CONVERGENCE 2023

30 November 2023
- 1 December 2023

Brussels

Trust at the Nexus of Cybersecurity & AI

Dr. Theodora Tsikrika

Information Technologies Institute – ITI

Centre for Research and Technology Hellas – CERTH

Information Technologies Institute

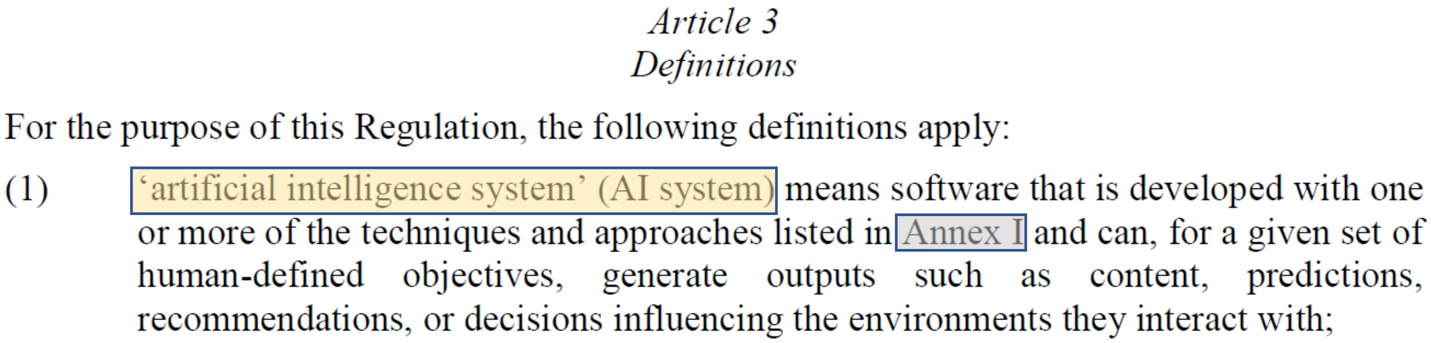
CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS

Supported by: IRIS

Draft AI Act

Proposal for a Regulation on Artificial Intelligence

Proposed in April 2021



Article 3
Definitions

For the purpose of this Regulation, the following definitions apply:

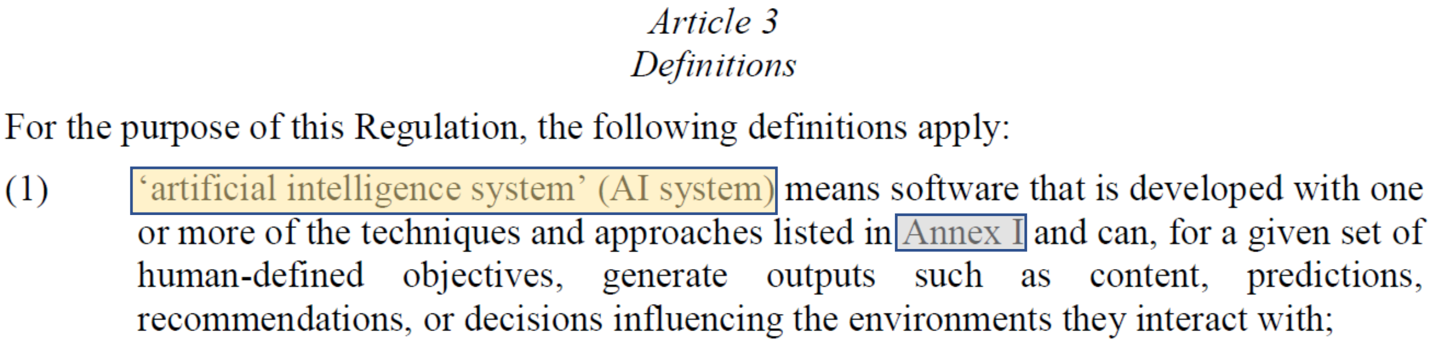
(1) ‘artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors’ view and European Commission is not responsible for any use that may be made of the information it contains.

