Școala Națională de Studii Politice și Administrative
Facultatea de Administrație Publică
Smart-EDU Hub

MACHINE INTELLIGENCE & SECURITY FOR SMART CITIES

Security Above All!

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON MACHINE INTELLIGENCE & SECURITY FOR SMART CITIES (TRUST)
5th Edition, 2024

COMITETUL ȘTIINȚIFIC

Școala Națională de Studii Politice și Administrative
Facultatea de Administrație Publică
Smart-EDU Hub

MACHINE INTELLIGENCE & SECURITY FOR SMART CITIES

Security Above All!

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON MACHINE INTELLIGENCE & SECURITY FOR SMART CITIES (TRUST)
5th Edition, 2024

Coordonator:

Conf. univ. dr. **Catalin VRABIE**

PRO UNIVERSITARIA 2024
BUCUREȘTI

Editat de **Pro Universitaria SRL**, editură cu prestigiu recunoscut.
Editura **Pro Universitaria** este acreditată CNCS în domeniul Științelor Umaniste și CNATDCU (lista A2-Panel 4) în domeniul Științelor Sociale.

ISBN 978-606-26-1946-6

ISSN 3061-2330

Redactor:	Elena Onea
Tehnoredactor:	Victor-Ovidiu Căpriceru
Copertă:	Aurelian Leahu

Smart-EDU Hub
Faculty of Public Administration, SNSPA
Expozitiei blvd. no. 30A, 6th floor,
Bucharest,
Postal code: 012104, Romania
Tel. 0372-177182
e-mail: catalin.vrabie@snspa.ro
www.administratiepublica.eu

Redacție:
tel.: 0732.320.664
e-mail: editura@prouniversitaria.ro
Editura Pro Universitaria
www.prouniversitaria.ro

Librăria UJmag:
tel.: 0733.673.555; 021.312.22.21
e-mail: comenzi@ujmag.ro
Ujmag.ro
www.ujmag.ro

CUPRINS

ICT Security in European Enterprises. Examples of ICT Security Solutions

Florin-Domnel GRAFU,
ROMATSA, Bucureşti, România
florin.grafu@romatsa.ro

Cristina LEOVARIDIS,
SNSPA, Bucureşti, România
cristina.leovaridis@comunicare.ro

Abstract
Lucrarea de față își propune să ofere o imagine statistică de ansamblu asupra situației actuale a digitalizării și a utilizării securității sistemelor TIC în companiile europene, cu accent pe propunerea implementării unor meteode de protecție a datelor companiilor. După o analiză secundară de date statistice recente furnizate de instituții europene cu privire la nivelul digitalizării, dar mai ales al preocupării pentru securitatea TIC în companile europene, în care se evidențiează comparația dintre România și media UE, articolul continuă cu o prezentare a unor soluții pentru implementarea tehnică a securității cibernetice în companii. Va fi expus un set de tehnologii care fac posibilă detectarea amenințărilor cibernetice și a atacurilor legate de securitatea IT.

Cuvinte cheie: securitate cibernetică, centru pentru operațiuni de securitate, digitalizare.

1. Introducere
În contextul digitalizării accelerate din ultimii ani a întreprinderilor europene, se pune din ce în ce mai acut problema asigurării securității sistemelor TIC. O imagine statistică de ansamblu asupra nivelului de digitalizare a întreprinderilor europene, urmată de o prezentare statistică a situației preocupărilor legate de securitatea cibernetică în cadrul companiilor, comparativ între țara noastră și media UE, am considerat a fi absolut necesare. Instrumentele tehnice utilizate pentru detecția și combaterea incidentelor la nivelul securității cibernetice sunt expuse sumar în cea de-a doua parte a acestui articol și constau în analiza SOC-ului și a SIEM-ului.

Dovadă a importanței acordate de instituțiile europene protecției datelor în format digital sunt și documentele oficiale europene elaborate în ultimii cinci ani, precum ”The EU's Cybersecurity Strategy for the Digital Decade” [1] și ”Network and Information Security 2 (NIS2) Directive” [2], ce includ recomandări și măsuri pentru un nivel comun ridicat de securitate a rețelelor și a sistemelor informatice în Uniune. Pe plan național, implementarea Legii NIS intră în responsabilitatatea Directoratului Național de Securitate Cibernetică (DNSC), autoritatea românească în competențele căreia intră securitatea rețelelor și sistemelor informatice [3].

2. Digitalizare și securitate TIC în companiile europene – o perspectivă statistică
În 2023, majoritatea covârșitoare (93,9%) a organizațiilor europene cu peste 10 angajați europene aveau conexiune în bandă largă fixă la Internet, peste trei sferturi (78.1%) aveau website și aproape două treimi (60.9%) foloseau cel puțin un tip de social media: peste jumătate (58.9%) foloseau rețelele de socializare (Facebook, LinkedIn etc.), o treime (31.5%) website-uri de multimedia content sharing (YouTube, Flickr, SlideShare,

Instagram, Pinterest, Snapchat), și doar 1 din 10 (10.2%) foloseau bloguri corporative (Twitter) [4].

Referitor la integrarea e-business (utilizarea TIC de către întreprinderi pentru a rula, integra și îmbunătăți procesele lor de afaceri, pentru a împărtăși și a face schimb de informații în interior, pentru a analiza date sau pentru a comunica cu partenerii de afaceri și clienții), peste 4 din 10 întreprinderi au apelat la aplicații software de planificare a resurselor (ERP - enterprise resource planning), un sfert (25.8%) au folosit aplicații de managementul relațiilor cu clienții (CRM - customer relationship management), iar peste 1 din 10 (15.3%) au folosit software-uri de Business Intelligence (BI). Puțin sub jumătate din organizațiile europene (45.2%) au apelat la serviciile de cloud (în loc să-și extindă propria infrastructură IT, întreprinderile pot cumpăra resurse de calcul găzduite de terți pe internet, iar aceasta include acces flexibil, la cerere, la servicii precum software, putere de calcul, capacitate de stocare etc.). În 2023, puțin peste un sfert (28,2 %) dintre întreprinderile europene peste 10 angajați au efectuat analize de date prin intermediul propriilor angajați (mai precis, au utilizat tehnologii, tehnici sau instrumente software pentru analiza datelor interne sau a datelor din surse externe și pentru a extrage modele, tendințe și perspective din date pentru a formula concluzii, predicții și a lua decizii mai bune cu scopul de a-și îmbunătăți performanța - creșterea producției, reducerea costurilor). Sub 1 din 10 (8%) din întreprinderile europene au apelat la inteligența artificială (sisteme care utilizează tehnologii precum: *text mining*, recunoașterea vorbirii, generarea limbajului natural, învățarea automată, învățarea profundă pentru a culege și/sau utiliza date pentru a prezice, recomanda sau decide, cu diferite niveluri de autonomie, cea mai bună acțiune de realizat anumite obiective specific etc.; sistemele de AI pot fi bazate exclusiv pe software sau încorporate în dispozitive) [4].

Ca urmare a necesității asigurării securității cibernetice, conform ultimelor date furnizate de Eurostat, în 2022, 92% dintre întreprinderile din UE cu mai mult de 10 angajați au utilizat cel puțin o măsură pentru a menține securitatea sistemelor TIC, mai precis integritatea, disponibilitatea și confidențialitatea datelor și a sistemelor TIC; valorile indicatorului în România sunt apropiate de media europeană – 86% din întreprinderi au apelat la cel puțin o astfel de măsură. Firmele pot implementa o varietate de măsuri de securitate TIC pentru a preveni incidentele și pentru a asigura integritatea, disponibilitatea și confidențialitatea datelor și sistemelor lor TIC. Cea mai frecventă măsură utilizată pe plan european a fost autentificarea cu parolă puternică (82%), urmată de backup-ul datelor într-o locație separată sau în cloud (78%) și de controlul accesului la rețea (65%) [5].

Sub jumătate (49%) dintre întreprinderi au utilizat rețele private virtuale (VPN) sau au avut fișiere jurnal pentru analize după incidente de securitate (45%). Cel mai rar au fost folosite tehnicile de criptare pentru date, documente sau e-mail-uri (36%), testele de securitate TIC (35%), evaluările riscurilor TIC (32%), combinații de două sau mai multe mecanisme de autentificare (31%) sau identificarea și autentificarea utilizatorului prin metode biometrice (13%). Frecvența utilizării acestor măsuri diferă în funcție de dimensiunea organizației. Mai precis, autentificarea cu parolă puternică a fost utilizată de aproape toate întreprinderile mari europene (96%), de 90% dintre întreprinderile mijlocii și de peste 8 din 10 întreprinderi mici (81%). Valori apropiate ale indicatorului au fost înregistrate și pentru

salvarea datelor într-o locație separată: 93% dintre întreprinderile mari, 88% dintre întreprinderile mijlocii și 75% dintre întreprinderile mici au apelat la ea. Diferențe între organizații în funcție de dimensiunea lor sunt mai evidente în cazul măsurilor de securitate mai rar folosite: autentificarea printr-o combinație de cel puțin două mecanisme a fost utilizată de două treimi (64%) dintre întreprinderile mari, în timp ce ponderea celor mici care utilizează această măsură particulară a fost de peste două ori mai mică (28%). Indiferent de dimensiunea întreprinderii, identificarea și autentificarea utilizatorilor prin metode biometrice a fost cea mai puțin utilizată măsură de securitate TIC: de către doar o treime (29%) dintre întreprinderile mari și de doar 1 din 10 dintre cele mici (12%) [6].

Peste o treime (37%) dintre întreprinderile europene dețin documente care pun în aplicare măsuri, practici sau proceduri privind securitatea TIC, România înregistrând la acest indicator valori peste media UE (45%) și situându-se astfel în prima treime a clasamentului. Mai mult, în aproape un sfert (24%) din întreprinderi aceste documente sunt foarte actuale, fiind elaborate sau revăzute în ultimul an [5]. În țara noastră, o și mai mare parte din aceste documente au fost actualizate în ultimele 12 luni, mai precis în 4 din 10 întreprinderi (40%).

Peste jumătate (58%) din întreprinderile europene și-au conștientizat angajații în legătură cu responsabilitățile pe care le au în aspectele referitoare la securitatea TIC, și aici valorile indicatorului pentru România (62%) depășind media UE, ceea ce situează România în prima jumătate a clasamentului țărilor europene după acest criteriu (pe primul loc situându-se Irlanda și Cehia cu 75%). Pentru a realiza aceasta, 42% dintre întreprinderile europene au oferit angajaților săi instruire voluntară în domeniul securității TIC, 21% au introdus cursuri obligatorii pentru salariați în acest domeniu, iar 32% au inclus obligații de securitate TIC în contractele de muncă ale angajaților lor [5]. Și aici există diferențe în funcție de dimensiunea întreprinderii: ponderea întreprinderilor mari care conștientizează angajații cu privire la obligațiile lor în domeniul securității TIC a fost foarte mare (91%), acestea fiind urmate de cele mijlocii (76%) și de cele mici (54%).

Mai mult de una din cinci (22%) întreprinderi europene a experimentat în anul anterior incidente de securitate legate de TIC ce au dus la diferite consecințe, precum indisponibilitatea serviciilor TIC, distrugerea sau coruperea datelor sau dezvăluirea datelor confidențiale, România și din acest punct de vedere având o situație mai bună decât media UE – doar 19% din întreprinderile românești experimentând astfel de incidente. Incidentele de securitate TIC pot fi cauzate de atacuri rău intenționate din exteriorul sau din interiorul întreprinderii, sau de cauze non-malițioase, obiective, cum ar fi: defecțiuni hardware sau software sau acțiuni neintenționate ale propriilor angajați, aceasta a doua categorie fiind cel mai des raportată. Cea mai des înregistrată consecință generată de incidentele de securitate TIC a fost indisponibilitatea serviciilor TIC din cauza defecțiunilor hardware sau software (19% dintre întreprinderi). Mult mai rar a fost menționată (de doar 4% dintre organizații) indisponibilitatea serviciilor TIC din cauza atacurilor din exterior (de exemplu, atacuri ransomware, atacuri Denial of Service. Distrugerea sau coruperea datelor ca urmare a defecțiunilor hardware sau software a fost raportată de 4% dintre întreprinderi, în timp ce infectarea cu software cu scop rău intenționat sau intruziunea neautorizată care a dus la distrugerea sau coruperea datelor a fost raportată de 2% dintre întreprinderi. Cel mai rar, întreprinderile au raportat dezvăluirea de date confidențiale din cauza intruziunii, atacurilor

pharming sau phishing sau acțiunilor intenționate ale propriilor angajați (1%) sau din cauza acțiunilor neintenționate ale propriilor angajați (1%) [6].

Deși există un consens general la nivelul organizațiilor europene (71%) că securitatea cibernetică reprezintă o prioritate, punerea în practică a măsurilor de securitate TIC rămâne relativ dificilă: referitor la principalele provocări ale companiilor europene când trebuie să recruteze angajați cu competențe în cyber security, mai mult de jumătate dintre companiile care au căutat astfel de candidați au întâmpinat dificultăți cum ar fi găsirea de candidați calificați (45%), lipsa de candidați (44%), lipsa de conștientizare a rolului cybersecurity (22%), a schimbărilor tehnologice prea rapide și a nevoii permanente de instruire (19%), dar și a constrângerilor bugetare inclusiv din cauza costurilor foarte ridicate ale echipamentelor (16%) [7]. Mai mult, pregătirea analiștilor din companii, specializați în cybersecurity, pentru a executa fluxurile de lucru cu viteză și consecvență, poate fi o sarcină consumatoare de timp [8].

Toate măsurile pentru menținerea securității sistemelor TIC enumerate mai sus, pot fi realizate în cadrul companiilor prin intermediul soluțiilor expuse în cele ce urmează.

3. Soluții de securitate cibernetică pentru companii

Centrul de operațiuni de securitate (SOC) este un centru al unei instituții sau organizații specializate în monitorizarea și gestionarea securității informațiilor. SOC este punctul de întâlnire al sistemelor, proceselor și tehnologiilor legate de securitatea cibernetică. SOC este format dintr-o echipă dedicată de analiști și ingineri specializați în domeniul securității informațiilor. Această echipă monitorizează continuu rețelele, sistemele și aplicațiile pentru a detecta amenințările și atacurile cibernetice. SOC se bazează pe instrumente și tehnologii avansate pentru a detecta, verifica și răspunde la amenințări.

3.1. Ce este un SOC?

Un Centru de operațiuni de securitate (SOC) este o echipă centralizată de analiști de securitate responsabilă de monitorizarea, detectarea și răspunsul la amenințările de securitate cibernetică. SOC-urile folosesc de obicei o varietate de instrumente și tehnologii de securitate pentru a colecta și analiza date din întreaga infrastructură IT a unei organizații. Aceste date pot fi folosite pentru a identifica potențialele amenințări, pentru a investiga incidente și pentru a răspunde la atacuri. SOC-urile joacă un rol critic în protejarea organizațiilor împotriva atacurilor cibernetice [9].

3.2. Unele dintre principalele beneficii ale SOC în domeniul securității cibernetice

Monitorizarea și detectarea amenințărilor: SOC monitorizează și detectează amenințările cibernetice avansate și atacurile de securitate asupra sistemelor și rețelelor. Tehnologiile avansate și instrumentele de securitate sunt utilizate pentru a detecta amenințările din timp și pentru a limita impactul acestora.

Răspuns eficient la incident: Sunt implementate strategii de răspuns rapide și eficiente pentru a face față incidentelor de securitate. SOC ajută la analiza și clasificarea incidentelor și la luarea de măsuri de răspuns imediat pentru a investiga, a limita atacurile și a reporni în siguranță sistemele afectate.

Detectare și analiză: Datele și jurnalele de securitate din diverse surse sunt agregate și analizate în continuare pentru a identifica modele, comportamente neobișnuite și potențiale amenințări. Acest lucru permite organizațiilor să ia măsuri corective și să își îmbunătățească măsurile de securitate.

Reducerea timpului de recuperare: Cu monitorizarea continuă a securității și analiza eficientă, SOC poate reduce timpul de recuperare în urma atacurilor cibernetice. Permite verificarea rapidă și răspunsul la urgență pentru a reduce impactul atacurilor și a reduce timpul de nefuncționare.

Îmbunătățirea deciziilor strategice: SOC oferă rapoarte și analize periodice care ajută la înțelegerea situației generale de securitate și la evaluarea eficienței strategiilor și măsurilor de securitate luate [9].

3.3. Un set de tehnologii utilizate pentru a detecta amenințările cibernetice

Sistem de informare și management al incidentelor (Incident Information and Management System - SIEM): Un sistem de informare și management al incidentelor este utilizat pentru a colecta, analiza și monitoriza înregistrările evenimentelor și datele de securitate din mai multe surse. Tehnicile avansate de analiză și monitorizare sunt utilizate pentru a identifica modele neobișnuite și alerte atunci când sunt detectate activități suspecte.

Advanced Threat Detection (APT): Această abordare implică utilizarea de instrumente speciale și tehnologii sofisticate pentru a detecta amenințările avansate și atacurile țintite. Aceasta include analiza comportamentului utilizatorului și monitorizarea traficului de rețea pentru activități suspecte sau neobișnuite.

Analiza comportamentală: Tehnicile de analiză comportamentală sunt utilizate pentru a crea modele ale comportamentului normal al sistemelor și utilizatorilor. Tiparele neobișnuite și comportamentul anormal sunt monitorizate pentru potențiale amenințări, cum ar fi infiltrarea *hackerilor* și încălcările de securitate.

Fig. 1. Reprezentare logică a unui SIEM
Sursa : https://grcico.com/

Informații despre amenințări și informații despre securitate: sunt utilizate mai multe surse pentru a obține informații despre amenințări și informații despre securitate, cum ar fi baze de date publice, platforme de schimb de securitate și colaborări cu alte organizații din domeniul securității cibernetice. Aceste informații ajută la identificarea activităților cibernetice rău intenționate [9].

3.4. Principala importanță a SIEM în SOC

Colectarea datelor: Sistemul SIEM colectează date de securitate și jurnalele de evenimente din diverse surse din infrastructura de rețea și din diverse sisteme. Aceste date sunt agregate într-un singur loc pentru o analiză și monitorizare cuprinzătoare.

Analiză și verificare: Datele colectate sunt analizate folosind reguli și criterii specifice pentru a identifica tipare anormale, activități suspecte și amenințări de securitate. Tehnici avansate de verificare și analiză sunt utilizate pentru a se asigura că amenințările sunt validate și clasificate corespunzător.

Alerte și alarme: Sistemul SIEM oferă alerte și alarme în timp real atunci când sunt detectate activități suspecte sau amenințări de securitate. Acest lucru ajută echipa SOC să răspundă rapid la amenințări și să ia măsuri pentru a reduce impactul negativ.

Investigarea și raportarea: SIEM facilitează investigarea și analiza ulterioară a incidentelor de securitate. Oferă un jurnal detaliat al tuturor activităților și incidentelor și permite crearea de rapoarte cuprinzătoare pentru a înțelege situația de securitate și a evalua performanța sistemelor de protecție.

Conformitate și audit: Sistemul SIEM ajută la respectarea standardelor și reglementărilor de securitate aplicabile, cum ar fi păstrarea datelor pentru o anumită perioadă, conform reglementărilor companiei [9].

4. Exemple de „Use Case-uri” în domeniul securității cibernetice

Problema 1: Răspuns la incident

O sală tehnică și echipamente actuale de înaltă tehnologie nu sunt necesare, nici măcar nu sunt suficiente pentru a preveni sau rezolva imediat incidentele. Elementele menționate sunt utile atunci când vine vorba de aspectele de „marketing” ale Centrului de operațiuni de securitate. Pentru a opera un SOC și în special în timpul răspunsului la un incident, oamenii joacă un rol cheie.

Soluția: oamenii

Abilitățile care se cer analiștilor SOC trebuie să corespundă cerințelor naturii muncii lor, care este foarte nișată. Abilitățile de bază care sunt necesare analiștilor SOC sunt următoarele:

Abilitati tehnice:

- Rețea (TCP/IP);
- Administrator de sistem (Linux, Windows);
- Analist malware (analiza comportamentală);
- Investigator criminalistic.

Abilități soft:

- Gândire analitică;
- Redactare;
- Comunicare;
- Lucru în echipă;
- Orientarea către client;
- Capacitatea de sinteză a rezultatelor obținute;
- Abilitatea de a lucra sub presiune.

Abilitățile de mai sus sunt cheia pentru stabilirea unei funcționalități de bază pentru SOC [10].

Problema 2: Cum să reacționați la evenimentele și incidentele de securitate

După cum s-a menționat mai sus, resursa umană este esențială. Dar organizația trebuie să fie sprijinită (în România, acest sprijin este oferit companiilor de către Directoratul Național de Securitate Cibernetică) pentru a putea răspunde în conformitate cu *Răspunsul* la incidentul de securitate IT.

Soluția: Procesul de răspuns la incident

Procesul de răspuns la incident urmează o abordare standard [10].

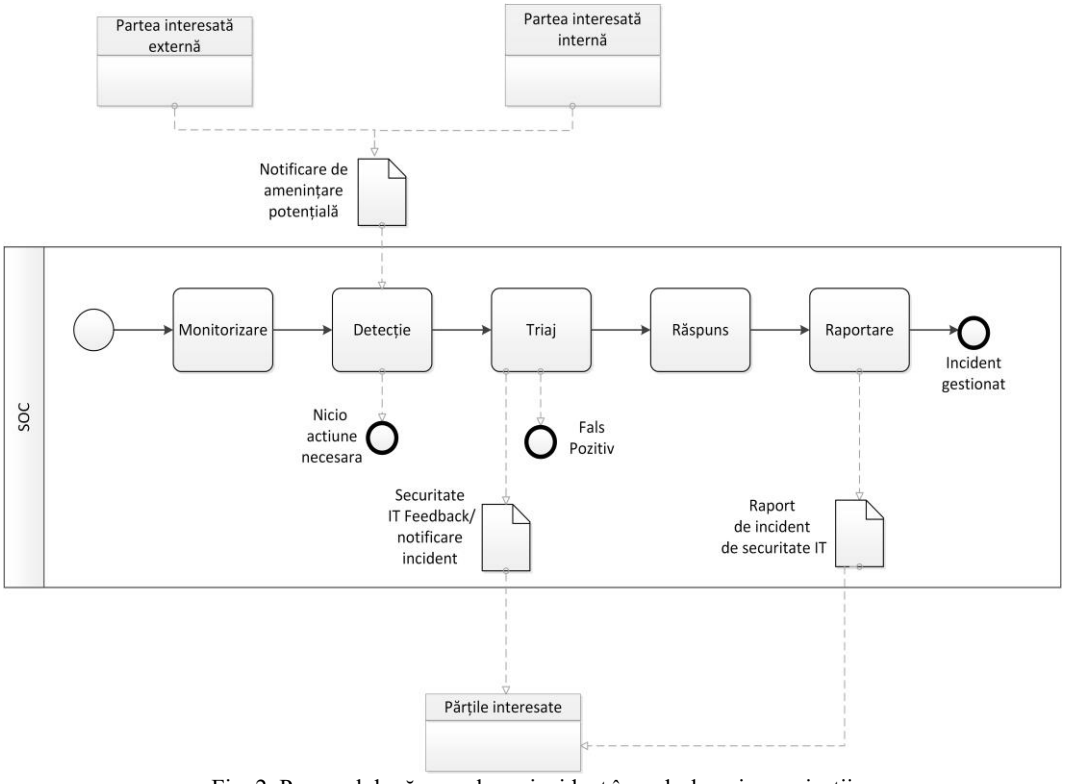

Fig. 2. Procesul de răspuns la un incident în cadrul unei organizații
Sursa : Eurocontrol, "ATM-ATM SOC Implementation v1.0"

5. Concluzii și discuții

Analiza secundară de date statistice realizată în prima parte a lucrării de față, pe baza datelor oficiale oferite de Eurostat și Comisia Europeană indică faptul că România se plasează, la majoritatea indicatorilor ce țin de securitatea cibernetică a companiilor, în prima parte a clasamentului țărilor europene, în cele mai multe cazuri valorile indicatorilor pentru România fiind peste media UE atunci când facem referire la aspecte pozitive (deținerea de către companii de documente care pun în aplicare măsuri, practici sau proceduri privind securitatea TIC, actualitatea acestora, conștientizarea angajaților în legătură cu responsabilitățile pe care le au vis-a-vis de securitatea TIC) sau sub media UE atunci când ne referim la aspecte negative (incidente de securitate cibernetică în companii, ce au dus la diferite consecințe).

