EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR MIGRATION AND HOME AFFAIRS

Directorate B: Borders, Interoperability and Innovation
Unit B.4 : Innovation and Industry for Security

AI AND SECURITY OPPORTUNITIES AND RISKS

Towards a trustworthy AI based on European values

PASAG report 3 -2020 – AI and security

CONTENTS

Artificial intelligence (AI) is finding its way into almost every application area where massive data sets are involved. It is a horizontal technology with use cases across very different sectors and not a "bringing a man to the moon" project.

The Commission's Communication "Coordinated Plan on Artificial Intelligence"[1] addresses security related aspects in section 2.7: "There is a need to better understand how AI can impact security in three dimensions: how AI could enhance the objectives of the security sector; how AI technologies can be protected from attacks; and how to address any potential abuse of AI for malicious purposes."

The first dimension highlights directly the opportunities for the application of AI in the security sector and the Societal Challenge 7 of Horizon 2020 "Secure Societies: Protecting the security of Europe and its citizens".

Deep domain knowledge is required for AI to be effective and to understand its potential, its limitations and possible side effects.

The ability to screen extremely large data sets in close to real-time, to identify patterns that would otherwise be missed, makes AI a powerful application to meet the needs of public security in border control, fighting terrorism, and multiple law enforcement requirements and other public security services where decisions need to be made at speed and with the support of extensive data analysis.

The two other dimensions, protecting AI technologies from attacks and addressing the potential malicious usage of AI, are relevant for almost all fields of application for AI, far beyond the security sector. Many application areas address the necessity of safety of AI but underestimate the challenge to secure AI against targeted attacks. Security is mostly implicitly assumed and covered by terms like "trustworthy", "robust", "reliable" or "resilient". To make AI secure is a major task for Horizon Europe.

This report is intended to provide an overview of the challenges in the adoption of AI for public security, the state of the art about the security of AI itself and concerns about the uncertainties of the use of AI.

[1] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Coordinated Plan on Artificial Intelligence (COM(2018) 795 final)

1. SUMMARY OF RECOMMENDATIONS

1. Basic research is necessary to make AI more secure, reliable, unbiased, and explainable. Current threats such as adversarial machine learning undermine the trustworthiness of AI and mitigations need to be researched. Assessments and metrics are needed to evaluate how reliable a given decision is.
2. AI's impact on innovation cultures and new business models related to digital economy requires further research and case studies to generate wider understanding of AI's infrastructural importance to the economy and society.
3. AI is pervasive and can have extensive application in public security and cyber security, if sufficiently large data sets are available. Research projects should explain why they expect significant progress and provide clear KPIs to measure success and error rates.
4. Current basic AI technologies are by default insecure by design and not trustworthy. This does not affect necessarily all use cases, but research projects should be aware of it and provide measures to mitigate these shortcomings where appropriate.
5. The Ethics Guidelines for Trustworthy AI should be used as guidance towards an AI based on European values.
6. Trustworthy AI requires trustworthy computing capabilities. Many AI applications are deployed into the cloud for learning and scalable production. The EU should promote cloud-computing services operating exclusively under EU legislation to protect data from non-EU access.
7. European data pools will make AI much more effective than national or regional ones. This will require responsible trade-offs between effectiveness of AI and fundamental rights such as privacy, especially in the public security sector. The data quality and homogeneity of merged data is crucial for success.
8. Development of defensive measures to detect and combat malicious use of AI. This includes also measures against fake news and deep fakes. This requires interdisciplinary understanding of attacks against AI and how AI can be used for attacks.
9. The talent pool for AI experts is very limited. Comprehensive education programs sponsored by the EU and member states are necessary to achieve competitiveness. The public security sector will need dedicated funding to successfully attract talent for a sustainable deployment of AI within the government sector. Interdisciplinary research is needed to understand the kinds of new skill sets that will be needed in the future not only to develop and operate new AI systems but to identify their potential societal impacts and how these need to be addressed.

