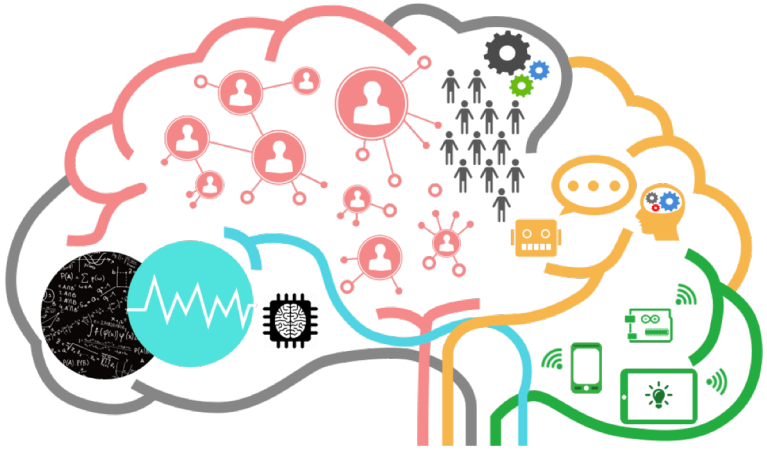
Software Autonomous Intelligent Agents

Internal Architecture and How They Fit Into the Internet of Battlespace Things Landscape



Alessandro Guarino

CEO, StAG Srl - NATO IST 152 Research Task Group Member

IET Cyber Security for ICSs – Cyber Defence Track - *London 8/2/2019*

Savoy Place – Mountbatten Room

The Speaker

20 Years in Information Security and Data Protection

StudioAG
Consulting & Engineering

Speaker / Author

CyCon International Conference on Cyber Conflict — 2013

ISSE 2016 — 2013-2017

E-PRIVACY XX — 2016 →

pirate security conference psc

IET Events The Institution of Engineering and Technology — 2017

Università di Roma Tor Vergata

NATO OTAN — 2018

12th INTERNATIONAL CONFERENCE CPDP DATA 2019 AND — 2019

Standards and Policy

UNINFO

CS·CG CEN/CENELEC/ETSI Cyber Security Coordination Group

ETSI

enisa THE EU CYBERSECURITY AGENCY

cen CENELEC

2011 →

Disclaimer

This presentation contains research results originating from the NATO IST-152 Research Task Group ("Intelligent Autonomous Agents for Cyber Defense and Resilience").

Views (and mistakes) expressed however are not those of the RTG or NATO, but are only ascribable to the expert member who is delivering this talk.

All results from the RTG are classified "Approved for public release; distribution unlimited"

Cyber Conflict Today

Armed conflicts commonly conducted in the cyber (fifth) domain.

- Russia-Georgia war of 2008
- Israeli attacks on Syria

Some purely cyber examples

- Olympic Games/Stuxnet, Black Energy

International law evolving to meet the challenges but State behaviour is still the main driver.

