out* of the loop? (remote ctrl, semi-autonomous, autonomous). → “Meaningful Human Control” (MHC) to maintain oversight/accountability in compliance with ethical, moral, legal constrains. Humans must make critical decisions concerning life & death, not machines
- Cyber-intelligence? Current algorithms not capable of human level reasoning. Presently employed to process & manage (sensor) data, monitor systems integrity, support vocal commands, navigate
- Support C4ISTAR system: Command, Control, Communications, Computing, Information, Intelligence, Surveillance, Targeting Acquisition & Reconnaissance

Unmanned systems

Vehicles used for various military operations with different levels of autonomy (remotely controlled ↔ fully autonomous). Advanced sensors & comms.

- **Aerial Vehicles (UAV):** long-range (stealth) missions for surveillance & reconnaissance, target acquisition, combat support, EW. Real-time data sharing with manned aircraft. (XQ-58A Valkyrie)
- **Ground Vehicles (UGV):** land robotic systems for reconnaissance, reducing risk in hazardous environments, bomb disposal, logistics. Autonomous navigation, obstacle avoidance, payload delivery. (M5 UGV)
- **Marine Systems (UMV):** underwater (UUV) drones & surface (USV) vessels for (coastal) surveillance, anti-submarine warfare, mine countermeasures. Autonomous operation, long-duration missions, data collection. (Orca XLUUV)

Trends & Challenges

➤ **Trends:** Increased autonomy. Fast integration of AI & (military) clouds for data collection/analysis & ML. Enhanced interoperability between unmanned & manned systems, Human/Machine Interface for robotic systems. Swarms of drones.

➤ Challenges:

- **Cybersecurity & vulnerabilities**
- **Legal issues:** compliance with international laws & regulations, including law of armed conflict & navigation rules
- **Ethical considerations**
 - ethical boundaries in autonomous decision-making in combat scenarios (use of lethal force)
 - clear lines of *accountability* for actions taken by unmanned systems (complex operational environments)
- **Meaningful Human Control (MHC)**





- **Ground Vehicles (UGV):** use in hazardous environments, autonomous navigation, obstacle avoidance, payload delivery
- **Marine Systems (UMV):** used for (coastal) surveillance, anti-submarine warfare, autonomous operation, long-duration missions



ent levels of autonomy & comms. is for surveillance & real-time data sharing with intelligence, reducing risk in autonomous navigation, obstacle & surface (USV) vessels for underwater measures. Autonomous XLUUV

Trends & Challenges

- Trends: Increased data collection/analysis, autonomous systems, Human-machine teaming
- Challenges:
 - Cybersecurity
 - Legal issues of armed conflict
 - Ethical concerns
 - ethical frameworks (use of force)
 - clear rules of engagement (comp)
 - Meaningful

Food for thought:

Are unmanned autonomous systems and drones reshaping military strategies and tactics? How? Consequences?

- Integration cyber/physical military assets with traditional warfare domains
- Intelligence, Surveillance, Reconnaissance (ISR), Situational Awareness and Electronic Warfare
- Targeting and Precision Strikes
- Cost-Effectiveness
- Operational Safety
- Asymmetric warfare

clouds for data shared between manned & unmanned systems.



Unmanned Systems Integrated Roadmap 2017-2042 (US DoD)

Drones' swarms

Multiple drones operating together as a coordinated/cohesive unit, controlled by AI.

Used in military operations for surveillance, attack missions, electronic warfare etc.

Recent developments: high-altitude operability, rough-weather performance, synchronized attacks. Also applied in civilian areas like disaster response & agriculture.

US, China, Russia, India, Israel (and a few more countries) actively developing & deploying drones' swarms.

Expected to revolutionize various sectors by enhancing efficiency, precision & operational capabilities.

- Autonomous swarms
- Collaborative Combat Aircraft (CCA, Kratos XQ-58 Valkyrie) to escort/support F-22/F-35 fighters in achieving air dominance and other roles. CCAs incorporating cutting-edge advancement in AI & autonomy (estimated ~\$20-28M\$). Deploying additional UAVs



- o [Future of Drone Swarm Proliferation](#) (Modern War Institute, West Point 2024)
- o [The US Navy wants swarms of thousands of small drones](#) (MIT Technology Review 2022)
- o → [Drones' swarm in dense forest](#) (video 2022)

Autonomous Weapon Systems (AWS)

Open Letter to the United Nations Convention on Certain Conventional Weapons by tech companies on AI & robotics (Future of Life Institute 2017)

“As companies building the technologies in Artificial Intelligence and Robotics that may be repurposed to develop autonomous weapons, we feel especially responsible in raising this alarm.

Lethal autonomous weapons threaten to become the third revolution in warfare. Once developed, they will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend. These weapons can be used against innocent populations and hacked to behave in undesirable ways. Not much time to act, once this Pandora’s box is opened, it will be hard to close.

International humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems”

Autonomous Weapon Systems (AWS)

Open Letter to the United Nations Convention on Certain Conventional Weapons by tech companies on AI & robotics (Future of Life Institute 2017)

REMINDER

In the cyber domain full autonomy is already operational

Attacks can be delivered at a non-human time scale

and hacked to behave in undesirable ways. Not much time to act, once this Pandora's box is opened, it will be hard to close.

International humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems”

Role of private ICT corporations in the defense sector

Anthropic and Palantir Partner to Bring Claude AI Models to Amazon Web Services for US Government Intelligence and Defense Operations
(Business Wire nov '24)

The AI Machine Gun of the Future Is Already Here - The Pentagon is pursuing every available option to keep US troops safe from the rising tide of adversary drones, including a robotic twist on its standard-issue small arms (Wired nov '24)

How Silicon Valley is prepping for War - As frontier models join forces with the Pentagon, corporations and the state are becoming too close

→ The case for targeted regulation - Anthropic warns of AI catastrophe if governments don't regulate in 18 months. real risks in the cyber & CBRN domains (Anthropic oct '24)

D.Eisenhower (Farewell Address 1961): “we must guard against the acquisition of unwarranted influence, whether sought or unsought, by the military-industrial complex. The potential for the disastrous rise of misplaced power exists and will persist”



PHOTOGRAPH COURTESY OF ALLEN CONTROL SYSTEMS

AI in nuclear domain: promises and realities

AI in Military Use: Recent advances in AI/ML increased interest in leveraging AI for military purposes, including nuclear deterrence.

Potential Impact: AI integration in nuclear-armed states* can affect missile early-warning systems, Intelligence, Surveillance, Reconnaissance (ISR) & Nuclear Command, Control, Communications (NC3). Evidence that some states made AI a strategic priority.

Challenges: AI adoption in nuclear domain faces challenges as limited training, unreliability of output, susceptibility to cyberattacks, lack of good-quality data, inadequate hardware.

Rule-based AI vs. ML: Traditional rule-based AI already used in NC3 systems but ML & DL offer more advanced capabilities.

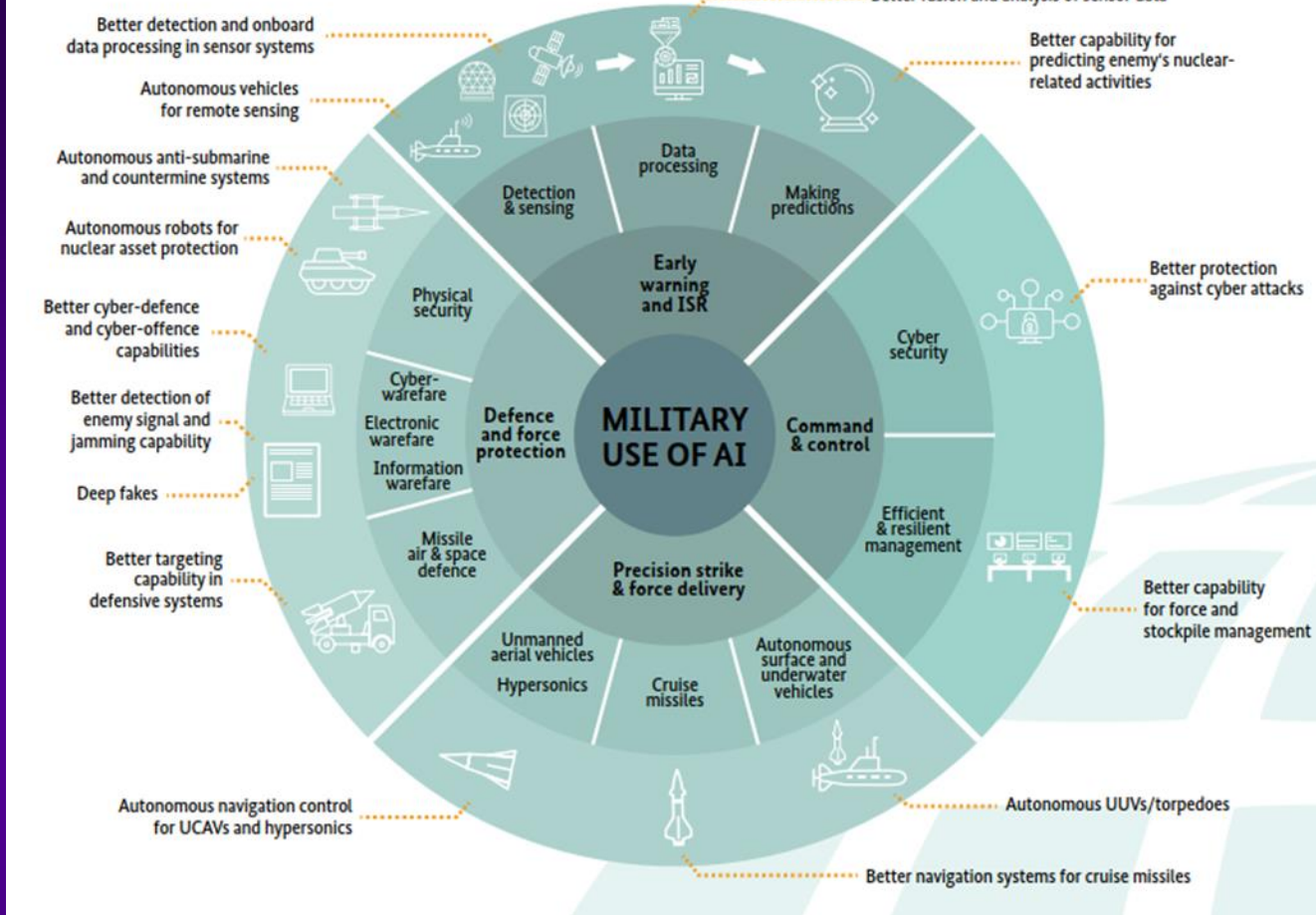
Foundation Models: large-scale AI models (GPT-like) can perform a wide range of tasks but require significant computational resources & high-quality data.

