



AI for Security
Topic Guide

Matilda Rhode

contact@cybok.org
www.cybok.org



© Crown Copyright, The National Cyber Security Centre 2023. This information is licensed under the Open Government Licence v3.0. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/opengovernment-licence/>.

When you use this information under the Open Government Licence, you should include the following attribution: AI for Security Topic Guide v1.0.0 © Crown Copyright, The National Cyber Security Centre 2023, licensed under the Open Government Licence <http://www.nationalarchives.gov.uk/doc/opengovernment-licence/>.

The CyBOK project would like to understand how CyBOK is being used and its uptake. The project would like organisations using, or intending to use, CyBOK for the purposes of education, training, course development, professional development etc. to contact it at contact@cybok.org to let the project know how they are using CyBOK.

[1] Introduction

1. Introduction
2. AI and ML
3. Why use AI
 1. Potential benefits
 2. Challenges
4. Applications
5. Common Pitfalls
6. Evaluation
7. AI Ecosystem
8. Key considerations for **implementing** AI (for security)
9. Key considerations for **procuring** AI (for security)
10. Conclusion

[2] Artificial Intelligence and Machine Learning



“Powered by AI”

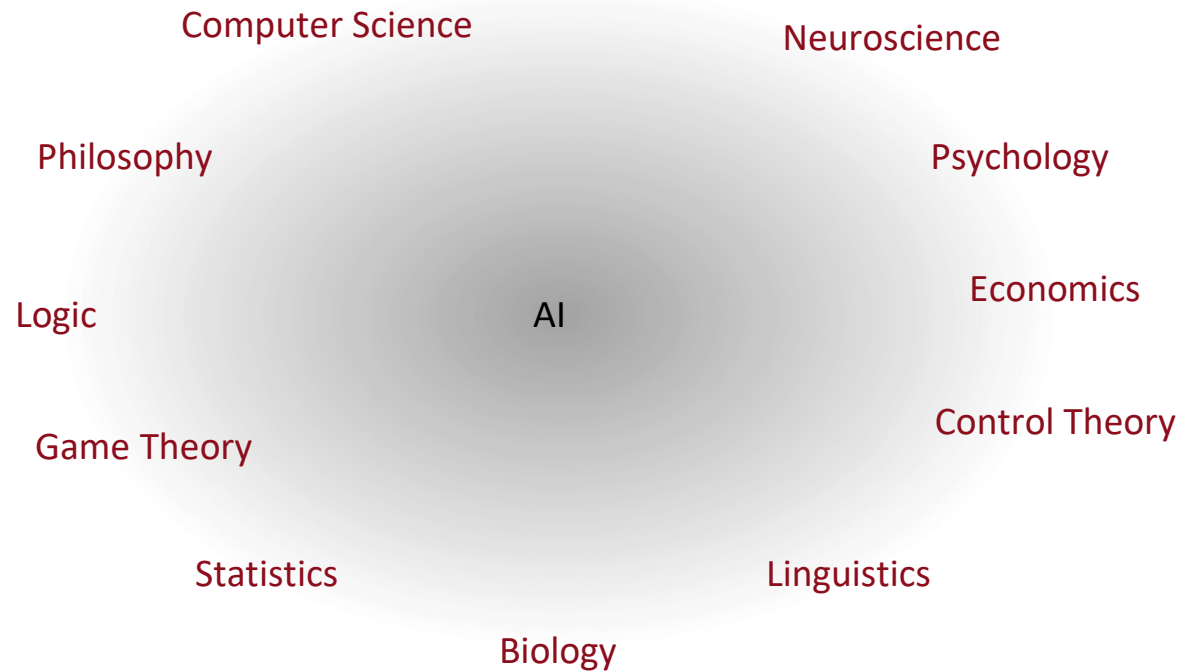
“State of the art deep-learning solution”

“Machine learning infused security software”

