The safeguards of the 1949 Geneva Conventions in the face of military use of AI

Outline

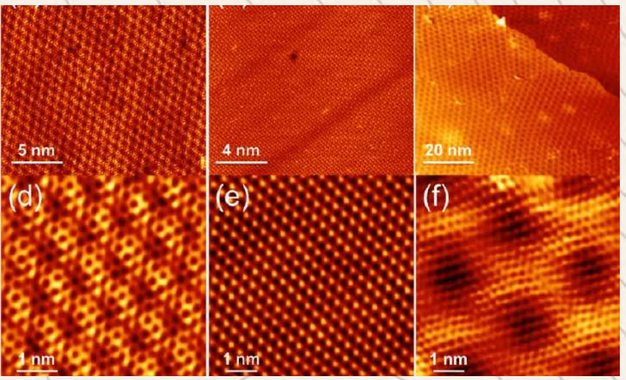
- SOTA: domains, systems in use
- Regulation: current state, initiatives, prospects
- Tension with AI and the GC? :
  discrimination, proportionality, dignity

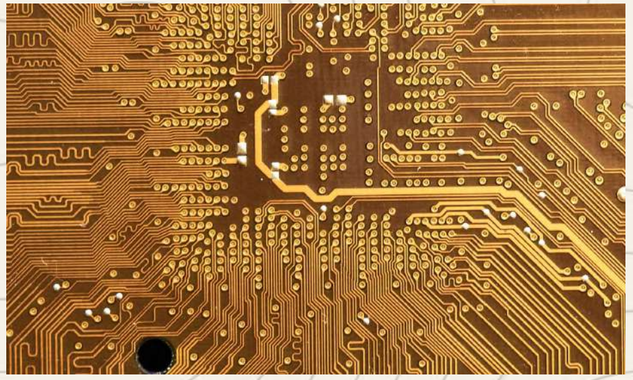
SOTA

AI in military domain: one of the EDTs

Change in warfare through **EDTs**:

development of big data, **AI**, biotechnology, robotics, nanotechnology, hypersonic systems and space technologies, quantum computing (for fast sensing and communications), novel materials (f.e. graphene)



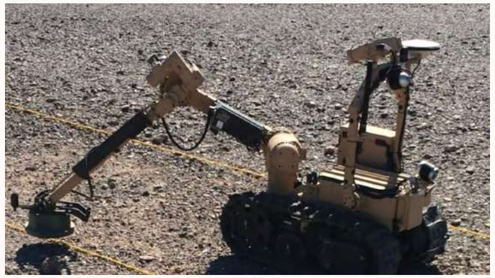


SOTA

Training

E.g. virtual simulations, tracking learning process,...

Diffusion of explosives

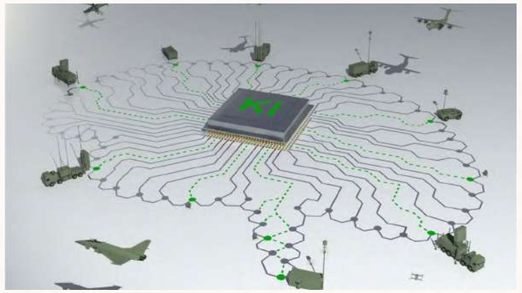


Targeting

Selection and engagement of military targets

Logistics

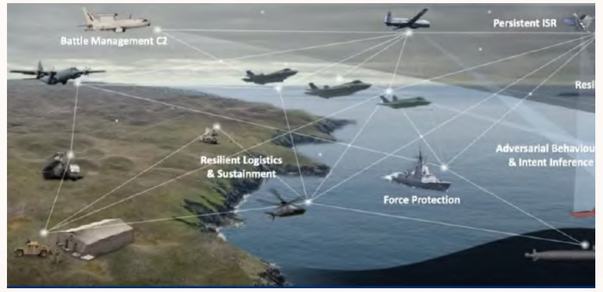
E.g. Recommendations routes and modes of transportation, organization of personnel rotation, predictive maintenance