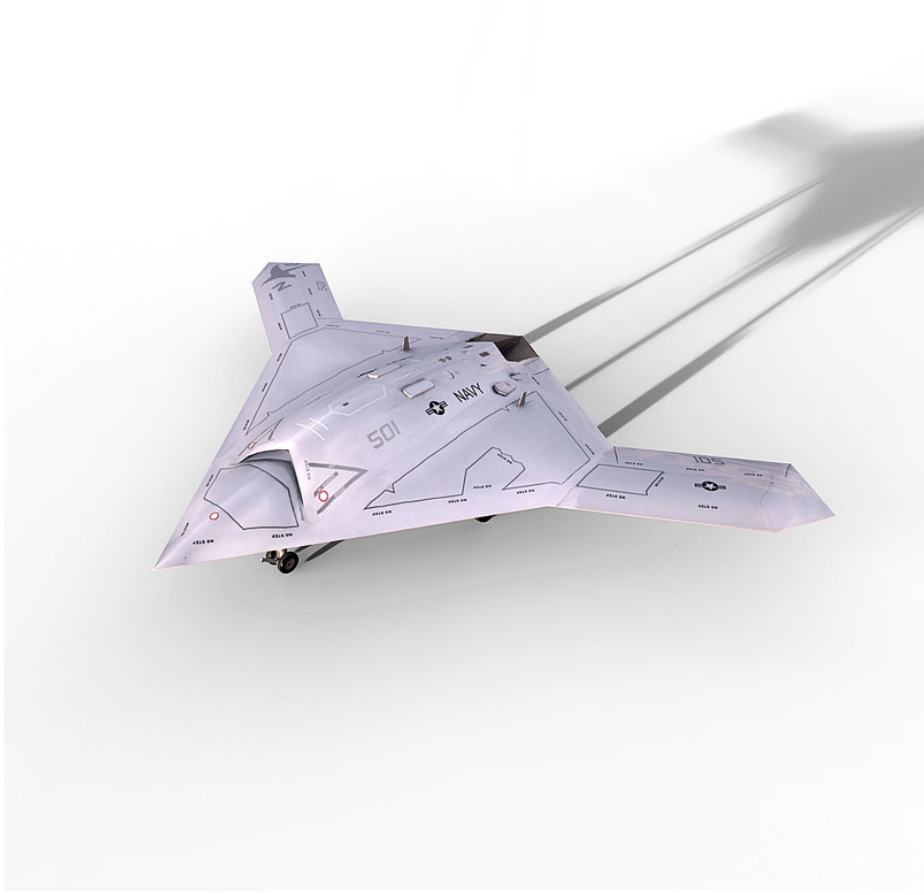
Autonomous Intelligent Agents

What makes an agent? And an AIA?

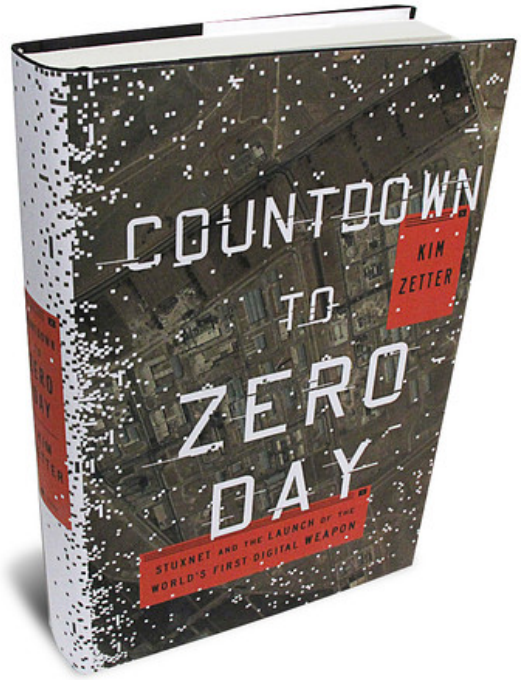
1 - An agent is strictly **associated with its environment**: an autonomous agent outside the environment it was designed for can be useless, or not an agent at all.

2- An agent **interacts** with the environment, via appropriate sensors providing input from it and appropriate actuators allowing the agent to act and influence that environment.

3 - An autonomous agent **acts towards a goal**, it has an 'agenda'. In particular, an Autonomous Agent developed for warfare operations is assigned a target.



Autonomous Intelligent Agents



4 - The activities of a truly autonomous agent are sustained ‘over time’, so it must have a **continuity of action.**

5 - An autonomous agent should possess an adequate **internal model of its environment**, including its goal together with some kind of **performance measure** or utility function that expresses its preferences.

6 - An agent must possess the **capability to learn** new knowledge and the possibility to modify over time its model of the world and possibly also its goals and preferences.

AIAs – Two-dimensional Taxonomy

Role:

Information gathering or military operations (or, in other words, defence/resilience or offence)

The main difference lies in the nature of offensive operations: usually intelligence-gathering does not cause damage to the targets and in fact tries to avoid detection in most instances.

Architecture

Monolithic or decentralised. **Monolithic agents** are constituted by a single piece of software or else by strictly coordinated elements without independent means of operation, for instance an executable file and libraries.
Decentralised intelligent agents are systems where intelligence is distributed among many simpler components, all similar or very similar, acting in concert.