Prin oferirea de cursuri de conștientizare a securității cibernetice, organizațiile trebuie să își responsabilizeze angajații să devină prima linie de apărare împotriva amenințărilor cibernetice. Angajații informați sunt mai bine pregătiți pentru a detecta și a răspunde eventualelor incidente de securitate. Existența unui SOC în interiorul unei companii crește rezistența acesteia la atacurile cibernetice; totodată, acest serviciu poate fi contractat și de la firme specializate în domeniu.

Anexă

Acronime

AI – Artificial Intelligence (Inteligență Artificială)
ATSEP – Air Traffic Safety Electronics Personnel (Personal electronist care asigură siguranța traficului aerian)
NIS - Network and Information Security (Securitatea Rețelei și a Informațiilor)
SIEM - Incident Information and Management System (Sistem de informare și management al incidentelor)
SOC – Security Operations Center (Centru de operațiuni pentru securitate)
TIC – Tehnologia Informației și a Comunicațiilor.

Referințe

[1] European Commission, "The EU's Cybersecurity Strategy for the Digital Decade," December 2020. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/eus-cybersecurity-strategy-digital-decade-0.

[2] European Commission, "Network and Information Security 2 (NIS2) Directive," Whitepaper, February 2023. [Online]. Available: https://www.dnv.com/cybersecurity/cyber-insights/nis2-directive/ .

[3] Directoratul Național de Securitate Cibernetică (DNSC), "Autoritatea competentă la nivel național pentru securitatea rețelelor și sistemelor informatice," 2024. [Online]. Available: https://dnsc.ro/pagini/ansrsi .

[4] Eurostat , "Digital economy and society statistics – enterprises. Statistics explained," January 2024. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digital_economy_and_society_statistics_-_enterprises.

[5] Eurostat, "Digitalisation in Europe – 2023 edition," 2023. [Online]. Available: https://ec.europa.eu/eurostat/web/interactive-publications/digitalisation-2023#ict-security.

[6] Eurostat , "Statistics explained. ICT security in enterprises," December 2022. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=ICT_security_in_enterprises#ICT_security_in_EU_enterprises .

[7] European Commission , "Eurobarometer survey on cyberskills.," May 2024. [Online]. Available: https://europa.eu/eurobarometer/surveys/detail/3176.

[8] T. Driggs, "New Charlotte AI Innovations Enable Prompt Collaboration and Demystify Script Analysis," Crowdstrike Blog, 2024.

[9] M. Abu-Fadaleh, "All About SOC (Security Operation Centers)," Green Circle, 2024.

[10] Eurocontrol, "ATM-ATM SOC Implementation v1.0," Brussels, 2024.

AI & Cybersecurity – connection, impacts, way ahead

Oana-Alexandra SARCEA (MANEA),
SNSPA University, Bucharest, Romania
oana.manea89@gmail.com

Abstract
Artificial intelligence (AI) and cybersecurity have a strong connection and impact each other in different ways. An overview is related to the following categories: detection and prevention, automated response, adversarial AI, data protection, risk assessment, privacy concerns. Looking ahead, the linkage between AI and cybersecurity will continue to evolve. Key areas of focus include: development of AI-driven security solutions, ethical AI-use, enhanced threat intelligence, human-machine collaboration. Overall, AI holds tremendous potential to revolutionize cybersecurity, but it also presents new challenges that must be addressed to ensure a secure and resilient digital environment. The relationship between AI and cybersecurity is multifaceted. AI technologies are increasingly being employed both to enhance cybersecurity defenses and to facilitate cyberattacks. In addition to the above key points, is to mention the cybersecurity skills gap: with the growing complexity of cyber threats, there is a shortage of skilled cybersecurity professionals. AI technologies can help bridge this gap by automating routine tasks and augmenting the capabilities of existing security teams. Behavioral analysis is another important element: AI-powered systems can analyze user and network behavior to identify anomalies that may indicate a security breach. By understanding typical behavior, AI can detect deviations that might signal an attack. Overall, the relationship between AI and cybersecurity is complex and evolving. While AI offers significant opportunities to enhance cybersecurity defenses, it also presents new challenges and risks that must be addressed. Ongoing research and development are essential to stay ahead of emerging threats in this rapidly evolving landscape.

Keywords: new technologies, AI benefits, cybersecurity challenges, AI-cybersecurity linkage.

1. Introduction
Digitalization has a big impact on all parts of a business and can alter fundamentally the way how an enterprise provides value and operates with its consumers [1]. Digital systems should be designed for peaceful use, but their potential for both beneficial and harmful applications has been explored extensively in the cybersecurity literature [2, 3]. One effective response to cyberattacks can be artificial intelligence (AI), which can automate threat detection and mitigation, ensuring a rapid and accurate response to various types of attacks. AI can analyze behavioral patterns and unusual activities in real-time, thereby identifying subtle indicators of attacks. Moreover, it can continuously adapt defense strategies based on automated learning from previous incidents. By incorporating AI within cybersecurity teams, organizations can gain an edge in combating cyber threats in an ever-changing and increasingly sophisticated environment. Artificial intelligence (AI) stands as a potent technology empowering cybersecurity teams to automate repetitive tasks, expedite threat detection and response, and enhance the precision of actions, thereby fortifying security against a spectrum of threats and cyberattacks. These undertakings align with the seamless integration of AI-based cybersecurity in today's landscape of digital transformation and multifaceted challenges [4].

Artificial intelligence and machine learning technologies play pivotal roles in this dynamic defense landscape, enabling organizations to detect anomalies, predict potential threats, and respond swiftly to emerging cyber risks.

AI is already a key technology in our economy, poised to bring transformative changes similar to those brought by the steam engine or electricity. However, concerns about potential loss of control in the human-AI relationship are increasing [5]. Issues such as autonomous driving and the opaque decision-making processes of vehicles, particularly in extreme situations just before a collision, have long been topics of public debate. Similar concerns arise regarding the extent to which AI should support or even make medical decisions independently. It will often be crucial to understand how a machine's decision was made and to evaluate the quality of its explanation [6].

In the age of digital interconnectivity, managing security vulnerabilities has become increasingly complex. Organizations face a growing number of potential vulnerabilities and often struggle to manage them effectively. Traditional vulnerability management approaches, which are typically reactive and address high-risk vulnerabilities only after they have been exploited [7], are inadequate in today's cybersecurity environment.

In this context, Artificial Intelligence plays a transformative role in vulnerability management. The combination of AI and Machine Learning (ML) offers a proactive and predictive approach. One significant method is User and Event Behavioral Analytics (UEBA), which enables AI systems to continuously analyze and learn from the baseline activities of an organization's user accounts, endpoints, and servers. This ongoing analysis helps identify abnormal behaviors that deviate from the established norms, potentially indicating zero-day attacks. Zero-day attacks exploit unknown vulnerabilities before developers can create and distribute patches, making them particularly dangerous [8]. AI and UEBA can detect these attacks much earlier in their lifecycle.

2. Artificial Intelligence and Cybersecurity connection

In an era marked by increasingly harmful and frequent cyberattacks, artificial intelligence (AI) adds an extra layer of complexity to an already entropic environment, for worse or better. While the recent debate revolves mostly around the challenges and security concerns posed by AI, the technology also provides the cybersecurity sector innovative ways to defend against hostile different actors. As a result, the market for AI in cybersecurity is expected to show considerable growth in the coming years, from around 24 billion U.S. dollars in 2023, to roughly 134 billion U.S. dollars by 2030.

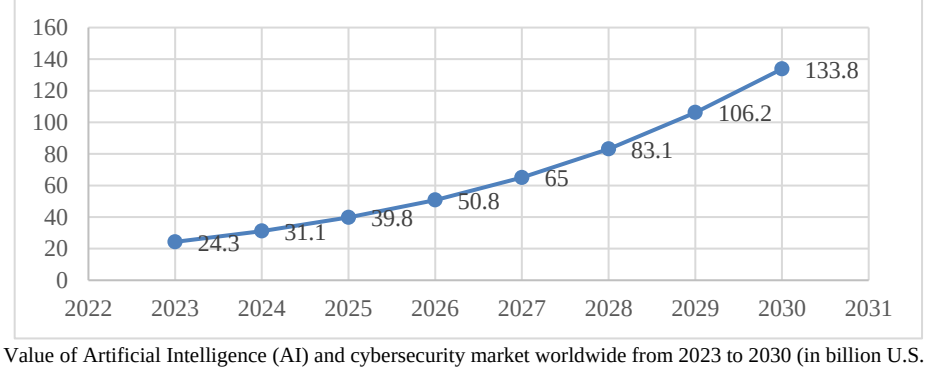

Fig. 1. Value of Artificial Intelligence (AI) and cybersecurity market worldwide from 2023 to 2030 (in billion U.S. dollars)
Source: Author's representation after A. Borgeaud, 18.03.2024, Statista
https://www.statista.com/statistics/1450963/global-ai-cybersecurity-market-size/

Many of the technologies influencing digital transformation are not novel innovations. The innovation lies in the integration of information, computing, communication, and connectivity technologies. The key technological domains enabling digital transformation are diverse and are commonly referred to as "general-purpose technologies" [9]. These encompass cyber-physical systems (CPS), the industrial internet of things (IoT), cloud computing (CC), big data (BD), artificial intelligence, and even augmented and virtual reality [10].

The significant impact of digital transformation on firms' processes and capabilities, as well as the alterations these technologies drive within industrial and organizational activities, has gained increasing academic attention in recent times. Digital transformation encompass a wide array of technological advancements, including the Internet of Things (IoT), Additive Manufacturing, Big Data, Artificial Intelligence, Cloud Computing, Augmented and Virtual Reality, and Blockchain, among others [11].

Fig. 2. Representations from craiyon.com after “Artificial Intelligence and Cybersecurity connection”

As companies traverse the path of digital metamorphosis, the very fabric of their operations undergoes a profound shift. This shift is characterized by the integration of cutting-edge technologies such as artificial intelligence, cloud computing, and the Internet of Things into daily business processes. The digital era heralds unparalleled opportunities for efficiency gains, enhanced customer experiences, and unprecedented innovation. However, this transformative journey is not without its perils.

The integration of artificial intelligence and machine learning in cybersecurity is a burgeoning area of research, as scholars explore how these technologies can enhance threat detection and response capabilities [12].

The integration of artificial intelligence and machine learning into business processes is a significant aspect of digital transformation. These technologies enable organizations to extract valuable insights from vast amounts of data, automate decision-making processes, and enhance the overall efficiency of operations. Predictive analytics powered by machine learning algorithms, for example, can help organizations forecast trends, identify potential risks, and make informed decisions.

In the realm of human resources, automation transforms traditional hiring processes. Intelligent recruitment tools, powered by artificial intelligence (AI), analyze resumes,

assess candidate suitability, and even conduct initial interviews. This not only accelerates the hiring cycle but also ensures a more data-driven and objective selection process, contributing to the acquisition of top-tier talent.

Technology, particularly data analytics and artificial intelligence, serves as the enabler of personalization. By leveraging customer data, businesses can unveil patterns, predict preferences, and tailor offerings to suit individual tastes. Recommendation engines, fueled by algorithms that discern purchase histories and browsing behaviors, elevate the customer experience by presenting relevant and personalized content [13].

Parallel to the cloud revolution is the integration of data analytics and artificial intelligence, reshaping how organizations derive insights, make decisions, and create value. The vast volumes of data generated in the digital age hold immense potential, but unlocking that potential requires advanced analytical tools and intelligent systems [14].

3. Methodology and impacts

3.1. Technical manner and explanations for the used VOSviewer method

Scopus (TITLE-ABS-KEY ("artificial intelligence") AND TITLE-ABS-KEY ("Cybersecurity")).

Subject area: Business, Management and Accounting: 197; Economics, Econometrics and Finance: 98; Total items: 224.

Web of science -> selected only business economics for research areas -> result 54 items. “artificial intelligence” (Topic) and cybersecurity (Topic); Manually merged duplicates in Zotero -> Result: 238 items; Type of analysis: Co-occurrence; Unit of analysis: keywords; Counting method: full counting.

Table 1. Thesaurus file

Label	Replace by
cyber security	cybersecurity
artificial intelligence (ai)	artificial intelligence
ai	artificial intelligence
machine-learning	machine learning
internet of things (iot)	internet of things
iot	internet of things
block-chain	blockchain
cyber-attacks	cyber threats
data privacy	data protection
data security	data protection
risk assessment	risk mitigation
risk management	risk mitigation
digitalization	digital transformation
intrusion detection	intrusion detection system

deep learning	machine learning
learning systems	machine learning
learning algorithms	machine learning

Minimum number of occurrences: 6; of the 1446 keywords, 27 meet the threshold.

Table 2.

Keyword	Occurrences	Total Link Strength
Cybersecurity	112	342
Artificial Intelligence	119	329
Machine Learning	52	159
Network Security	24	112
Internet of Things	27	103
Blockchain	22	88
Cyber Threats	19	69
Intrusion Detection System	15	58
Computer Crime	12	55
Digital Transformation	19	55
Risk Mitigation	12	47
Information Management	8	37
Security	11	37
Data Protection	13	35
Decision Making	8	32
Big Data	8	31
Sustainable Development	6	31
Denial-of-Service Attack	6	30
Sustainability	7	29
Crime	7	28
Automation	7	25
Malware	6	25
Behavioral Research	7	23
Industry 4.0	12	22
Security of Data	8	21
Fintech	8	16

Excluded elements: cybersecurity, artificial intelligence (search terms); security, crime, ‘current (general terms):

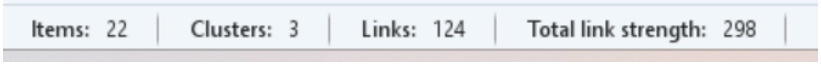

Fig. 3.

The network consists of 22 items, organized into a focused set of keywords. These keywords are grouped into 3 distinct clusters, indicating that the topics covered fall into three main thematic areas. There are 124 links, suggesting numerous connections between these keywords, creating a rich web of relationships and co-occurrences. The total link strength of 298 further emphasizes the robustness of these connections, indicating that the keywords frequently appear together in the literature, reflecting strong thematic interdependencies.

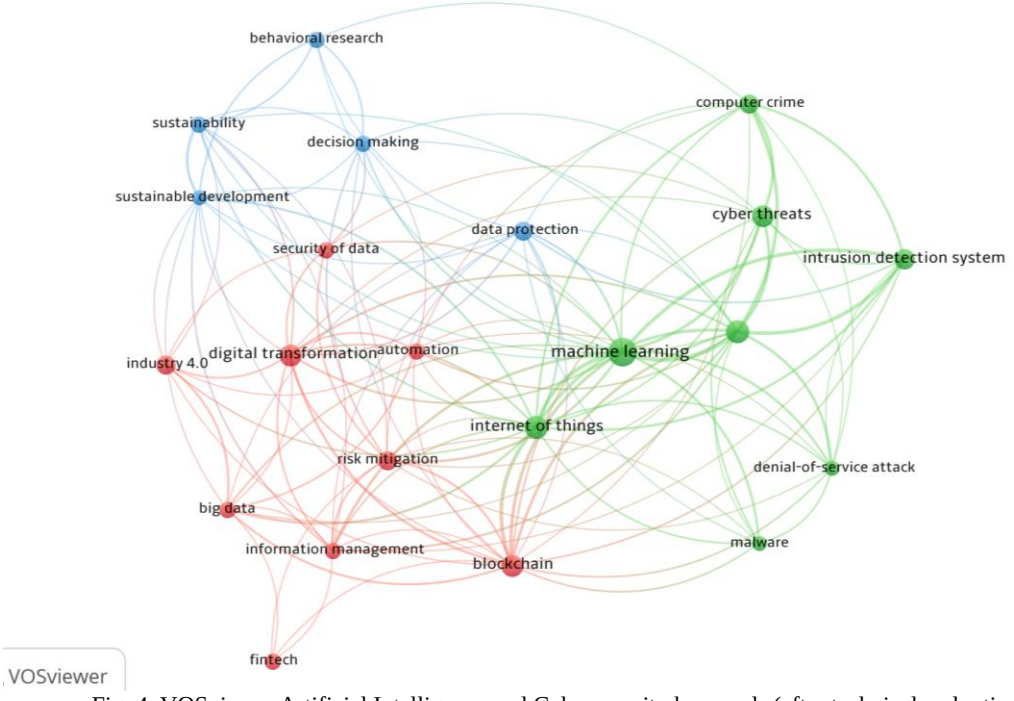

Fig. 4. VOSviewer Artificial Intelligence and Cybersecurity keywords (after technical and rationale methods applied)

Table 3.

Cluster	Keywords
Red Cluster	automation, big data, blockchain, digital transformation, fintech, industry 4.0, information management, risk mitigation, security of data
Green Cluster	computer crime, cyber threats, denial-of-service attack, internet of things, intrusion detection system, machine learning, malware, network security
Blue Cluster	behavioral research, data protection, decision making, sustainability, sustainable development

The red cluster emphasizes the transformative power of technology in industries, particularly fintech, and Industry 4.0. Fintech is reshaping financial services with innovations like blockchain, which ensures secure and transparent transactions. Automation and big data analytics are revolutionizing how financial institutions operate, making processes faster and more efficient. Industry 4.0 incorporates sophisticated technologies like the IoT and robotics into manufacturing, creating smarter factories where machines optimize production through communication. Digital transformation in these

sectors involves more than just adopting new technologies; it also requires changing business models and processes to remain competitive. Managing information effectively and mitigating risks is crucial in this tech-driven landscape, ensuring that data security remains a top priority.

The green cluster focuses on cybersecurity, emphasizing how machine learning and artificial intelligence (AI) enhance security systems. Machine learning algorithms are able to verify extensive amounts of data to find distinctive patterns and potential threats in real time. For instance, AI-powered intrusion detection systems can identify and respond to cyber-attacks faster than traditional methods. These technologies learn from past incidents to improve their accuracy, making it harder for malicious activities to go unnoticed. The IoT is expanding the attack surface, making robust cybersecurity measures even more critical. By leveraging AI and machine learning, cybersecurity systems become more adaptive and proactive, effectively safeguarding against evolving cyber threats such as malware, denial-of-service attacks, and other forms of computer crime.

The blue cluster, delves into sustainability, decision-making, and data protection, presenting a holistic approach to modern challenges. Behavioral research in this cluster examines how people's actions and decisions impact sustainability efforts and data security. Understanding these behaviors helps in designing better policies and practices that promote sustainable development. Data protection is another critical aspect, ensuring that personal and sensitive information is secure from breaches. This cluster also emphasizes the importance of informed decision-making in achieving sustainability goals. Decision-making processes need to be grounded in reliable data and consider long-term environmental, social, and economic impacts. By integrating sustainability into every decision, from policy-making to daily operations, organizations can contribute to a more sustainable future. This approach balances immediate needs with the well-being of future generations, ensuring that development is both responsible and resilient.

The infusion of emerging technologies such as artificial intelligence (AI) and machine learning (ML) facilitates the creation of intelligent systems capable of learning, adapting, and making informed decisions autonomously. This not only augments operational efficiency but also empowers businesses to glean actionable insights from vast troves of data, steering them towards more informed and strategic decision-making.

The integration of artificial intelligence (AI) into communication systems enhances efficiency and personalization. AI-powered chatbots, for instance, facilitate instant responses to routine queries, freeing up human resources to focus on more complex tasks. Machine learning algorithms analyze communication patterns, providing insights into team dynamicArtificial intelligence (AI) injects a layer of sophistication into communication systems. AI-driven chatbots, for instance, offer instantaneous responses to routine queries, enhancing the efficiency of information retrieval. Machine learning algorithms sift through communication patterns, unveiling valuable insights into team dynamics and areas where communication can be refined. The synergy between human intuition and AI precision augments the efficacy of communication processess and identifying areas for improvement in collaboration and workflow [15].

AI-powered platforms, and networks that generate vast amounts of data. Network Security is a critical element and a subfield of cybersecurity and is closely linked to it. The Internet of Things (IoT) is also deeply tied with both digital transformation and cybersecurity, indicating the rapid development and growing concerns related to security challenges presented by IoT devices. Artificial Intelligence (AI) has been a highly researched area in recent years and has the potential to disrupt business practices by enhancing capabilities in technology systems. Its role is becoming increasingly important in the digital economy and in IT practices within the field of cybersecurity.

Artificial intelligence and machine learning technologies play pivotal roles in this dynamic defense landscape, enabling organizations to detect anomalies, predict potential threats, and respond swiftly to emerging cyber risks.

As organizations migrate their operations to cloud environments, the paradigm of cybersecurity undergoes a fundamental shift. Cloud computing introduces a new dimension of flexibility and scalability, allowing companies to scale resources on-demand. However, this flexibility also raises concerns about data sovereignty, regulatory compliance, and the security of shared cloud infrastructure. Cybersecurity strategies must align with the nuances of cloud environments, encompassing robust identity and access management, encryption protocols, and continuous monitoring to ensure the integrity of data stored in the cloud [16].

The human element in cybersecurity remains a critical factor that cannot be overlooked. Employees, often unwittingly, become vectors for cyber threats through social engineering, phishing attacks, or unintentional data breaches. Establishing a culture of cybersecurity awareness becomes imperative. Companies must invest in comprehensive training programs, educating employees about the evolving tactics of cyber adversaries and instilling a sense of responsibility for safeguarding sensitive information.

3.2. Way ahead

Artificial intelligence needs rich and accurate knowledge, just like authentic human experiences. At the same time, they must have a deep understanding of users, including their psychographic characteristics. [17].

I refers to the ability of a machine or computer to mimic the capacities of the human mind, which often learns from past experiences to respond and understand the language, complex problems and decisions. When incorporated into cybersecurity operations, AI is expected to improve vulnerability management and threat detection and accelerate incident response times. More than this, AI could also contribute to easing talent shortage issues in cybersecurity operations. Overall, improved security constituted the main benefit of AI initiatives in enterprises in 2023. Nevertheless, the integration of AI into cybersecurity processes is not without risks as AI, particularly generative AI, can be used to advance adversarial capabilities, such as malware development, phishing and deepfakes [18].

AI is gaining attention in most quadrants of economy and the society, as it can plays a key role in the ongoing digital transformation and impact people's daily lives through its

automated decision-making capacities. AI is also seen as an important enabler of cybersecurity innovation for two main reasons: its ability to respond and detect the cyber threats and the need to secure AI-based applications.

4. Conclusion

As a conclusion, artificial intelligence significantly enhances cybersecurity capabilities through automation, predictive analytics, advanced threat detection and adaptive learning. It also introduces new challenges and demands ongoing management to ensure ethical use and effectiveness. AI has become a driving force for cyber threats, enabling adversaries to launch more effective and sophisticated attacks at a larger scale and faster. Cybersecurity has evolved along with information and technology systems to become an important aspect of the contemporary world; the development of artificial intelligence opens up new ideas for overcoming cybersecurity difficulties [19]. The field of cybersecurity has been very much impacted by artificial intelligence. It is an essential tool for information protection due to its identification' capacity, evaluate, and avoid cyber risks. Artificial intelligence can scan huge volumes of data, anticipate weaknesses and spot anomalies, assisting companies and people in more effectively and quickly defending against attacks. Successful applications of AI in cybersecurity already present the technology's advantages and promise. To overcome opposition in a successful manner and reach a high degree of cybersecurity maturity, it is decisive to be advanced in technology continuously and have abilities and knowledge in this area.