Integration Realities: Despite the potential, integrating advanced AI in nuclear systems is complex and requires addressing technical, integration and resource access challenges.

Stabilizing/destabilizing effects at strategic level?

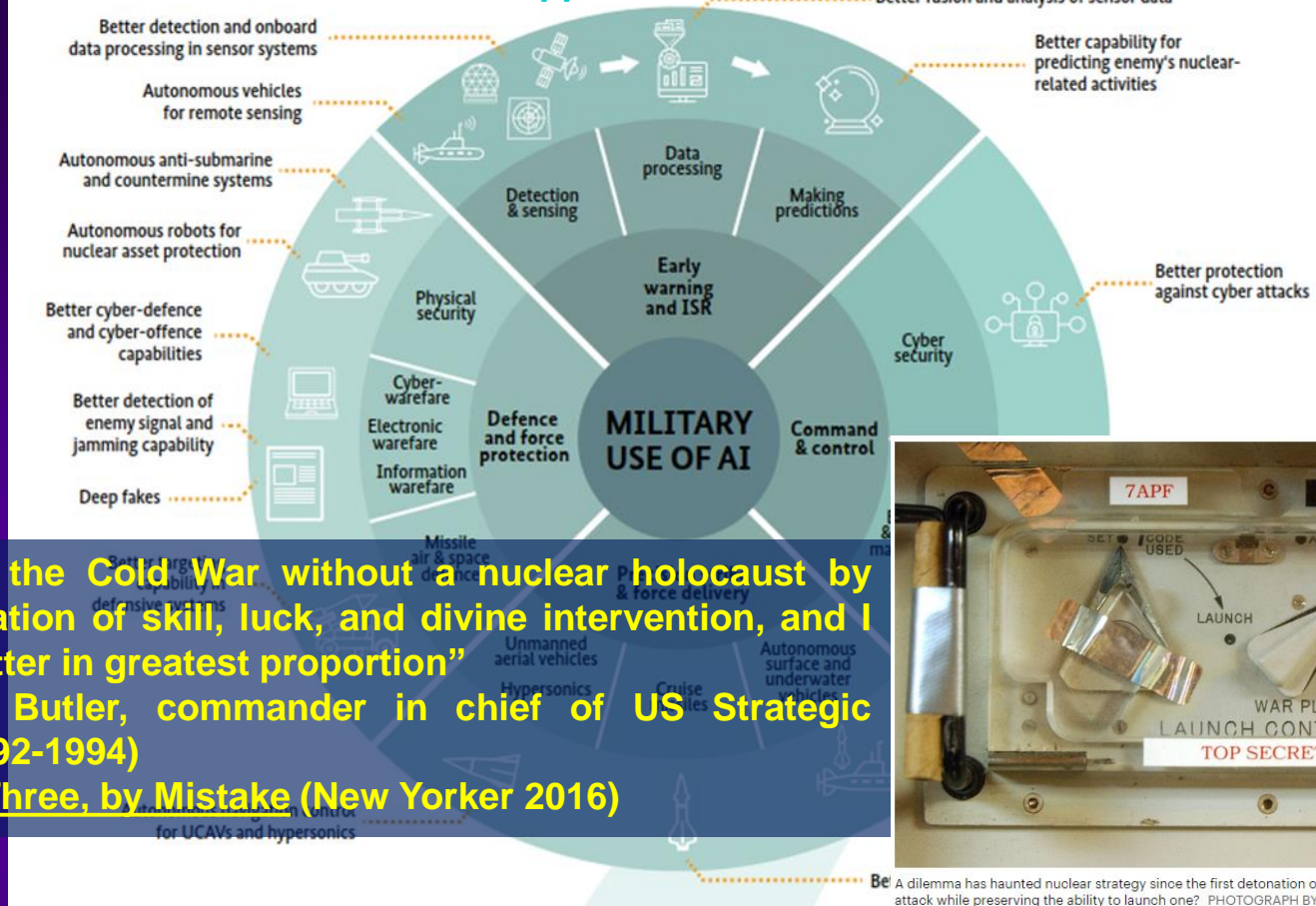


Foreseeable AI application in nuclear deterrence



"Artificial Intelligence, Strategic Stability and Nuclear Risk" (SIPRI 2020)

Foreseeable AI application in nuclear deterrence



Be A dilemma has haunted nuclear strategy since the first detonation of an atomic bomb: How do you prevent a nuclear attack while preserving the ability to launch one? PHOTOGRAPH BY ANDY CROSS / THE DENVER POST VIA GETTY

“We escaped the Cold War without a nuclear holocaust by some combination of skill, luck, and divine intervention, and I suspect the latter in greatest proportion”

Gen. G. Lee Butler, commander in chief of US Strategic Command (1992-1994)

→ World War Three, by Mistake (New Yorker 2016)

➤ Consequences:

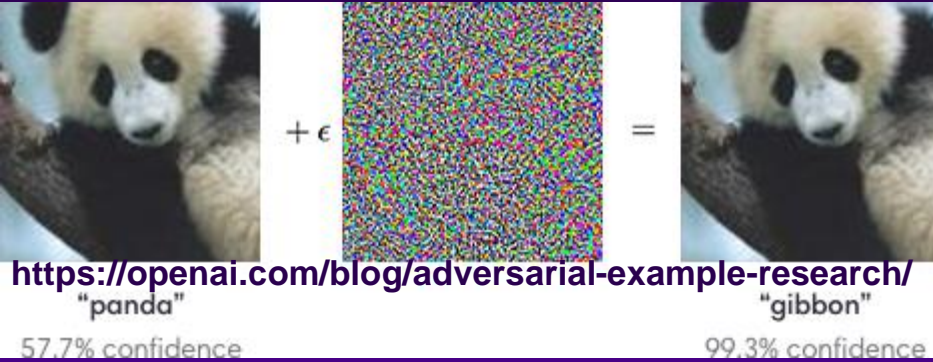
- Investment in AI (even non-nuclear-related) by the adversary could threaten a state's future second-strike capability, generating insecurity, decreasing strategic stability and increasing risk of a nuclear conflict
- AI could fail or be misused in ways triggering an accidental or inadvertent escalation of a crisis or conflict into a nuclear conflict
- Support awareness-raising measures helping relevant stakeholders (governments, industry & civil society) to understand challenges posed by AI in the nuclear arena
- Support transparency & CBMs to reduce misperception/misunderstanding among nuclear-armed states
- Discuss and agree on concrete limits to the use of AI in nuclear forces/infrastructures

AI intrinsic vulnerabilities

- Growing pervasiveness gives rise to “adversarial AI”: exploiting machine learning models to misinterpret inputs into the system and behave in a way that’s favorable to the attacker
- Produce unexpected behavior: attackers create “adversarial examples” often appearing as normal inputs but instead meticulously optimized to break the model (instability & inaccurate predictions). → Hallucinations