2. OVERVIEW

The EC has been taking a pro-active approach to a concerted European policy on the development and deployment of AI in Europe. There are several initiatives ongoing, including the development of a strategy document "Communication of the European Commission about Artificial Intelligence for Europe"[2], which proposes a human-centric AI; the "Coordinated Plan on Artificial Intelligence", intends to maximize the impact of investments at the EU level; the "High Level Expert Group on AI"[3] supports the implementation of the European AI Strategy with the development of guidelines for trustworthy and ethical AI.

AI is a generic term for a variety of techniques. This paper makes no attempt to distinguish analytically between AI, deep learning, machine learning, statistical approaches, data science, big data and related technologies. The use cases and research questions examined here have in common that they refer to automated learning, reasoning and decision-making processes on large datasets using more complex methods as predictable programs or rule engines. The recommendations made here are basically independent from the underlying technologies implementing AI.

AI research is currently dominated by approaches using machine learning (ML) theory. Its objective is the inference of general, probabilistic rules from a data set of real-world observations. This works well with large data sets across diverse areas where ML can be applied to identify consistent and meaningful patterns, which would otherwise require extensive manual analysis or bespoke computer programming to achieve. However, the focus on ML has been at the expense of its security and the algorithms developed to enable it are not currently designed to prevent their malicious use. They are effectively insecure by design. Additionally, individual choices proposed by ML systems lack technical transparency. If ML systems are trained with biased data, the systems will tend to reinforce that bias as they generate more data, which is flawed at its origin. Also, errors like misclassification are an intrinsic property of machine learning systems: "Machine learning promises to find rules that are *probably* correct about the *most* members of the set they concern."[4] There is an important need for scoring methods to determine how reliable a single decision of an ML system is. Additionally, there is an intrinsic bias in the use of the English language in software development and particularly, when supervising and training algorithms to learn, where the interaction defaults almost exclusively to English. This may not pose problems for many applications but may raise flags where data sets are not English based. China is an exception given its large Chinese language data sets and its sizeable investment in the technology.

These deficiencies are not caused by individual implementation or training flaws, rather by systemic issues that could lead to significant challenges, especially in the security

[2] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Artificial Intelligence for Europe {SWD(2018) 137 final}

[3] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

[4] „Deep Learning" p.114 by Bengio e.a., MIT Press 2016

area. This is of importance in the public security sector as its end users are likely not experts in these technologies.

Obviously, there is great potential to derive intelligence from large datasets or even from all available data collected by Law Enforcement, Border Control, network security etc. This will likely lead to completely new approaches to addressing security and enable new and efficient system architectures. A major feature of AI is the real-time capability. While the learning or training process is time consuming, decisions can be made in real-time. For some use cases this opens up some kinds of autonomy, on the other hand, this requires very high standards in the security and safety of AI based systems, if the capabilities for human intervention or correction are limited.

Additionally, AI can be too vulnerable to targeted attacks, if the attackers have direct access to the system and can feed it with spoofed or manipulated data. For example, researchers could fool a well-trained face recognition system by special crafted eyeglass frames and impersonate another person.[5]

Conversely, attackers can also apply AI to overwhelm classic security defences or spoof digital media or evidence, which can also be hard to detect.

Automated decision-making based on data that identifies individuals can seriously undermine fundamental rights of citizens or violate ethical values. This requires guidelines and regulations based on recommendations from policy makers.

As mentioned above, several initiatives and expert groups with a focus on AI have been launched, however it is recommended that cybersecurity expertise is incorporated in all of them.

3. APPLICATION AREAS FOR AI IN THE SCOPE OF THE "SECURE SOCIETIES" CHALLENGE

There are countless possible applications for AI in the security area. Typical topics include the processing of vast amounts of sensor data as in biometrics, image analysis, object detection and classification, CCTV, surface observation from space, speech recognition and similar sources. AI is applicable for finding specific patterns in large datasets, augmenting data. It can be applied to forecasting, planning and scheduling tasks, which are for example prerequisites for predictive policing. Autonomous systems and robotics need AI to interact with their environment, for navigation and control. In the Appendix A, we give a list of typical use case areas for security. However, this list should not be considered exhaustive.