References

[1] B. Hinings, T. Gegenhuber and R. Greenwood, "Digital innovation and transformation: An institutional perspective," *Information and organization*, 2018 March 2018.

[2] J. Nye, "How Will New Cybersecurity Norms Develop?," *Strategist*, 12 March 2018.

[3] T. Riebe and C. Reuter, "Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment," 2019. [Online]. Available: https://link.springer.com/content/pdf/10.1007/978-3-658-25652-4_8.pdf.

[4] R. Kaur, D. Gabrijelčič and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," 2023.

[5] A. Holzinger, G. Langs, K. Zatloukal and H. Müller, "Causability and explainability of artificial intelligence in medicine," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 4, 2019.

[6] R. R. Hoffman, S. T. Mueller, G. Klein and J. Litman, "Metrics for explainable AI: Challenges and prospects," Cornell University, 2018.

[7] S. Kumar, U. Gupta and A. K. Singh, "Artificial Intelligence: Revolutionizing Cyber Security in the Digital Era," *Journal of Computers, Mechanical and Management*, vol. 2, no. 3, pp. 31-42, 2023.

[8] K. Al-Dosari, N. Fetais and M. Kucukvar, "Artificial intelligence and cyber defense system for banking industry: A qualitative study of AI applications and challenges," *Cybernetics and Systems*, 2022.

[9] H. Hirsch-Kreinsen and M. T. Hompel, "Digitalisierung industrieller Arbeit: Entwicklungsperspektiven und Gestaltungsansätze," 2017.

[10] X. Cheng, J. Sun and A. Zarifis, "Artificial intelligence and deep learning in educational technology research and practice," *British Journal of Educational Technology*, 2020.

[11] A. Rindfleisch, M. O'Hern and V. Sachdev, "The Digital Revolution, 3D Printing, and Innovation as Data," *J PROD INNOV MANAG*, 2017.

[12] M. Ettredge, F. Guo and Y. Li, "Trade Secrets and Cybersecurity Breaches," *Journal of Accounting and Public Policy*, vol. 37, no. 6, 2018.

[13] A. Kuzior, P. Brożek, O. Kuzmenko, H. Yarovenko and T. Vasilyeva, "Countering Cybercrime Risks in Financial Institutions: Forecasting Information Trends," *Journal of Risk and Financial Management,* 2022.

[14] D. Teece, "Business models and dynamic capabilities," *Long Range Planning,* 2018.

[15] M. Malatji and A. Tolah, "Artificial intelligence (AI) cybersecurity dimensions: a comprehensive framework for understanding adversarial and offensive AI," *AI Ethics,* 2024.

[16] D. Rammanohar and S. Raghav, "Artificial Intelligence in Cyber Security," *Journal of Physics: Conference Series,* 2021.

[17] C. Vrabie, "De la idee la implementare," *Traseul sinuos al inteligentei artificiale catre maturitate,* vol. 1, 2024.

[18] A. Borgeaud, "Artificial intelligence (AI) in cybersecurity - statistics & facts," [Online]. Available: https://www.statista.com/topics/12001/artificial-intelligence-ai-in-cybersecurity/#topicOverview.

[19] R. Fazley, "Artificial Intelligence in Cyber Security," *SSRN,* 2024.

Inclusion: access & interaction of people with disabilities with the physical & digital environment

Monica-Mihaela FRANGULEA,
Architect PhD, CEO Juxta Foundation
monica_frangulea@yahoo.com

Alexandru LINCA,
IT engineer, Founder & CEO at AdManager.ro
alex@wph.ro

Abstract
It is crucial to ensure that people with disabilities have equal access and interaction opportunities with both physical and digital environments. This includes providing accessible facilities, technologies and platforms that accommodate different types of disabilities, such as visual, hearing, mobility, or cognitive impairments. The issue of making it easier for people with disabilities to access and interact with their environment: living spaces, working spaces and public spaces has been a major priority for designers, architects, engineers and even society in general for many years. Disability is the experience of any condition that makes it more difficult for a person to do certain activities or have equitable access within a given society. Disabilities may be cognitive, developmental, intellectual, mental, physical, sensory, or a combination of multiple factors. For every single type of disability we will see that there is a different set of measures that need to be taken into consideration when designing accessible tools, furniture, transportation, communication means, spaces, a.s.o. We have created ways to communicate with people that are unable to speak or hear, we are enabling individuals with impaired mobility to access all types of structures and pathways, to drive cars and manipulate complex machinery. In this context, the access to the digital world becomes more and more a necessity for any individual in our society in order to be able to communicate, to work, to get access to information, private and public services, a.s.o. By creating inclusive environments, we can promote independence, empower individuals, and enhance their overall quality of life.

Keywords: social inclusion, disabilities, accessibility, handicap, digital service access, digital skills.

1. Introduction
The issue of making it easier for people with disabilities to access and interact with their environment: living spaces, working spaces and public spaces has been a major priority for designers, architects, engineers and even society in general for many years.

To a certain degree, physical impairments and changing mental states are almost ubiquitously experienced by people as they age. Aging populations are often stigmatized for having a high prevalence of disability.

To the number of persons with disabilities from birth or acquired during a person's lifetime we will have to continously add a big procentage of the aging population as well, as it is for certain that getting older, we will all be disabled eventually.

For every single type of disability we will see that there is a different set of measures that need to be taken into consideration when designing accessible tools, furniture, transportation, communication means, spaces, a.s.o.

We have created ways to communicate with people that are unable to speak or hear, we are enabling individuals with impaired mobility to access all types of structures and pathways, to drive cars and manipulate complex machinery.

In this context, the access to the digital world becomes more and more a necessity for any individual in our society in order to be able to communicate, to work, to get access to information, private and public services, a.s.o.

2. Basic physical mobility

In 1936 the German architect Ernst Neufert published the well-known work *Architects' data*, which contains standardisation rules for all elements of the built space, rules that are taught in all schools of architecture in the world, are respected, applied and constantly optimised.

First published in 1936, its 39 German editions and translations into 18 languages have sold over 500,000 copies worldwide [1].

This manual is an essential reference for the initial design and planning of a building project. It provides, in one single volume, the core information needed to form the framework for the more detailed design and planning of any building project. Organised largely by building and space type, it covers the full range of preliminary considerations, and with over 6200 diagrams it provides a mass of data on spatial requirements.

Architects all over the world are able to design the same way spaces, pathways and furnished environment in order to be easy accessable to people with mobility difficulties.

To improve accessibility for mobility disabled people in the city, a comprehensive approach can be taken. This includes installing ramps and elevators in public buildings, pedestrian crossings, and transportation hubs, as well as providing accessible parking spaces. Additionally, cities can implement audio signals at pedestrian crossings, tactile markings on sidewalks, and audible announcements at train stations to help visually impaired individuals navigate. Furthermore, cities can also promote the use of assistive technologies such as wheelchairs, scooters, and mobility aids, and provide training for emergency responders and healthcare professionals to ensure they can effectively respond to the needs of mobility disabled individuals.

Assistive technology is a generic term for devices and modifications (for a person or within a society) that help overcome or remove a disability. The first recorded example of the use of a prosthesis dates to at least 1800 BC. The wheelchair dates from the XVII-th century [2].

Fig. 1. Concept of futuristic mechanical exoskeletons for mobility-disabled persons created with A.I. assisted software
Source: Monica-Mihaela Frangulea

But mobility difficulty is just one type of disability.

Disabilities may be cognitive, developmental, intellectual, mental, physical, sensory, or a combination of multiple factors [3].

3. Visual impairments

Impaired vision can range from poor vision to blindness. Visual disability requires a totally different set of measures in order to facilitate access to the physical environment than in the case of other sensory or physical impairments.

A brilliant inventor, Louis Braille have developed a system of tactile code that could allow blind people to read and write quickly and efficiently. He presented his work to his peers for the first time in 1824, when he was fifteen years old, opening the possibility of reading to blind people and set the fundamentals visual impaired people access systems in our society.

Louis Braille's system of embossed type is now used by blind and partially sighted people for reading and writing all over the world and it has been adapted to almost every known language [4].

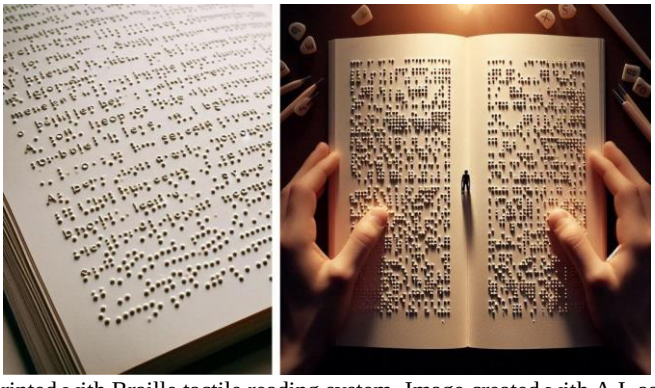

Fig. 2. Books printed with Braille tactile reading system. Image created with A.I. assisted software
Source: Alexandru Linca

Today we are developing much more advanced technology to assist people with this type of condition, such as screen readers and print reading devices, which convert text into spoken words.

To address the challenges faced visual impaired individuals in urban spaces, various access systems have been developed and implemented. One such system is the installation of tactile paving on sidewalks and public spaces, providing tactile cues to guide visually impaired individuals safely. Additionally, audible signals at pedestrian crossings have been introduced to assist blind people in safely navigating busy intersections. Moreover, the development of mobile applications specifically designed for visual impaired people has revolutionized the way they access information and navigate urban environments. These applications provide real-time navigation assistance, information on nearby points of interest, and even indoor navigation support in complex buildings such as shopping malls or airports.

Enhancing access systems for blind individuals in urban areas has a profound impact on their independence, safety, and overall quality of life. By improving accessibility features such as tactile paving and audible signals, blind individuals can navigate urban spaces with greater ease and confidence, leading to increased independence and mobility. Furthermore, enhancing access systems contributes to the safety and security of blind individuals in urban environments, reducing the risk of accidents and improving their overall well-being. Ultimately, promoting inclusivity and equal opportunities for all individuals in urban spaces, including those with visual impairments, is essential for creating a more accessible and welcoming environment for everyone.

In conclusion, the challenges faced by blind individuals in urban spaces necessitate the development and enhancement of access systems tailored to their specific needs. By incorporating tactile paving, audible signals, and mobile applications, urban environments can become more inclusive and accessible for individuals with visual impairments. The impact of these enhanced access systems extends far beyond the individual level, contributing to the safety, independence, and inclusivity of urban communities as a whole. It is imperative that urban planners, policymakers, and communities work together to ensure that access systems in urban spaces are designed with the needs of all individuals in mind, fostering a more inclusive and equitable urban environment for everyone.

Fig. 3. Concept of automatic stairs access to buildings from the street for mobility-disabled persons created with A.I. assisted software
Source: Monica-Mihaela Frangulea

In April this year the students at the University of Architecture and Urban Planning „Ion Mincu" in Bucharest were given the task of testing in real life action the accessibility of interior public spaces by trying to access the university lobby and the other surrounding areas wearing blindfolds. Such an exercise brings awareness to the future architects of how difficult such a condition is (visual impairment) and what they have to take into consideration when designing interior and exterior structures and spaces.

Fig. 4. Students wearing blindfolds in order to test the accessibility of the lobby space of the University of Architecture and Urban Planning „Ion Mincu" in Bucharest
Source: Monica-Mihaela Frangulea

We also have to mention the major role the urban public lighting is playing in our society. During night time, without artificial lighting to highlight the environment, everybody is blind!

Without artificial light we would not be able to move around at night, to find our way around and to navigate urban routes - streets, pavements, alleys, paths, parks, squares, a.s.o.

A proper amount and distribution of artificial light in the urban areas as well as in the interior spaces insures the safety and security for everybody at the level of urban circulation and functionality, because in addition to the street lighting strictly necessary for this safety, there is also a need for visible urban landmarks on a larger scale (buildings, monuments, landmarks), in order to facilitate the correct perception of the position the observer is having in the urban environment, can calculate the correct path and distance to its destination or just be aware of the location the observer is at all times.

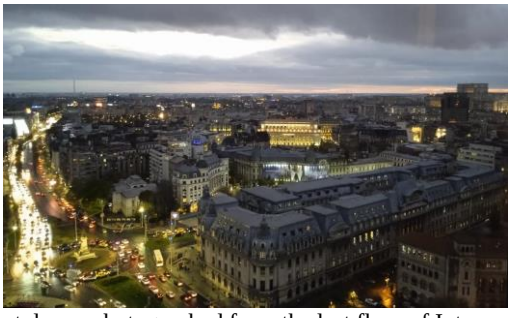

Fig. 5. University Square at dawn, photographed from the last floor of Intercontinental Hotel, Bucharest
Source: Monica-Mihaela Frangulea

4. Hearing impairment
By 2050, nearly 2.5 billion people are projected to have some degree of hearing loss, and at least 700 million will require hearing rehabilitation.

Over 1 billion young adults are at risk of permanent, avoidable hearing loss due to unsafe listening practices [5].

Along with the efforts to promoting safe listening to reduce the risk of recreational noise-induced hearing loss, innovation is taking also steps in creating new technologies to help people with hearing disabilities.

Hearing impaired individuals in urban areas grapple with limited access to auditory information and communication, which are essential components of daily life. The absence of sound cues can hinder their ability to perceive important announcements, warnings, or instructions in public spaces. Additionally, navigating public transportation systems poses a significant challenge for deaf persons due to the reliance on auditory cues for boarding announcements and route information. This can lead to feelings of isolation, frustration, and even safety concerns. Furthermore, accessing public services and facilities, such as government offices, healthcare facilities, and educational institutions, can be daunting for those persons without adequate support or accommodations in place. To enhance accessibility for hearing impaired individuals in urban spaces, cities can implement various systems and technologies. For instance, public transportation systems can be equipped with visual indicators, such as LED displays or audio-to-text systems, to provide real-time information to passengers. Additionally, auditory signals can be replaced with visual alerts, such as flashing lights or vibrating seats, to notify individuals of arrivals, departures, and stops. Furthermore, city squares and public spaces can be designed with visual cues, such as Braille signage and audio loops, to facilitate navigation and communication.

5. Invisible disability
Invisible disabilities, also known as Hidden Disabilities or Non-visible Disabilities (NVD), are disabilities that are not immediately apparent, or can be noticed. They are often chronic illnesses and conditions that significantly impair normal activities of daily living. Invisible disabilities can hinder a person's efforts to go to school, work, socialize, and more. Some examples of invisible disabilities include intellectual disabilities, autism spectrum disorder, attention deficit hyperactivity disorder, fibromyalgia, mental disorders, asthma, epilepsy, allergies, migraines, arthritis, and chronic fatigue syndrome.

6. Digital access for persons with disabilities
Many of these community members face communication challenges and while technology has become more ubiquitous in people’s lives, those with such challenges face a digital divide that is present due to a lack of accessibility considerations within the digital ecosystem. Affordability and accessibility of technology products as well as digital literacy are the main barriers affecting their digital access and inclusion [6].

In this context, access to the digital world is becoming more and more a necessity for any individual in our society in order to be able to communicate, to work, to get access to information, private and public services, a.s.o.

Accessibility systems to digital platforms are crucial for people with disabilities, enabling them to fully participate in the digital world. These systems include features such as text-to-speech, screen reader software, and closed captions, which provide equal access to information and opportunities for individuals with visual, hearing, motor, or cognitive impairments. Additionally, accessibility features like keyboard-only navigation, high contrast modes, font size adjustments, screen readers, speech recognition software, alternative input devices, and captioning tools are just a few examples of the innovative technologies that have been developed to improve accessibility and can help individuals with mobility or visual impairments navigate and interact with digital platforms with ease.

Companies like Apple and Microsoft have also integrated built-in accessibility features into their operating systems, making it easier for users to customize their digital experience according to their needs.

For example, Apple has recently launched the “Eye Tracking” system powered by artificial intelligence, that gives users a built-in option for navigating iPad and iPhone with just their eyes. Designed for users with physical disabilities, Eye Tracking uses the front-facing camera to set up and calibrate in seconds, and with on-device machine learning, all data used to set up and control this feature is kept securely on device, and isn’t shared with Apple.

The same company has developed the “Music Haptics” system as a new way for users with hearing disability to experience music on iPhone. With this tactile accessibility feature turned on, the Taptic Engine in iPhone plays taps, textures, and refined vibrations to the audio of the music.

A range of “Vocal Shortcuts" was also developed, giving users an option for enhancing speech recognition for a wider range of speech. Designed for users with acquired or progressive conditions that affect speech, such as cerebral palsy, amyotrophic lateral sclerosis (ALS), or stroke, these features provide a new level of access, customization and control, building on features introduced in iOS 17 for users who are non speaking or at risk of losing their ability to speak [7].

The work on accessibility at MSR India has spanned the range of new access systems from spatial audio with HoloLens to the use of feature phones to reach children with vision impairments and a spectrum of tangible toys to enhance numeracy for them, to a quiz platform for the Deaf or Hard of Hearing community, with an overarching new methodology called “Ludic Design for Accessibility”.

Microsoft Ludic Design for Accessibility is a novel methodology that puts playfulness at the center of any design for accessibility. The key idea was that a solution for accessibility designed with this methodology will be in the form of an engaging and inclusive game. By

extended and joyful play with the game the players can acquire the designed in skills purely as a side effect [8].

The same company has developed SEEDS (Scalable educational experiences with digital scaffolding), a project that builds on top of the work over the past four years with Vision Empower Trust, a DPO that has now reached about one hundred schools for the blind across a dozen states of India, codesigning the solutions with Microsoft and taking them to the end users. The goal of the project is to introduce digital technologies to children in schools for the blind from the primary stage onwards but the long-term vision is to reach every one of the estimated 1–2 million children with vision impairments in India [9].

By incorporating these accessibility features, digital platforms can promote inclusivity, diversity, and social equality, ensuring that everyone has equal access to education, employment, and social opportunities.

7. Impact of Accessibility on Security for People with Disabilities

The integration of accessibility measures for individuals with disabilities into both physical and digital environments significantly impacts security, fostering a safer and more inclusive society. Ensuring accessibility not only enhances the independence and quality of life for people with disabilities but also contributes to overall societal security in various ways:

Physical Security: Technologies such as advanced driver-assistance systems (ADAS) in cars, which include features like lane departure warnings and adaptive cruise control, support drivers with disabilities and improve overall road safety. By assisting drivers in maintaining control and awareness, these systems reduce the risk of accidents, contributing to safer road conditions for all users.

Cybersecurity: The increasing reliance on digital assistive technologies, such as smart home devices and wearables for health monitoring, necessitates robust cybersecurity measures. Protecting these devices from cyber threats ensures that people with disabilities can rely on them without compromising their personal safety and privacy.

Fig. 6. Abstract images of Multi-Factor Authentication for people with disabilities.
Image created with A.I. assisted software
Source: Alexandru Linca

Ensuring that Multi-Factor Authentication (MFA) options are accessible to all users, including those who may have motor or cognitive difficulties, is crucial. For example, using mobile authentication apps or physical tokens can be adapted to meet diverse needs.

In conclusion, the integration of accessibility measures for people with disabilities profoundly enhances security across physical, digital, and social dimensions. By fostering an inclusive environment, society not only supports the independence and well-being of individuals with disabilities but also enhances overall safety and resilience, benefiting all members of the community.

References

[1] „Ernst Neufert," [Interactiv]. Available: https://architectuul.com/architect/ernst-neufert. [Accesat 15 May 2024].

[2] R. A. Cooper, Hisaichi Ohnabe și Douglas A. Hobson, în *An Introduction to Rehabilitation Engineering*, CRC Press, 2006, p. 131.

[3] L. Francis și A. Silvers, „Perspectives on the Meaning of "Disability"," *AMA Journal of Ethics*, vol. 18, nr. 10, pp. 1025-1033, 2016.

[4] Royal National Institute of Blind People (RNIB), *Invention of Braille*, 2017.

[5] Worlds Health Organisation, „Deafness and hearing loss," Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss#:~:text=A%20person%20who%20is%20not,conversational%20speech%20or%20loud%20sounds. [Accesat 18 May 2024].

[6] The UN RefugeeAgency (UNHCR), „Digital Access and Inclusion of People with Disabilities," [Interactiv]. Available: https://www.unhcr.org/innovation/wp-content/uploads/2021/03/Digital-Access-and-Inclusion-of-People-with-Disabilities.pdf. [Accesat 19 May 2024].

[7] Newsroom, Press Release, „Apple announces new accessibility features, including Eye Tracking, Music Haptics, and Vocal Shortcuts," Available: https://www.apple.com/cm/newsroom/2024/05/apple-announces-new-accessibility-features-including-eye-tracking/. [Accesat 2 June 2024].

[8] Microsoft, „Accessibility and Assistive Technology," Available: https://www.microsoft.com/en-us/research/project/accessibility-and-assistive-technology/. [Accesat 2 June 2024].

[9] Microsoft, „Scalable Early Education with Digital Scaffolding (SEEDS)," Available: https://www.microsoft.com/en-us/research/project/scalable-early-education-with-digital-scaffolding-seeds/. [Accesat 3 June 2024].

The education system, the way to fight fake news

Andreea Florentina RADU,
Babes Bolyai University Cluj Napoca, Romania
andre.radu@gmail.com

Ioana PETCU,
National Institute for Research and Development in Informatics – ICI Bucharest, Romania
ioana.petcu@ici.ro

Abstract
To form broad and sound opinions on various issues, we must be able to access a large amount of information, including online. However, given the increasing prevalence of fake news and disinformation, the information received must be reliable and verifiable. Recognizing questionable or malicious content can be very challenging in today's interconnected world, especially for youngsters. For these reasons, sociologists and teaching staff have voiced the need for policymakers to combat the spread of this phenomenon among youngsters and so, they came together in testing and, later on, in gradually introducing, in the education system, lectures the phenomenon of fake news and disinformation. Lectures increasingly incorporate debunking and digital literacy and are mostly focused on raising awareness of the dangers generated by fake news and disinformation among pupils and students and on learning various techniques for identifying and thus stopping the spread of this type of malicious information. The present article maps the real and potential impact of fake news and disinformation among pupils and students and their healthy mental and social evolution. Therefore, it highlights some recommendations meant to prevent the spread of this phenomenon from the early stages of the individual's evolution, respectively starting in the classroom. The key message of the article could be considered that education is an effective weapon to combat fake news and disinformation.

Keywords: new trends, information sabotage, misinformation, disinformation, digital literacy.

1. Introduction
False information, also known as rumor, deceit, distraction, and propaganda, has been a persistent issue since the inception of news. Technological advancements and the widespread reach of mass communication platforms have led to the widespread dissemination of misinformation. The emergence of digital 'fake news' presents a significant threat due to its increased accessibility for creation, dissemination, and consumption, extending beyond traditional print media to digital channels.

The advent of digital communications technologies has facilitated the extensive creation, dissemination, and consumption of fabricated news and disinformation, hence posing challenges in discerning between genuine and deceptive information.

1.1. Fake news - features and challenges
Social networks provide a conducive setting for the widespread creation and distribution of false information. Two possible reasons for this phenomenon are the rapid dissemination of content and the emergence of specialized areas of thinking.

Social media platforms such as Facebook, Instagram or Twitter are frequently mentioned as enabling the dissemination of false information, highlighting its swift propagation and possible harm [1]. Entities involved in creating of false information encompass corporations, government entities, and people, each driven by various motives to spread

such content. For instance, the impact of false information on voters during election campaigns, such as the 2018 Brazilian elections, the 2019 Romanian elections, or the UK's BREXIT referendum, which led to the country's departure from the European Union and not last, in the field of public health - by disseminating false information about the effectiveness and probable risks of anti-Covid vaccines.

Regrettably, and quite concerning, both informed and uninformed audience members are frequently subjected to misleading material, such as fake news and disinformation. This can doubt the beliefs and choices of consumers of such content, including scholars and students.