- Exploiting a particular behavior in AI internals (Neural Networks) unknown to developers
- Opacity of AI / Machine Learning / Deep Learning internals (NN level). Black box model (even designers cannot explain why an AI system reaches a specific result)
- “Poisoning attacks” during *supervised learning* phase (wrong, noisy, manipulated, non balanced data). “Garbage In / Garbage Out”. Biases (intellectual, ethical)
- Slowly time drifting conditions in *unsupervised learning*
- Backdoors
- Various classes of vulnerabilities (“Failure Modes in Machine Learning” 2019)
- AI algorithms currently lack of interpretability, predictability, verifiability, reliability
- NN used to hide/store/transfer (EvilModel) and/or trigger (DeepLocker) payload/malware
- Inserting a backdoor into a ML system: “ImpNet: Imperceptible and blackbox-undetectable backdoors in compiled neural networks”
AI = Artificial Intelligence or Automating Ignorance??
- XAI (eXplainable Artificial Intelligence): evolution of AI such that results can be understood by humans
(“Peeking inside the black box: a survey on explainable artificial intelligence” 2018)

Securing AI supply chain: 1.secure AI infrastructure 2.secure algorithms 3.secure training process & data 4.identify & manage external data dependencies. Assurance of provenance along the entire technical pipeline (data, model architecture, compiler, h/w specification)

AI intrinsic vulnerabilities



➤ **Opacity of AI / Machine Learning / Deep (even designers cannot explain why an AI**

➤ **“Future(?) developments in AI (also effects on nuclear strategy?):**

- Detection, tracking & targeting adversary forces
- Decision support system, trusted adviser in decision making

➤ **Possible to control/influence/degrade/neutralize AI by controlling sensor data, attacking/poisoning training data, associated sensors & communication networks, humans or processing resources.**

➤ **Intrinsic AI vulnerabilities (adversarial attacks), defense algorithms?! AI as the new cyberattack surface??**

➤ **Pressure to use AI/ML before it is technologically mature?!**

Cyber vulnerabilities & threats evolve following digital technology evolution.