Mission planning

Command and control; information management

E.g. weapon selection recommendations based on target and environmental data within a given area of operation, suggestions actions based on CDE and weapon capability, pattern recognition, sensor data analysis



SOTA

(1) AI uses similar to civilian applications

= adapted for military use: related to the daily running of the military infrastructure; navigation and transportation; logistical and organizational planning; training; system robustness and resilience

e.g. HR and payroll management; navigation and transportation; predictive maintenance, route recommendation; cybersecurity tools

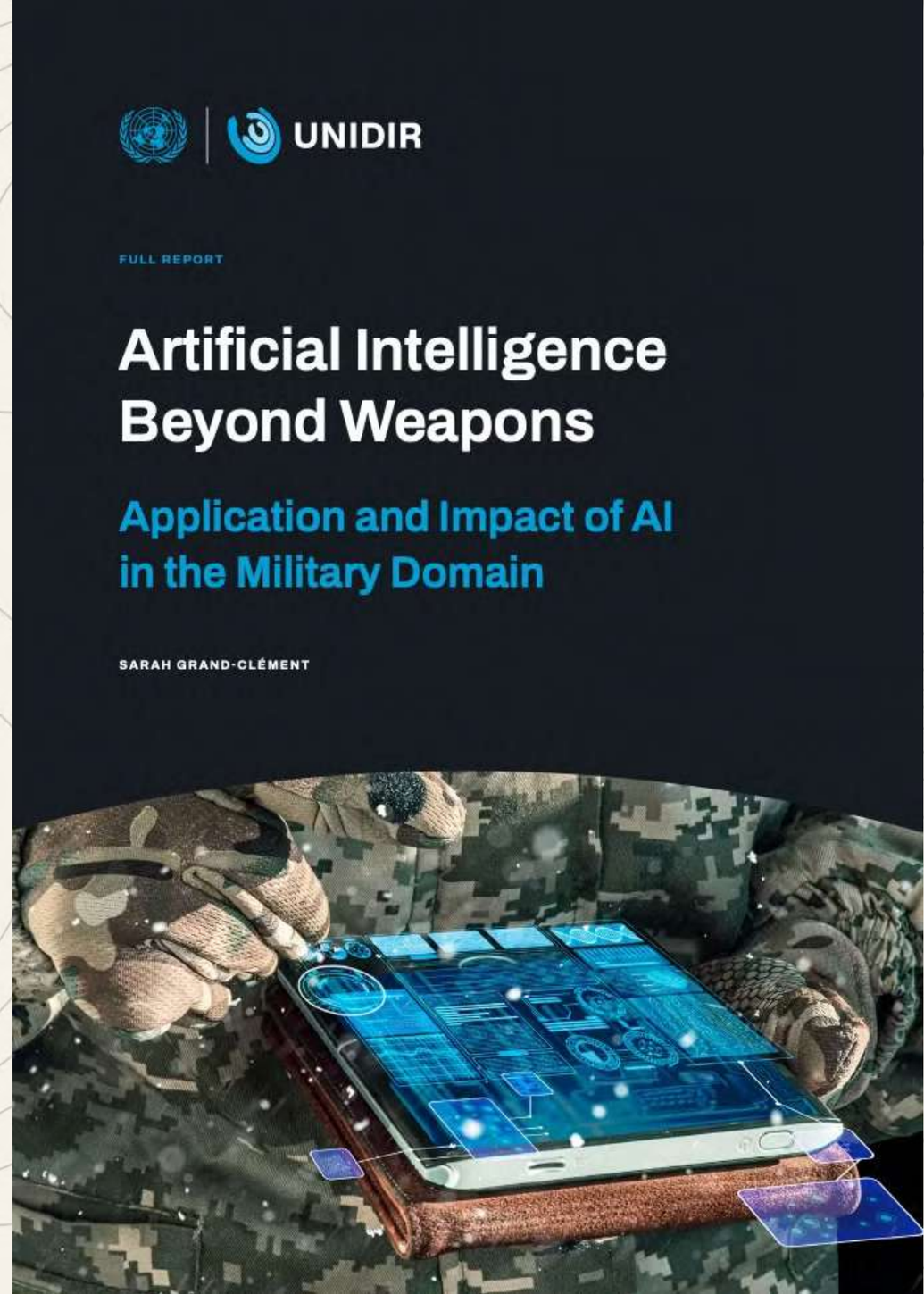
(2) Specific to military environment but prior to engagement