1.2. The impact of fake news among teenagers

While acknowledging the significance of the issue of false news, both students and teachers find that the current education system cannot effectively address this problem. This is mostly due to the heavy workload on teachers and a lack of professional assistance and experience. School leaders, instructors, and education professionals are inadequately equipped to handle the issue of fake news in schools. According to studies, the majority of teachers believe that secondary school children are exposed to fake news to a greater extent than adults and the impact goes much behind a greater extent. Therefore, their skills, in identifying fake news (e.g.: critical thinking, media literacy or the ability to analyze sources) should be strengthened and developed to increase their immunity to fake news.

The exposure of primary, secondary and high-school students to fake news presents new challenges for the education system, which is deeper and more serious than many might think.

Teenagers are overwhelmed with information from diverse channels, including educators, relatives, peers, advertisements, television shows, and the Internet. This material may possess varied levels of accuracy, objectivity, and ethical legitimacy, but it can also be highly erroneous and biased.

A poll conducted at Stanford University revealed that young individuals exhibit more susceptibility to misinformation, making them more susceptible to manipulation through the dissemination of fake news, paid content, and biased "expert opinions", often without their awareness.

1.3. The early fight against fake news at the high school education system level

Education is crucial in shaping the future and represents our intentions for the preparation of future generations. The type of education we currently possess will greatly influence the future we will experience.

The primary objective of the educational system is to provide young individuals with the necessary skills and knowledge to actively participate in society and take on responsibilities with a sense of commitment. In the absence of this, intricate and diverse cultures that are rapidly progressing cannot endure.

Students must cultivate a robust sense of skepticism in order to discern falsehoods and proficiently leverage internet information in a multifaceted society. The primary objective is to cultivate a populace proficient in reading, enabling them to comprehend information efficiently within the transformed informative environment, hence impeding the swift dissemination of falsehoods and guaranteeing protection.

A best practice example is Finland, Europe's most resilient nation against fake news, as can be seen in the figure below, which is actively addressing the global challenge of false information including by incorporating it into the curriculum of primary schools, as part of a broader national strategy.

In this regard, unfortunately, Romania is among the European states with a very low rate of media literacy, occupying the penultimate place before Bulgaria, according to The European Media Literacy Index 2023 [2] - an instrument for assessing and ranking societies in their potential for resilience in the face of "fake news" phenomena.

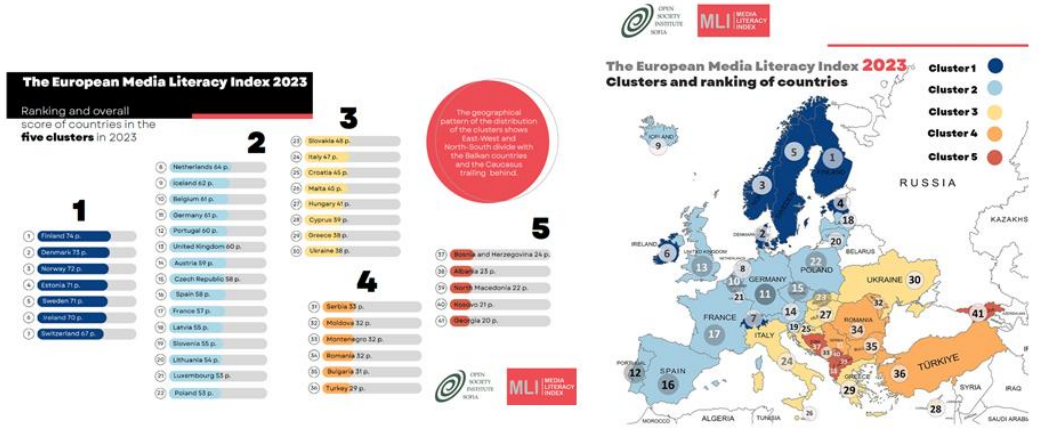

Fig. 1. The European media Literacy Index 2023 [2]
Source: Marin Lessenski, Media Literacy Index 2023 report entitled: Bye, Bye Birdie: The Challenges of Disinformation, 2024

According to The Media Literacy Index 2023, the top-performing countries in terms of media literacy in the first cluster are primarily in Northern and Western Europe, including Scandinavian nations, Estonia, and Ireland (Fig. 1).

The second cluster covers Western and Central Europe, while the third includes Southern and Central Europe, including Ukraine and Greece.

The fourth cluster includes Romania, Serbia, Bulgaria, Turkey, and Moldova. So, the 2023 statistics indicate a notable disparity in media literacy, with the Balkan countries and Caucasus region lagging. These countries are close to the Russia-led conflict in Ukraine and are more vulnerable to malicious disinformation. The Balkan countries, which continue to experience internal instability and tensions between states, are ranked among the countries with the highest risk in the index.

Thus, after comprehensive research on the topic, we strongly believe that the early fight against fake news starts from still high school. Therefore, there is a need for an educational system that aims to train individuals to recognize false information. Furthermore, we identified the necessity for a "pre-regulation education" strategy towards addressing the challenges of fake news.

According to our research, potential educational approaches to combat fake news include media literacy, news literacy, and information literacy [3]. Generally, these literacies seek to enhance an individual's interaction with the media by imparting information that can encourage responsible consumption of various media.

2. Tools used in fighting fake news

2.1. Media literacy

Experts from diverse disciplines are collaborating to counteract the dissemination of false information by integrating fact verification into the practice of journalism. This technique evaluates the truthfulness of news or statements made by public personalities on the internet. Media literacy is a multidisciplinary field that centers on educating people about the role of media in society. It entails developing an independent connection with various forms of media, analyzing them thoughtfully, and recognizing that content is created within a particular framework. News literacy is a specialized form of media literacy that focuses specifically on journalistic media [4]. Information literacy highlights the fact that information can be examined from multiple perspectives, resulting in diverse interpretations and conclusions. This strategy facilitates the development of an independent mindset towards the media and its impact on society.

Literacy skills are intricately linked to the Common National Curriculum Base, which governs the curricula in both public and private schools in Brazil. The curriculum places a strong emphasis on cultivating critical thinking skills grounded on factual information, as well as the significance of pupils analyzing material from many sources, with a special focus on internet-based resources. Engaging in literacy practice entails employing lateral reading, which means expanding the reading process beyond relying solely on information from a single source. This method facilitates the identification of significant metadata, such as the individual accountable for the publishing and editorial preferences, and it also allows for the comparison of the source with other media platforms. This strategy differs from vertical reading, which exclusively concentrates on the original text and takes into account visual elements. Lateral reading entails the assessment of material obtained from the primary website, completing further research, and comparing the source with other sources to evaluate the credibility and any biases. This strategy enables inquiries regarding the veracity of material on professional websites and the dependability of sources.

2.2. The "Mapping" method

Determinig who is behind the news

- Research on the author of the news: By emphasizing this information, the user will be motivated to verify if the individual responsible for the writing is an established expert in the field of journalism and, subsequently, to explore the professional's potential personal inclinations.

- Research to determine the entity responsible for the domain hosting the news. If possible, utilize the information about the individual or legal entity to assess the trustworthiness and potential ideological biases of the news outlet.
- Conducting a Wikipedia search to gather information about the website where the news is published: While Wikipedia is often seen as an untrustworthy source, particularly for academic research, it can be used as an initial reference for verifying facts, especially when seeking information on the source of news articles.

Evidence checking for the news content

- Fact-Checking Websites

The Google Search API (Googio) enables unrestricted searches on Google's search engines, delivering links, descriptions, and websites. Users can personalize parameters such as language, country, and geographic region. The API additionally provides access to searches utilizing Google Images, Google News, and Google Scholar.

Fact Check Explorer is a tool developed by Google that assists journalists and researchers in analyzing news articles that have been verified by fact-checking websites. It offers the latest findings released by various websites, enabling users to assess the accuracy of a specific topic. The sample space of Fact Check Explorer encompasses Fact Check Agencies that adhere to Google's requirements, which encompass data layout, correction processes, and transparency regarding sources and techniques.

Two notable fact-checking websites are AFP fact-checking (https://factcheck.afp.com/), which is managed by the international media organization France-Presse Agency, and Reuters Fact Check (https://www.reuters.com/fact-check/), which is maintained by the multinational media firm Reuters.

- Publication date

The phenomenon of old, out-of-context articles being disseminated on social networks to spread disinformation is widespread [5]. Furthermore, negligent users frequently fail to allocate sufficient attention to these dates. Hence, in certain instances, highlighting the publication date is enough to demonstrate the utilization of outdated material for misinformation intentions.

- Checking behind the website

It is essential to verify the individual or legal body who registered the website where the news is hosted. When a domain name is acquired from a specific provider, the purchaser's information is transmitted to ICANN (Internet Corporation for Assigned Names and Numbers), a non-profit organization that oversees the DNS protocol (Domain Name System) and other related tasks.

ICANN manages the WHOIS system, which allows public access to certain data on these names. For instance, website.informer.com offers a concise and comprehensive overview of various information available on the internet regarding a specific website or domain. This includes details such as the number of daily visitors, safety status, Alexa rank, owners, and other relevant data.

2.3. Recommendations for different types of fake news education initiatives

Many government and educational institutions worldwide are confronted with the task of incorporating supplementary curriculum content into existing curricula as a tool to address the issue of false news. The process is complex and time-consuming, but worth implementing we consider. Students must develop the ability to critically evaluate information and effectively utilize online resources to thrive in an ever more intricate society.

Summarizing what we mentioned earlier in this paper, we recommend that those involved in developing and implementing education policies keep in mind the purpose of pedagogical working with different stage students, namely to provide them with the necessary information in the field of fake news identification and in the same time assist them in addressing three primary inquiries: ”who is responsible for the news?”, „what evidence supports the statements?” and "What have other public open sources have to say about the same news/topic”.

As stated previously in this paper, more advanced European countries in terms of media literacy have targeted programs included in the curricula for the first two years of elementary school, where students get involved in various literacy practices, such as lateral reading, a technique which enhances individuals' ability to critically analyze material they come across on the internet.

Some potential educational projects to consider implementing in the elementary, secondary, and high school system include: Collaborating with colleagues and professionals in the sector to exchange experiences; offering educators easily available materials for immediate use in their teaching; developing educational materials for kids in elementary, middle, and high school that focus on the subject of disinformation and ensuring that they are easily accessible; offering easily accessible and highly effective methods, facilitating cooperative endeavors focused on countering misinformation and associated concerns, etc.

Fighting against fake news equals fighting inoculation and students should learn to develop a critical way of thinking to build active resistance, capable of protecting them against disinformation.

The initiatives vary among various educational [6] and cultural systems, but they mainly refer to some constants: courses combined with applied learning on media literacy and technological/digital tools, optional lectures based on gamification, and diverse educational practices.

More specific policy recommendations, as advised by pedagogical experts, envisage:

- Incorporating media literacy and fake news education into the curriculum during early childhood. Although there are individual efforts, they are typically limited in scope and only impact a small number of kids. Schools should form collaborative alliances with other entities and specialists to cultivate novel strategies and proficiencies in combating misinformation. Education leaders and institutional leaders should promote collaboration with these external actors to develop a comprehensive strategy for combating misinformation.
- To guarantee comprehensive intervention across the entire system, teachers must undergo extensive training on the subject of false news. This training should include the integration of essential skills into existing training programmes, in collaboration with non-governmental organizations (NGOs) and experts. Enhancing teachers' capabilities and alleviating their workload is of utmost importance. The curriculum reform should prioritize the cultivation of practical life skills and the development of social competencies necessary for active engagement in social and public spheres. These competencies include fostering a culture of discussion, empathy, media literacy, resilience against misinformation, and critical thinking. This initiative aims to enhance teachers' social competencies, hence fostering a more inclusive education system.
- Developing age-appropriate educational resources for different areas, such as media studies, business, lifestyle, history, citizenship knowledge, and foreign language schools, to tackle fake news and misinformation. Teachers should be provided with readily available resources such as pre-made materials, worksheets, tools, toolsets, and class plans. Utilizing interactive techniques, such as employing analogous programmes in different subjects, can aid in the cultivation of skills about counterfeit information. Creating a platform for the exchange of experiences and sharing of teaching resources on the topic of false news could serve as an engaging and beneficial tool.
- Classroom sessions should prioritize discussing subjects that captivate pupils, such as misinformation about individuals, actions, or influential figures [7]. This activity aids in the cultivation of aptitudes for discerning and validating misinformation. Enhancing students' abilities, such as their engagement with current affairs, analytical reasoning, understanding of media, capacity for empathy, teamwork, communication skills, and fostering a culture of discussion, can bolster their resistance to misinformation, heighten their recognition of potential dangers, and encourage their active participation in the community.
- Successful media instruction necessitates strong cooperation between families and schools. Parents need to participate in training programmes that specifically target their cognitive abilities and talents. Programmes should foster collaboration among parents, teachers, and students, while also offering support to both homeroom and subject-matter teachers.

3. Conclusion

In a society filled with multifaceted challenges, false information, including rumor, deceit, distraction, and propaganda, has become a significant issue due to technological advancements and mass communication platforms. Digital 'fake news' presents a threat due

to its accessibility and rapid dissemination. Social networks, corporations, government entities, and individuals contribute to the spread of misleading content, posing challenges for informed and uninformed audiences. Fake news affects teenagers, but also the education system struggles to address it due to big workloads and lack of professional assistance. Secondary school children are exposed to more fake news than adults, requiring the strengthening of critical thinking and media literacy skills. The early fight against fake news [7] should start at the pre-university education system level. Education is crucial for shaping the future and preparing future generations. It aims to provide young individuals with skills to participate in society and take on responsibilities. Finland is a resilient nation against fake news, incorporating it into primary school curriculums. However, Romania has a low media literacy rate, with Balkan countries and Caucasus regions lagging.

Research suggests pre-university education is crucial in combating fake news, with potential approaches including media literacy, news literacy, and information literacy aimed at promoting responsible media consumption. Responsible authorities in the field and educational institutions must incorporate supplementary curriculum content to address false news, which enables the young generation to develop critical evaluation skills and utilize online resources to thrive in complex societies. Some of the tools we have recommended in identifying fake news at the level of primary school, secondary school, and high school level, include media literacy and mapping, which refers to checking who is behind the news and checking the evidence for the news. Furthermore, potential educational projects include collaborations, accessible materials, and cooperative efforts to counter disinformation. Initiatives include courses, gamification, and diverse practices.

It is recommended the examination of successful methods of other states that have already implemented similar measures, yielding excellent outcomes. These practices should be tailored to the Romanian education system and the unique issues faced by the new generation in the current context of social-economic and political.

References

[1] M. Aldwairi și A. Alwahedi, „Detecting fake news in social media networks," *Procedia Computer Science,* vol. 141, pp. 215-222, 2018.

[2] Marin Lessenski, Open Society Institute – Sofia, *Media Literacy Index 2023 report entitled: Bye, Bye Birdie: The Challenges of Disinformation,* 2024.

[3] A. F. Radu și I. Petcu, „Digital technologies used to combat disinformation and fake news," *Romanian Cyber Security Journal,* vol. 6, nr. 1, 2024.

[4] Farmer Lesley, „News Literacy and Fake News Curriculum: School Librarian Perceptions of Pedagogical Practices," *Open Information Science,* nr. 3, pp. 222-234, 2019.

[5] C. Ireton și J. Posetti, „UNESCO's handbook for journalism education," *Journalism, Fake News and Disinformation,* 2018.

[6] Luan Imeri, „The role of education in the fight against fake new," *Project Fighting False News Narratives,* 2018.

[7] Pappas Stephanie, „Fighting fake news in the classroom," vol. 53, nr. 1, p. 87, 2022.

The role of the Interreg Programmes in strengthening cyber security within the regions of the Republic of Moldova

Anatolie BABIN,
MBA, Academy of Economic Studies of Moldova
anatolii.babin@ase.md

Sergiu TUTUNARU,
Academy of Economic Studies of Moldova
tutunaru@ase.md

Ion COVALENCO,
Academy of Economic Studies of Moldova
covalenco@ase.md

Abstract
To explore the role of Interreg programmes in strengthening cyber security in the regions of Moldova, which is an increasingly important aspect of national security, especially in regions prone to cyber threats. The work is based on the authors' recent results published in international publications in previous years. The research is based on innovative concepts of digital economy aimed at developing smart villages. Drawing on European experience, accumulated good practices and case studies, the study highlights the importance of international cooperation and capacity building initiatives to reduce cyber risks and build cyber defence capabilities in Moldova. The results of the study highlight the importance of a framework for cross-border cooperation to promote a holistic approach to cybersecurity, integrate best practices and facilitate the sharing of expertise and resources between neighbouring countries. In the context of smart specialisation approaches, digital innovations in rural areas will reach consumers – local action groups – faster, creating synergies between national funds and international programmes at the regional level. Recommendations are made for policy makers, practitioners and stakeholders to further leverage European cross-border programmes to improve cybersecurity resilience and ensure a more secure digital environment in the regions. Implications: academics, researchers, practitioners, EU Commission, JRC. The value of this paper is derived from its examination of the influence of transregional programmes on the advancement of cyber resilience, the facilitation of information sharing, and the stimulation of interregional collaboration. Furthermore, it offers recommendations based on the analysed best practices for addressing cyber security challenges in the context of the specific development of the Republic of Moldova.

Keywords: cybersecurity, interreg, cross-border security operations centers.

1. Introduction
In the context of accelerating digitalization and the growing number and impact of cybersecurity incidents, the European Commission adopted the “EU Cybersecurity Strategy for the Digital Decade” in December 2020. Among other goals, the Cybersecurity Strategy aims to improve capacity and collaboration in detecting cyber threats before they can cause large-scale damage, in order to detect more threats and do so much faster. Russia's incursion into Ukraine further highlights and reinforces the need to urgently build cybersecurity capabilities at the national and Union levels, including through increased information sharing and improved detection of cybersecurity threats to help improve situational awareness and inform preventive and responsive actions.

Cross-border cooperation promotes integrated regional development between neighboring regions having sea and land borders in two or more Member States, or between neighboring regions in, at least, one Member State, and a third country on the external borders of the

Union other than those concerned by the Programs in the field of the Union's external financing instruments [1]. Interreg VI (2021–2027) covers all 27 EU Member States, six accession countries and thirteen neighbouring countries. The participation of Russia in Interreg programmes is suspended following Russia's aggression against Ukraine [2].

The EU Cyber Security Strategy proposes to establish, strengthen and integrate security operations centers (SOCs) and cyber threat intelligence (CTI) capabilities (monitoring, detection and analysis) across the European Union (EU) to support the detection, prevention of cyber threats and timely warning to authorities and all relevant stakeholders' sides. Such cybersecurity capabilities are typically provided by security operations centers (SOCs) of public and private organizations in combination with computer emergency response teams/computer security incident response teams (CERT/CSIRT), supported by external specialized sources of cybersecurity threat intelligence. To implement this strategy, the EU DIGITAL funding program has allocated to "Building the Capacity of Security Operations Centres". One of the key actions envisaged is the creation of cross-border platforms to pool data on cybersecurity threats between multiple Member States.

2. Understanding European Cross-border and Interreg Next Programmes

Cross Border Cooperation (CBC) is a key element of the EU policy towards its neighbours. It supports sustainable development along the EU's external borders, helps reducing differences in living standards and addressing common challenges across these borders.

Territorial cooperation under Interreg is built around four strands [3]:

- Interreg A – cross-border cooperation between adjacent regions (which should in principle be located along land or sea borders separated by up to 150 km of sea) to tackle common challenges identified jointly in the border regions and to exploit the untapped growth potential in these areas.
- Interreg B – transnational cooperation over larger transnational territories or around sea basins with a view to achieving a higher degree of territorial integration.
- Interreg C – interregional cooperation through four specific programmes to boost the effectiveness of cohesion policy by promoting:
    a) The exchange of experiences, innovative approaches and capacity building with a view to identifying and disseminating good practices and implementing them in regional development policies, including the „investment for jobs and growth goal" programmes;
    b) The exchange of experiences, innovative approaches and capacity building with a view to identifying, transferring and capitalising on good practices on integrated and sustainable urban development (the Urbact programme);
    c) The exchange of experiences, innovative approaches and capacity building with a view to improving and simplifying the implementation of Interreg programmes and cooperation actions, along with the setting up and operation of European groupings of territorial cooperation;
    d) The analysis of development trends in relation to territorial cohesion goals the European Spatial Planning Observation Network (ESPON) programm [4].

- Interreg D – cooperation between outermost regions to facilitate the development and integration of outermost regions and OCTs (for example, Caribbean regions) in their neighbouring environment. Interreg NEXT programmes [5] foster cooperation and tackle shared challenges among Member States and non-EU countries from East and South Neighbourhood region of the EU.

Cross-border infrastructure projects are fixed-asset investments that physically link two or more countries via infrastructure, including digital infrastructure, enabling the flow of people, goods, commodities or data [6]. Regional innovation governance refers to all processes of interaction between different actors that together determine the priorities, strategies, activities and outcomes of research and innovation at the regional level. This governance involves adopting institutional arrangements that facilitate systemic interactions between different innovation actors in the region, for example through the triple helix model of innovation or between policy hierarchies with improved policy coordination through multi-level governance.

Multilevel innovation governance [7] can be defined as a complex process of co-operation between different levels of government (supranational, national, regional, local) and/or agents to promote innovation for territorial development strategies. It aims to open up regional innovation strategies, such as the Smart Specialisation Strategy (S3), to other actors in production and knowledge systems simultaneously at different scales. Thus, it is important to implement practices that promote collaboration and alignment across levels of government and territorial entities in defining and developing S3 and other regional strategies to promote more effective multilevel governance. The Partnership for Regional Innovation (PRI) [8] focuses on developing more effective multi-level governance and strengthening synergies between policies and between different funding instruments such as the Cohesion Policy and Horizon Europe, while enhancing stakeholder engagement and co-governance. Encouraging innovative and partnership-based methods of multi-level management [9] of economic, technological and social development is forcing a change in the mentality and practice of populated areas in the regions Republic of Moldova. The Community practices learned should be enriched by innovative and experimental local practices, based on the experience and knowledge of local and regional elected representatives who are most often required to implement common policies and apply Community legislation. In this regard, experimentation is a good governance tool that allows actions to be taken on a small scale to test their impact with a view to wider adoption if the results are convincing, and allows policymakers to base their decisions on evidence that has already been tested at the level of their territorial impact.

2. The Role of Regions in Cross-border Cooperation

European Cross-Border cooperation, known as Interreg A, supports cooperation between NUTS III regions [10] from at least two different Member States lying directly on the borders or adjacent to them. It aims to tackle common challenges identified jointly in the border regions and to exploit the untapped growth potential in border areas, while enhancing the cooperation process for the purposes of the overall harmonious development of the Union. Regions have the privilege of operating close to local businesses, academia,

education and training players. Innovation, economic growth and social development, all happen at the local level. S3 approach invites regions Republic of Moldova to identify their key sectors for generating investment, discovery entrepreneurial potential at the local level and the required actions to reach excellence and development. Thus, the S3 aims to increase synergies between different European, national and regional policies, as well as public and private investments. To maximize its global impact, the EU needs more regions having a more solid knowledge of their own specialization. In 2017 the Joint Research Centre counted 18 Regions prioritizing cyber security in their S3.

