

Autonomous weapon systems and the impact on strategic stability

Moa Peldán Carlsson

20 February 2020

Impact on strategic stability

Discussions focus on legal, ethical and operational implications

What about strategic stability???

Does autonomy in weapon systems

... undermine strategic stability?

... lead to an arms race and escalation?

... increase the risk of a nuclear launch?



SIPRI mapping study

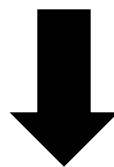


- 3 reports, last one out in 2020
- Find on SIPRI's website



Definitions: AWS

Autonomous weapon systems



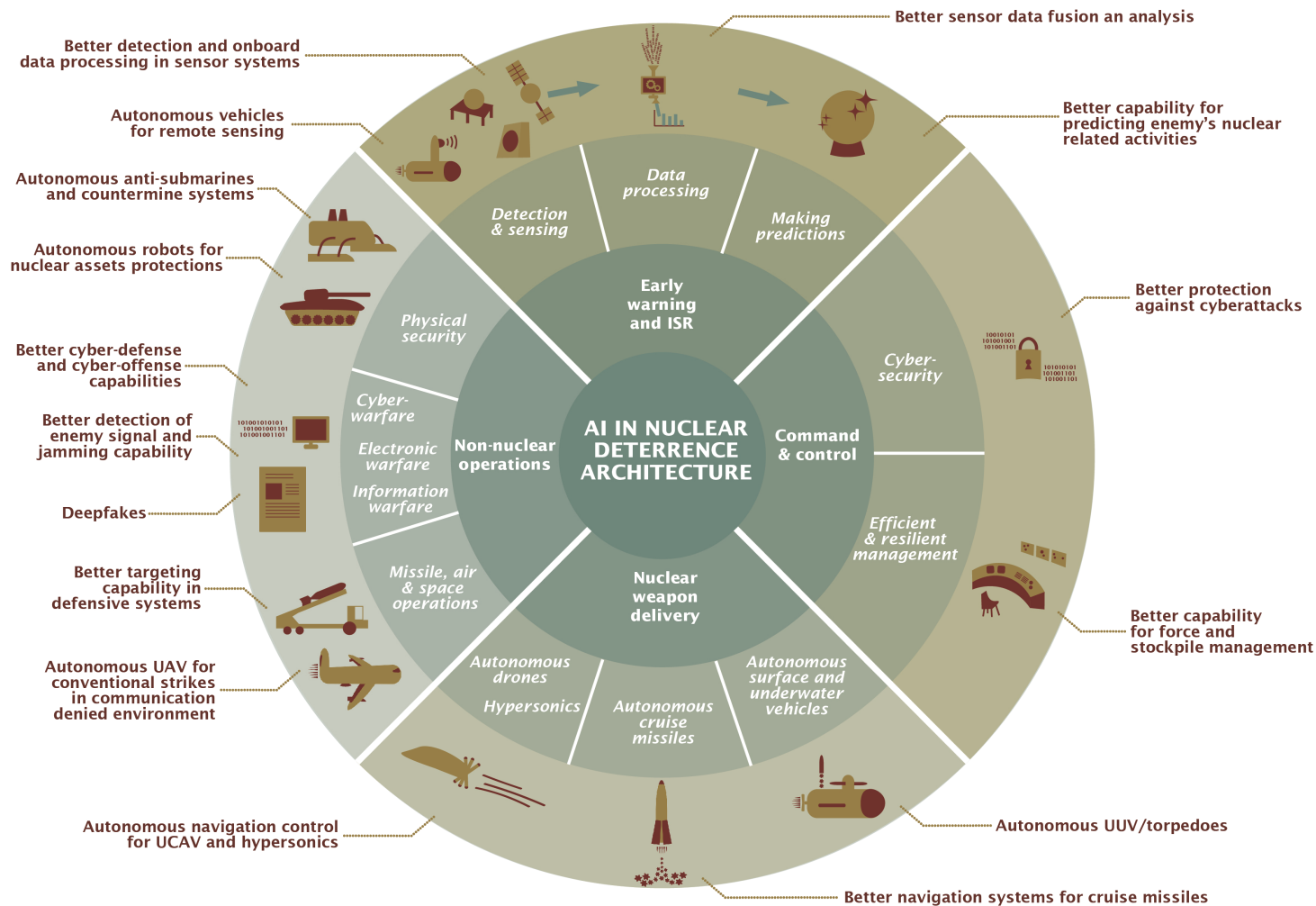
Autonomy *in* weapon systems

Definitions: strategic stability

- "A state of affairs in which countries are confident that their **adversaries would not be able to undermine their nuclear deterrent capability** using nuclear, conventional or other non-conventional means"
- Achieved by mutually assured destruction (MAD)
- Depends on
 - 1) the possession of second-strike capability
 - 2) that the capability are credible, effective and survivable