NATO IST-152 Research Task Group

Timeline 2016-2019

Vision: stealthy software AIAs monitoring and defending the network(s) installed on a weapon system (land vehicle, UAV…)

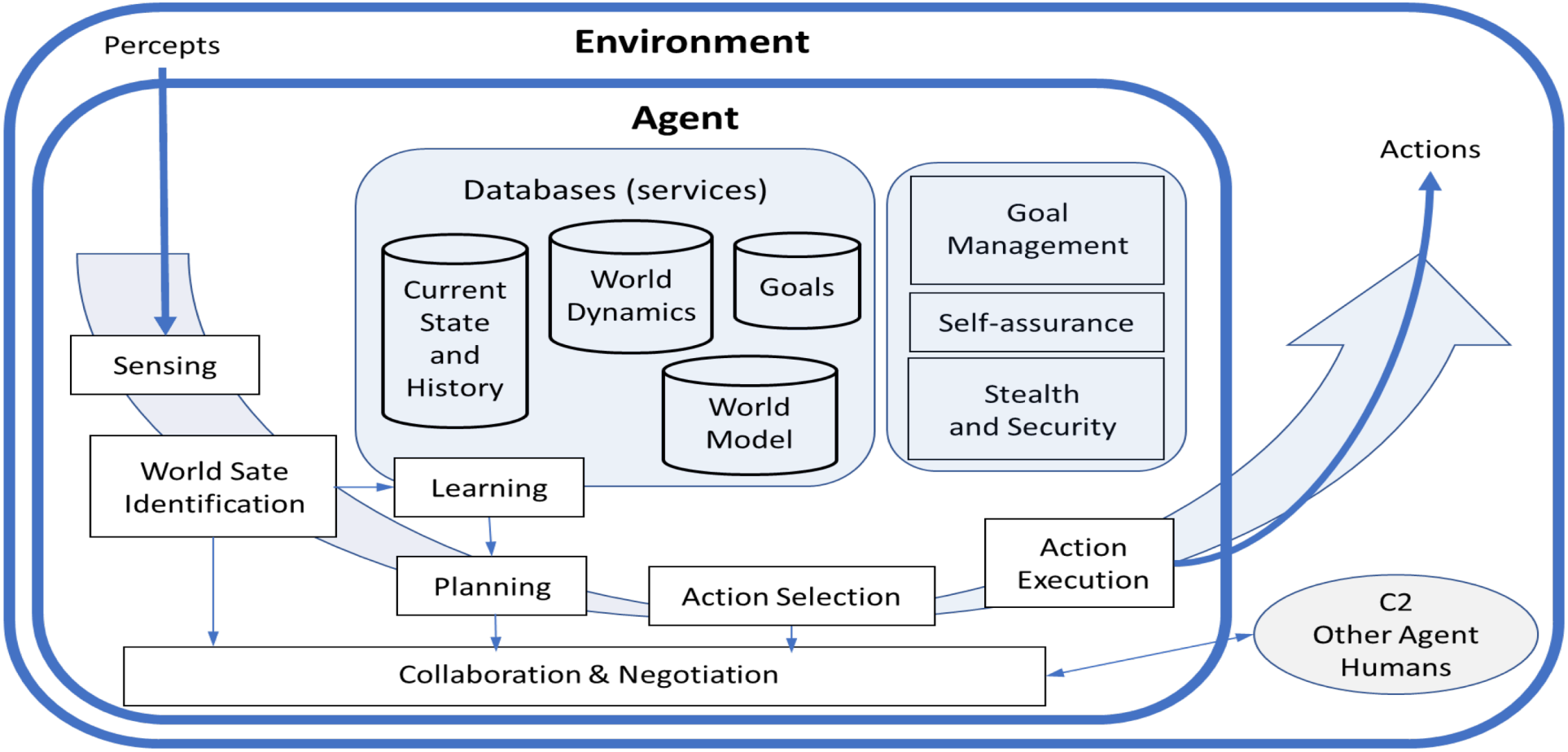
Agents are capable of learning to counter evolving threats

Agents are capable of operating autonomously without communication if needed

Reference Architecture aims to be a base for future NATO procurement.