Draft AI Act

Proposal for a Regulation on Artificial Intelligence

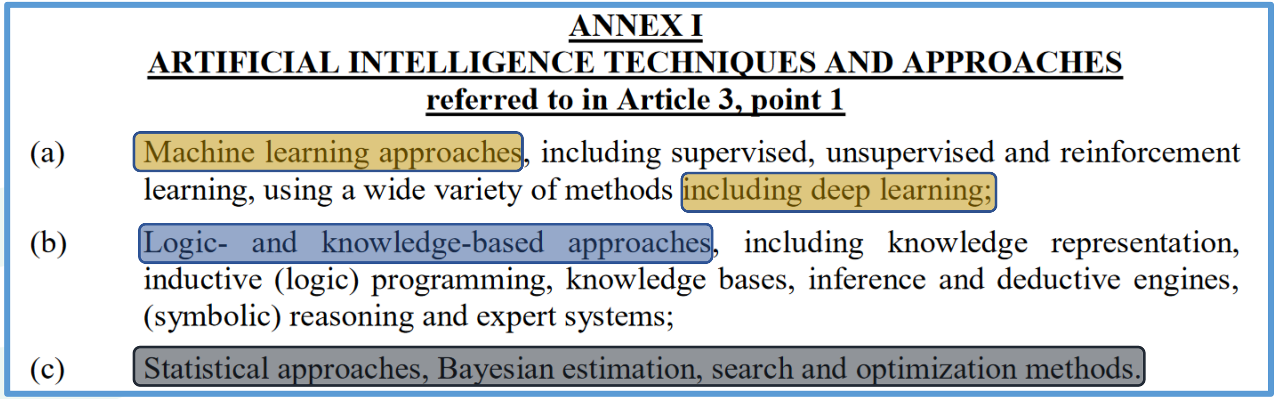
Proposed in April 2021



Article 3
Definitions

For the purpose of this Regulation, the following definitions apply:

(1) ‘artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;



ANNEX I
ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES
referred to in Article 3, point 1

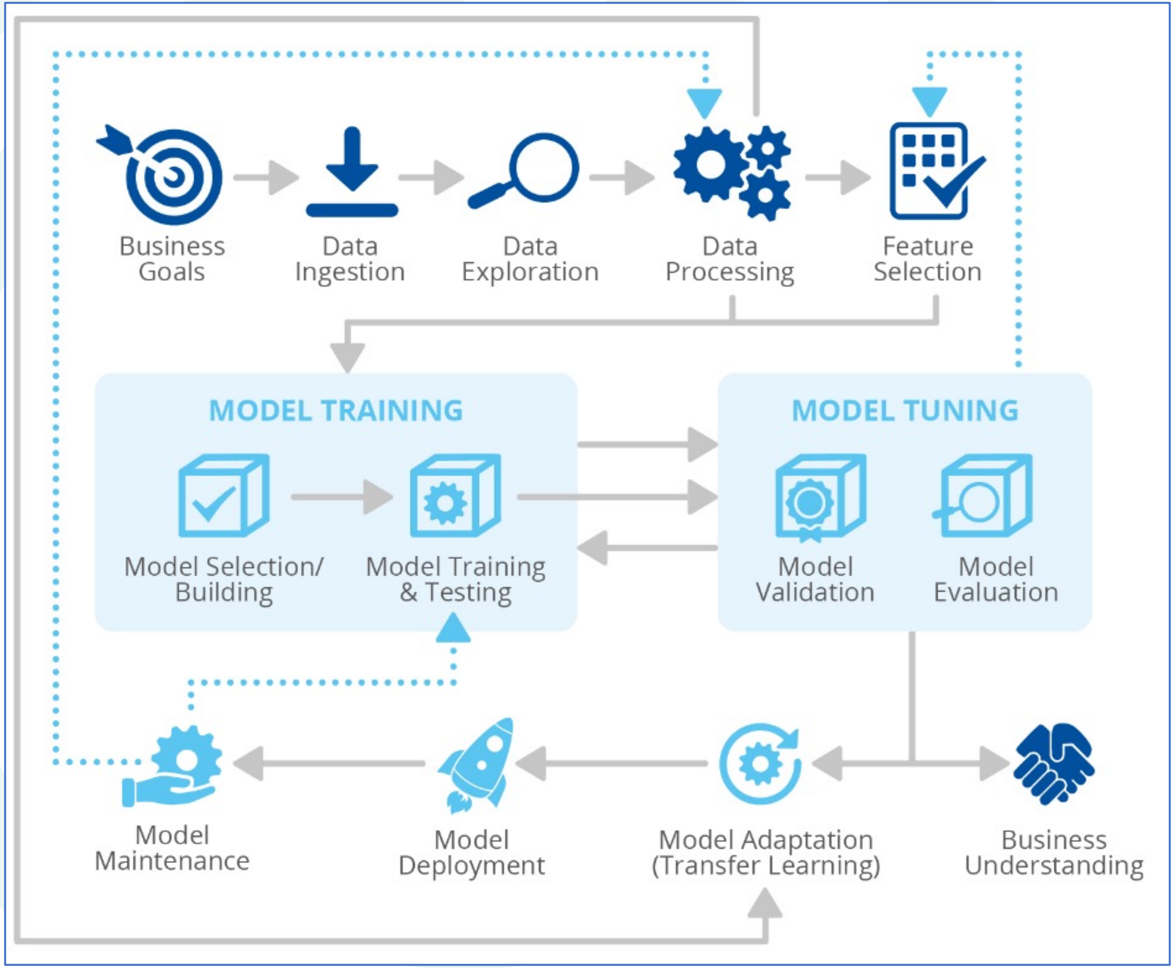
(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