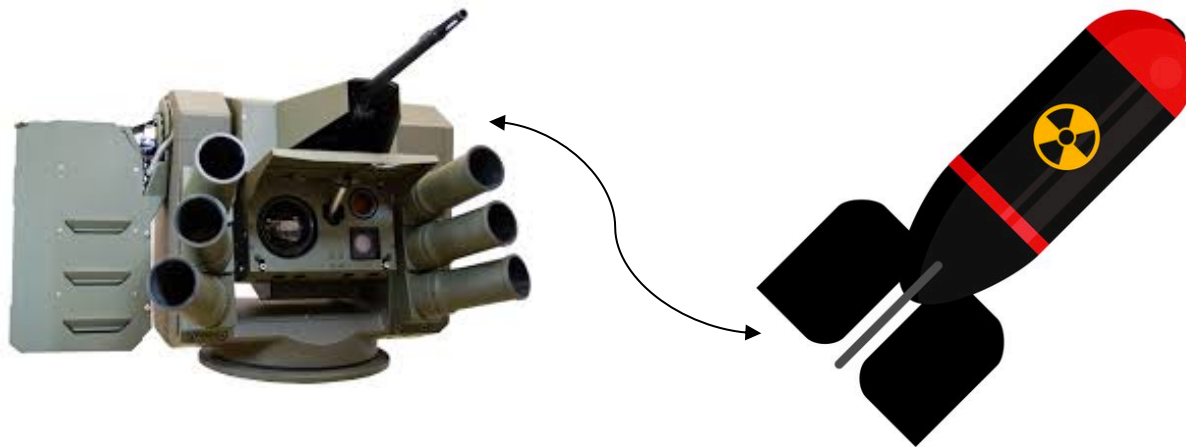


Application of autonomy in nuclear deterrence architecture



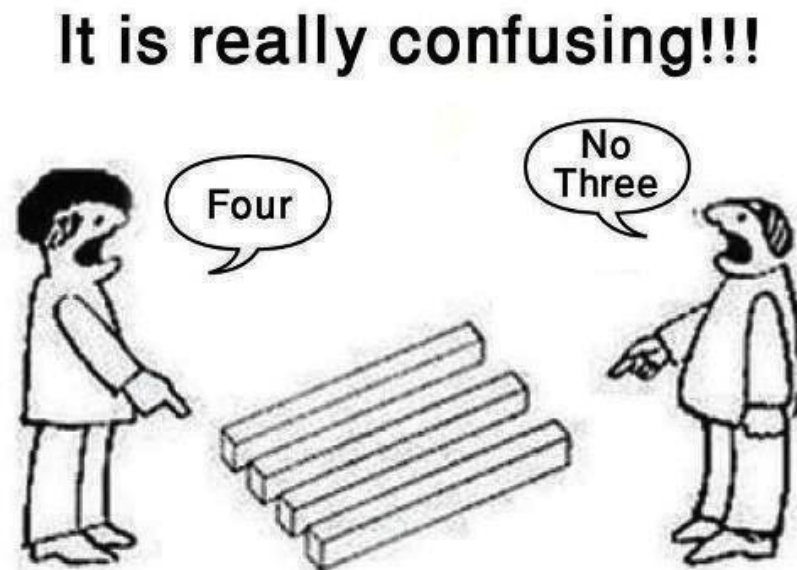
Entanglement

- Connection between conventional and nuclear arms
- Autonomy in weapon systems driver
- Easier to hold nuclear assets at risk
- Willingness to counter conventional attacks with nuclear