NATO IST-152 Research Task Group

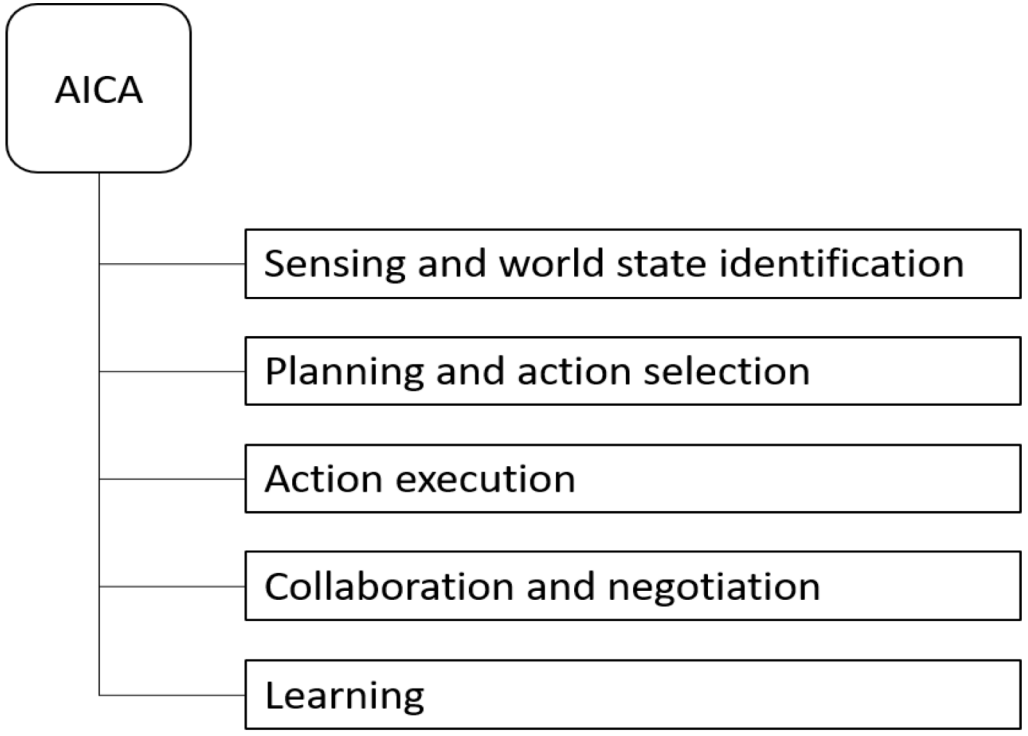
The AIA Reference Architecture



The AIA Reference Architecture

Assumptions:

- Single platform (e.g. a vehicle)
- One or more networks to defend
- Comms could be negated
- Centralised cyberdefence impossible
- Limited human cyber expertise available
- Limited time available (combat situation)
- Provision for remote controller



The AIA Reference Architecture - Sensing

Sensing

Collects and interprets data about:

current state of the world

the agent itself

adversarial events,

anomalies in data

changes of all the above since the last observation

Subsystems:

Self (collects data about the agent integrity)
System (Collects data about the system the agent is tasked to defend)
Environment (Monitors data coming from the outside the agent)

Data is normalized, collated/fused then passed to World State Identification

The AIA Reference Architecture - WSI

World State Identification

Assesses the State of the World in respect to the world model

Performs environment identification (Am I running in a sandbox/honeypot?)

IFF (Identification, Friend or Foe) – Tags processes and files

Anomaly identification – id anomalies in data from Sensing

The AIA Reference Architecture – Planning and Execution

Planning creates a set of possible plans of action in order to guide the AIA from the current state of the world to a desired state of the world.

These sets are sent to the action selection module for execution.

The AIA Reference Architecture – Planning Redux

Other possibilities for planning are considered:

(other then tree exploring / pruning)

Game theoretic approach

- Models (cyber)security as a game between defenders and attackers

Reinforcement learning

- The agent learns (before deploying) to adapt to the possible scenarios and to minimize risks to the defended systems

The AIA Reference Architecture - Collaboration

An agent can **individually perform** one or multiple tasks but **also choose to cooperate** with other agents to perform coordinated actions.