Often new technologies bring in a new spectrum of vulnerabilities.

What are the security threats unique to NN & DL algorithms? Do we know them all?

backdoors/

lware
ackbox-

can be

When Artificial Intelligence Goes Wrong

→ OODA Loop special report (May 2023)

➤ Summary of AI key concerns:

- **Self-Corruption:** algorithms that can teach themselves can corrupt themselves
- **Hallucination** (→ “Why ChatGPT answered queries in gibberish on Tuesday” ZDNET Feb 2024)
- **Inscrutability:** no human can understand what (many/most/all?) ML/DL algorithms are doing
- **Deceivability:** most AI systems assume the trusted training data, what if...
- **New attack vectors**
- **Data protection**
- **Bias** (Google had to pause the image generation feature of its new Gemini AI model yesterday... Feb 2024)
- **Adversary use of AI**

Most of these challenges can be mitigated, and those that cannot might be better understood and dealt with in other creative ways

➤ AI “weaponization”

➤ **Inability to ensure security and compliance, complicated by the unexplainability of complex AI. Issues of fairness and ethics also require deep scrutiny.**

➤ **Some regulations start impacting AI deployments**

➤ **Entering/(jumping into?) the era of geopolitical AI: *cyber-diplomacy***

When A

- Summary of AI
 - Self-Corrupt
 - Hallucinations
 - ZDNET Feb
 - Inscrutable algorithms
 - Deceivability
 - New attacks
 - Data protection
 - Bias (Good model yest)
 - Adversary
- Most of these understood and

- AI “weaponization”
- Inability to ensure complex AI. Issues
- Some regulations
- Entering/(jumping)

OWASP Top 10 for LLM Applications

LLM application security

LLM01: Prompt Injection

This manipulates a large language model (LLM) through crafted inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02: Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03: Training Data Poisoning

This occurs when LLM training data is tampered, introducing vulnerabilities or biases that compromise security, effectiveness, or ethical behavior. Sources include Common Crawl, WebText, OpenWebText, & books.

LLM04: Model Denial of Service

Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.

LLM05: Supply Chain Vulnerabilities

LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre-trained models, and plugins can add vulnerabilities.

LLM06: Sensitive Information Disclosure

LLMs may inadvertently reveal confidential data in their responses, leading to unauthorized data access, privacy violations, and security breaches. It's crucial to implement data sanitization and strict user policies to mitigate this.

LLM07: Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control. This lack of application control makes them easier to exploit and can result in consequences like remote code execution.

LLM08: Excessive Agency

LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.

LLM09: Overreliance

Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

LLM10: Model Theft

This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.

Wrong

“Prompt themselves
“fish on Tuesday”

“(most/all?) ML/DL

“a, what if...

“ts new Gemini AI

“ot might be better

“nexplainability of

Information Warfare & PSYOPs

- Consider Internet as a global communication “medium”, forget about bits
- Information Operations (IO): info manipulation for (counter)propaganda, disinformation, consensus building, discrimination, defamation, delegitimation, censorship/content filtering, Deception, influence attitudes, manipulate target's values, perceptions, beliefs, emotions, reasoning and behavior. Counter-intelligence, ops security

Traditional techniques (centuries old) on a new medium

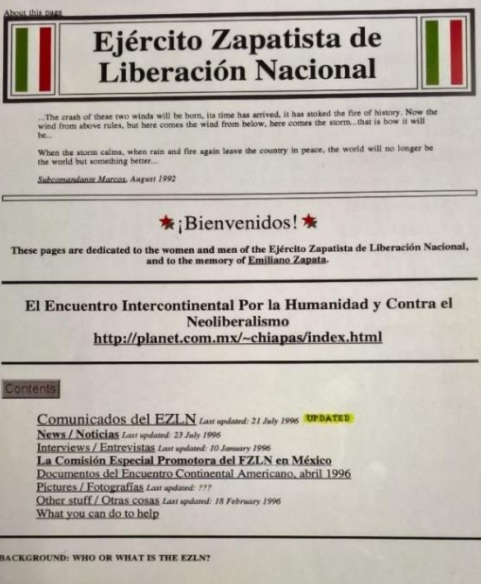
“Nihil est quod videtur” “..Cicero..”