(c) Statistical approaches, Bayesian estimation, search and optimization methods.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors’ view and European Commission is not responsible for any use that may be made of the information it contains.

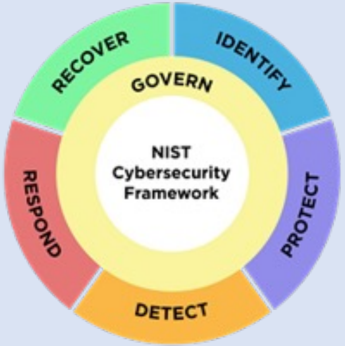
AI Lifecycle



Source: ENISA (December 2020) AI CYBERSECURITY CHALLENGES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

AI for Cybersecurity



AI-powered cybersecurity solutions

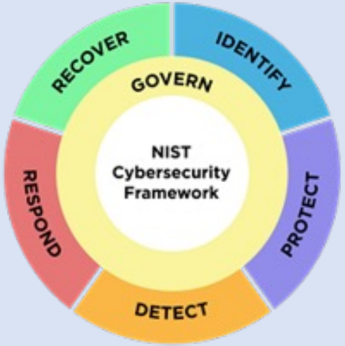
e.g.

- vulnerabilities & cyber risk assessment
- penetration testing
- anomaly detection / behavioural analysis
- intrusion/malware/phishing & identification
- spam filtering
- reporting
- forecasting
- etc.

* https://www.nist.gov/cyberframework

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors’ view and European Commission is not responsible for any use that may be made of the information it contains.

AI for Cybersecurity



AI-powered cybersecurity solutions

e.g.

- vulnerabilities & cyber risk assessment
- penetration testing
- anomaly detection / behavioural analysis
- intrusion/malware/phishing & identification
- spam filtering
- reporting
- forecasting
- etc.

* https://www.nist.gov/cyberframework

AI vs Cybersecurity

AI-powered attacks

AI-powered cyberattacks

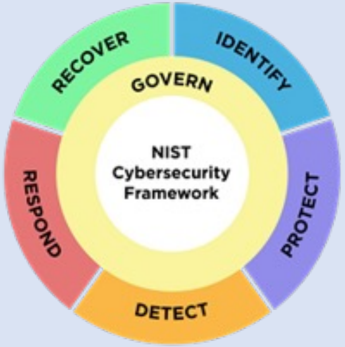
- improved efficiency, effectiveness, scale, adaptability, persistence, cost, etc.

AI-based deepfakes generation

Advanced spear-phishing emails

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

AI for Cybersecurity



AI-powered cybersecurity solutions

e.g.

- vulnerabilities & cyber risk assessment
- penetration testing
- anomaly detection / behavioural analysis
- intrusion/malware/phishing & identification
- spam filtering
- reporting
- forecasting
- etc.

* https://www.nist.gov/cyberframework

AI vs Cybersecurity

AI-powered attacks

AI-powered cyberattacks

- improved efficiency, effectiveness, scale, adaptability, persistence, cost, etc.