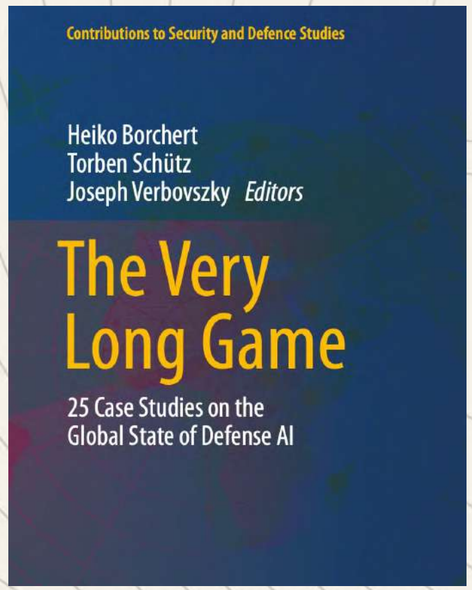
e.g. strategic planning, coordination and decision-making to prepare for military operations; suggest weapons choices; recommendation courses of action

(3) Related to target selection and engagement

= support human-decision making and AWS

e.g. target detection, classification, selection, engagement recommendation

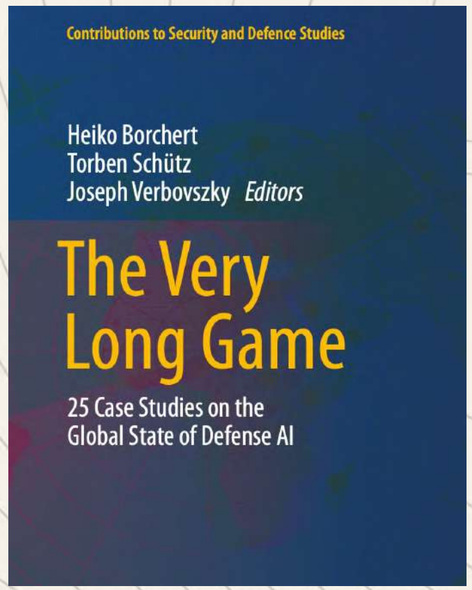




Useful: DAIO - Defence AI Observatory (https://defenseai.eu/english) by Helmut Schmidt University in Hamburg that monitors and analyzes the use of artificial intelligence by armed forces