In the collaborative agent system of AICARA, each individual agent should be able to solve the problem autonomously and only start to collaborate with other agents to improve the common plan of actions or extend the individual capacities of plan

The AIA Reference Architecture - Collaboration

Collaboration:

Allows an individual agent A to interact with other agents to make plans of actions more effective or solve a task that is beyond the agent's capabilities.

Negotiation:

The goal of the negotiation is to reach an agreement within a set of agents regarding a goal or a plan execution.

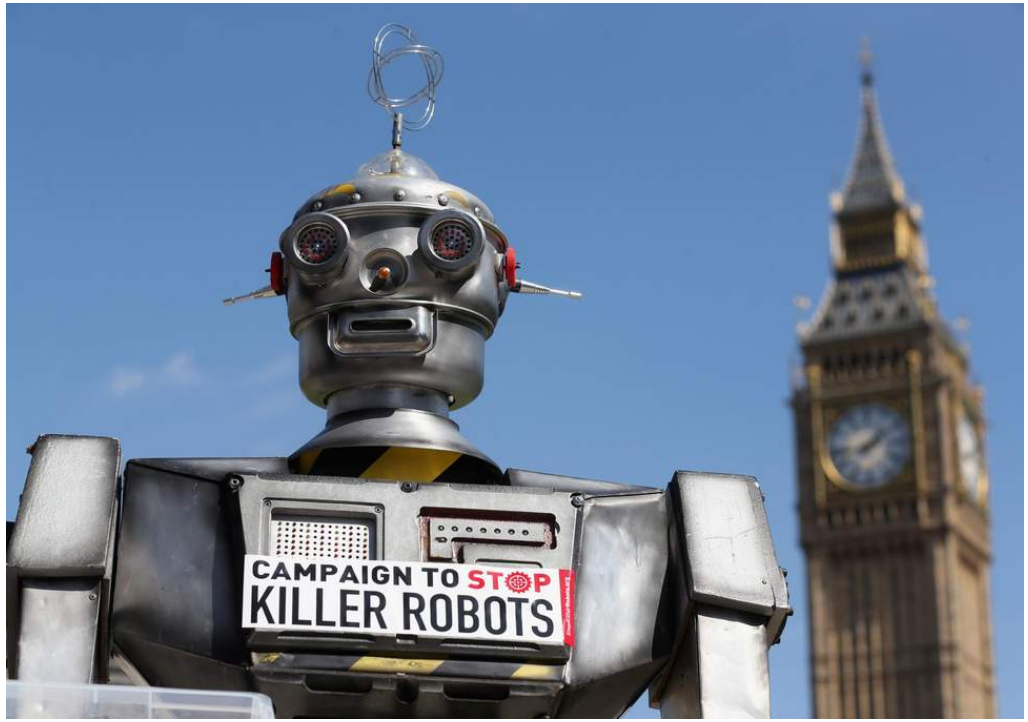
The AIA Reference Architecture - Learning

Necessary to adapt to environmental changes (sixth trait)

Wide range – in AICARA the focus is on experience gained about details of the effectiveness (reward) for specified actions

Horizon: next decade

Ethical / legal issues for software and hardware AIAs – aka…



Internet of Battlefield Things (R)Evolution

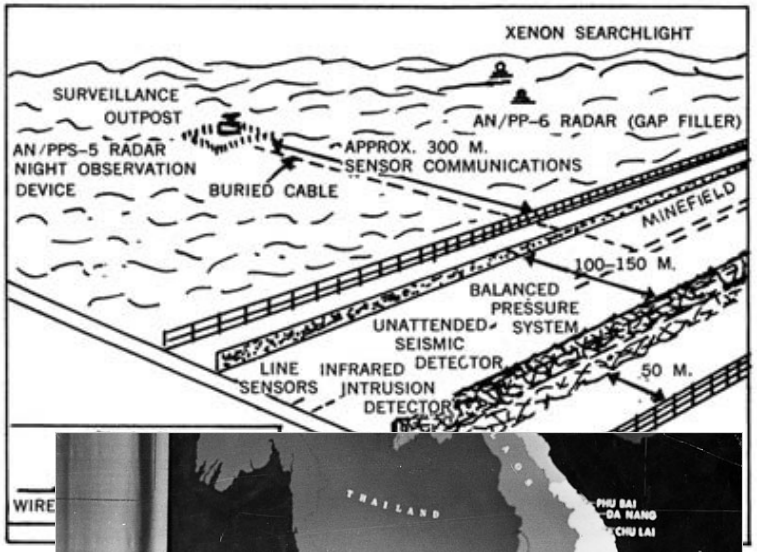
The term IoBT was invented by Alexander Kott (Chief Scientist of ARL and chair of the IST-152 RTG).