AI-based deepfakes generation

Advanced spear-phishing emails

Cybersecurity for AI

Defending AI-powered solutions against

Adversarial attacks

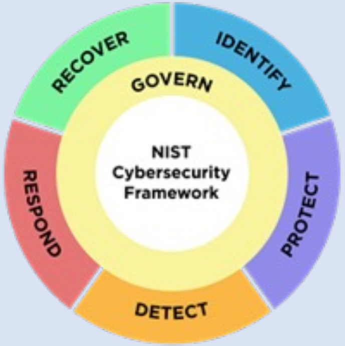
Training data poisoning attacks

Exploiting vulnerabilities in widely used (open source) libraries

Reverse engineering

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors’ view and European Commission is not responsible for any use that may be made of the information it contains.

AI for Cybersecurity



AI-powered cybersecurity solutions

e.g.

- vulnerabilities & cyber risk assessment
- penetration testing
- anomaly detection / behavioural analysis
- intrusion/malware/phishing & identification
- spam filtering
- reporting
- forecasting
- etc.

* https://www.nist.gov/cyberframework

AI vs Cybersecurity

AI-powered attacks

AI-powered cyberattacks

- improved efficiency, effectiveness, scale, adaptability, persistence, cost, etc.

AI-based deepfakes generation

Advanced spear-phishing emails

Cybersecurity for AI

Defending AI-powered solutions against

Adversarial attacks

Training data poisoning attacks

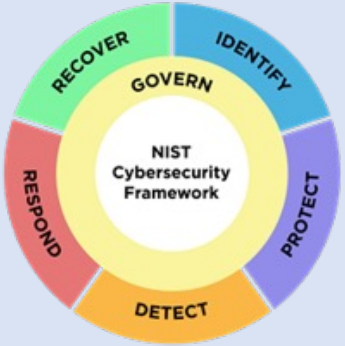
Exploiting vulnerabilities in widely used (open source) libraries

Reverse engineering

Technology

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

AI for Cybersecurity



AI-powered cybersecurity solutions

e.g.

- vulnerabilities & cyber risk assessment
- penetration testing
- anomaly detection / behavioural analysis
- intrusion/malware/phishing & identification
- spam filtering
- reporting
- forecasting
- etc.

* https://www.nist.gov/cyberframework

AI vs Cybersecurity

AI-powered attacks

AI-powered cyberattacks

- improved efficiency, effectiveness, scale, adaptability, persistence, cost, etc.

AI-based deepfakes generation

Advanced spear-phishing emails

Cybersecurity for AI

Defending AI-powered solutions against

Adversarial attacks

Training data poisoning attacks

Exploiting vulnerabilities in widely used (open source) libraries

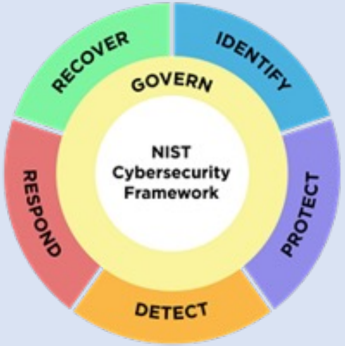
Reverse engineering

Technology

Training

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors’ view and European Commission is not responsible for any use that may be made of the information it contains.

AI for Cybersecurity



AI-powered cybersecurity solutions

e.g.

- vulnerabilities & cyber risk assessment
- penetration testing
- anomaly detection / behavioural analysis
- intrusion/malware/phishing & identification
- spam filtering
- reporting
- forecasting
- etc.

* https://www.nist.gov/cyberframework

AI vs Cybersecurity

AI-powered attacks

AI-powered cyberattacks

- improved efficiency, effectiveness, scale, adaptability, persistence, cost, etc.

AI-based deepfakes generation

Advanced spear-phishing emails

Cybersecurity for AI

Defending AI-powered solutions against

Adversarial attacks

Training data poisoning attacks

Exploiting vulnerabilities in widely used (open source) libraries

Reverse engineering

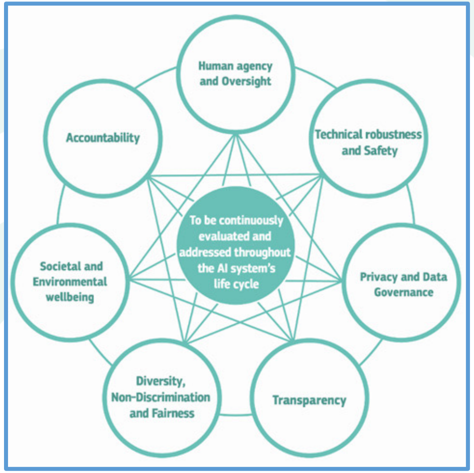
Technology

Training

Certification

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors’ view and European Commission is not responsible for any use that may be made of the information it contains.