Table 9 International defense AI fielding use cases

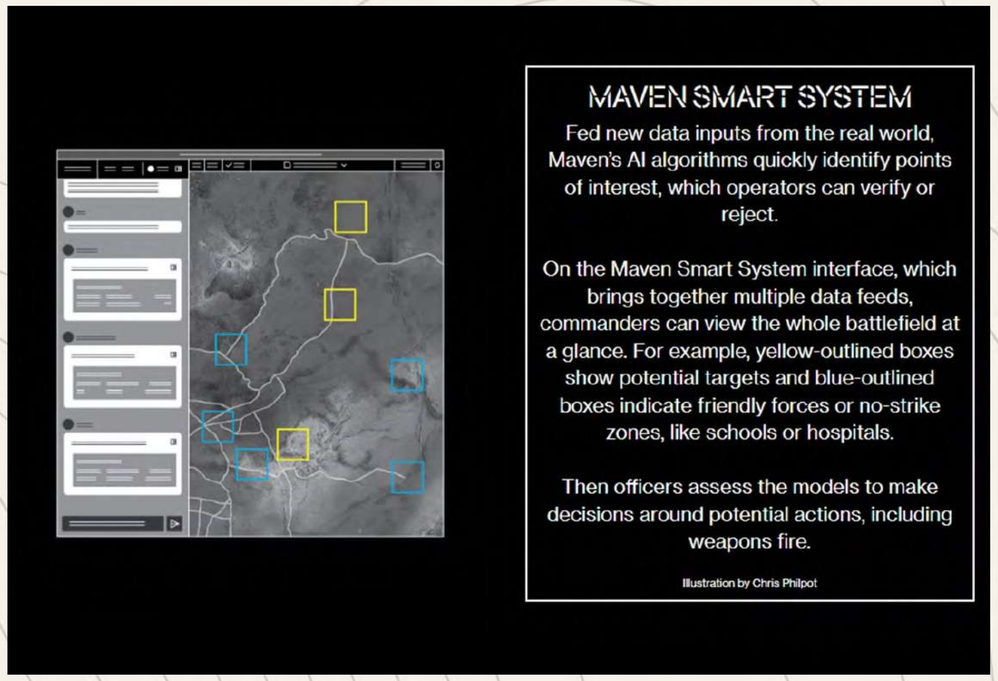
Use case	USA	CAN	GBR	SWE	FIN	EST	DNK	DEU	NLD	FRA	ESP	ITA	GRE	TUR	RUS	UKR	ISR	IRN	IND	CHN	JPN	KOR	TWN	SGP	AUS
Air/missile defense			■				■	■		■							■	■	■				■	■	
Air traffic management							■			■															
Battle/combat management			■		■		■	■		■									■					■	
Border security			■										■	■				■	■					■	
Close-in weapon systems			■				■																		
Command and control							■					■			■		■			■	■			■	
Common operational picture	■							■		■			■		■	■			■					■	
Cyber/CNO	■		■				■			■					■	■	■				■			■	
Data analytics and data management		■		■	■		■	■	■	■	■			■		■	■					■	■	■	■
Decision and planning support											■			■	■	■			■					■	
Fire support (e.g., artillery)															■	■								■	
Force protection								■									■					■		■	
Human performance modification											■														
Intelligence, surveillance, reconnaissance	■	■	■	■		■	■	■		■				■	■	■	■		■	■	■	■		■	
Influence/ information operations			■							■	■				■	■				■				■	
IOD/MCM/mine clearance/UXO		■			■		■												■					■	
Loitering munition			■											■											
Manned-unmanned teaming			■									■		■	■				■			■		■	
Medical services		■						■														■	■		
Mission planning					■		■																	■	
Modelling, simulation, red teaming, Wargaming			■																					■	■
Precision effects			■		■			■		■			■		■	■		■							
Predictive maintenance, logistics, MRO	■				■		■	■		■							■		■	■			■	■	■
Swarming	■		■																■						
Training (incl. Simulation-based training)			■							■		■	■	■					■	■	■		■	■	
Target detection, classification, identification	■	■	■					■	■	■		■	■	■		■	■	■	■	■	■			■	
Uncrewed systems	■		■		■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■
Uncrewed systems: Counter solutions																								■	■



Useful: DAIO - Defence AI Observatory (https://defenseai.eu/english) by Helmut Schmidt University in Hamburg that monitors and analyzes the use of artificial intelligence by armed forces

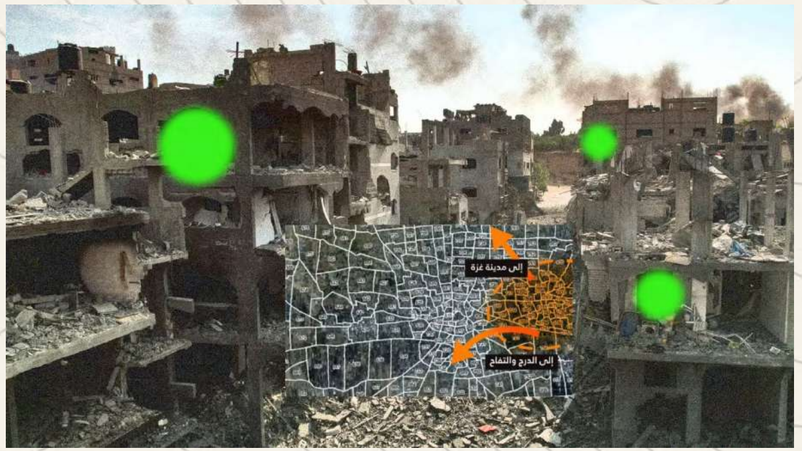
Table 9 International defense AI fielding use cases