- Real world examples: support to dissident groups, recruitment campaigns & proselytism, use/manipulation of social media/networks for disinformation on wide scale. EZLN ('90 & →today) The Zapatista "Social Netwar" in Mexico (RAND 1998), “Strategic information warfare: a new face of war” (RAND 1996), US Army document (2005), Wikileaks (2010, Assange), NSALeaks (2013, Snowden), CIALeaks (2017), Cambridge Analytica (2018)
- Network is an ubiquitous surveillance environment
- Info war: primary political (strategic) value. “c” contribute to political and social instability of a country between military and civilian domains/infrastructures
- US Psychological operations (strategic, operational) PSYOPs units in US Army to convey selected info to target audiences to influence emotions, motives, objectives of governments, organizations, groups, individuals (many other countries have similar activities)

“Service member influencers are helping DOD recruit, Per

➤ What, when no one will trust anything anymore?





U.S. Army War College

U.S. Army War College
Dept. of Military Strategy, Planning, and Operations

November 2006
AY07 Edition

Information Operations Primer



CNN let army staff into newsroom (The Guardian 2000)



recruitment campaigns & works for disinformation on social Netwar" in Mexico
pace of war" (RAND 1996),
ange), NSAleaks (2013.

➤ **Real world example of proselytism, use of wide scale. EZLN (RAND 1998), "Str US Army document Snowden) CIA e**

➤ **Network** The Information Support Force is a newly created strat
➤ **Info w:** the armed forces (China's military network 2024)
contribu

Deepfakes: created by AI and currently undetectable by AI (Not anymore? "OpenAI secret tool revealed", AI photos manipulation detection summer 2024)

Hidden watermarking/steganography?

Is AI destroying credibility of/on Internet?

An update on disrupting deceptive uses of AI (OpenAI oct 2024)



➤ **What, when no one will trust anything anymore?**

Cyber diplomacy



Going forward: from a Single Track (GGE) to two Parallel Tracks (GGE + OEWG).

In 2018 established two parallel (hopefully converging & complementary) processes to discuss ICT security in 2019-2021 - Outcome: VERY difficult & hard convergence reached in 2021

- GGE of 25 members (chaired Amb. Guilherme de Aguiar Patriota BR). Final report in 2021. Chair (Amb. G. de Aguiar Patriota BR) holding consultations with the wider membership in between sessions. Consultations with regional organizations (AU, EU, OAS, OSCE, ASEAN)
→ Final report (march 2021)
- Open-Ended Working Group (OEWG) open to all Member States. Report to GA in 2020. OEWG holding inter-sessional consultative meetings with private sector, civil society*, NGOs, academia. New wider multistakeholder approach ("The Value of Multistakeholder Engagement", hypothetical form of new UN governance?!)

Cyber-OEWG should refer to shared conclusions of previous GGEs (2015 A70/174) and represents by itself a sort of CBM

Current process: OEWG 2021-2025 (Chair Amb. B. Gafoor, SG)

(*) GPS representing Pugwash Conferences on Science and World Affairs in cyber-OEWG

- "The right to privacy in the digital age", A/RES/68/167 (2013)
- Proposed universal code of conduct for Information Security (2015)
- CT Declaration on responsible state behaviour in cyberspace (2017)

UN GGE¹ on LAWS²

in the context of CCW (Convention on Conventional Weapons)

AI, robotics, and weapons

Intergovernmental
negotiations

- Mandate: group is to consider proposals and elaborate (by consensus) possible measures related to normative & operational framework on emerging technologies in the area of LAWS, bringing in expertise on legal, military & technological aspects
- International law, in particular the UN Charter and IHL (International Humanitarian Law), as well as relevant ethical perspectives should guide the work of the Group
- Several proposals:
 - Legally VS non legally binding instruments under the framework of the CCW
 - Clarify implementation of existing obligations under international law (in particular IHL)
 - Discussion on prohibition VS regulation and whether further legal measures are needed
 - Limitation of types of targets, duration & scope of operations with which weapon systems can engage
 - Adequate training to human operators
 - In cases where the weapon system based on technologies in the area of LAWS cannot comply with international law, the system must not be deployed



UNIDIR on LAWS (2021): defining AWS, characteristics, human element, responsibility & accountability, moral & ethical considerations, legal reviews

What happens when a lethal autonomous weapon relies on an incomplete dataset? (video ZDNET 2021)

Overview & 2022 GGE/LAWS Report (Digital Watch Observatory)

CCW GGE on LAWS Documents and working papers (2022)

Advance version of 2023 report (CCW/GGE.1/2023/2)

2024 CCW GGE on LAWS resources (Reaching Critical Will)

(1) Group of Governmental Experts

(2) Lethal Autonomous Weapon Systems

IT seminar, CERN - November 2024

Gian Piero Siroli

**A Diplomat's Guide
to Autonomous
Weapons Systems**

UN SG report on AWS (jul 2024)

Views of states, civil society, academia, industry & others.