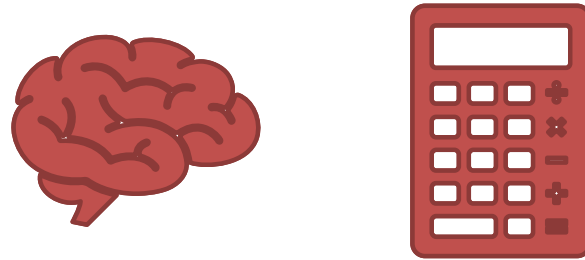
“Intelligent cyber solution”

“AI-backed security”

[2] Artificial Intelligence and Machine Learning

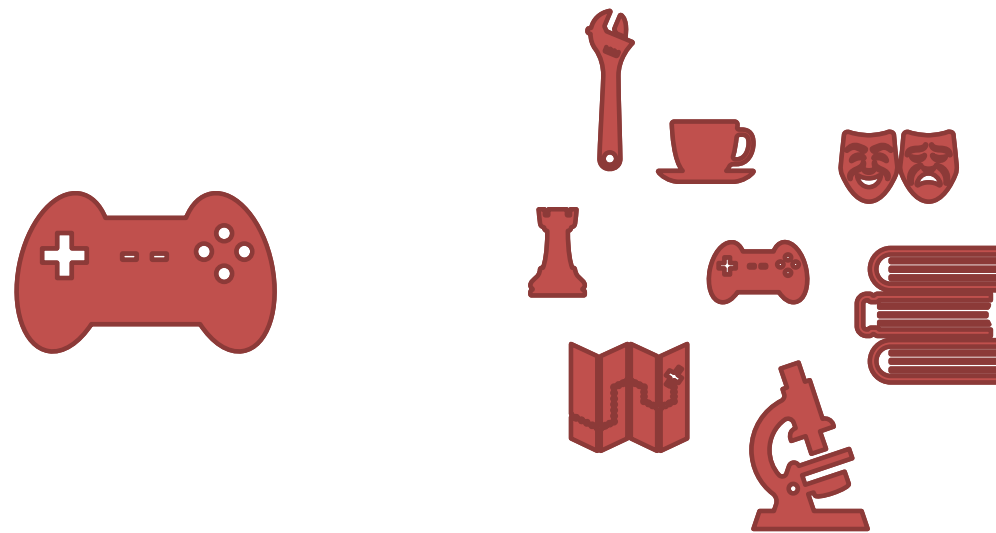


[2] Artificial Intelligence and Machine Learning



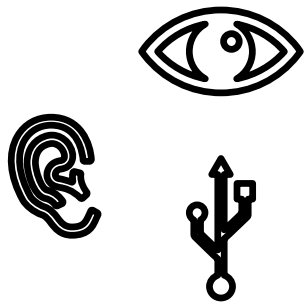
human vs. rational
intelligence

[2] Artificial Intelligence and Machine Learning

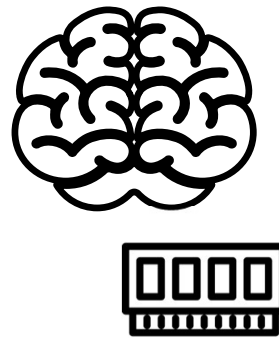


narrow vs. general
intelligence

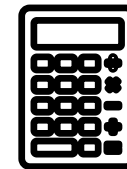
[2] Artificial Intelligence and Machine Learning