In accordance with the objectives of the Council of Europe Convention No. 106 of May 21, 1980 (concluded in Madrid), cross-border cooperation is understood as any concerted action aimed at strengthening and promoting relations between neighboring territorial communities and authorities under the jurisdiction of two or more Contracting Parties, as well as the conclusion of any agreements and arrangements necessary to achieve the above objectives. Cross-border cooperation is carried out within the powers of territorial communities (Regions) and authorities determined by the internal legislation of each of the Parties. The added-value of the regional engagement for cyber security is highlighted through key examples of existing regional initiatives and related policy and financial tools [11], namely:

- STRATEGIC PLANNING & DECISION MAKING: Regions elaborate their own economic and innovation strategy, which is known as Smart Specialisation Strategy (S3);
- LONG-TERM INVESTOR & SUPPORTER OF THE "MADE IN EUROPE": Regions are the managing authority for the European Structural and Investment Funds (ESIF);
- THE TRIPLE HELIX SYSTEM: Regions host a triple helix system offering a truthful environment for innovation to the market;
- PROXIMITY WITH END-USERS & CRITICAL INFRASTRUCTURES: Regions are accessible territories of experimentation and connection with end-users;
- REDUCING CYBER SECURITY SKILLS SHORTAGE: By supervising education and training, regional authorities play a key role in addressing cyber security skills shortage.

In this context, European Cyber Security Organisation (ECSO) [12] has mapped 25 sustainable local strategies (e.g. Brittany, Central Finland and North Rhine Westphalia, "Connecting European Cyber Valleys"). The Joint Research Centre (JRC) databases (e.g. Eye@RIS3 and Digital Innovation Hubs catalogue) are the main sources on the related-data to DIHs and S3. The keywords used for the research on the data base are "cybersecurity" addressing the emerging challenges posed by cyber security and the increasing digitalisation of the society.

The 2017 EU Joint Communication "Resilience, Deterrence and Defence: Building a strong cyber security for the EU" [13] underlined that cyber security requires a comprehensive cross-policy approach involving the whole economy and all levels of government, including local and regional authorities. The EU strongly promotes the

position that international law, and in particular the UN Charter, applies in cyberspace. As a complement to binding international law, the EU endorses the voluntary non-binding norms, rules and principles of responsible State behaviour that have been articulated by the UN Group of Governmental Experts it also encourages the development and implementation of regional confidence building measures [14], both in the Organisation for Security and Co-operation in Europe and other regions.

3. Multi-level Governance in Cybersecurity

Regions has the biggest potential to connect the technology with the end users, assist local small and medium enterprises (SMEs), and provide them with business support and access to innovative technologies. In order to achieve an effective cybersecurity posture, the National government are not enough and that more structured inclusion of the regional and local authorities to the definition of they strategic role and needs in cybersecurity in the way to integration in Digital Single Market of EU. Therefore, we argue that the multi-level governance model needs to be established to include regions Republic of Moldova in the EU cybersecurity policy implementation. One of the principal advantages of regional entities is their proximity and capacity to cultivate trust among cybersecurity stakeholders at a local level. Unlike national governments, whose vantage point is often constrained by the necessity of maintaining a comprehensive overview of the nation's cybersecurity posture, regions enjoy more proximate connections with local cybersecurity stakeholders, encompassing a range of actors, including end users, integrators, research and innovation (R&I) centers, and product and service providers. The French Ministry of Armed Forces and the Brittany Region launched a noteworthy initiative, the "Pole d'excellence cyber", to enhance the country's cyber resilience through a collaborative approach with regional authorities on cybersecurity matters. This initiative exemplifies a state's potential to benefit from regional expertise in cybersecurity. It also underscores the value of regional involvement in cybersecurity innovation and industry development, especially when these issues are challenging to address by national authorities due to their limited understanding of regional cybersecurity contexts and dynamics.

The Governments of the Republic of Moldova and the French Republic concluded March 7, 2024 in Paris an agreement on cooperation in the field of defense. The document comes in the context of strengthening bilateral cooperation relations and adjusting the legal framework to the current security situation [15]. In frameworks of Partnerships for Regional Innovation (PRI), there is four main pillars when developing multi-level governance:

- Complexity (dealing with conflicts);
- Emergence (learning through the process);
- Context specificity (different regional experiences);
- Reciprocity (recognising each other level of governance).

Like large companies and administrations, the Ministry of the Armed Forces has launched its digital transformation which should enable it to offer 100% digitalized business processes (dematerialized services) and to have a suitable digital ecosystem with the expectations of its employees and partners. Defined within the Digital Ambition of the Armed Forces, the objectives of the digital transformation of the Armed Forces are to:

- Guarantee operational superiority and control of information in theaters of operation;
- Strengthen the efficiency of support and facilitate the daily lives of staff;
- Improve the relationship with citizens and the attractiveness of the ministry.

Broken down into broad guidelines, these objectives aim in particular to develop an innovation ecosystem as well as the use of emerging technologies for the benefit of the ministry's systems and applications. By choosing a strong technological transformation, the ministry offers itself the possibility of becoming a leading digital player.

If new information technologies [16], as shown in the Figure 1, allow companies to have certain competitive advantages, they can also allow the Armed Forces to benefit from a primordial operational advantage.

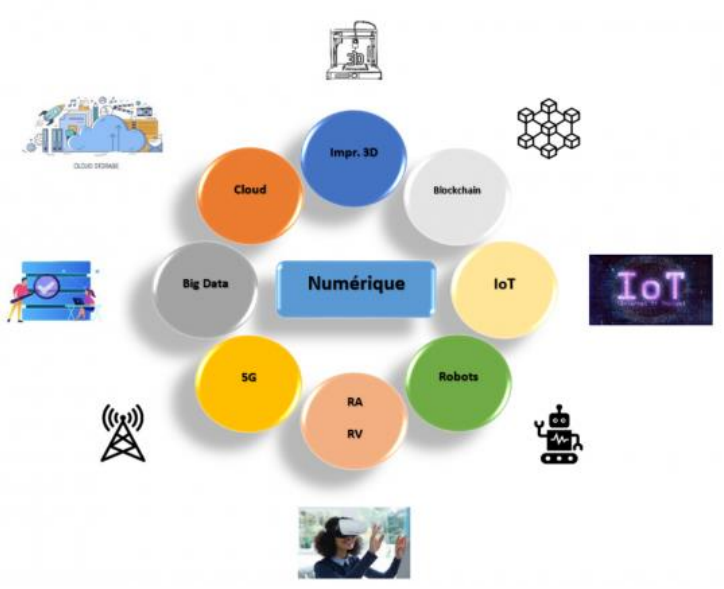

Fig. 1. Overview of innovative technologies.

AI can, for example, pilot aircraft systems supported by combat drones (SCAF), data analysis using Big Data techniques can help identify areas for improvement in our operational functioning, the cloud secure which can make it possible to have operational data available at any time and in any place within extremely short deadlines, autonomous robots can carry out risky tasks for humans (e.g.: mine clearance), the blockchain can immutably store sensitive data from our activity while ensuring their total traceability.
The European Network of Defence-related Regions (ENDR) was created in 2016 by the European Commission. The ENDR has three objectives [17]:

- Bringing regional organisations and clusters together to exchange best practices on how to develop dual use (defence-related) development strategies, including the integration of defence industrial and research assets into smart specialisation strategies, and support defence-related SMEs;

- Facilitating the flow of information about funding opportunities (such as the European Defence Fund and European Structural and Investment Funds); and
- Promoting the development of regional clusters of excellence.

One of the best EU practicies in frameworks of this network, which present a role of regions in Multi-level Governance, can be „European Cluster of Ceramics", that operate at the regional, national, european and international levels. The mission of the European Cluster of Ceramics is to boost the activity of the ceramics sector through innovation. Thus, it animates its network to increase its capacity:

- Accelerate innovation thanks to the active support of the engineering team;
- Boost growth through networking and visibility activities;
- Dvelop business on the European and international scale.

The "Association for the development and promotion of the European ceramic center" also known as European Cluster of Ceramics is a competitiveness cluster specialised in ceramics and related materials and processes. The cluster represents 119 manufacturers, 36 laboratories and 5 RTOs from France, Portugal, Switzerland, Belgium, Italy, and Germany. The cluster is part of other associations including ERMA (European Raw Material Advanced) and Cerame-Unie. Furthermore, the European Ceramic Cluster actively participates and influences the regional roadmap on advanced materials. With the aim of fostering innovation within its sector, the cluster's strategy is defined around four strategic markets:

- Luxury and Design (Tableware, jewelry, cosmetic);
- Energy systems (Mechanical components, buildings and refractories);
- Electronics, Electrical and optical components;
- Protection of people and environment (bioceramics, filters, armouring).

In the field of dual applications, the defence cluster covers high-performance materials such as transparent ceramics for ballistic defence and aeronautical applications, surface treatment, etc. The objective of the cluster is to facilitate collaborative R&D projects across the three components, Figure 2, which will bring together companies, public research laboratories and academic institutions in order to create new products and services on the market.

Fig. 2. Three Transitions in Innovation Ecosystem of "European Cluster of Ceramics".
Source: https://cerameurop.com/en/the-cluster/

Additionally, the cluster aspires to assist enterprises in the advancement of their innovation and to stimulate their economic development, particularly with regards to the export sector. In this regard, the Sirena'Mic project, financed by the Nouvelle Aquitaine region, represents a significant avenue for ceramic companies. This programme is specifically designed to facilitate the expansion of these enterprises into international markets, with a particular focus on the aeronautics, space and defence sectors in Europe and the USA, among other regions.

Republic of Moldova is a new candidate country and member of Interreg Europe programme. In this context multi-level innovation governance can be an important topic for many Interreg Europe projects, with partipation regions and local authoroties. As an example of building multi-level governance in regions and between them, can be among which: ECORIS3, aiming to foster policies and measures to support local and regional innovation ecosystems; RELOS3 which promotes the deployment of S3 at the local level; S34GROWTH that aims to enhance interregional collaboration through new industrial value chains; COHES3ION integrating a regional and sub-regional element into S3 and looking into how regions can develop their multi-level governance [7]. The Orkestra Basque Institute of Competitiveness shared six insights on multi-level governance:

- The urgency of developing multi-level governance;
- Its systemic nature (responding to complexity);
- The need for a new role for citizens in multi-level governance;
- The apparent dilemma between efficiency and democratisation;
- The importance of complementarities and collaboration in multi-level-governance;
- The hybrid nature of multi-level governance (responding to its emergent nature).

4. Strategies for Effective Cross-border Cooperation

Cross Border Cooperation (CBC) is a key element of the EU policy towards its neighbours. It supports sustainable development along the EU's external borders, helps reducing differences in living standards and addressing common challenges across these borders.

The developed overall strategy for sustainable management of the transboundary area between Bulgaria and Romania within the framework of the SPATIAL [18] project aims at a common integrated approach to solving development problems in the border areas of neighboring countries, overcoming the limitations imposed by national borders. Taking a holistic approach and taking into account economic, social and environmental elements, the strategic nature of the project is emphasized by the proposed activities:

- Identification of local capacity within the framework of S3 approaches;
- Identification of key issues/industry areas of analysis;
- Development of an integrated and harmonized database of border settlements of neighboring countries;
- Development of a strategy covering transboundary territory;
- Finding ways of cooperation between authorities and citizens, as well as between all interested parties in border areas;

- A unified approach to the implementation of investment projects within the framework of the Danube Strategy of the European Union and the European Structural and Investment Funds for the period.

The SPATIAL project has defined and established a framework interoperability Web-GIS systems, Fugure 3, for cooperation to facilitate the use of potential capital/local assets of the border areas of the countries in order to increase competitiveness and innovation along the entire common border, as well as to protect and improve the environment in the context of the development priorities of the European Union.

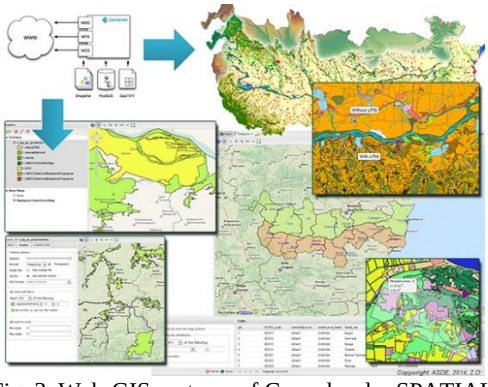

Fig. 3. Web-GIS systems of Cross-border SPATIAL
Source: https://cbc171.asde-bg.org/

B-solutions [19] is a pilot initiative aimed at removing legal and administrative border barriers along the EU's internal land borders. This initiative could well complement the methodology of the SPATIAL project, if it is adapted to the section of the Moldavian-Ukrainian and Moldavian-Romanian borders, within the framework of one project. B-solutions is promoted by the European Commission's Directorate-General for Regional and Urban Policy (DG REGIO) and managed by the Association of European Border Regions (AEBR) as one of the actions proposed in the Communication "Boosting growth and cohesion in EU border regions". The fundamental objective of b-solutions is to identify and implement solutions to legal and administrative border obstacles [20]. This entails enabling border-adjacent municipalities and regions, as well as cross-border entities, to submit information about border-related legal or administrative obstacles they face. Successful candidates are then assigned support from the European Commission to remove these border-related obstacles. In this new context for the Republic of Moldova, as an EU candidate country, we highlight the best practices of neighboring countries in the joint planning "Common Strategy for Sustainable Territorial Development of the Romania-Bulgaria Border Territory".

In our publications [21], we have repeatedly emphasized the role of Living Labs, the concept of which effectively contributes to the innovative development of settlements and organizations participating in its functioning. Living Labs are open innovation ecosystems in real-world settings that use iterative feedback processes throughout the innovation lifecycle to ensure sustainable impact. They focus on co-creation, rapid prototyping and

testing, and scaling innovation and business by delivering (various types of) shared value to stakeholders. In this context, living laboratories can also act as intermediaries/organizers between citizens, research organizations, companies and government agencies of border areas and Euroregions. The "European Parliament resolution of 15 September 2022 on EU border regions: living laboratories of European integration" [22] welcomes the 2021 Commission communication [23] which places emphasis on the following cooperation topics [24]:

- Barriers faced by EU border regions;
- Specific characteristics of border regions;
- Sustainability through closer institutional cooperation;
- More and better cross-border government services;
- Dynamic cross-border labor markets;
- Border regions for the European Green Deal.

In accordance with the Council Conclusions on the EU's Cyber Posture of May 2022 and as previously announced in the Joint Cyber Defence Communication, the Commission has proposed the EU Cyber Solidarity Act [25]. This Act encompasses a series of measures designed to reinforce solidarity and enhance coordinated EU detection and situational awareness. At the same time, it aims to bolster Member States' preparedness and response capabilities to significant or large-scale cybersecurity incidents. This is achieved through [26]:

- The European Cyber Shield will comprise a pan-European infrastructure of Security Operation Centers (SOCs), which will be utilized to construct and enhance coordinated detection and situational awareness capabilities;
- The Cybersecurity Emergency Mechanism will be employed to provide support to Member States in the preparation for and response to major or large-scale cybersecurity incidents;
- The Cybersecurity Incident Review Mechanism to review and assess significant or large-scale incidents;
- The European Cyber Shield will consist of a pan-European infrastructure that connects Security Operations Centres (SOCs) spread across the EU.

Procedure for creating a cross-border SOC platform [27] with funding support are follow:

- Following a request for expressions of interest, the European Commission will select applicants intending to establish a cross-border SOC platform;
- All parties to the request nominate their national coordinators, with whom the European Cybersecurity Competence Center (ECCC) [28] enters into a joint agreement for deployment and use, for example in border areas. This agreement sets out practical arrangements for managing the deployment and use of tools and infrastructure jointly owned by the ECCC and participating national SOCs following joint procurement. The coordinator is usually the national SOC of one of the EU Member States participating in the consortium;
- Each selected coordinator will participate in joint procurement of goods and services with ECCC;

- Individually participating partners in each selected consortium may apply for an additional grant to cover eligible costs such as the creation and launch of a cross-border SOC platform.

The EU contribution will cover up to 75% of the cost of purchasing instruments and infrastructures. The remaining procurement costs will be covered by the member states participating in each cross-border SOC platform. The ECCC Financial Rules set out the conditions under which the ECCC may engage in procurement, including joint procurement with member states.

A technical group, composed of experts from EU Member States, has been created to support the process and, in particular, to help identify the main types of goods and services that need to be jointly procured and to discuss the overall plan at a high level. Typical examples of procured goods and services identified by the technical team include (indicatively):

- Hardware: servers, micro data center racks, high-speed switches, firewall switches, GPUs, HSMs, sensors;
- Software: visualization tools, SIEM tools, vulnerability managers, aggregation tools, incident reporting tools, situation awareness correlation tools, AI/ML tools, PKI tools, orchestration systems;
- Services: CTI channels, AI/ML feature updates, dedicated virtual phone line, cloud storage, software development and customization services, consulting services.

The goal is to promote convergence between different platforms and, to the extent possible, to use procurement to acquire goods and services that can benefit all platforms. If properly justified, procurement may also include specific types of goods and services for individual platforms.

5. Policy Recommendations:

- One of the principal objectives of the "Initiating cross-border and interregional cooperation" initiative is to identify the characteristics and key services that a cross-border innovation ecosystem designed to support Small and Medium-sized Enterprises (SMEs) in the cybersecurity sector should provide.
- For potential partner public administrations, it is advisable to focus on identifying the main barriers to success. These include a lack of coordination between relevant actors, market fragmentation, and a lack of skills. For each identified barrier, a SWOT analysis should identify the strengths, weaknesses, opportunities, and threats at the regional and cross-border levels.
- In light of the varying levels of cyber development among partners, it is recommended that they identify best practices that align with their respective strengths and potential solutions to the needs of other partners. The European best practices identified in the article can be classified into two distinct categories of policies that facilitate multi-level governance of cyber defense. The first category encompasses policies that support the structure of the cyber innovation ecosystem, while the second category encompasses policies that support the advanced services

provided within the ecosystem. Examples of the latter category include labeling, access to public and private funding, capacity building, and so forth.

- As a consequence of inter-regional exchange processes, partners are able to select and adapt best practices and solutions, which are then reflected in regional Action Plans. The Action Plans serve to provide a concrete and tangible roadmap for interested regional authorities, delineating a pathway through which they may develop and channel greater levels of funding in order to enhance the competitiveness of SMEs in the field of cybersecurity. Moreover, their relevance is crucial in the context of Moldova's EU integration, as they provide a contribution that can be made to the European Investment for Growth and Jobs Program and the European Territorial Cooperation Program. Additionally, they can be utilized to address cybersecurity issues through the lens of the recently proposed NIS2 Directive. The aforementioned cross-border action plans, developed by the partners, serve as pivotal documents for both regional cooperation in Europe and for policy.

6. Conclusions

Regions can play a pivotal role in the advancement and dissemination of cybersecurity products and services in Europe, reducing the EU's dependence on third-country and non-European solutions. The near-future European cybersecurity landscape will be influenced by initiatives with a direct impact on regional ecosystems, including the European Cybersecurity Competence Centre Network, European digital innovation hubs, and the renewed smart specialization strategy in each region. Consequently, interregional collaboration is of paramount importance for the identification of solutions and the advancement of a more integrated cybersecurity market. Due to their privileged connection with the local ecosystems, the EU regions are playing a fundamental role in structuring the still young European cyber security ecosystem:

- Regions should be important part of the triple helix model of economic development, which involves governments, academia and business.
- Regions facilitate the development of the local cyber security ecosystems by involving Regional Technology Offices, training centres, services operators, incubators, SMEs and assist to establishe clusters initiatives and large companies.
- Regional level should be significant in disseminating good practices and establishing preventive measures and immediate response services due to its proximity with the end-users.
- Regions should play a key role in addressing cyber security skills shortage, as the demand for cyber security providers and operators are increasing at the EU local level.
- Due to the potential cyber attacks, local public administrations should be the direct users of cyber security products when developing their sectoral policies (e.g. health, energy or transport).

The Cross-border SOC platforms must enable and encourage the exchange and consolidation of large volumes of cybersecurity threat data from multiple sources in a trusted environment, and provide high quality, actionable information to their members through expert analysis and the use of state of the art information modern tools and

infrastructure. This should serve to improve detection capabilities and ultimately prevent and respond to cyber threats and incidents.

References

[1] European Commission, "Directorate-General for EU regional and urban policy," May 2024. [Online]. Available: https://ec.europa.eu/regional_policy/policy/cooperation/european-territorial/cross-border_en# .

[2] TESIM, "Neighbourhood External Cooperation Programmes Interreg VI (2021–2027)," EU, 2024. [Online]. Available: https://interregtesimnext.eu/about-interreg-next/ .

[3] EUR-lex, "Interreg – Supporting cooperation across borders" (2021–2027)," [Online]. Available: https://eur-lex.europa.eu/EN/legal-content/summary/interreg-supporting-cooperation-across-borders-2021-2027.html .

[4] European Commission, "European Observation Network for Territorial Development and Cohesion (ESPON)," [Online]. Available: https://www.espon.eu/espon-2030/espon-2030-programme.

[5] European Commission, "Directorate-General for EU regional and urban policy," [Online]. Available: https://ec.europa.eu/regional_policy/policy/cooperation/european-territorial/next_en.

[6] European Invest Bank, "Cross-border infrastructure projects," May 2023. [Online]. Available: https://www.eib.org/en/publications/20230107-cross-border-infrastructure-projects .

[7] Interreg Europe, "Multi-level Governance for Innovation," 24 January 2023. [Online]. Available: https://www.interregeurope.eu/find-policy-solutions/stories/multi-level-governance-for-innovation.

[8] European Commission, "Partnerships for Regional Innovation: 63 regions, seven cities and four Member States selected for Pilot Action," 17 May 2022. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_3008.

[9] EUR-lex, "Encouraging innovative and partnership-based methods of governance," Publications Office of the European Union, 2024. [Online]. Available: lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A52009IR0089.

[10] European Commission, "Interreg A - Cross-border cooperation," Directorate-General for EU regional and urban policy, [Online]. Available: https://ec.europa.eu/regional_policy/policy/cooperation/european-territorial/cross-border_en .

[11] EU Network EUROBITS, "The Role of the Regions in strenghtening the European Unionțs cyber security Position Paper," 15 March 2019. [Online]. Available: https://www.eurobits.de/wp-content/uploads/20190320_Regions_Position_Paper_approved.pdf.

[12] European Cyber Security Organisation (ECSO), "ECSO to lead ECCO, the European Cybersecurity Community Support project," 20 December 2022. [Online]. Available: https://ecs-org.eu/ecso-to-lead-ecco-the-european-cybersecurity-community-support-project/.

[13] European Commission, "JOINT COMMUNICATION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL. Resilience, Deterrence and Defence: Building strong cybersecurity for the EU," 2017. [Online]. Available: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52017JC0450.

[14] CCDCOE, "OSCE Expands Its List of Confidence-Building Measures For Cyberspace: Common Ground on Critical Infrastructure Protection," [Online]. Available: https://ccdcoe.org/incyder-articles/osce-expands-its-list-of-confidence-building-measures-for-cyberspace-common-ground-on-critical-infrastructure-protection/ .

[15] Republic of Moldova, Minisrty of Defence, "Intergovernmental Defense Cooperation Agreement signed in Paris," 7 March 2024. [Online]. Available: https://www.army.md/?lng=3&action=show&cat=122&obj=8847 .

[16] Ministry of Defence of France , "Digital transformation in the Army," [Online]. Available: https://www.defense.gouv.fr/terre/nos-materiels-nos-innovations/nos-innovations/pole-numerique-coordination-linnovation/numerique-0# .

[17] "The European Network of Defence-related Regions (ENDR)," [Online]. Available: https://endr.eu/about-us/ .

[18] Agency of Sustainable Development and Eurointegration – Ecoregions (ASDE), "Common Strategy for Sustainable Territorial Development of the cross-border area Romania-Bulgaria 2012-2014," [Online]. Available: https://cbc171.asde-bg.org/ .

[19] Association of European Border Regions (AEBR), "b-solutions," January 2022. [Online]. Available: https://www.aebr.eu/projects/b-solutions/ .

[20] Association of European Border Regions (AEBR), "b-solutions," [Online]. Available: https://www.b-solutionsproject.com/about.

[21] A. A. Babin, I. V. Covalenco, S. A. Tutunaru and E. A. Babina, "Some Aspects of the Formation of an Innovation Ecosystem for the Sustainable Development of Smart Villages in the Republic of Moldova," *INstrumentul Bibliometric Național,* p. 52, 2023.