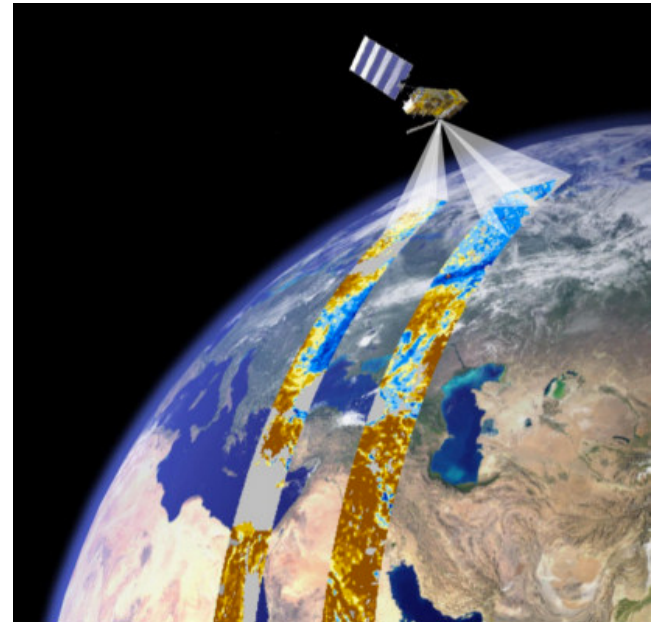
Destabilising effects on strategic stability

- Reinforce asymmetry between states
- Incentives to respond with destabilising measures
 - arms race
 - modernize nuclear arsenals
 - renounce NFU policy
 - increase alert status
 - automate launch
- **Issue of perception**
- Speed of warfare
- Lower threshold of war



Escalation

- Accidental escalation
 - less time for decision-making
 - brittleness in systems
 - operation unknown
- Inadvertent escalation
 - issue of perception
- Deliberate escalation
 - AI generated information



(Stabilising effects on strategic stability)

- Mutual vulnerability
- Better prepared to deal with crisis



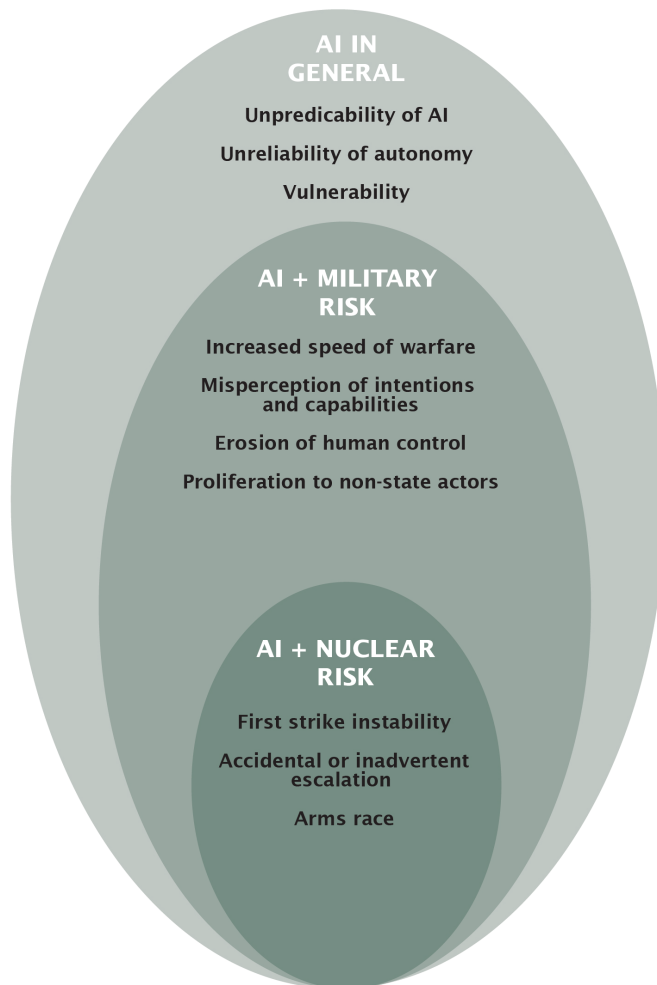
How serious is the risk?

- Autonomy in weapon systems alone is not enough to trigger escalation
- Other key factors



How serious is the risk?