No matter what application, AI requires large data sets with adequate data quality. There is a strong need for research on the evaluation and standardization of merged data sets from multilateral (EU-wide) sources. Challenges are the inhomogeneity of data merged from different sources and collected under various circumstances and legal bases. The quality of data used for training an AI system is essential. Training data can contain

[5] https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf

implicit racial, gender or social context information and result in biased decisions of systems. See Section 3 for details.[6]

Data sets that utilise information on individuals must be anonymized for privacy reasons. This is a special challenge for merging data sets, which require referential integrity. Research on the prepossessing and anonymization of real, sensitive data could contribute to a more effective AI.

Innovation in the development of AI systems that are useful and operational is not possible without proper knowledge and data management. Data sets and models of high quality (and quantity) are, therefore, one of the fundamental elements to achieve efficient use of AI.

The large amount of data available, with multiple formats, sources of origin, generated at high speeds, high performance processing requirements, are today one of the main constraints in AI research and development. Therefore, because data structures and models are the raw material of AI, it is mandatory to establish recommendations that facilitate their usability and interoperability to exploit them.

In this aspect, some lines of work could be:

- Strengthen and promote the use of common and shared information banks that combine government, industrial and scientific data, as well as promoting the interoperability of platforms, systems and devices that operate with that data.
- Promulgate adequate and rapid access to data, facilitating its standardization and quality, complying with the principles of their location, accessibility and possibility of reuse, identifying the gaps that must be corrected in order to provide compatibility between different systems / platforms.
- Encourage the collection and storage of data through pipelines that provide solid data structures and models, as well as reliable performance and with little tolerance to cracks and inconsistencies in the information contained.

A very promising area is the use of AI in software and cyber security research itself. Manual code reviews to find programming bugs are exhausting and expensive, but code reviews by Machine Learning systems could significantly improve software security.[7] This contributes to cyber security and improves safety and reliability of security critical software. There are already competing undertakings like the DARPA funded project CHESS[8] aiming to discover vulnerabilities of all types by combinations of "automated program analysis techniques with support for advanced computer-human collaboration".

6

7 See https://www.schneier.com/blog/archives/2019/01/machine_learnin.html for examples like https://arxiv.org/pdf/1807.04320.pdf and https://dspace.ou.nl/bitstream/1820/9725/1/Kronjee%20J%20IM9906%20AF%20scriptie.pdf

8 "Computers and Humans Exploring Software Security (CHESS)", https://www.fbo.gov/utils/view?id=c4a5f718356316bc219d585541a2b961

Malware analysis with ML is already established in the industry but has still a lot of potential for improvements.

The detecting of manipulated digital evidence is an aspiring research topic. Important examples are the revealing of manipulated audio tracks, digital images or videos including deep fakes[9], a technology for human image synthesis, which can be used to disseminate fake news.

Cyber security is considered an important application area for AI.[10] Incident detection and automated incident response ("self-healing systems") using automation and robotics are also considered as suitable applications for anomaly detection by AI. New attack vectors unknown to the AI system and very rare events are difficult to detect. Therefore, as a first stage we can expect assisted or augmented AI and not fully autonomous systems. It can also be used for abstracting experiences from single incidents to improve defence and attack capabilities, preparedness and resilience.

With regards to cyber security, the security challenge facing the EU, named "Hybrid Threats" is well recognized by MS policy makers, especially in the last three years. Offering an overview of resilience-based decision making through an approach that integrates the threats and dependencies related to infrastructural, informational, and social considerations, could contribute to more resilient societies, governments and public and private institutions.

Use of AI in traffic management will meet citizen needs by reducing the number of victims and traffic accidents, ensuring mobility through proper traffic management and providing better management of all the procedures associated with traffic management.

Video and image IA analysis, using of open sources, which allows the recognition of objects or certain patterns of behaviour, as well as the prediction of the evolution of certain phenomena (spread of a fire, gas, fluid, etc.). Additionally, applying these analysis techniques together to image and audio will enable the identification and monitoring of certain phenomena in social networks.

Surveillance and control of public spaces through AI techniques in order to predict which geographical areas, time slots, associated events (shows, events, concentrations, etc.) and their possible combinations are more prone to Incidents / crime or even which people are more likely to be involved in a crime, both as perpetrators and as victims. These predictions will assist in a better practice in making decisions for the deployment of officers in these areas, or on the preventive identification of suspects.