[22] European Parliament, "European Parliament resolution of 15 September 2022 on EU border regions: living labs of European integration (2021/2202(INI))," 15 September 2022. [Online]. Available: https://www.europarl.europa.eu/doceo/document/TA-9-2022-0327_EN.html.

[23] European Commission, "EU Border Regions: Living labs of European integration," Brussels, 2021.

[24] European Commission, "Delivering the European Green Deal," 14 July 2021. [Online]. Available: https://commission.europa.eu/publications/delivering-european-green-deal_en .

[25] European Commission, "Commission welcomes political agreement on Cyber Solidarity Act," 6 March 2024. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1332.

[26] European Commission, "A European Cyber Shield to step up our collective resilience | Opening of the International Cybersecurity Forum | Speech by Commissioner Thierry Breton," 5 April 2023. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/speech_23_2145 .

[27] European Commission, "Cybersecurity: EU launches first phase of deployment of the European infrastructure of cross-border security operations centres," 24 November 2022. [Online]. Available: https://digital-strategy.ec.europa.eu/en/news/cybersecurity-eu-launches-first-phase-deployment-european-infrastructure-cross-border-security .

[28] European Commission, "Regulation (EU) 2021/887 of the European Parliament and of the Council of 20 May 2021 establishing the European Cybersecurity Industrial, Technology and Research Competence Centre and the Network of National Coordination Centres," 20 May 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021R0887.

Impact Zones: How cybercrime disrupts and shapes the landscape of data security

Claudia Alecsandra GABRIAN,
Babeș Bolyai University, Doctoral School of International Relations and Security Studies, Cluj-Napoca, Romania
claudia.gabrian@ubbcluj.ro

Abstract
In a highly networked digital world, cybercrime and data breaches are increasing together, putting up a strong front and proving to be a challenge for everyone: individuals, organizations, and governments. Cybercrime is associated with financially motivated attacks and one type of cyberattack that is one of the most prolific is ransomware. Objectives of the study: to analyze the disruption which destabilizes and breaks the system's normal functioning. The overall horizontal objective is to identify the activities that can breach data, cause financial losses, and also can influence the cybersecurity landscape. The paper is important in this topic area because knowing about cybersecurity is the key to knowledge, advancement of research, and practical application of solutions for society. Approach: strategies and regulations are always changing, and have the main role in protecting the data and responding to cyber-attacks; but unfortunately, cyber-attacks are more present and more advanced every day, and this involves different perspectives for this type of cyberattacks, in this paper the main methods are netnography, document analysis and the case study. The results show that a large number of interconnected devices in an Internet of Things landscape enhances vectors for attacking and amplifies the complexity of data security challenges. The following research will demonstrate the fact that ransomware attacks are a part of a huge disrupting type of cyberattack and represent a critical threat to data security and cybercrime has a major role in this topic. Also, an example of data leaks is a specific case study about one of the NSA officers who tried to send classified defense information to Russia, explaining that the Snowden case is not unique in the cybersecurity landscape.

Keywords: cyberattacks, ransomware, data leaks, darknet and social media marketplaces.

1. Introduction
Data security and cybersecurity in a particular way represent a suite of strategies and technologies aimed at protecting data from unauthorized access, in this case, cybercrime has a very important role because involves breaching the systems to access personal and sensitive data. Also, data leaks are closely related to data security because some people want to disclose information by intention. Firstly, cybercrime includes a lot of illicit activities, some of these activities are financially motivated attacks such as ransomware. In order to exploit at least one vulnerability in data security measures, ransomware in 2023 according to the ENISA report, was the first type of cyberattack will the most scale of attacks [1].

The most recent challenge in cybersecurity is to identify what type of cyberattacks increase their sophistication and malicious activity because represent a major challenge to experts, governments, and society. The interconnection between cybercrime and data security can be identified in various contexts, and one of these is the proliferation of interconnected devices using the Internet of Things (IoT) which represents a complexity challenge for cybersecurity. For a collective defense and to prevent these types of cyberattacks that are constantly evolving, is necessary to have a comprehensive approach that integrates technical defense, threat intelligence, and cybersecurity culture among people [2].

Cybersecurity has developed mechanisms to prevent and respond to the attacks. Categories of this have included a systemic procedure, for instance, an application security model to all the systems, whereby all best practices are included. On the other hand, data information and security levels are at stake since many cyberattacks seek to obtain all the data through unauthorized access. Securing data is another area entire of challenges regarding threats; security measures go a long way. Applied to these, data analytics play other roles in learning from existing threats in developing solutions for unknown threats toward these networks, infrastructures, data, and information. It would be collected in massive amounts, leading to the popular term "big data," signifying large datasets not only in size but also generated at a high rate, having heterogeneity, and that first and foremost is, in this complex environment, it will give valid findings or patterns. In each of these systems, is necessary to monitor the infrastructure for accurate functioning and to prevent, detect, and recover from cyber threats. There is a varying degree of supervision and management of such data with the varying degrees of prevention, detection, or recovery expected in the domain. Some domains are very preventive while others are very detective or recovery-based. In both cases, multiple types of datasets can be collected to provide intelligence about the cyber threats and evaluate user behaviors to prevent future threats or even to identify some insider propagating the threats [3].

2. Modern market sales for cybercrime

With Telegram becoming the most critical messaging application for very many people in the world, it is also becoming a hub for several cybercrime activities like sales and leaks of stolen personal or corporate data organization and operation of cybercrime gangs, distribution of hacking tutorials, hacktivism, and more. The Telegram messaging application has gained many users, thus making it a big challenge for security researchers in the war against cybercrime. The features that make Telegram appealing to cybercriminals are its purported built-in encryption and the ability to create channels and huge private groups. Tracing and monitoring criminal activities within the platform, though, becomes a challenging task for any law enforcement or security researcher because of these two characteristics. Furthermore, cybercriminals use coded language and alternative spellings to discuss their activities on the platform, making it quite difficult for security agencies to decipher their conversations. It is also used for the sale of stolen data and illicit goods through the recruitment of new members. The ability to remain anonymous while on Telegram is one of the main features that attract hackers to the platform. Registration of accounts without personal information makes it possible for most users to create multiple identities and easily engage in conversation without using their identity [4].

Many cybercriminals, including Discord, Jabber, Tox, and Wickr use several more chat applications. Each one offers a specific set of features and characteristics, but they all offer some level of furtiveness and protection that cybercriminals find attractive. It is a decentralized, secure application for messaging where one doesn't need to register or provide personal information; data is also encrypted through peer-to-peer technology and the NaCl library, with users identified by a Tox ID. Indeed, the Telegram channel is also the platform of preference for cybercriminals to sell and share just about any PII, from social security numbers, driver's licenses, and passports to dates of birth and physical and email addresses. Cybercriminals can then exploit this information to engage in fraud that

leverages stolen identities, including taking up bank loans, for instance, and opening bank accounts. Many of the ransomware and data extortion groups are cybercrime gangs, using their net offense experiences to steal private data from organizations while threatening to publish it in return for ransom money from victims. Ransomware gangs encrypt information using ransomware, while data extortion groups, on the other hand, only steal the data [4]. When the war between Russia and Ukraine started, a lot of telegram groups appeared, one of these groups that posted important data leaks was “Data1eaks” in Russian.

Darknet is another market where data are posted and sold, these data contain full names, birthdates, social security numbers, credit card information, bank account details, email addresses, and passwords. Cybercriminals also trade medical records, driving licenses, and passport details. For example, payment card data costs around $10 and ranks as the most commonly found item on the darknet market. Mobile phone numbers and online accounts cost around the same amount. Cryptocurrency wallets and account login details attract more interest than bank accounts. Passport copies topped the list as the most expensive item, averaging around $600. While the prices for stolen information may seem high, the repercussions for individuals whose data is sold can be higher. People whose data gets sold online may face financial losses, damage to their credit scores, and identity theft [5].

3. Ransomware attacks

Ransomware is an extortion attack in which an attacker deprives a victim of their valuable organizational data until the attacker is paid. The ransomware groups are upping the ante now with an assortment of extortion tactics: posting sensitive information online if not paid. RaaS (Ransomware as a service) provides user-friendly tools for performing this act amateurishly and essentially widens the scope of ransomware for more would-be bad actors [6].

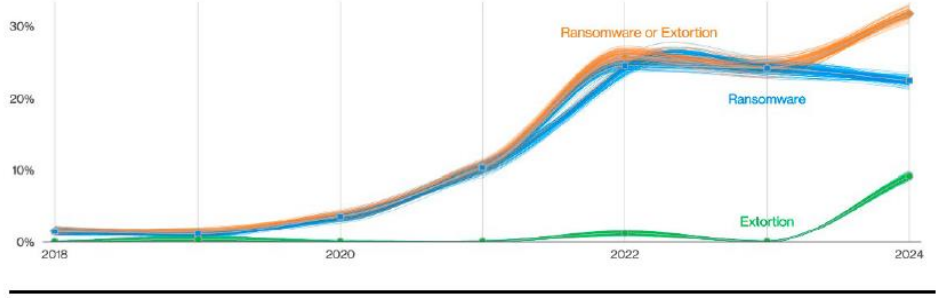

Fig. 1. Ransomware and Extortion breaches over time
Source: Verizon Business, "2024 Data Breach Investigations Report", 2024.

The chart indicates a steep increase in incidents with ransomware, beginning in the middle of 2019, cresting up through 2021, and stabilizing at a high level by the end of 2024. Extortion remained low until mid-2021 before it grew sharply, particularly in 2022. Both concatenated categories, "Ransomware or Extortion", show a similar trend from the Year 2021 and further increase until the Year 2024. This conveys the evolving and changing space of cyber threats: a somewhat scarily high level of ransomware attacks, and a significant rise in extortion instances that call for better security measures.

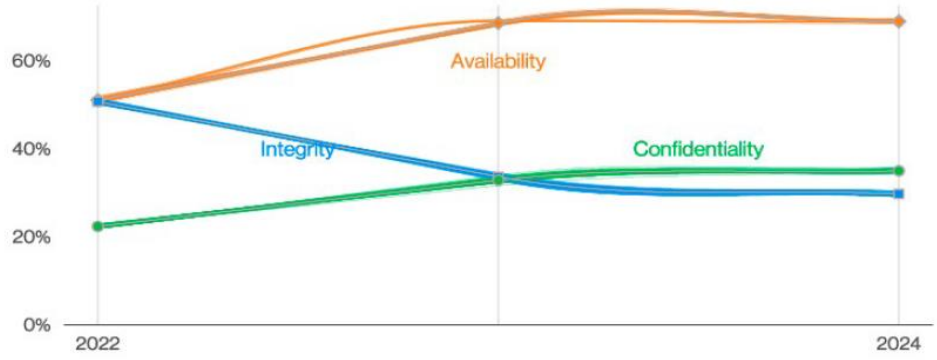

Fig. 2. Attributes over time in incidents
Source: Verizon Business, "2024 Data Breach Investigations Report", 2024.

The graph shows that incidents of availability rise sharply between 2022 and 2024, peaking above 60% by mid-2023 and then declining, while incidents of integrity decrease steadily from close to 40% to around 30%; at the same time, incidents of confidentiality rise from around 20% to 30%, finally overtaking integrity by the middle of 2023. It represents an increase and continued focus on the disruption of service availability, a declining trend of threats to data accuracy, and quite an increase in data privacy risks—reflecting an evolving threat landscape in cybersecurity.

	Incidents				Breaches			
Industry	Total	Small (1–1,000)	Large (1,000+)	Unknown	Total	Small (1–1,000)	Large (1,000+)	Unknown
Total	30,458	919	1,298	28,241	10,626	617	986	9,023
Accommodation (72)	220	16	9	195	106	16	9	81
Administrative (56)	28	7	7	14	21	6	4	11
Agriculture (11)	79	5	0	74	56	4	0	52
Construction (23)	249	17	6	226	220	12	5	203
Education (61)	1,780	82	630	1,068	1,537	56	618	863
Entertainment (71)	447	16	2	429	306	10	1	295
Finance (52)	3,348	75	122	3,151	1,115	54	87	974
Healthcare (62)	1,378	54	21	1,303	1,220	41	18	1,161
Information (51)	1,367	79	62	1,226	602	49	19	534
Management (55)	22	4	1	17	19	4	1	14
Manufacturing (31–33)	2,305	102	81	2,122	849	62	49	738
Mining (21)	30	1	2	27	20	1	1	18
Other Services (81)	462	13	5	444	417	8	5	404
Professional (54)	2,599	205	102	2,292	1,314	124	73	1,117
Public Administration (92)	12,217	56	115	12,046	1,085	39	27	1,019
Real Estate (53)	432	35	5	392	399	29	2	368
Retail (44–45)	725	90	47	588	369	55	32	282
Transportation (48–49)	260	21	38	201	138	17	12	109
Utilities (22)	191	17	11	163	130	12	6	112
Wholesale Trade (42)	76	22	21	33	54	17	14	23
Unknown	2,243	2	11	2,230	649	1	3	645
Total	30,458	919	1,298	28,241	10,626	617	986	9,023

Fig. 3. Number of security incidents and breaches by victim industry and organization size
Source: Verizon Business, "2024 Data Breach Investigations Report", 2024.

This table classifies cybersecurity incidents and breaches by differences between industries, total incidents and breaches, and small (1-1,000 employees) and large (1,000+ employees) organizations. The total number of incidents reported is 30,458, with the vast majority falling under unknown size at 28,241, indicating serious under-reporting or a lack of data on the organizational sizes affected. Public Administration (12,217), Finance

(3,348), and Professional Services (2,599) are the first three in line to have major incidents. In terms of breaches, unknown organization sizes account for 10,626 and 9,023 incidents. Here, too, Public Administration with the highest number—1,085—is followed by Finance at 1,115 and Education at 1,537.

Data shows that the Public Administration, Finance, and Education sectors are at higher risk of cyber incidents and breaches, reflecting the great value of sensitive information and essential services within these sectors. More broadly, smaller organizations in all sectors report fewer incidents and breaches. This may reflect lower targeting by cyber threats, or it may also relate to possible under-reporting. Large organizations also reported fewer incidents compared to organizations of unknown size, but they still account for significant numbers of breaches. Notably, in the Education sector, there are 618, and Professional Services account for 124. This shows that while all businesses must take security very seriously, those working at scale need even better cyber protection for the more considerable possible risks.

LockBit was known in the first place as “ABCD” ransomware, and it has evolved into a specific threat over the years. LockBit is another subclass of ransomware; this means that it is a 'crypto virus' for the apparent reason that it has crafted its ransom demands around financial gain from those affected in return for the decryption of the data. It targets mainly enterprises and government organizations rather than individuals [7].

The organizations big and small across the globe were negatively affected by the LockBit RaaS and its affiliates. In 2022 and 2023, LockBit was declared as the most active global ransomware group and RaaS provider, going with the number of victims claimed on their data leak site. A Ransomware as a Service (RaaS) criminal cyber-group operates a particular strain of ransomware and rents that ransomware out to one or more other individuals or groups of actors (commonly referred to as "affiliates"). It helps affiliates distribute their ransomware by charging fees up front, operating subscription services, sharing profits, or some combination of these three remuneration models [8].

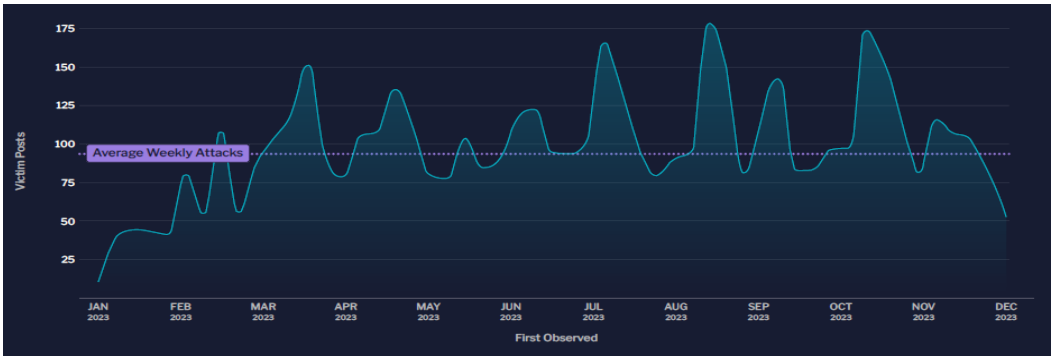

Fig. 4. Public Victim Posts by Week in 2023
Source: Flashpoint, "2024 Global Threat Intelligence Report," 2024.

The ransomware landscape has changed considerably since the apparition of the LockBit group in 2019. After that, since 2022 and continuing in 2023, LockBit remains one of the

most prolific ransomware, just with an exception of 3 months when CL0P ransomware targeted a lot of companies with two zero-day campaigns. LockBit is recognized as a more evolved and ruthless version of ransomware and LockBit 3.0 is still active but not sure for how long. The 4.0 variant seemed to be in the making after the operation Cronos when LockBit was taken down in February of 2024. To add to this, it all leaves the victims with yet another aggressive form of negotiation along with a triple-extortion scheme.

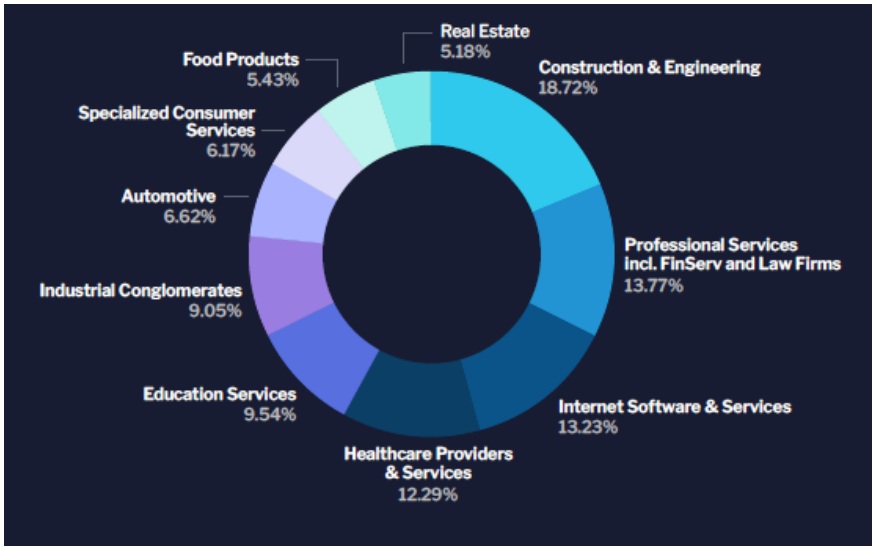

Fig. 5. Top Ten Industries Targeted by Ransomware in 2023
Source: Flashpoint, "2024 Global Threat Intelligence Report," 2024.

The construction and engineering sector was the most targeted in 2023, with 416 publicly reported incidents. Other highly targeted sectors were professional services, internet software and services, and healthcare, reinforcing the cross-industry impacts ransomware can have and the vital need for industry-tailored defense strategies. The most targeted with 57 public attacks is the construction and engineering sector, only in the first two months of 2024. Following behind are the manufacturing and healthcare sectors, with 49 public attacks in the first two months of 2024.

4. Case study about data leaks

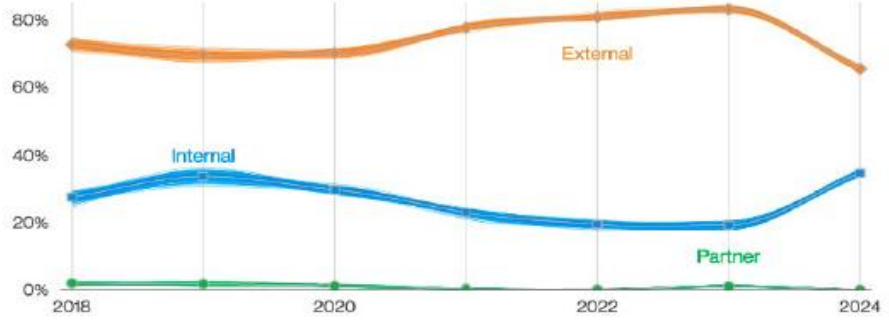

Fig. 6 Threat actors in breaches over time
Source: Verizon Business, "2024 Data Breach Investigations Report", 2024.

In 2022, one of the former employees of the United States National Security Agency was arrested for interfering in the sale of classified information to a spy operating for Russia, who was an operative for the Federal Bureau of Investigations. Jareh Sebastian Dalke, presently 32 years old, had worked at NSA for just under a month from June 6 to July 1, 2022, contracted in Washington D.C. while on a temporary assignment to NSA from his employer. During his employment with the NSA, he emailed three classified documents to a non-government entity on his encrypted email account: one at a secret classification level and two additional documents at the top-secret classification level. He also scheduled the transmission of an extra quantity of National Defense Information subject to his control to be transferred to the undercover FBI agent and scheduled payment for such transmission via cryptocurrency. The proposed payment was for compensation in exchange for transmitting the information. The files also contained a letter from Dalke in which he stated, "Dear friends! I am thrilled to have this opportunity finally to present this information to you. I welcome our friendship and mutual benefit. If there are desired documents that you would like for me to locate, please let me know, and I shall try when returning to the main office." He further indicated that he had attempted to establish contact using a submission to the SVR TOR site. Dalke's arrest followed within days of the Russian government conferring Russian citizenship on former U.S. intelligence contractor Edward Snowden, who is wanted on charges of espionage after leaking tens of thousands of documents listing a plethora of surveillance programs operated by members of the UKUSA community. He also said that his revelations would have meaning for Russia and urged them to keep him in touch, promising to provide more documents later. This man was convicted and sentenced to nearly 22 years in prison for attempting to transfer classified documents to Russia [9].

Jareh Sebastian Dalke's act, therefore, was a significant violation of ethics and general legal and constitutional provisions regarding national security. Jareh's act of trying to dispose of classified data to a foreign body of interest seriously removed the integrity and safety of the United States as a whole; innumerable lives and a slew of strategies of national defense would be set on the line. There is a reason for the classification of documents: they contain sensitive information important to national security, defense, and intelligence activities. Enemies might use the release of such information without due authority for several disastrous consequences – espionage, sabotage, and even terrorism. This is a relevant example that data leaks are no just from external, but also internal, and this example is relevant because when we analyze all the types of data leaks, human error intentional or unintentional is present every day.

Further, when classified information is divulged in the purview of gaining monetary benefit and that, too, in terms of a cryptocurrency, Dalke inflicts a blow on the name and value of the institutions responsible for national security. This erodes the confidence that the public and other nations might have in the United States to guard secrets and maintain global stability. In betraying his country, Dalke did not just place the current operations at risk, but he set an example that would give other people plans on how to exploit their access to sensitive information for personal gain.

The relevance of cases such as Jareh Sebastian Dalke and Edward Snowden contributes significantly, underlining, in one way or another, the risk level that insider threats pose to national security through unauthorized disclosure and data leakage of classified information. Furthermore, while Dalke's failure to sell susceptible data to foreign power put on the table some of the weaknesses in security protocols, Snowden's massive leaks exposed large government surveillance programs, leading to ongoing debates on government surveillance and changes that privacy and intelligence practices and policies will face in the process. They both emphasized the necessity of security measures, legal frameworks, and oversight mechanisms strong enough to guard sensitive information, prevent leaks, maintain trust, and maneuver in international relations and espionage.

5. Conclusions

In the contemporary digital world represents a challenge to face cybercrime, data leaks, and the ransomware ecosystem. Cyberattacks such as ransomware have evolved and at present have turned into a very organized and money-making enterprise. Cybercriminal groups have developed the Ransomware as a Service (RaaS) business model, affiliate networks, and very sophisticated methods to demand ransom from their victims and earn a lot of money. Such attacks against businesses, critical infrastructures, and public entities are seriously increasing around the globe, emphasizing the multi-domain impact of incidents on the global security paradigm.