Stop Killer Robots stressed the urgency for international preventive action to address humanitarian, legal, security, technological & ethical issues related to AWS. Opportunity to take meaningful action by next UN GA.

Support towards negotiations of an international legally binding instrument banning AWS.

“Time is running out for the international community to take preventive action on this issue. I therefore reiterate my call for the conclusion, by 2026, of a legally binding instrument to prohibit lethal autonomous weapons systems that function without human control or oversight and that cannot be used in compliance with international humanitarian law”.

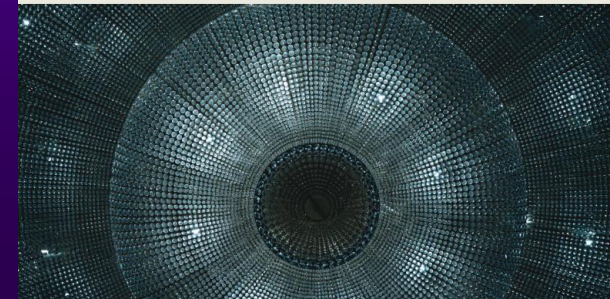
Earlier Vienna Conference on Autonomous Weapons with Chair’s summary on regulation of AWS supported by 35 states (as of report’s writing)

**AWS DIPLOMACY
REPORT**

Civil society perspectives on the
Vienna Conference on Autonomous Weapon
Systems “Humanity at the Crossroads”
29-30 April 2024

VOL.1 NO.1

7 May 2024



UN Security Council discussion on AI

(Geopolitics of AI)

- On July 4th 2023 UN Security Council held its first-ever meeting on the potential threats of AI to international peace and security, organized by UK. AI has tremendous potential but also major risks about possible use for example in autonomous weapons or in control of nuclear weapons.

China said the technology should not become a “runaway horse” and US warned against its use to censorship or repression.

The UN Secretary General emphasized the potential of AI to accelerate human development while also cautioning against the malicious use of AI.

➔ Secretary-General's remarks to the Security Council on Artificial Intelligence (UN Secretary-General) (July 18th 2023)

- UN council to hold first meeting addressing threats of AI to global peace (AI Arabiya)
- Guterres calls for AI ‘that bridges divides’, rather than pushing us apart (UN News)
- International Community Must Urgently Confront New Reality of Generative, Artificial Intelligence, Speakers Stress as Security Council Debates Risks, Rewards (UN Press)



UN Security Council discussion on AI

(Geopolitics of AI)

A.Guterres (UN SG): "...AI has tremendous potential but also major risks about possible use for example in autonomous weapons or in control of nuclear weapons...shocked by the newest form of generative AI, a radical advance in its capabilities (probably a step function in AI evolution)...even its own designers have no idea where their stunning technological breakthrough may lead...consider the impact of AI on peace and security, already raising political, legal, ethical & humanitarian concerns..."

An example:

"Dual use of artificial-intelligence-powered drug discovery" (Nature 2022) - How AI can be used to design toxic/deadly biochemical weapons. "In less than 6 hours...on our in-house server, our model generated 40,000 molecules...AI designed not only VX (nerve agent), but also many other known chemical warfare agents...predicted to be more toxic...The genie is out of the medicine bottle when it comes to repurposing our machine learning".

Implications for Chemical and Biological Weapons Conventions.

- AlxBio White Paper 1: Introduction to AI and Biotechnology (US Senate, NSCEB, jan 2024)
- GPT-4 'mildly useful' in creating bioweapons, says ChatGPT (feb 2024)

UN Security Council discussion on AI

(Geopolitics of AI)

UN adopts first global artificial intelligence resolution

(Reuters march 2024)



A.Guterres (2024): “We cannot sleepwalk into a dystopian future where the power of AI is controlled by a few people - or worse, by opaque algorithms beyond human understanding. We need rules. Safety. Universal guardrails. How we act now will define our era”.

An example of how things can go fast at UN



United Nations cyber-OEWG (Open-Ended Working Group on Information and Communication Technologies (2021-2015))

July 2024: 3rd Annual Progress Report (APR) adopted by consensus.


Chairman: “this APR is an important step towards the future UN single-track mechanism to drive ICT security”

...and also a CBM by itself...