IA monitoring and tracking of migratory flows, to allow the identification of key routes from the point of view of illegal traffic, as well as predicting mass movements.

---

[9] https://en.wikipedia.org/wiki/Deepfake

[10] https://www.ieee.org/content/dam/ieee-org/ieee/web/org/about/industry/ieee_confluence_report.pdf?utm_source=lp-link-text&utm_medium=industry&utm_campaign=confluence-paper "Artificial Intelligence and Machine Learning Applied to Cybersecurity"

4. VULNERABILITY OF AI AGAINST ATTACKS AND RELATED DEFENCE TECHNIQUES AND OTHER WEAKNESSES

Like the early internet, AI research and development is assumed to be benign in its applications, where all inputs are bona-fide data and trustworthy. This implies that input or training data are assumed to be authentic and not maliciously exploited with the objective of biasing the system. Together with the property that machine learning is trained to achieve best results for the average case and not for every possible input, including malformed or spoofed data, this offers a wide field of attack against ML. Specific attacks against ML like "adversarial machine learning" can harm reliability, integrity, and security of AI systems.

Recent research has produced a number of surprisingly simple attacks, which illustrate the vulnerability of AI: physically realisable and inconspicuous attacks against state-of-the-art face recognition systems[11], "one pixel attacks" show that small perturbations in the input can often result in misclassification by the AI system.[12] Attacks are also possible as black-box attacks without detailed knowledge of internal information about the AI system[13] or can be implemented to extract internal private or sensitive data from the system[14]. Most of these attacks do not require deep technical or specialist knowledge, as would be required for the development of "zero-day exploits"[15]. A person who can develop a ML system can write attack code as well.

Currently, neither the full range of possible attacks nor feasible defences are well understood. Also, the recent ENISA report "Analysis of the European R&D priorities in cybersecurity - Strategic priorities in cybersecurity for a safer Europe"[16] addresses these issues in appendix B1.

Another weakness of AI is its susceptibility to bias. Many AI systems are trained with inherently biased data, based on past human also biased decisions, and can contain gender, racial or social class bias and could make decisions affected by these biases. A prominent example with serious impact on individuals is COMPAS, a software used to predict the probability of which criminals are most likely reoffending. A study by ProPublica claimed that "defendants were twice as likely to be incorrectly labeled as higher risk than white defendants."[17] COMPAS' actual bias disposition has been discussed prominently in the community. [18] Managing the bias problem of AI is one of the most urgent practical problems. See for example "Tackling bias in artificial

[11] https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf

[12] https://arxiv.org/pdf/1710.08864.pdf

[13] https://arxiv.org/pdf/1602.02697.pdf

[14] https://www.cs.cmu.edu/~mfredrik/papers/fjr2015ccs.pdf

[15] A zero-day exploit is an exploitation of a software vulnerability, which is unknown to the public or even the software vendor. https://en.wikipedia.org/wiki/Zero-day_(computing)

[16] https://www.enisa.europa.eu/publications/analysis-of-the-european-r-d-priorities-in-cybersecurity

[17] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[18] See https://medium.com/thoughts-and-reflections/racial-bias-and-gender-bias-examples-in-ai-systems-7211e4c166a1,

https://epic.org/algorithmic-transparency/crim-justice/ and www.crj.org/assets/2017/07/9_Machine_bias_rejoinder.pdf.

intelligence (and in humans)" [19], McKinsey Global Institute, 2019, or "Responsible AI Practices"[20], Google AI Research, 2020.

Transparency of decision-making processes is a fundamental requirement in democratic societies. A usual requirement is the disclosure of the data categories used as input and the rules with which they are processed, at least on an abstract level. Currently, AI has not evolved sufficiently to provide a clear understanding of how decisions are arrived at. Especially in the case of false decision ("false positives") by complex AI systems, it is currently impossible to understand the rationale for the erroneous decision.