Having personal data published on Telegram and the Dark Web exposes people to a large of potential threats: from targeted phishing, social engineering, and extortion schemes. Cybercriminals manage to impersonate their victims, take over their accounts, and commit fraud or even blackmail. Moreover, after exposure to these platforms, personal data can quickly multiply and be next to impossible to control or eliminate, raising risks of identity and reputational compromise. The combination of factors under which exponential data growth in digital format, the sophistication of cyber threats, and expanding attack surfaces driven by cloud services, IoT devices, and mobile technology in return actually facilitate data leaking.

Cybercriminals wish to acquire and disclose data for many reasons, whether financial, ideological, or personal vendettas. In the digital age, data is among the most valuable things around, and cybercriminals take advantage of systems and network vulnerabilities to steal sensitive information like intellectual property and other private information—personal and financial data and trade secrets. They might want to commercialize such data by resorting to identity theft, extortion, or even selling the data on underground markets. Further reasons may lie in political motivations, notoriety, or personal beliefs against organizations or individuals. Data are published as a control mechanism for causing damage or for manipulating public opinion; this, indeed, is the multi-motivation behind cyber activities.

The future of LockBit ransomware, especially after Operation Cronos, will be nothing short of uncertain, and indeed, the 4.0 variant that now gets admission to the cyberspace of this underworld will most probably indicate new trends of evolution and adaptation in the ransomware space. It is thus that Operation Cronos managed to prevent the ransomware—at least for now—as police forces from around the world and cyber-researchers worked

unceasingly to access the infrastructure belonging to the LockBit syndicate. However, above all, cybercriminals are best known for their stubborn resilience and adaptability. The creation of yet another LockBit 4.0 variant shall only be one more indication in the process of continuous evolution of the ransomware technical capabilities, with new features, encryption methods, and evasion techniques likely to be added that will effectively facilitate targeting new victims.

References

[1] ENISA, "ENISA Threat Landscape Report 2023," European Union Agency for Cybersecurity, 2023.

[2] T. Holt and A. Bossler, "Cybercrime and Digital Forensics," *Routledge*, 2022.

[3] V. P. Janeja, Data Analytics for Cybersecurity, 2022, pp. 8-29.

[4] Kela Cyber, "Telegram: The Cybercriminal's Toolkit," 2023.

[5] V. Lyskoit, "Darknet Markets: The Complete Guide," NordVPN , 2024. [Online]. Available: https://nordvpn.com/blog/darknet-market/. [Accessed 7 May 2024].

[6] Unit42, "Stages of a Ransomware Attack," 2022.

[7] Cybersecurity and Infrastructure Security Agency (CISA), "Cybersecurity Advisory: AA23-165A," CISA, 2023.

[8] Kaspersky, "LockBit Ransomware," [Online]. Available: https://www.kaspersky.com/resource-center/threats/lockbit-ransomware. [Accessed 10 May 2024].

[9] U.S. Department of Justice, "Former NSA Employee Sentenced to Over 21 Years in Prison for Attempted Espionage," 1 February 2024. [Online]. Available: https://www.justice.gov/opa/pr/former-nsa-employee-sentenced-over-21-years-prison-attempted-espionage. [Accessed 10 May 2024].

Risk management, protection, and security of personal data in Romania

George-Loredan POPA,
Politehnica National University for Science and Technology, Bucharest
georgepopa1986@yahoo.ro

Abstract
In recent years, the digitization of institutions has gained momentum as a result of the coronavirus pandemic. The pandemic period imposed the digitization of services in an aggressive and forced way, many authorities had to apply immediate measures, without having a test base behind them. People were forced to interact with a computer, thereby giving up physical contact with colleagues, which in one way or another increased the exposure of companies, thus allowing cyber-attacks easier access to target targets. The human factor, education, and balance are important elements when it comes to the area of data security. Any digitization process is meant to provide operational efficiency, productivity, and information security. Analyzing cyber attacks we will notice that they evolve with the help of AI, attackers use this technology to be able to produce personalized messages, customer information, etc. At the moment, Romania is at an average level, or even below the average level in terms of the functioning of the Internet and data security. In this article, we will present the results regarding data security, as well as the measures that were taken to protect them, both at the level of Romania and other EU member states after the coronavirus pandemic period.

Keywords: coronavirus, data security, digitization, management.

1. Introduction

IT risk [1] management represents all efforts made to reduce threats, vulnerabilities, and consequences as a result of unprotected data. The risk analysis starts with its vulnerabilities, the evaluation of a potential computer attack, and the identification of the necessary measures to prevent or interrupt it.

Data protection [2] aims to protect information from unauthorized access, destruction, or modification. The main component of a process within an institution is represented by information. The information system represents a set of processes of the organization, the object of realization of the information being represented by technology. To protect data, at the level of each institution it is necessary to implement a very well-structured set of procedures, practices, functions, IT equipment, software applications, etc.

Data protection and security include a fairly wide range of activities such as risk analysis, best practices guide, management in situations (coronavirus pandemic), development, responsibility, and liability.

The crisis caused by the coronavirus pandemic [3] has put cyber security to the test [4], becoming a very important topic both at the level of the European Union and in its member states. The public sector had to quickly face new challenges in terms of IT, especially in the transition from working with physical presence to working from home. Since then, institutions have stepped up their activity in terms of data protection, paying a lot of attention to the preparation of systems against cyber attacks, especially on data protection.

However, fast digitization has had the role of offering the world new opportunities bringing a plus to the IT side, but also many threats in terms of increasing risks [5], many public institutions have faced cyber-attacks which have led to a social and economic impact.

2. Data Use Strategy

Following the coronavirus pandemic, the European Commission through its digital strategy has acquired enormous importance [6].

The approval of the strategy through digital tools was aimed at monitoring and limiting the virus, supporting research and development of new diagnostic strategies, treatments, and vaccines, and more than that, assuring the population about data protection as a result of the transfer of work online.

The imposition of restrictions, social distancing, and the working environment have become much more digitized, and public institutions and people have relied heavily on the internet and connectivity.

Governments took swift action, ensuring the continuity and availability of public services through e-government and e-health, while security systems protected online identity.

EU member states applied social distancing measures to combat the COVID-19 pandemic, and the demand for internet capacity increased massively, regardless of whether it was the provision of remote activities, e-learning, or entertainment, that led to network tension.

But as a result of these activities, new dangers have emerged, and cyber security has had a lot to do from the online safety of consumers, and the normal functioning of hospital facilities, to the management of existential energy and water supplies.

The coronavirus pandemic has highlighted digital skills for work and how to interact with others, while also demonstrating deficiencies in IT knowledge and the importance of digital education. In 2021, according to the EUROSTAT database, just over half of the workforce aged 16-74 had at least minimal IT knowledge.

According to this ranking, Romania ranks last, with a percentage of 28%, and at the opposite pole with the highest percentage are the Netherlands and Finland with 79%. As we can see from Fig. 1, an equally low percentage is also found in the case of Bulgaria 31% and Poland 43%.

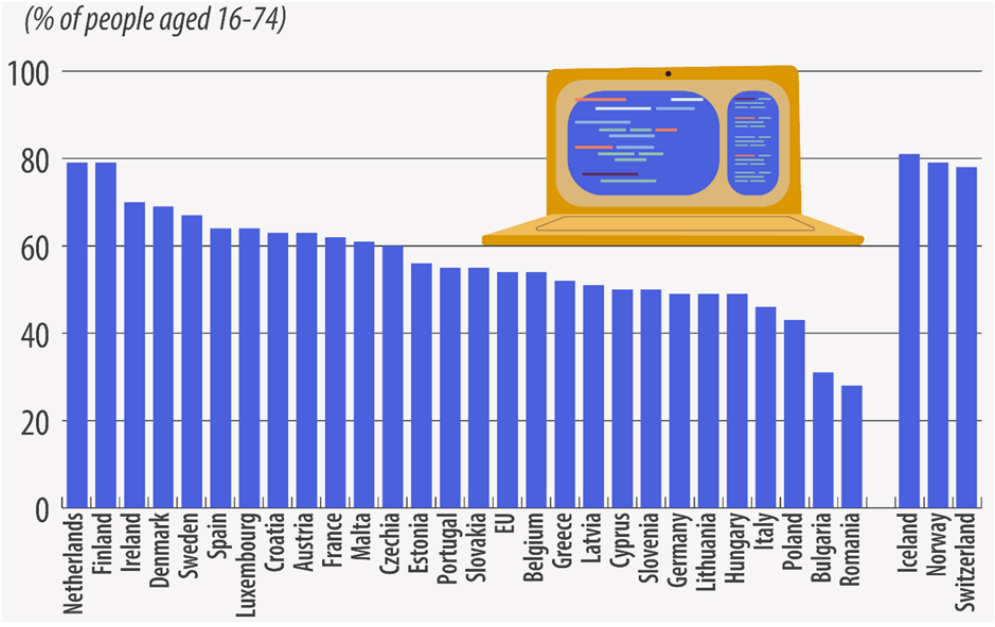

Fig. 1. People with at least basic overall digital skills in 2021
Source: EUROSTAT

Making a comparison, in the year 2023, according to EUROSTAT, of the analyzed population, the highest percentage is found in the Netherlands, and the lowest level of minimum knowledge is still occupied by Romania.

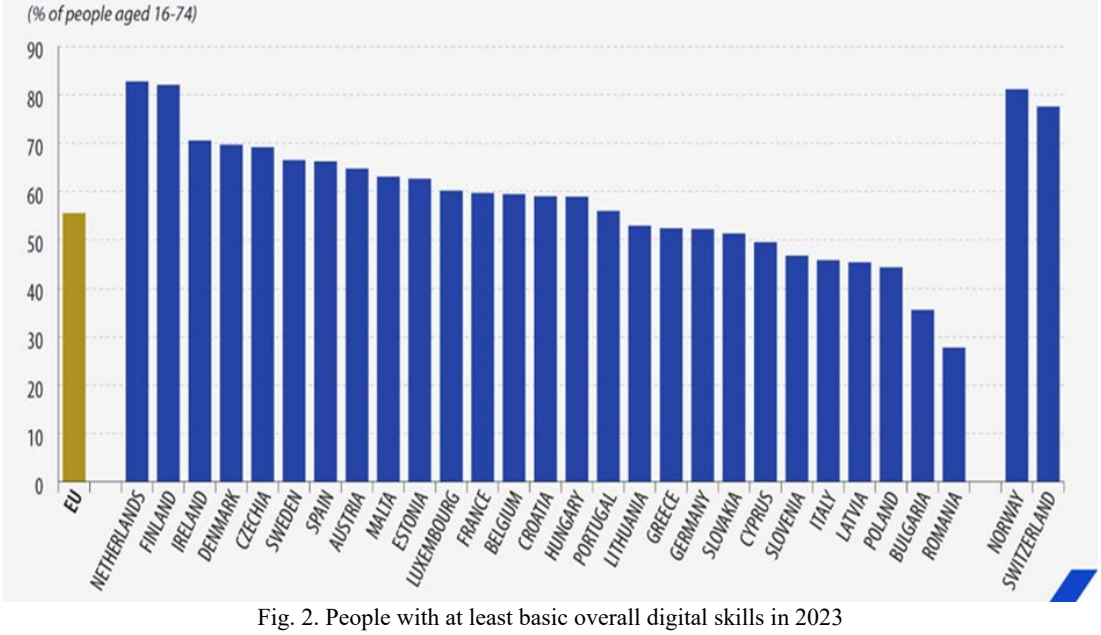

Fig. 2. People with at least basic overall digital skills in 2023
Source: EUROSTAT

Following what was presented by EUROSTAT, at the beginning of March, the European Commission proposed to transform the results of digital skills by 2023, as follows:

- Digitally literate staff and highly qualified digital professionals, 80% of adults with basic digital skills by 2030;
- Secure digital infrastructures;
- Digitization of companies, three out of four companies use cloud computing services, big data systems, and artificial intelligence;
- Digitization of public services, the possibility of accessing and processing all public services in the online system.

This Compass proposed by the European Commission for the digital dimension establishes a robust governance structure, shared with Member States, based on an annual monitoring system in the form of color codes.

3. Data Security

Data Security refers to the process of protecting digital information against potential threats. These include cyber attacks, hacking, phishing, and malware.

To prevent, detect, and respond to cyber threats, it is necessary to use physical, administrative, and technical measures. Data is a valuable asset and the document cycle is a key success factor.

Data protection is not only a legal and ethical responsibility but also a strategic necessity. A data security incident can have serious consequences, possible actions, or financial losses.

Online platforms are the important side of life and the economy.

Data confidentiality refers to the guarantee that data is accessible only to authorized personnel. Integrity assures that data is accurate and complete and has not been tampered with, and availability conveys that data is accessible when it is needed.

Another important role belongs to Technology, which has a vital part in Data Security.

In the age of digitization, the use of specialized software and hardware is becoming increasingly important. Firewalls, intrusion detection systems, encryption, multi-factor authentication, and backup solutions are some examples of technologies used to protect data and prevent security incidents.

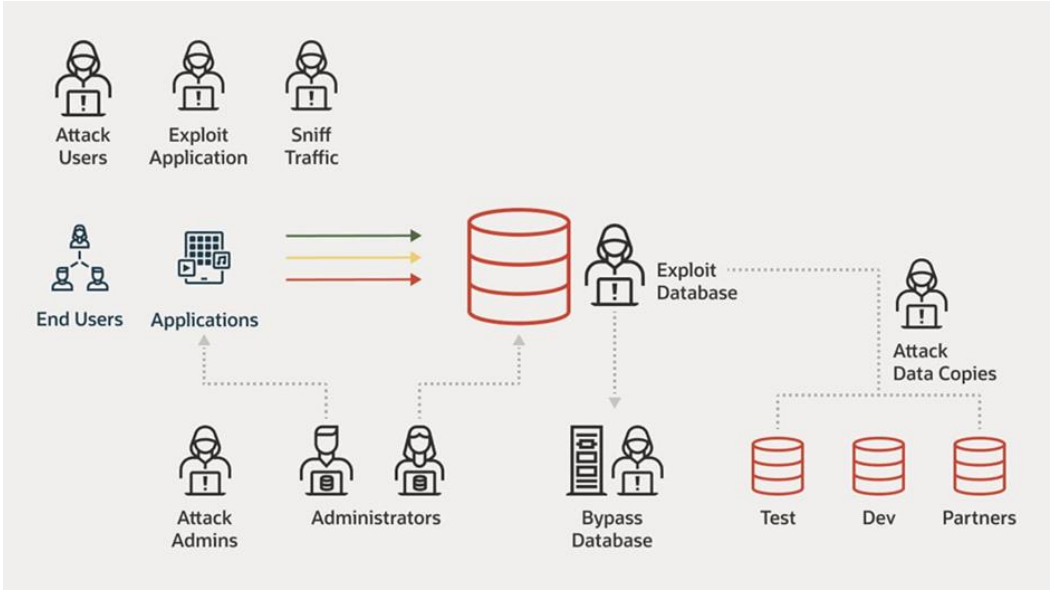

Fig. 3. What is data security?
Source: ORACLE

Data encryption is an effective solution for protecting information. This involves transforming the data into an unreadable format without using a decryption key. Thus, the data remains inaccessible to unauthorized persons, but the costs related to this procedure are very expensive and can slow down the performance of the systems.

Antivirus and antimalware software is an important solution for Data Security. These programs detect and remove malware, such as viruses, trojans, or spyware, that can compromise data security. Antivirus software often has low detection rates or may be ineffective against new threats.

Firewalls control network traffic and protect the network and data from unauthorized access. Firewalls filter data packets and can block unsafe or suspicious connections. Be careful though, they can be overcome by sophisticated attacks and require constant configuration and updates.

Cloud security services provide a scalable and efficient solution for protecting data. They include continuous monitoring of network activity, detection, and prevention of cyber attacks and secure data storage. There are also security risks associated with cloud services, such as unauthorized access to data. Consider Cloud ERP solutions in which to integrate Data Security components, for a complex and efficient IT ecosystem.

Multi-factor authentication involves using at least two authentication methods, such as password, fingerprint, or two-factor authentication. This solution adds an extra layer of security because even if an authentication method is compromised, there is still one step to verify the user's identity.

Blockchain technology can be used to secure transactions and data decentralized and transparently. This provides a method of recording and verifying information without the need for a central authority. Blockchain technology is still in development and may have limitations in terms of scalability and efficiency.

4. Data security risks

The most important risk is misusing or disclosing personal data to third parties without users' consent. The risks must be taken into account when using especially artificial intelligence for data processing and the implementation of appropriate measures to minimize them. Thus, data protection must be guaranteed regardless of the circumstances, and artificial intelligence systems must mature, adapt, and guarantee this right of citizens [7] [8].

Giving up artificial intelligence and the benefits it has in daily activities would be regrettable, it's all about risk assessment and prevention.

According to a study carried out by IPSOS regarding the degree of understanding of artificial intelligence, in Figure 4 we will find the following values:

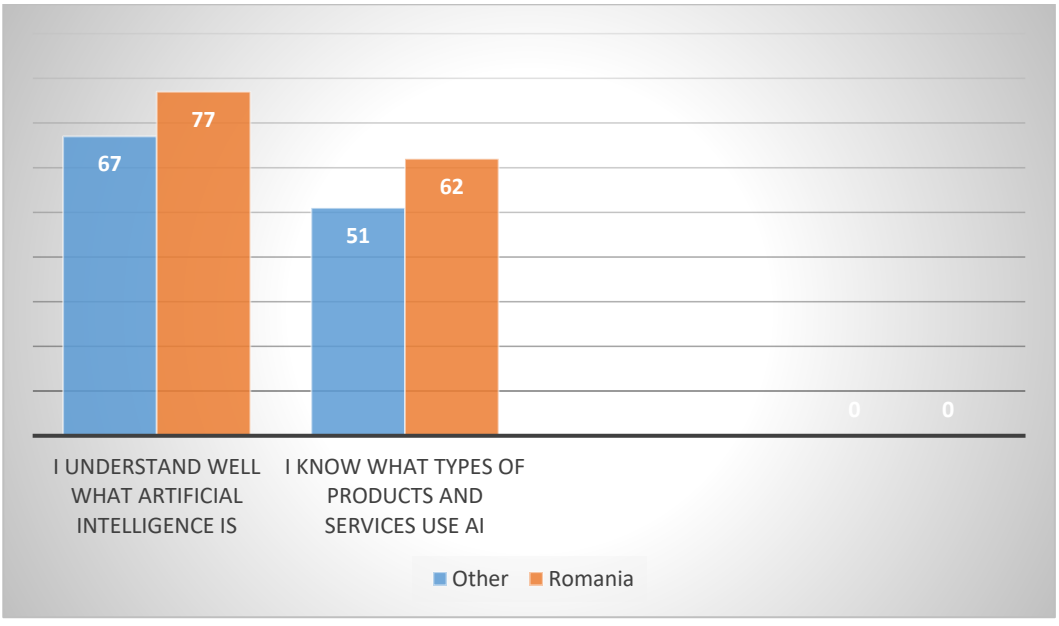

Fig. 4. What is data security?
Source: IPSOS

A total of 31 countries participated in this sampling, two-thirds 67% think they understand what artificial intelligence is, but only 51% also know what products and services use AI. This knowledge of products and services is increasing among adults, the employed, the educated, and the better-off.

Romanians think they know enough about AI, so Romania ranks 4th, with 77% of citizens claiming to know what artificial intelligence is.

In 2022, at least one Romanian out of twenty was the direct or indirect target of cyber attacks. According to INSSE, approximately 82% of households have Internet access, meaning that approximately 775,000 Romanians have received emails, messages, and calls or have been directly targeted by groups of hackers specializing in phishing, social engineering, or scamming campaigns.

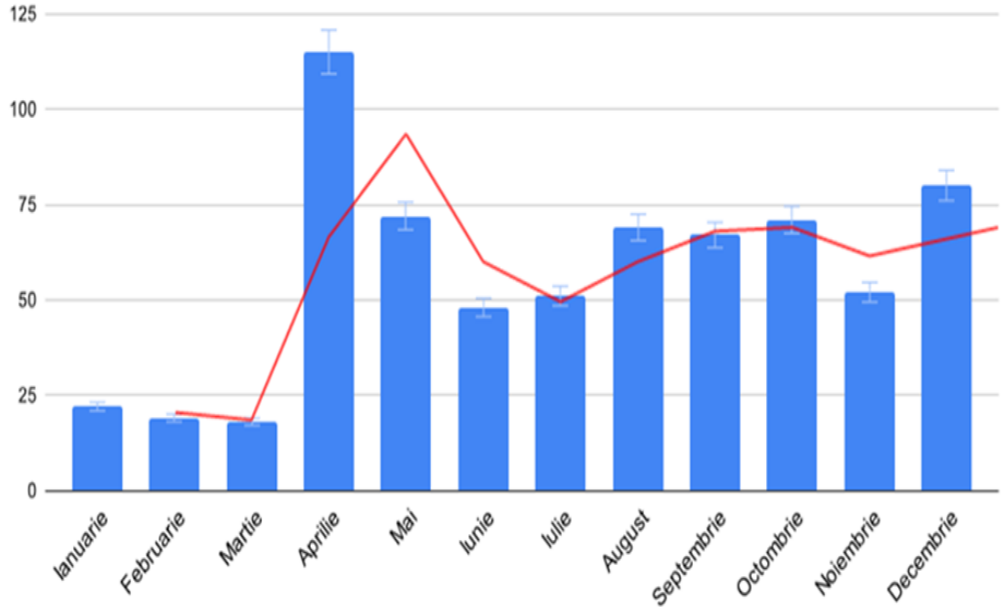

Fig. 5. Evolution of cyber attacks in 2022
Source: Hackout

Taking into account the fact that communication infrastructures are interconnected worldwide, cyber attacks [9] can be carried out from anywhere in the world, and at the same time, they can take place over very long time intervals and sometimes even without interruption, depending on the resources and technical capabilities and human that the attackers possess.

The last period has presented massive challenges regarding attempts to compromise email accounts. Such attacks are aimed at stealing access credentials (username and password) either to change them, block users' access to their own e-mails, or use the account to launch new attacks or obtain information available in these accounts.

Most of the time, these attacks present repeated attempts to identify the password or urgency, the obligation to do something in 30 seconds or minutes, which can send us an alarm signal. That is why it is advisable to have backups for important things in the digital environment, using external personal storage devices or in public clouds.

As published and recommended by the Special Telecommunications Services, an additional measure of protection is that access to various applications is protected with a second authentication factor. Second-factor authentication would make it impossible for the attacker to quickly access the account. It should only be available on the user's phone and should be changed at regular intervals.

Optimizing data protection at the individual level will also have positive consequences for the organization in which the individual works. Self-protection will be extremely useful and will provide long-term support in carrying out the activity.

5. Conclusions

SARS-COV-2 coronavirus has caused intense political, economic, and legal effects. The decree promulgated by the Government, which instituted the state of emergency, did not limit the right to the protection of personal data, at least not directly, as it would have had consequences on the right to private life according to Human Rights.

The actions of data protection supervisory authorities were put in different ways: most offered guides, and guidelines on the application of data protection rules in the context of labor relations, others made collages with the legal rules adopted during the emergency, and certain states have sanctioned operators for non-compliance with data protection rules.

Organizations must implement minimum protection measures against cyber threats. They consist of:

- Updating the IT and communication systems used;
- Implementation of authentication using 2 factors (2FA);
- Securing and monitoring services that present risks;
- Promotion of awareness campaigns and training of own users.

At the moment there are no miracle solutions that ensure 100% availability, integrity, and confidentiality of data.

Data security is a shared responsibility, it requires cooperation at the institutional level, between several categories of experts. This must be effective and materialize with the dissemination of information in real-time, as well as the exchange of knowledge and information.

References

[1] "Open security," [Online]. Available: http://www.opensecurityarchitecture.org/cms/definitions/it-risk.

[2] "Data security," [Online]. Available: https://www.microsoft.com/ro-ro/security/business/security-101/what-is-data-protection.