Use case	USA	CAN	GBR	SWE	FIN	EST	DNK	DEU	NLD	FRA	ESP	ITA	GRE	TUR	RUS	UKR	ISR	IRN	IND	CHN	JPN	KOR	TWN	SGP	AUS
Air/missile defense			■				■	■		■							■	■	■				■	■	
Air traffic management							■			■															
Battle/combat management			■		■		■	■		■									■					■	
Border security			■										■	■				■	■					■	
Close-in weapon systems			■				■																		
Command and control							■					■			■		■			■	■			■	
Common operational picture	■							■		■			■		■	■			■					■	
Cyber/CNO	■		■				■			■					■	■	■				■			■	
Data analytics and data management		■		■	■		■	■	■	■	■			■		■	■					■	■	■	■
Decision and planning support											■			■	■	■			■					■	
Fire support (e.g., artillery)															■	■								■	
Force protection								■									■					■		■	
Human performance modification											■														
Intelligence, surveillance, reconnaissance	■	■	■	■		■	■	■		■				■	■	■	■		■	■	■	■		■	
Influence/ information operations			■							■	■				■	■				■				■	
IOD/MCM/mine clearance/UXO		■			■		■												■					■	
Loitering munition			■											■											
Manned-unmanned teaming			■									■		■	■				■			■		■	
Medical services		■						■														■	■		
Mission planning					■		■																	■	
Modelling, simulation, red teaming, Wargaming			■																					■	■
Precision effects			■		■			■		■			■		■	■		■							
Predictive maintenance, logistics, MRO	■				■		■	■		■							■		■	■			■	■	■
Swarming	■		■																■						
Training (incl. Simulation-based training)			■							■		■	■	■					■	■	■		■	■	
Target detection, classification, identification	■	■	■					■	■	■		■	■	■		■	■	■	■	■	■			■	
Uncrewed systems	■		■		■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■
Uncrewed systems: Counter solutions																								■	■



Manson, K. (2024, 28 February) AI Warfare is Already Here. Bloomberg.

Project Maven, an AI development contract for the US military: assisting people to view ISR images



System used for target selection in Gaza for bombing , recommendations and ex post assessments of intelligence analysts

(accelerated: 50-100 targets/year → 100 targets/day of which 50% are attacked)

Cf. Habsora ('the Gospel'), Lavender, Where's daddy

https://www.972mag.com/lavender-ai-israeli-army-gaza/

Key developers

Numerous countries have recognised the potential of AI for strengthening their military capabilities in defence and innovation strategies

→ Countries commonly considered leaders:

5 permanent members of the UN Security Council (US, China, Russia, UK, France) and Israel, Republic of Korea, Turkey, Australia

1) Clear intention (e.g. China, Russia)

2) Robust hardware infrastructure (e.g. Taiwan, South Korea)

3) Advanced expertise in AI (e.g. US)

Putin says the nation that leads in AI 'will be the ruler of the world' / The Russian president warned that artificial intelligence offers 'colossal opportunities' as well as dangers

By James Vincent, a senior reporter who has covered AI, robotics, and more for eight years at The Verge.