5. ETHICAL CONSIDERATIONS ON AI

The Commission states in its Communication "Artificial Intelligence for Europe" that "The guiding principle of all support for AI-related research will be the development of "responsible AI", putting the human at the centre[21]. The "Draft "Ethics Guidelines for Trustworthy AI" of the High Level Expert group on Artificial Intelligence (HLEG) places trustworthiness of AI at the centre: "Trustworthy AI has two components: (1) it should respect fundamental rights, applicable regulation and core principles and values, ensuring an "ethical purpose" and (2) it should be technically robust and reliable since, even with good intentions, a lack of technological mastery can cause unintentional harm."[22]

Summary of the guideline's major requirements for AI systems:

1. Accountability
2. Data Governance
3. Design for all
4. Governance of AI Autonomy (Human oversight)
5. Non-Discrimination
6. Respect for (& Enhancement of) Human Autonomy
7. Respect for Privacy
8. Robustness
9. Safety
10. Transparency

[19] https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans

[20] https://ai.google/responsibilities/responsible-ai-practices/

[21] see the Commission's "Responsible Research and Innovation" workstream: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation"

[22] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57112, p. 4.

The Guidelines of the HLEG should provide a common ground for security research programs. Especially (but not exclusively), the requirements for non-discrimination, privacy, robustness, safety and transparency should be the basis of trustworthy European AI applications in security. However, this is a challenging goal in view of the current systemic security weaknesses of AI.

The recent PASAG report "Achieving synergies between security and information-related fundamental rights (IRFR) in a digital intensive environment"[23] addresses the seemingly competing fundamental rights of security and privacy.

Regarding the use of AI based automated decision systems and the associated large databases, there is a special responsibility to protect the fundamental rights as mentioned in that report, including:

- The right of freedom of expression
- The right of freedom of information
- The right to private and family life
- The right to privacy and protection of reputation
- The right to a fair trial including the right for a final judgement by humans

These rights should be characterised by the following features:

Fairness: AI should not breach rights, such as the right of freedom of expression.

Accountability: a culture of accountability must be established at an institutional and organizational level.

Transparency: the path taken by the system to arrive at a certain conclusion or decision must not be a 'black box'.

Explainability: the decisions and actions of a system must be comprehensible to human users.

These rights can be impacted in many AI use cases, even with good intentions. For example, a system for the automated detection of hate-speech can impact the freedom of expression and the freedom of information. Another example is predictive policing, where AI is used to identify potential criminal activity but can also harm the privacy of innocent or uninvolved persons.

Besides the fundamental rights (UN Universal Declaration of Human Rights[24], the UN International Covenant on Civil and Political Rights[25], the European Convention on Human Rights[26], the Charter of Fundamental Rights of the European Union[27]), the

[23] http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=38224&no=1

[24] http://www.un.org/en/universal-declaration-human-rights/

25 https://www.ohchr.org/en/professionalinterest/pages/CCPR.aspx

26 https://www.echr.coe.int/Documents/Convention_ENG.pdf

27 www.europarl.europa.eu/charter/pdf/text_en.pdf

European General Data Protection Regulation (GDPR) and the European Police directive (EU) 2016/680 regarding the processing of personal data by competent authorities are of particular relevance. As the use of AI by law enforcement becomes more pervasive, touching ever more upon the lives of citizens, it becomes increasingly important for law enforcement to ensure that the use of these technologies is ethical.

Part of the challenge to deciphering the ethical use of AI is that law enforcement and civil society come at this from different perspectives. The primary role of law enforcement is, in essence, to protect the community and its citizens from harm and, in doing so, it must find a balance between security and privacy.

Law enforcement is, at the same time, not detached from either the community or its citizens, meaning that, should it overstep its boundaries through an alleged unethical behavior or action, it exposes itself to be held accountable by the citizens it serves. Accordingly, law enforcement must carefully consider the use of, and particularly the placement of, sensors and the use of the collected data.

AI systems and the knowledge of how to design them can be put to both civilian and military uses, and more broadly, to beneficial and harmful ends. Many tasks that would be beneficial to automate are themselves dual use. For example, systems that examine software for vulnerabilities have both offensive and defensive applications, and the difference between the capabilities of an autonomous drone used to deliver packages and the capabilities of an autonomous drone used to deliver explosives, needs not be very great.