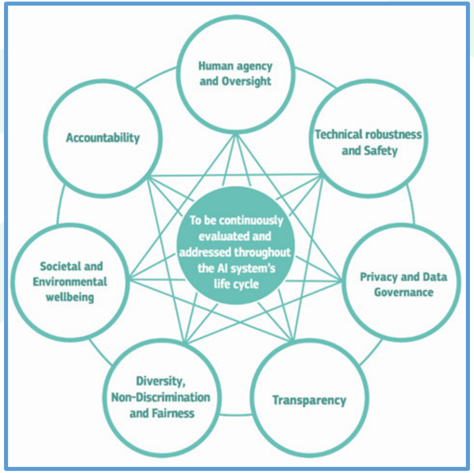
Trustworthy AI


Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

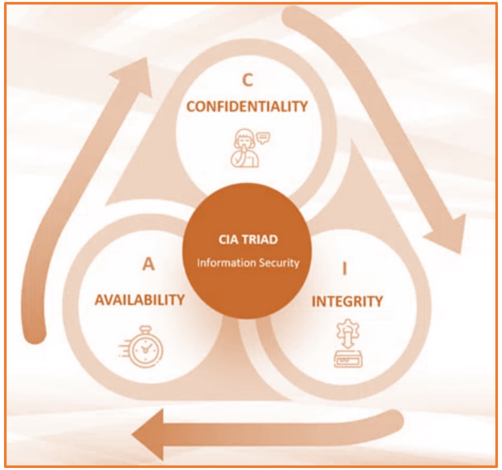
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

Trustworthy AI


Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

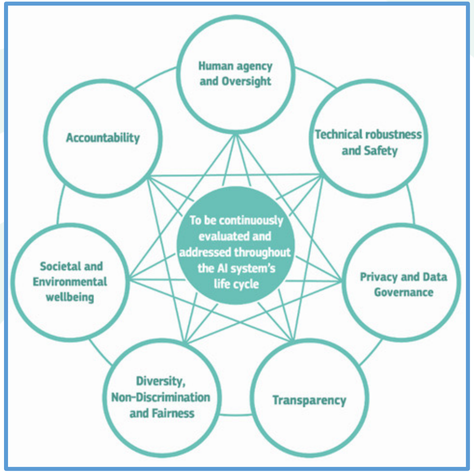
Trustworthy Cybersecurity



https://www.i-scoop.eu/cybersecurity/cia-confidentiality-integrity-availability-security/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

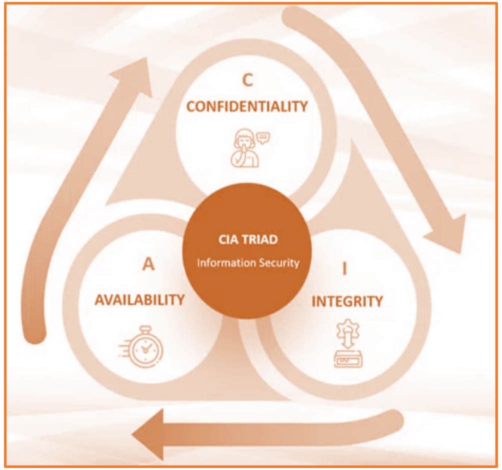
Trustworthy AI



Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Trustworthy Cybersecurity



https://www.i-scoop.eu/cybersecurity/cia-confidentiality-integrity-availability-security/

Cybersecurity for AI Trustworthiness & Cybersecurity Trustworthiness for AI

Trustworthy Cybersecurity ➔ correct implementation of Trustworthy AI ➔ Trustworthy Cybersecurity

- Need for common understanding and what the trustworthiness characteristics are
- Need for coherence between the (draft) AI Act & the (draft) Cyber Resilience Act

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.

Dr. Theodora Tsikrika

https://m4d.iti.gr/
theodora.tsikrika@iti.gr

Supported by: IRIS

https://www.iris-h2020.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101021727.
This material reflects only the authors' view and European Commission is not responsible for any use that may be made of the information it contains.