3rd APR Main topics:

- Existing and Potential Threats
- Rules, Norms and Principles of Responsible State Behaviour
- International Law
- Confidence-Building Measures
- Capacity-Building
- Regular Institutional Dialogue
- Point of Contact (PoC) directory at technical and diplomatic level (already being operationalized)
- Program of Action (PoA) for permanent, global, inclusive mechanism to coordinate ICT domain in the context of international security

United Nations A/AC.292/2024/CRP.1

 **General Assembly** 12 July 2024

English only

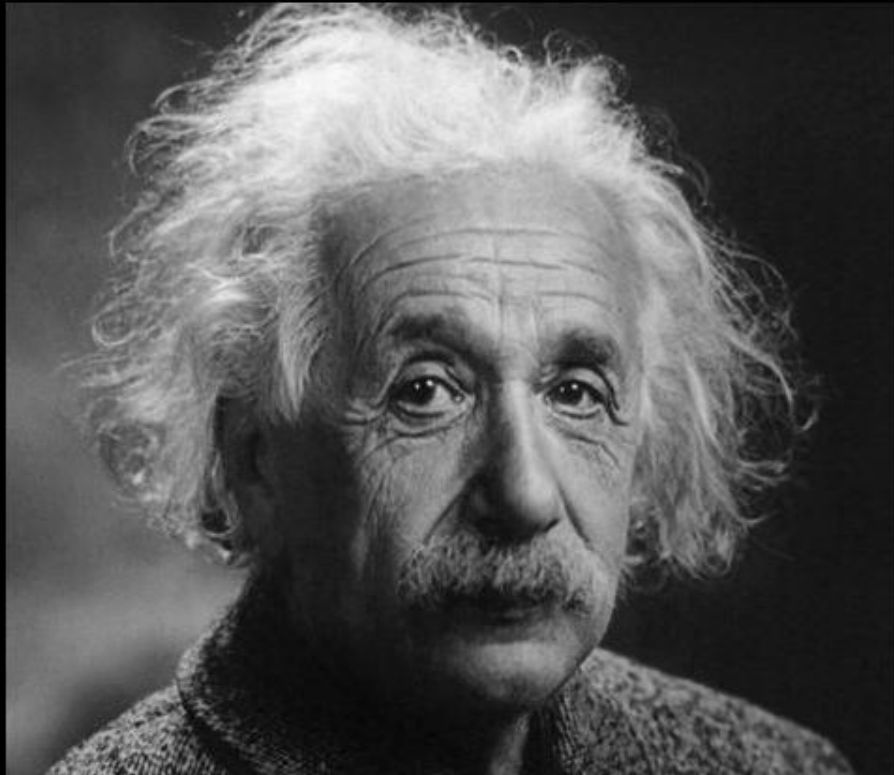
Open-ended working group on security of and in the use of information and communications technologies 2021-2025
Eighth substantive session, New York
8-12 July 2024

Draft Annual Progress Report

A. Overview

1. The sixth, seventh and eighth formal sessions as well as the dedicated intersessional meetings of the Open-ended Working Group (OEWG) on the security of and in the use of Information and Communications Technologies (ICTs) 2021-2025 took place in a geopolitical environment that continues to be challenging, with rising concerns over the malicious use of ICTs by State and non-state actors that impact international peace and security.
2. At these sessions, States recalled the consensus decisions and resolutions of the General Assembly in which States agreed they should be guided in their use of ICTs by the OEWG and GGE reports.¹ In this regard, States further recalled the contributions of the first OEWG, established pursuant to General Assembly Resolution 73/27, which concluded its work in 2021, through its final report agreed by consensus,² as well as noted the Chair's summary and list of non-exhaustive proposals annexed to the Chair's summary, and recalled the contributions of the sixth Group of Governmental Experts (GGE), established pursuant to General Assembly Resolution 73/266, which concluded its work in 2021, through its final report agreed by consensus.³
3. Furthermore, States reaffirmed the consensus first and second annual progress reports (APRs) of the current OEWG,⁴ the consensus report of the 2021 OEWG on developments in the field of ICTs in the context of international security and the consensus reports of the 2010, 2013, 2015, and 2021 GGEs.⁵ States recalled and reaffirmed that the reports of these Groups “recommended 11 voluntary, non-binding norms of responsible State behaviour and recognized that additional norms could be developed over time”, and that “specific confidence-building, capacity-building and cooperation measures were recommended”. States also recalled and reaffirmed that “international law, in particular the Charter of the United Nations, is applicable and essential to maintaining peace, security

¹ GA decisions 77/512 and 75/564, GA resolutions 70/237 and 76/19.
² A/75/816.
³ A/76/135.
⁴ A/77/275 and A/78/265 respectively.
⁵ A/65/201, A/68/98, A/70/174 and A/76/135.



Solution is not at the ICT technical level only

“The importance of securing international peace was recognized by the really great men of former generations. But the technical advances of our times have turned this ethical postulate into a matter of life and death for civilized mankind today, and made it a moral duty to take an active part in the solution of the problem of peace, a duty which no conscientious man can shirk” (A. Einstein 1934)

Russel-Einstein Manifesto (1955) (Pugwash founding charter)

LIFE AFTER ICT (AND AI)