6. MALICIOUS USE OF AI

An interdisciplinary team Researchers from different institutions released a report about "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation"[28]. We mention here their three main domains of consideration.

Digital security

The use of AI to automate tasks involved in carrying out cyberattacks (training machines to hack); will erase the existing trade-off between the scale and efficacy of attacks. This will likely expand the threat associated with labour-intensive cyberattacks - such as spear phishing. These attacks use personalized messages to extract sensitive information or money from individuals, with the attacker often posing as one of the target's friends, colleagues, or professional contacts. The most advanced spear phishing attacks require a significant amount of skilled labour, as the attacker must identify suitably high-value targets, research these targets' social and professional networks, and then generate messages that are plausible within that context. Spear phishing is more effective than regular phishing, which does not involve tailoring messages to individuals, but is relatively expensive and cannot be carried out en masse.

28 https://maliciousaireport.com/

Novel attacks that exploit human vulnerabilities are also expected to benefit from AI (e.g. the use of speech synthesis for impersonation), existing software vulnerabilities (e.g. through automated hacking), or the vulnerabilities of AI systems (e.g. through adversarial examples and data poisoning).
AI can also be used to produce „deep fake" imagery. Existing images or video are superimposed with other sources to forge pictures or videos.[29]

1. Physical security

The use of AI to automate tasks involved in carrying out attacks with drones and other physical systems (e.g. through the deployment of autonomous weapons systems) may expand the threats associated with these attacks. The proliferation of lethal autonomous weapons may lead to a global arms race, because unlike nuclear weapons, they require no costly or hard-to-obtain materials and are inexpensive to mass-produce. It will only be a matter of time until they appear on the black market and in the wrong hands where they could be used in crowd control and population surveillance, or as biological or chemical weapons delivery systems. Autonomous weapons are ideal for tasks such as assassinations, destabilizing nations, subduing populations and selectively attacking ethnic groups. Novel attacks that subvert cyber-physical systems are also expected (e.g. causing autonomous vehicles to crash) or involve physical systems that would be infeasible to direct remotely (e.g. a swarm of thousands of micro-drones).
Automated risk in the use of IoT assisted by AI systems should be considered and analysed as new attack methods.

2. Political security

The use of AI to automate tasks involved in surveillance (e.g. analysing mass-collected data), persuasion (e.g. creating targeted propaganda-fake news), and deception (e.g. manipulating videos) may expand threats associated with privacy invasion and social manipulation. Novel attacks are also expected that take advantage of an improved capacity to analyse human behaviours, moods, and beliefs based on collected data. These concerns are most significant with authoritarian regimes but may also undermine the ability of democracies to sustain truthful public debates.

7. AI SECURITY RESEARCH AS A HORIZONTAL TOPIC

Security awareness for the AI implementation strategy

Security, robustness, reliability, and transparency are key success factors for a trustworthy, human centric AI. There is a strong imperative to connect the AI implementation strategy with AI security research.

29 https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

The AI HLEG sets out a framework for trustworthy AI and offers guidance for the implementation, realization, and the assessment for AI to be trustworthy. AI research and implementation projects should all have the requirement to consider security issues and assess whether a project is potentially affected by them.

Enabling requirements for trustworthy AI for the security sector

Powerful AI systems need three key resources: large data pools for training, qualified personnel for engineering and operation and trustworthy and secure computing capabilities, very often as a cloud-based solution.

The European Union should actively support the creation of European data pools for research and operation. These must comply with fundamental regulations like the General Data Protection Regulation (GDPR) and the European Police directive (EU) 2016/680. This will require responsible trade-offs between effectiveness of AI and fundamental rights such as privacy, especially in the public security sector. However, it is evident that these necessary trade-offs may lead to databases that are not sufficiently sizeable to enable effective machine learning, as is possible in some countries which are not governed by the same regulations that protect European values. One way to ensure European competitiveness in this domain is to improve the effectiveness of AI by applying research to make it more effective with smaller datasets. This could also be linked to increasing quality of data through examination and addressing of identified biases.