RISK PICTURE



RISK MITIGATION MEASURES

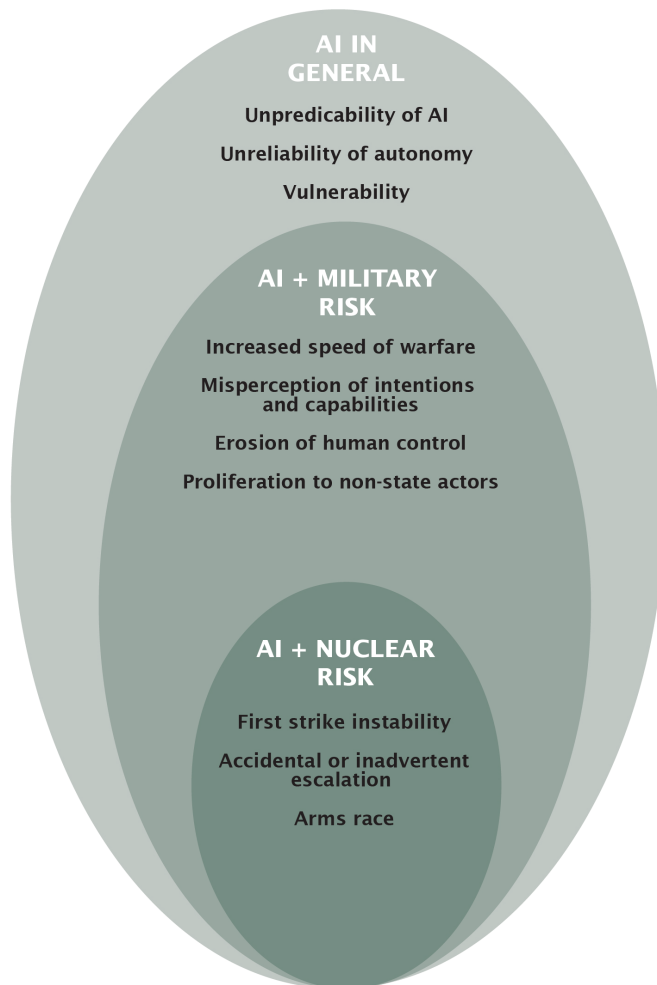
WHAT	HOW	WHO
Information sharing and cooperation on testing and verification of AI systems		
National policy on response development and use of AI technology		
National certification systems for AI safety		
Robust testing and evaluation of AI in safety critical systems		
Information sharing and cooperation through expert and military to military contact		
Commitment on responsible use of AI technology (e.g. on human control)		
National policy and strategy on development and use of military AI		
Transparency measures on existing AI-related military R&D		
Identification of military point of contact		
Robust testing and evaluation of AI enabled military systems		
Robust training of human operators		
Information sharing on developments and use of AI in nuclear weapons systems		
Commitment to maintain human control under nuclear launch decision		
No first use policy		
Lower alert status of nuclear weapons/demating		
Expert dialogue on risk posed by AI in nuclear sphere		
Keep AI-enabled early warning and ISR separate from C2		
Require human verified intelligence		
Do not make launch decisions based on single source of information		

KEYS

- Unilateral level
- Multilateral/bilateral/trilateral level
- Scientific community
- Military
- Private sector
- Government
- Government (nuclear armed state)

How serious is the risk?

RISK PICTURE



RISK MITIGATION MEASURES

WHAT

HOW WHO

Information sharing and cooperation on testing and verification of AI systems		
National policy on response development and use of AI technology		
National certification systems for AI safety		
Robust testing and evaluation of AI in safety critical systems		
Information sharing and cooperation through expert and military to military contact		
Commitment on responsible use of AI technology (e.g. on human control)		
National policy and strategy on development and use of military AI		
Transparency measures on existing AI-related military R&D		
Identification of military point of contact		
Robust testing and evaluation of AI enabled military systems		
Robust training of human operators		
Information sharing on developments and use of AI in nuclear weapons systems		
Commitment to maintain human control under nuclear launch decision		
No first use policy		
Lower alert status of nuclear weapons/demating		
Expert dialogue on risk posed by AI in nuclear sphere		
Keep AI-enabled early warning and ISR separate from C2		
Require human verified intelligence		
Do not make launch decisions based on single source of information		

KEYS

- Unilateral level
- Multilateral/bilateral/trilateral level
- Scientific community
- Military
- Private sector
- Government
- Government (nuclear armed state)

Key takeaways

- Autonomy in weapon systems a driver for entanglement
- Perception can be destabilising
- Key to agree on regulations + share information

4 STEPS TO DEAL WITH THE NEGATIVE IMPACT OF AI ON STRATEGIC STABILITY AND NUCLEAR RISK



Thank you!

Website: www.sipri.org

Email: moa.peldan@sipri.org