Capturing information



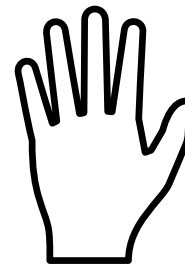
Storing information



Processing information including planning, learning, decision making

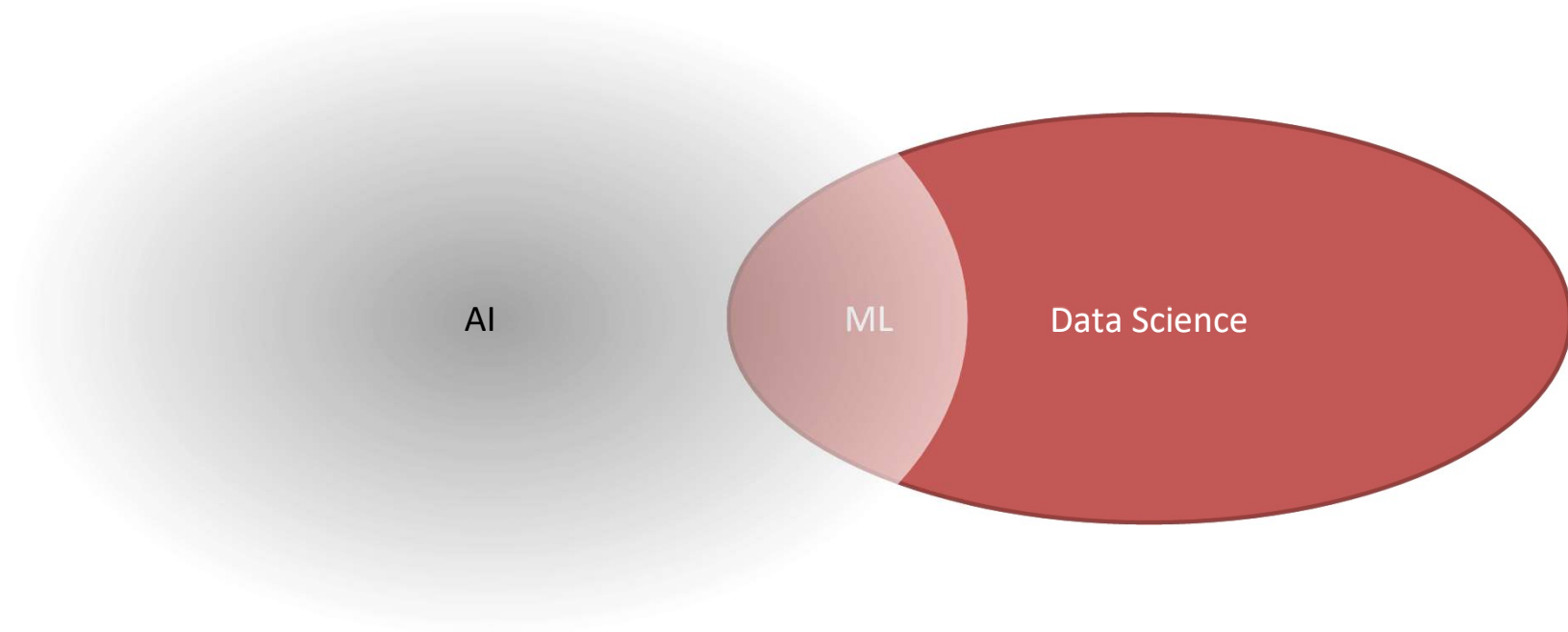


Communication

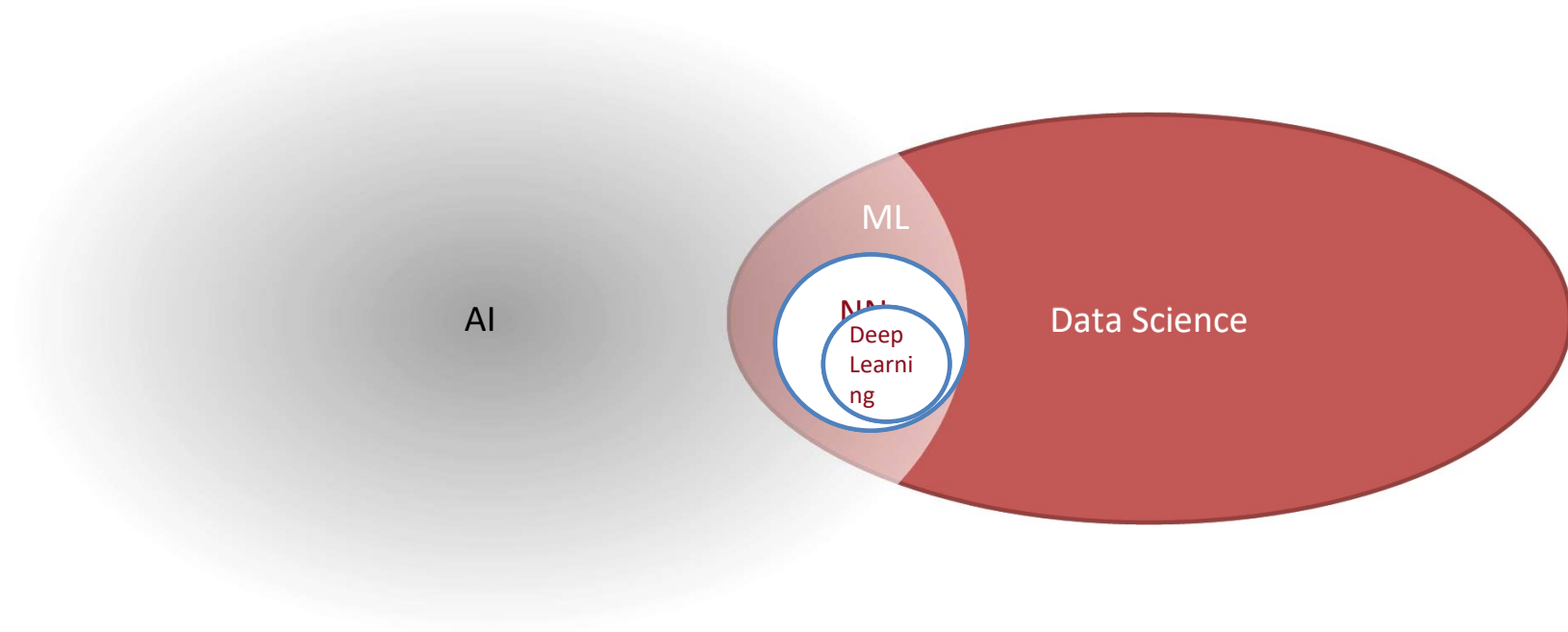


Action



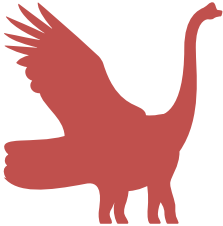

[2] Artificial Intelligence and Machine Learning



[2] Artificial Intelligence and Machine Learning



[2] Artificial Intelligence and Machine Learning

- Supervised learning 
- Unsupervised learning 
- Semi-supervised learning 
- Reinforcement learning 

[3] Why use AI?

Potential benefits

- (Partial) automation of tasks
- Automatic retraining
- Big data analysis
- Harnesses (latent) information

[3] Why use AI?

Potential benefits

- (Partial) automation of tasks
- Automatic retraining
- Big data analysis
- Harnesses (latent) information

Challenges

- Reliance on large datasets
- Cost of labelling data
- Lack of benchmark datasets
- Data privacy
- Infrastructure cost
- Opacity, robustness, and security of ML
- Incoming regulation

[4] Applications

CyBOK



The Five Functions, NIST Cyber Security Framework

<https://www.nist.gov/cyberframework/online-learning/five-functions>

[4.1] Identify

- Vulnerability assessment



[4.1] Identify

- Vulnerability assessment
- Vulnerability criticality



[4.1] Identify

- Vulnerability assessment



- Vulnerability criticality



- Automated red-teaming



[4.1] Identify

- Vulnerability assessment



- Vulnerability criticality



- Automated red-teaming

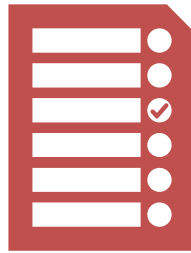


- Governance and compliance



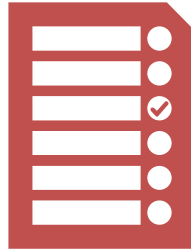
[4.2] Protect

- Access control



[4.2] Protect

- Access control



- Authentication



[4.3] Detect

- Intrusion detection systems
- Malware detection and analysis
- Phishing detection
- User Entity Behaviour Analytics
- Data aggregation and System Information and Event Management Systems