Competitions launched to exploit AI on given data sets are popular and have proven their worth in some ML research areas to measure the performance of different approaches. Research teams can compete and learn from other teams to improve their own methods. The best-known platform for such competitions is Kaggle[30], operated by Google. An EU funded platform like Kaggle for research with public or even "restricted" data from the public security sector could foster competition and collaboration, as well as new research projects.

The rapid rise of AI research and technology implementation results in a high demand for experts. Currently the talent pool in this area is limited, including in Europe. There is no concerted programme at both national and EU level to develop the talent pool that would be required in Europe. Government agencies, such as those in Law Enforcement, will be at the bottom of the priority list for the development of AI capabilities because commercial ventures will pay more and be more attractive. There is a need for programmes for the sustainable training of the public sector staff in AI. This includes not only AI engineers and researchers but also the end users, which are not experts but must work with AI-based technology, and their benefits and shortcomings. Education initiatives are necessary to improve the competitiveness of Europe versus the US and China. In particular, the public security sector will need dedicated funding.

Trustworthy AI demands trustworthy computing capability. Applications are very often deployed in the cloud. The training of machine learning is done with dedicated hardware

30 www.kaggle.com

which is available on demand in the cloud. These singular load peaks do not justify expensive investments in on-premise hardware. In addition, AI production is more effectively executed within cloud services due to their scalability and flexibility.

Europe needs competitive cloud services regulated under EU legislation, to protect the privacy and confidentiality of data of European citizens, companies, and government agencies.[31]

[31] See the recent discussion about the storage bodycam videos of German police officers in the Amazon cloud https://www.dw.com/en/german-police-storing-bodycam-footage-on-amazon-cloud/a-47751028

8. APPENDIX A: USE CASE AREAS FOR AI IN THE SCOPE OF "SECURE SOCIETIES"

Due to the sharp growth and application of AI in nearly all conceivable domains, it is not effective to recommend selected use cases for research funding calls. Instead, we have identified some security areas that would benefit from AI and listed some exemplary use cases.

Evident applications leading to potential high returns involve the examination of large data sets, biometrics, image analysis, real time video analysis, modelling, planning and decision-making processes in each area and use of more automated actions and systems to help practitioners.

AI use cases for LE

- Interpreting massive data sets (images, videos, audio records, geospatial intelligence, communication data, …) and supporting decision making processes and prediction of criminal behaviour;
- Movement profiles of suspicious or stolen vehicles;
- Web analytics and open source intelligence from social media, dark web;
- Detection of forged digital evidence;
- Face and soft biometrics to detect and identify criminals, and search for persons of interest;
- Attention to Victims of Gender Violence

AI use cases for Border Control

- Enriching biometrics and identity techniques by machine learning;
- Identifying suspicious behaviour and events;
- More efficient methods to balance the facilitation of free movements of goods and people and security;
- Disease detection;
- Translation of foreign language conversation (collaboration between Border controls, conversation with citizens);
- Patrol drones for borders surveillance;

AI use cases for network security / cybersecurity

- Detection of network attacks like "advanced persistence threats" by nation state actors or cyber criminals;
- Monitoring and protection by AI of networks in critical infrastructures;

AI use cases for the protection of Smart Cities / Critical Infrastructures

- Improving situation awareness in real time, generating qualified alerts;
- Designing new architectures of detection & protection systems and level of resilience, leveraging the complex systems behaviour simulation capability enhanced by AI;

AI use cases for disaster recovery

- Monitoring of damage assessment to landscapes;
- Use of social media information to better assess the situation and decide;
- Modelling resilience;
- Disease detection at borders, airports;

AI use cases for first responders

- Automated Incident Response for level 1 cybersecurity incidents. AI could help, to anticipate or improve detection capabilities, based on experience and/or data sets. This is related to Cybersecurity Threat Intelligence (CTI);
- Automated Incident Detection and Prevention. The increasing volume of cyber security incidents managed by CSIRTs together with the increasing benefit of using AI could be used as a basic response to incidents: blocking, filtering, enriching, information gathering, escalation, domain cancellation, etc. Related to Cyber Threat Intelligence (CTI);
- Augmented /virtual vision of the scene;
- Offering “serious games” opportunities to analyse past events and simulating possible complex accidental situations with potential cascading effects;