Sep 4, 2017 at 10:53 AM GMT+2

0 Comments (0 New)

Photo by Lintao Zhang/Pool/Getty Images

Table 2 Three strategic drivers of defense AI

Fear of missing out (FOMO)	AI as a capability multiplier	Threat-based thinking
DNK, FRA, GRE, ITA, TWN	AUS, CAN, DEU, DNK, ESP, EST, FIN, FRA, GBR, GRE, IRN, ISR, ITA, JPN, KOR, NLD, RUS, SGP, SWE, TUR, TWN, USA	CHN, GRE, IND, IRN, ISR, JPN, KOR, RUS, TWN, UKR, USA

Country Code: AUS Australia, CAN Canada, CHN China, DEU Germany, DNK Denmark, ESP Spain, EST Estonia, FIN Finland, FRA France, GBR United Kingdom, GRE Greece, IND India, IRN Iran, ISR Israel, ITA Italy, JPN Japan, KOR South Korea, NLD The Netherlands, RUS Russia, SGP Singapore, SWE Sweden, TUR Turkey, TWN Taiwan, UKR Ukraine, USA United States

Regulation

IHL
Geneva Conventions 1949 and AP I and II, art. 36 AP I, Customary IHL

European Parliament resolution
20 sept 2018 on autonomous weapon systems (2018/2752(RSP))

No technology specific regulation
e.g. Ottawa Convention, BWC, CWC,...

UN debate: GGE LAWS
-11 guiding principles (2019)
-UN General Assembly Resolution (2023)

Recent initiatives
REAIM Summit (the Hague 2023, Seoul 2024), Political declaration US, conferences in Vienna 2024, Costa Rica, Luxembourg...

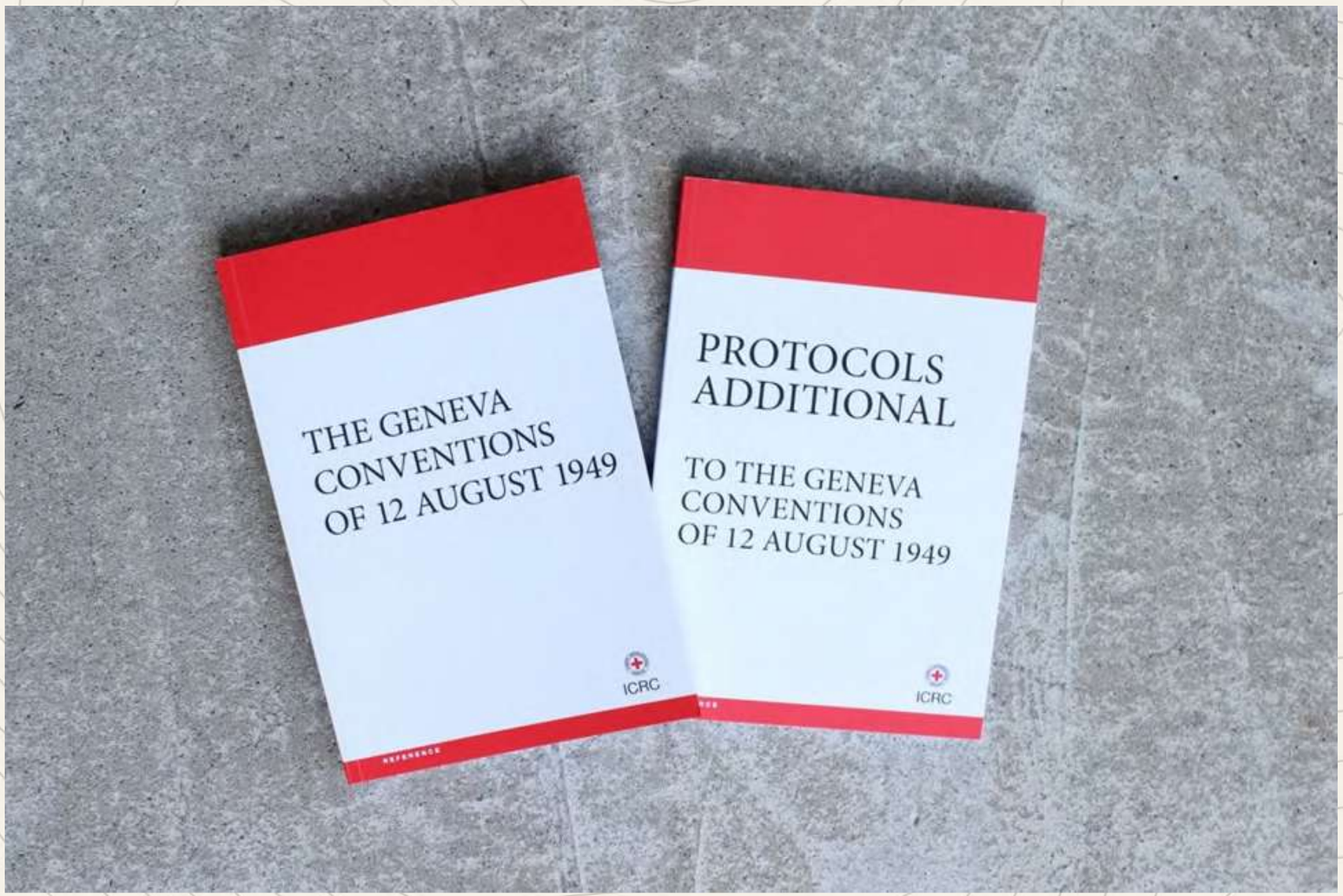
Other initiatives
Certification (min. skills and knowledge) NATO's DARB working on responsible AI certification standard to help industries and institutions make sure new AI and data projects in line with int. law and NATO norms and values (incl. quality controls),...