[4.4] Respond

- Attribution and attacker profiling



[4.4] Respond

- Attribution and attacker profiling



- Automated response



[4.4] Respond

- Attribution and attacker profiling



- Automated response



- Forensic investigation



[4.5] Recover

- Possible data-driven automated restoration of systems

[5] Common Pitfalls

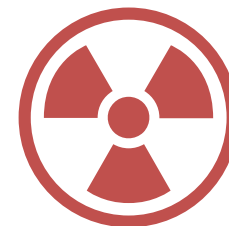
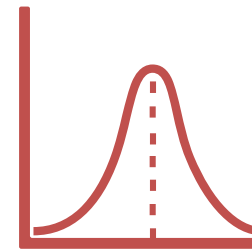
- sampling bias
- data snooping
- lab-only evaluation [1]

[1] Daniel Arp, Erwin Quiring, Feargus Pendlebury, Alexander Warnecke, Fabio Pierazzi, Christian Wressnegger, Lorenzo Cavallaro, and Konrad Rieck. Dos and don'ts of machine learning in computer security

[5] Common Pitfalls

- sampling bias
- data snooping
- lab-only evaluation [1]

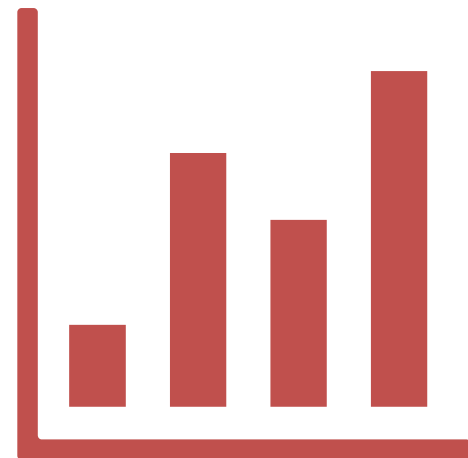
- baselining
- failing to create a failure mode



[1] Daniel Arp, Erwin Qiring, Feargus Pendlebury, Alexander Warnecke, Fabio Pierazzi, Christian Wressnegger, Lorenzo Cavallaro, and Konrad Rieck. Dos and don'ts of machine learning in computer security