AIAs and “less intelligent” systems talk to each other **“to create a kind of society of things”** on the battlefield.

Sensors, weapons systems, vehicles, UAVs…

Connection of all devices is happening fast (despite the risks).

Actually the concept is a return back to the origins of IoT, to the “McNamara line”, a concept that was proposed in the late sixties during the Vietnam war.



Internet of Battlefield Things (R)Evolution

How AIAs fit?

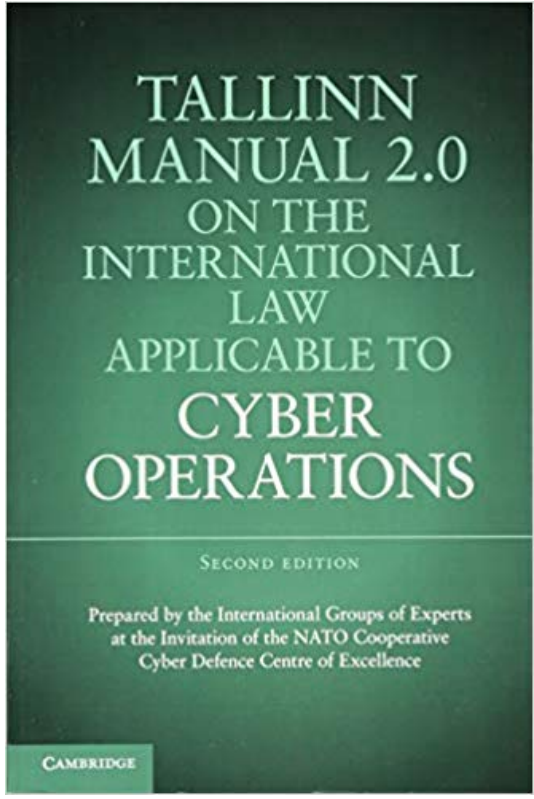
– In a world of machines, AIA could theoretically become king…

– Agents can operate now beyond the confines of the AICARA, in support of the soldier or autonomously.

- The human in the loop could only ask for a high level objective without needed to know the details of how an AIA (or a swarm of AIAs).

– Future battles could be fully fought by AIAs?

– What abut the laws of war? (Formally the International Humanitarian Law).



Thank You for Your Time!
What are your questions?

Contacts:

a.guarino@studioag.eu

 **@alexsib17**

Slides online on:
www.studioag.pro

StAG – Information Governance
www.stagcyber.eu

References

S. Franklin and A. Graesser. 'Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents', in Proceedings of the Third International Workshop on Agent Theories, Springer-Verlag, 1996.

S. Russell and P. Norvig. Artificial Intelligence: a modern approach 3rd edition, Pearson, 2010.

M. Schmitt (ed.), Tallinn Manual on the International Law Applicable to Cyber Warfare - Cambridge University Press 2013.

M. Schmitt (ed.), Tallinn Manual 2.0 on the International Law Applicable to Cyber Warfare - Cambridge University Press 2017.

K. Zetter, Countdown to Zero Day, Crown 2014

A. Guarino, Autonomous Intelligent Agents in Cyber Offence - in "5th International Conference on Cyber Conflict – Proceedings", K. Podins – J. Stinissen – M. Maybaum (eds.), IEEE 2013

vv.aa. Proceedings of the NATO IST-152 Workshop on Intelligent Autonomous Agents for Cyber Defence and Resilience" – Prague, Czech Republic, October 18-20, 2017, CEUR-WS Vol. 2057

A. Guarino , Hello World Autonomous Agent - in "Toward Intelligent Autonomous Agents for Cyber Defense: Report of the 2017 Workshop by the North Atlantic Treaty Organization (NATO) Research Group IST-152-RTG" – US Army Research Laboratory ARL-SR-0395 – April 2018