New binding rules?

(1) Multilateral legally-binding treaty

(2) Two-tier approach: ban on certain systems, regulations for other systems

(3) Current international framework (incl IHL and customary int law) is sufficient and at most needs to be clarified


THE GENEVA CONVENTIONS OF 12 AUGUST 1949
ICRC
PROTOCOLS ADDITIONAL
TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949
ICRC

Tensions between AI and IHL?

- ~~Principle of distinction~~
- Principle of proportionality
- Human dignity

Existing problems in military **exacerbated due to technology?**

Tensions between AI and IHL?

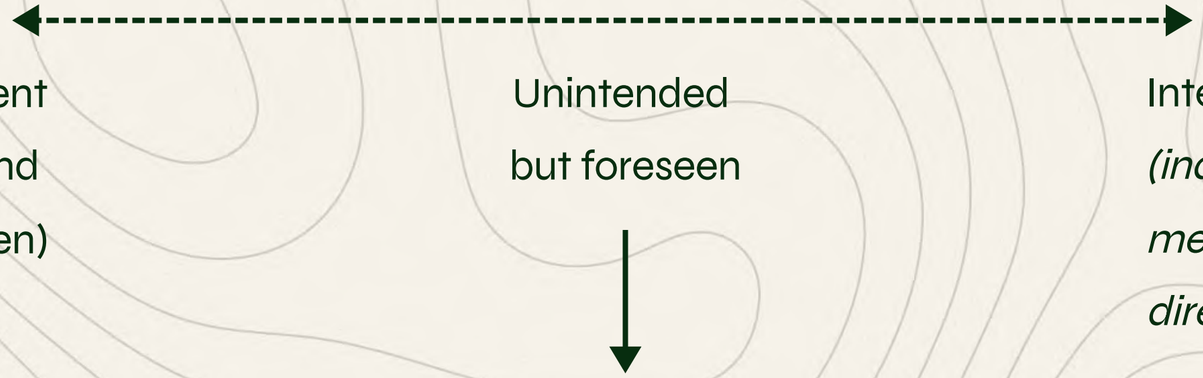
Distinction

- Combatants must distinguish between civilians and combatants and direct their attacks only against the latter (Art. 48)
- Indiscriminate attacks are prohibited, and are defined as:
  - those not directed at a military objective (51.4.a)
  - employing a method or means whose effects cannot be so directed (51.4.b)
  - or which may be expected to cause incidental harms to civilians which would be excessive in relation to the military advantage gained (51.5.b)
- Distinction provides protection to enemy combatants who are to be deemed *hors de combat (Art. 41.1)*

Tensions between AI and IHL?