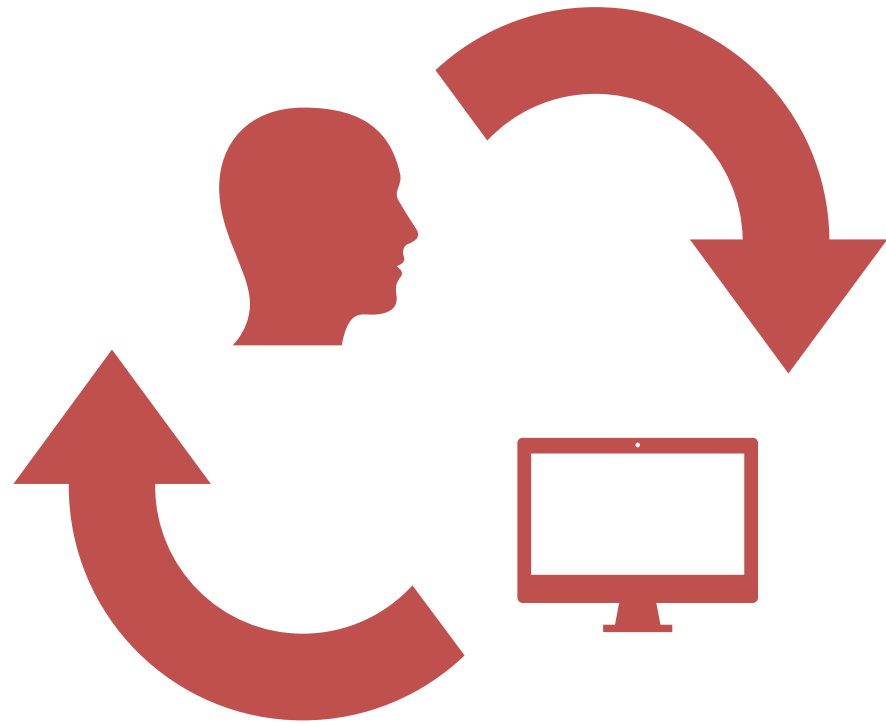
[6] Evaluation

- Performance metrics
- Analysis of datasets
- Testing with multiple datasets
- Lifecycle maintenance



[7] AI Ecosystem

- Humans in the Loop
 - Labelling and label dynamics



[8] Implementing AI

- Model design // Feature Engineering, Algorithm Selection

[8] Implementing AI

- Model design // Feature Engineering, Algorithm Selection
- Certification and Compliance // AI Act (EU), NIST Framework (USA) and others

[8] Implementing AI

- Model design // Feature Engineering, Algorithm Selection
- Certification and Compliance // AI Act (EU), NIST Framework (USA) and others
- Privacy // federated learning, secure multiparty computation, homomorphic encryption

[8] Implementing AI

- Model design // Feature Engineering, Algorithm Selection
- Certification and Compliance // AI Act (EU), NIST Framework (USA) and others
- Privacy // federated learning, secure multiparty computation, homomorphic encryption
- Robustness and Concept Drift // evaluation, drift-detection, formal methods

[8] Implementing AI

- Model design // Feature Engineering, Algorithm Selection
- Certification and Compliance // AI Act (EU), NIST Framework (USA) and others
- Privacy // federated learning, secure multiparty computation, homomorphic encryption
- Robustness and Concept Drift // evaluation, drift-detection, formal methods
- Bias mitigation and explainability // inherently explainable vs. post-facto explainability

[9] Procuring AI

- Performance metrics // independent benchmarking

[9] Procuring AI

- Performance metrics // independent benchmarking
- Baselineing // stress testing the anomaly threshold

[9] Procuring AI

- Performance metrics // independent benchmarking
- Baselineing // stress testing the anomaly threshold
- ML/AI infrastructure // resilience to cloud connectivity issues

[9] Procuring AI

- Performance metrics // independent benchmarking
- Baselineing // stress testing the anomaly threshold
- ML/AI infrastructure // resilience to cloud connectivity issues
- Data privacy assurance // how will your (valuable and sensitive) data be handled?

[9] Procuring AI

- Performance metrics // independent benchmarking
- Baselineing // stress testing the anomaly threshold
- ML/AI infrastructure // resilience to cloud connectivity issues
- Data privacy assurance // how will your (valuable and sensitive) data be handled?
- Robustness // who, what, how of regular evaluation

[9] Procuring AI

- Performance metrics // independent benchmarking
- Baselineing // stress testing the anomaly threshold
- ML/AI infrastructure // resilience to cloud connectivity issues
- Data privacy assurance // how will your (valuable and sensitive) data be handled?
- Robustness // who, what, how of regular evaluation
- Robustness, bias and explainability // in-house testing, explanations, failure mode map

[10] Conclusion

- Majority of AI products use machine learning, data-driven technology
- Widely used in cyber security applications, especially for attack detection
- Data driven detection can make
- Models must be built with rigorous testing and tested in situ regularly to avoid 'hidden' weaknesses causing security weaknesses
- Key legislation may impact the use of ML in some applications in the near future

СyBOK