Distinction

However, war can yield adverse outcomes for civilians and *hors de combat:*



Adherence to the principle of dist is crucial for accurately categorising these situations, as it directly affects the determination of culpability

Tensions between AI and IHL?

✕ Distinction

→ Distinction in practice

In **conventional war**, distinction is a straightforward affair:

recognise combatants attempting to surrender or those who are incapacitated

↓

Internationally recognised ways to signal surrender:

white flags, hands raised, throwing down weapons, aircraft waggling wings

In **irregular war and modern conflicts**, more challenging: lack clear battlefield, states and non-state actors engaged in asymmetric urban warfare, insurgents may not display clear military insignias, blending in with civilian population

↓

Anyone directly participating in hostilities loses their protection against attacks, but in case of doubt whether a person is a civilian, person shall be considered to be a civilian (AP I, art. 50.1)

Tensions between AI and IHL?

✕ Distinction

Introduction AWS worsen problem?

"Compliance with principle of distinction and rules protecting combatants hors de combat already presents formidable challenges. The introduction of AWS to target persons can only increase these challenges (…) It is difficult to envisage realistic combat situations where AWS use against persons would not pose a significant risk of IHL violations" (ICRC 2020, p. 1345)

→ Challenge in relation to AI systems extends beyond the ability of algorithms to identify visual markers such as uniforms

Tensions between AI and IHL?

✕ **Distinction**

→ Challenge = how to train systems to be as effective/more effective than humans in making situational judgements regarding an individual's liability to be targeted?

↓

'intent' to decide legitimate target dependent on **'proxy features'**, but currently no fixed list for these proxies (i.e. correlates with: location, age, communication with suspicious persons, carrying a weapon but based on context, intelligence, experience humans in the field)

(1) Establishing such a list to create parameters for a rule-based model

(2) ML-based system: dealing with uncertaity of which proxies the model considers based on training dataset

Tensions between AI and IHL?

✕ Distinction

:. No one currently advocates for autonomous systems to independently make distinctions in scenarios involving humans, at best AI-DS as tools to assist humans in their assessment

? unclear whether relying on these systems for advice could satisfy legal requirements (Holland 2024)
? autonomous attacks against clearly identifiable military objects/means of war by nature such as tanks, missiles

"most participants tolerated the scenario involving a loitering munition designed to direct attacks against enemy tanks in communication-denied environments (i.e. wit no option for direct supervision or intervention once the system is activated), free of civilians and with a loitering time of two hours"
(Bruun 2024)

Tensions between AI and IHL?

Proportionality

Distinction does not preclude one from carrying out attacks which can be expected to cause incidental harm to civilians

However, collateral costs must be balanced with the concrete and direct military advantage (art. 51.5 (b) AP I) → "extensive but never excessive"

Tensions between AI and IHL?

✕ Proportionality

→ Challenge = how to design system capable of re-assessing situations in real-time at tactical level (i.e. integrating active learning capabilities where models can continue to update after deployment)

↓

Currently decided case-by-case and on subjective judgement and states and military lawyers have so far refused quantifying the value of civilian lives in relation to military advantage, the boundary between extensive and excessive harm remains imprecise

? Does context-dependence require human determination?

Tensions between AI and IHL?

✕ **Human dignity (= deontological)**

Underpins IHL, ethical imperatives permit **only humans to decide over life and death** of other humans

-Delegating life-and-death decisions to AI systems is dehumanising as it turns humans into data points, variables in a computational equation
-Lack of respect towards humans (civilians and combatants), fundamental asymmetry between humans and machines

? Fundamental difference with other modern weapons systems?
(e.g. remote controlled drones, precision-guided missiles etc)

Thank you!

ann-katrien.oimann@mil.be
ann-katrien.oimann@kuleuven.be