[3] World Health Organization, "Coronavirus disease (COVID-19) pandemic," 2020. [Online]. Available: www.who.int/emergencies/diseases/novel-coronavirus-2019.

[4] "Cybersecurity of EU institutions, bodies and agencies: Overall, the level of preparedness is not proportionate to the threats," [Online]. Available: https://op.europa.eu/webpub/eca/special-reports/hack-proofing-eu-institutions-05-2022/ro/.

[5] "Managementul riscului informatic," [Online]. Available: https://en.wikipedia.org/wiki/IT_risck.

[6] European Commision, [Online]. Available: https://commission.europa.eu/strategy-and-policy/coronavirus-response/digital-solutions-during-pandemic_ro.

[7] "Artificial intelligence in the context of data privacy," [Online]. Available: https://issuemonitoring.eu/inteligenta-artificiala-in-contextul-confidentialitatii-datelor/.

[8] "Artificial intelligence and GDPR - the impact of the use of AI on the protection of personal data," [Online]. Available: https://gdprcomplet.ro/inteligenta-artificiala-si-gdpr/.

[9] M. Simoes, M. Elmusrati, T. Vartiainen, M. Mekkane, M. Karimi, S. Diaba, W. Lopes and others, ""Enhancing data security against cyberattacks in artificial intelligence based smartgrid systems with crypto agility. arXiv preprint arXiv:2305.11652," 2023.

Unauthorized access control in water utility computer networks

Ioan Florin VOICU,
ING Hubs, Bucharest, Romania
ioan-florin.voicu@ing.com

Dragos Cristian DIACONU,
Bucharest University of Economic Studies, Bucharest, Romania
diaconudragos23@stud.ase.ro

Daniel Constantin DIACONU,
University of Bucharest, Bucharest, Romania
daniel.diaconu@unibuc.ro

Abstract
Virtual tampering in water utility systems can lead to highly dangerous real-world situations such as shortages and permanent damage to infrastructure. While cybersecurity guidelines do exist for Romanian companies like ApaNova, they are inadequate for protecting the water supply chain. Evaluating the potential vulnerabilities such systems have and presenting open-source methods to improve them is critical for the cybersecurity sustainability of utility services. Building on previous research regarding network cybersecurity, Kali Linux was used as a penetration testing platform in conjunction with an OPNSense-based network configuration. Initially the test included just the Apa Nova-mandated security settings (focusing on ransomware & database access protection), after which additional protective layers were added. The first extra layer was VLAN network segmentation, in compliance with Environmental Protection Agency (EPA)'s America's Water Infrastructure Act (AWIA) guidelines. Afterwards, additional settings were added, such as: Intrusion Detection Systems (IDS) & Intrusion Prevention Systems (IPS); Employee access only via Virtual Private Network (VPN) and Medium Access Control (MAC) address filtering for all employee Wi-Fi devices. A monitoring solution in OPNSense was also implemented, in order to be informed of any suspicious activity on the network. In conjunction with this, a patching strategy was created, which would minimize downtime, while ensuring the system is kept up to date. This is facilitated by the open-source nature of OPNSense, which does not need costly license upgrades to remain secure. The results showed that while protection against ransomware/viruses is important and relatively easy to implement, testing confirmed the findings of previous articles that malicious internal actors are an even greater threat than viruses. This requires constant protection and monitoring against privilege misuse by even authorized personnel. A wider view is offered on how easy it is to gain access to current systems and several off-the-shelf open-source software solutions are highlighted that can prevent water utility shutdown or misuse by malicious actors.

Keywords: Pen Testing, OPNSense, VPN, water management.

1. Introduction
Due to current private & state-sponsored hacking attempts on water, power & other such utility infrastructure, it's become ever more important to protect these assets from unauthorized remote access and ensure that any virtual threat is dealt with proactively by proper server patching & employee access procedures.

The hypothesis of this case study is that a water utility network with minimal cybersecurity infrastructure is relatively easy to penetrate by malicious actors and adding extra layers of network & authorization security greatly increases the difficulty of unauthorized access.

Whereas in the past obtaining proper network security required expensive licenses and specialized hardware, open-source projects such as OPNSense have lowered the cost and complexity of obtaining such benefits, while virtual machine platforms like Proxmox VE have enabled easier configuration and improved uptime for any of the operating systems it hosts.

Such platforms have standardized both the testing & roll-out of necessary systems, while ethical hacking Linux distributions like Kali have enhanced the capabilities and ease of penetration testing, which has been proven to improve the overall security evaluation of Wi-Fi networks [1].

Notable is also the fact that with newer hardware and OPNSense being able to use regular x86 processor platforms, the cost and performance penalty of IDS & IPS protection has significantly decreased, making enabling it less of a tradeoff than in previous years [2]. It is also important, however, to calibrate the monitoring & alerting capabilities appropriately to not have either "notification overload" or miss potentially important network events [3].

Ultimately, though, the security culture of the utility company is one of the best guarantors of its continued proper operation, with the incident and reputational cost of not applying best practices in this field far outweighing the occasional operational savings that could be made. This is especially important for smaller water utility operators, that may have an oversized impact, but do not have enough resources allocated for such purposes [4]. Open-source software would be an ideal way for these operators to have a widely-supported security infrastructure with crowdsourced threat identification.

2. Methodology

This case study attempted to simulate on a small scale the computer network layout of a water utility provider, creating a complete sample network infrastructure, as well as ways to access this infrastructure both locally and remotely.

The attempts to access the infrastructure initially mimicked the behavior of company employee devices, in order to not trigger anti-malware systems within the network.

The ease of unauthorized access was then compared at each step during the addition of multiple extra layers of security. The network thus progressed from being relatively trivial to penetrate from a nearby physical location to requiring separate VPN access rights, blocking access to the most critical areas by default and triggering the IDS/IPS notifications that had been added.

Just as importantly, the patching & downtime strategy that was implemented led to a significant increase in planned availability and a reduction in the number of downtime-causing incidents.

2.1. Case Study Hardware Setup

For this case study there needed to be several virtual servers, a Wi-Fi access point, a PC operating system client and a mobile operating system client. The choices made in terms of hardware were:

- An Intel Core i5 (7th generation) server/router/firewall with virtual machine support & 2 network cards
- A Ubiquiti Unifi AC Pro Wi-Fi access point
- A Lenovo ThinkPad T480 for both wired & Wi-Fi access from a PC operating system
- A Google Pixel 7 device for both Wi-Fi & 5G access from a mobile operating system

2.2. Case Study Software Setup

In terms of software setup, the architecture was designed around a server running Proxmox VE. This server would be the host for multiple virtual machines. One of these virtual machines was running OPNSense for router/DHCP/firewall/VPN duties, another was running the Unifi Network application for Wi-Fi access point management, while yet another was the target utility company application server for pen testing/exploiting, running the latest version of Ubuntu Server. The VPN plugin that was used in OPNSense was Wireguard, widely seen as the best-performing open-source VPN solution currently available.

For pen testing, the mobile ThinkPad T480 machine was used, running Kali Linux, the standard Linux distribution for such purposes. The machine's mobility allowed for testing the effectiveness of the security measures both in physical proximity to the simulated utility company's offices, as well as from within the network itself.
The testing was done as a multi-tier process, as the utility company's software defenses were also ramped up.

Initially, the testing was performed with a regular Unifi Wi-Fi configuration, with no MAC filtering & a shared WPA2 Wi-Fi access password. This was then enhanced to a Wi-Fi configuration with MAC filtering that only allowed all known employee devices.

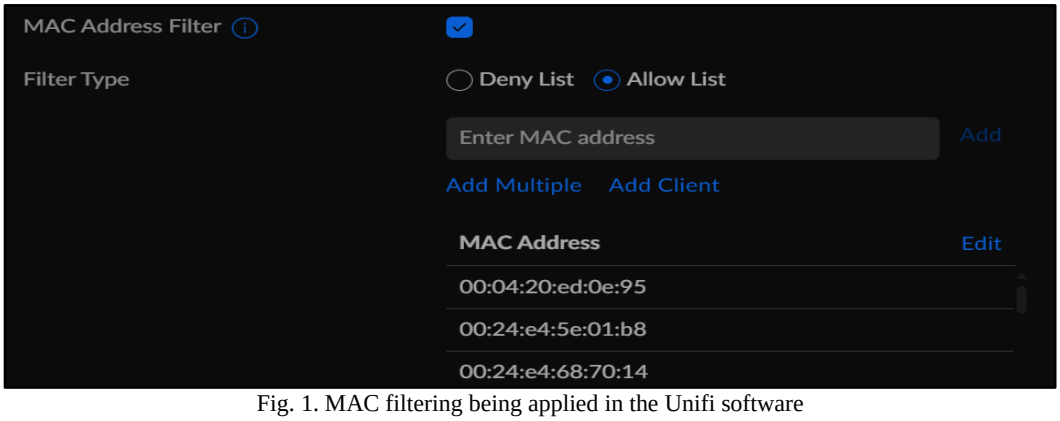

Fig. 1. MAC filtering being applied in the Unifi software
Source: author's testing

After the wireless pen testing results were obtained, OPNSense was deployed as a router, DHCP server & firewall. The first stage was without VLAN traffic separation, after which VLANs were created for the employee & server infrastructure networks.

This still allowed for unencrypted remote access, so a VPN solution was installed and made mandatory for all mobile employee devices.

The next stage after securing access was to monitor network traffic, for which OPNSense has a Suricata-based solution for IDS/IPS. To implement this, appropriate rule lists were downloaded & installed.

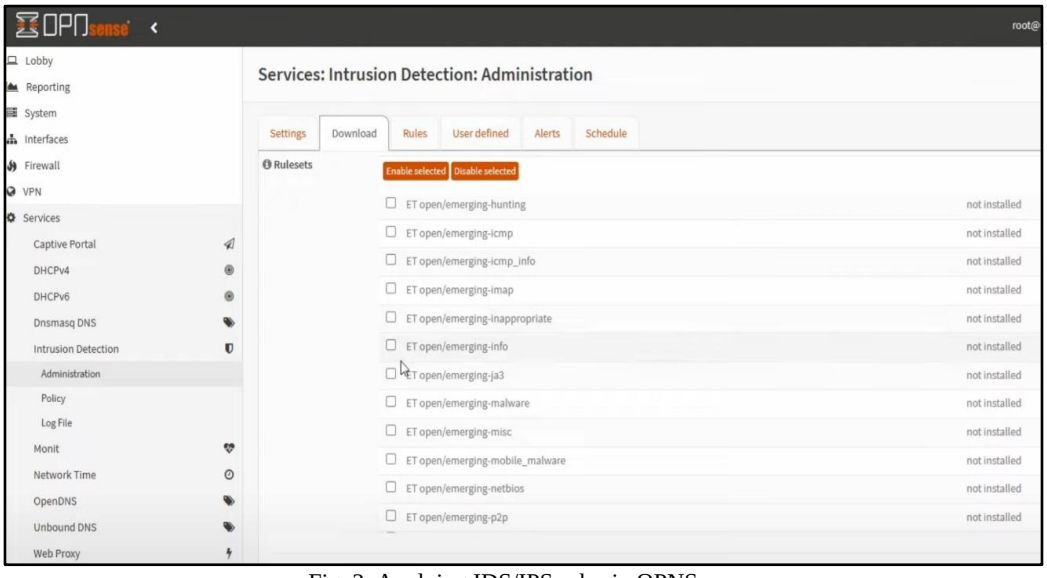

Fig. 2. Applying IDS/IPS rules in OPNSense
Source: author's testing

Additionally, a setup and procedure were created for keeping the servers up to date with the latest FreeBSD & Linux security patches, while minimizing downtime. This included the setup of a high-availability server cluster that would be load-balanced by the open-source solution Traefik (fig. 3). The result of such a setup would be the possibility of patching servers one by one without creating overall downtime for the water utility's network services.

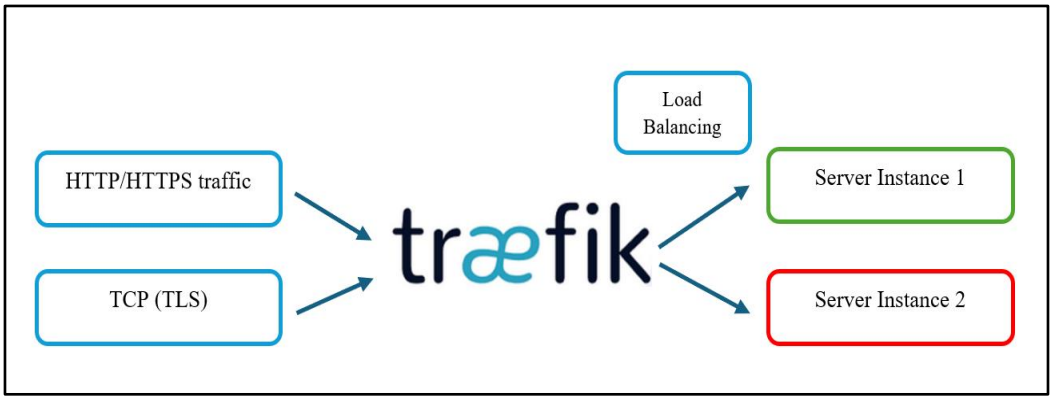

Fig. 3. Traefik load balancer architecture
Source: author's architecture

3. Results

Using Kali Linux on a mobile device in physical proximity to the simulated utility company's HQ, it was possible to scan for the access point SSID, identify it by signal strength and force-disconnect potential target devices that had been connected to it (fig.4).

Fig. 4. Identifying the SSID in Fluxion
Source: author's testing

While scanning for reconnect attempts, a connection hash was then obtained from one of these devices using the Fluxion software, which was afterwards used to compare to the inputted WPA password via the Aircrack-NG method (fig.5).

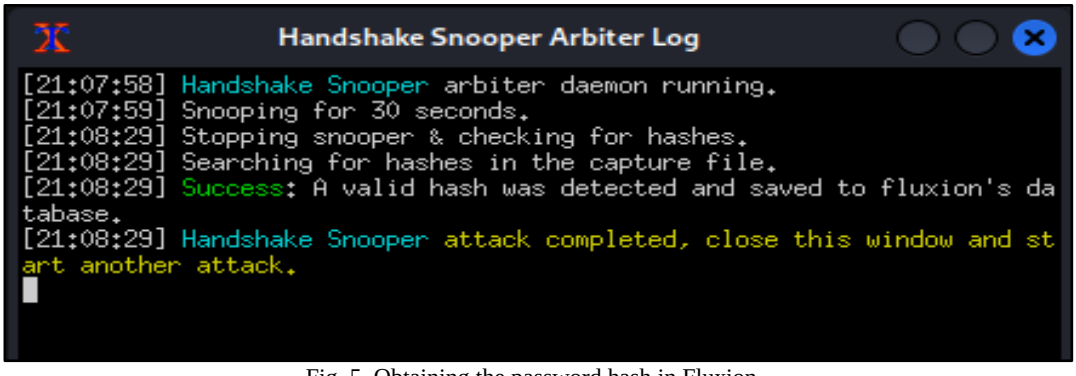

Fig. 5. Obtaining the password hash in Fluxion
Source: author's testing

Aircrack-NG was used to disconnect multiple devices via signal jamming, then spoof the SSID with the expectation that at least one person would succumb to social engineering. This was done by creating a new pop-up in which for them to input their WPA2 key (Fig.

6), which would appear on their device when trying to reconnect to the now malicious network.

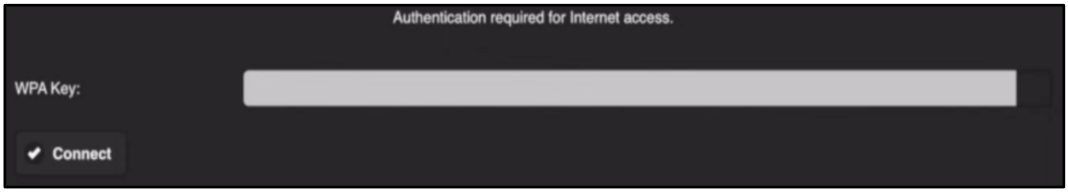

Fig. 6. Malicious popup created by Aircrack-NG for WPA password input
Source: author's testing

A user did input the WPA2 password there for our simulation's purposes, but even if it would not have worked, the hash obtained with Fluxion could be used offline for a brute-force password search until a match would be made, after which the resulting password could be used to connect to the network.

This result revealed the need for MAC address filtering, which was implemented on the network. However, Fluxion can also reveal the MAC addresses of nearby devices, so this only created another step which needed to be done in order to enter the network, i.e. MAC address spoofing for the malicious device. Once a correct MAC address was identified, the device was still able to access the network by spoofing the address.

Once able to access the network, the malicious device could browse across any part of it as a legitimate employee, including the IP address ranges reserved for the servers. VLAN creation in OPNSense managed to restrict this, segregating employee access from the critical network infrastructure (fig.7).

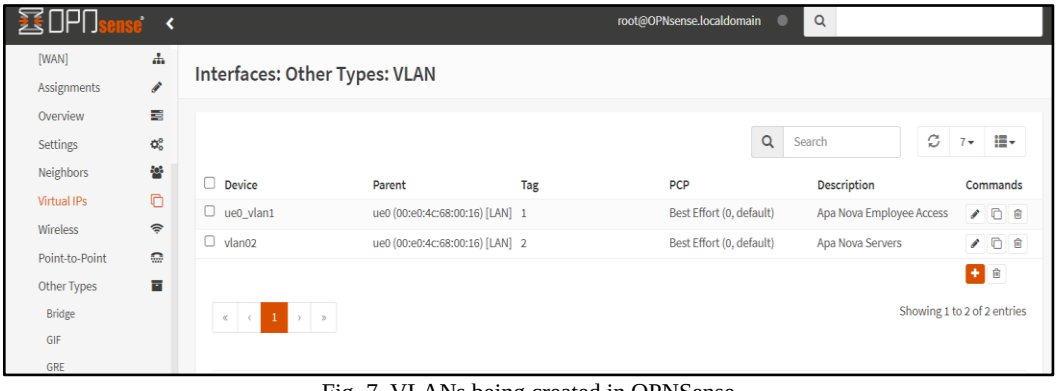

Fig. 7. VLANs being created in OPNSense
Source: author's testing

However, once inside the network, the malicious device was still able to use the packet-sniffing tool Wireshark to obtain access to the unencrypted data that was being transmitted from other users of the network. After making VPN usage mandatory for all remote users of the network, this issue was also mitigated, as connecting via Wi-Fi without also having a Wireguard VPN configuration file was no longer possible (fig.8).

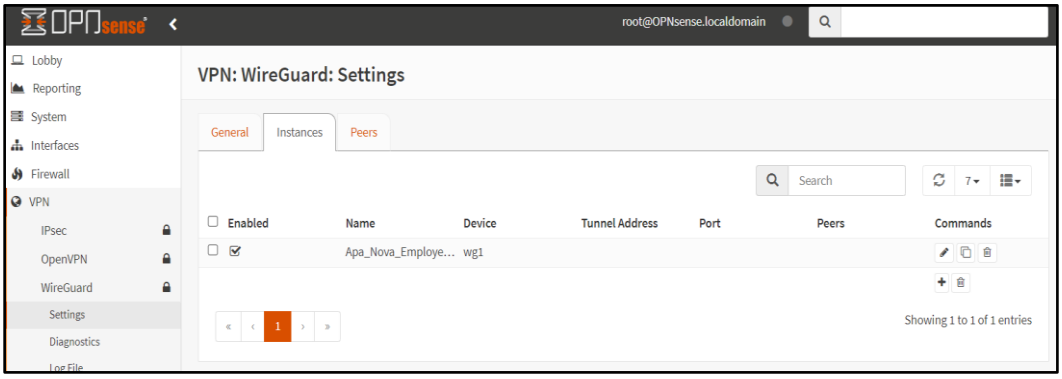

Fig. 8. Wireguard VPN setup in OPNSense
Source: author's testing

Moving on to the results of the shift to the high-availability server solution, over a 30-day period the improvement in downtime from going from the non-redundant server (Non-HA in Fig. 9) to the high-availability cluster (HA in Fig. 9) meant having just 3 downtime events instead of 11, cumulating 46 non-operational minutes instead of 102.

Moreover, the non-operational minutes were solely due to unforeseen technical incidents, not planned outages.

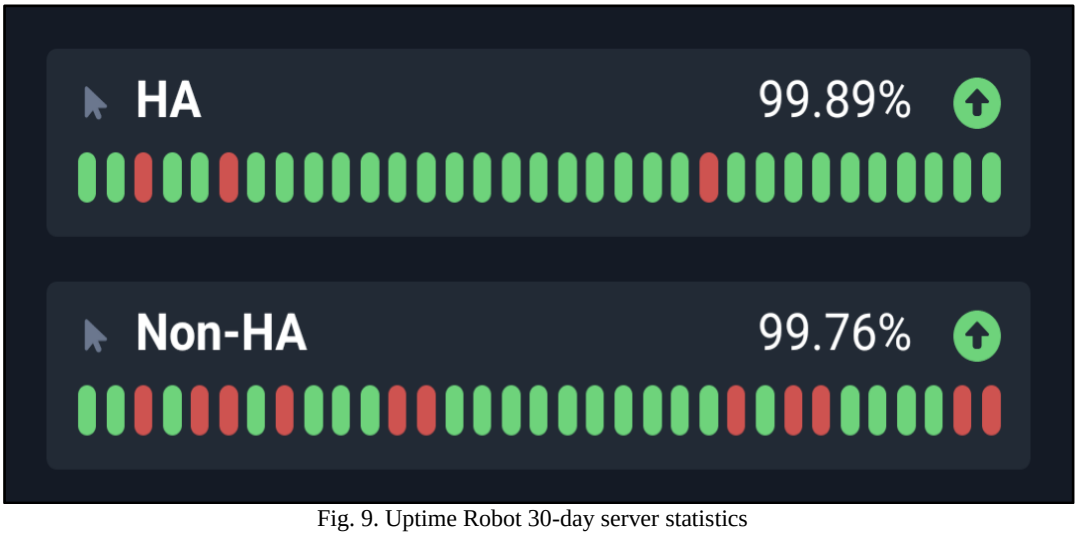

Fig. 9. Uptime Robot 30-day server statistics
Source: author's testing

4. Discussions

4.1. Applicability

The lessons learned during this case study can be applied to any utility or general computing networks that require increased security, decreased downtime & better monitoring of the implemented solutions.

The open-source nature of nearly all software involved also significantly lowers the Total Cost of Ownership and allows for more investment to be made in hardware (such as in adding redundant servers in a cluster) [5].

It should, however, be added that the implementation of any such solution without a constant process of maintenance in order to keep up to date with the latest types of threats would decrease its effectiveness with each passing day.

4.2. Limitations of case study

This case study was performed with real-world hardware & software, but in a simulated situation which only took into account the declared minimum compliance requirements as set forth by the EPA's America's Water Infrastructure Act (AWIA) guidelines [6] and Apa Nova's self-declared 2022 Sustainability Report [7].

While it is possible that water utility companies have higher level of protection or mitigations in place, even as of 2024 the EPA found that 70% of utilities in the USA inspected by federal officials over the past year violated standards meant to prevent breaches or other intrusions [8], being particularly vulnerable to state-sponsored cyberattacks.

Mentioned breaches include basic measures such as changing default passwords or cutting off system access to former employees, so protective measures as described in the case study still seem far from being the norm within these entities.

5. Conclusions

Given that the possible impact of cyberattacks may not just be limited to water service interruptions, but also alter the chemical balances in water treatment plants to dangerous levels, as was the case in Florida in 2021 [9], it is highly important that computer network access control in water utility companies is both properly implemented to begin with and monitored constantly.

It has been proven in the past that even a briefly unpatched server or an account left active can be used as an attack vector, so having a proper policy in place is crucial. Open-source software is a valid solution for lowering the costs and ease of implementation of such a policy and can also be used for post-incident mitigation [10].

It should also be mentioned that when using DPI for identifying potentially dangerous network traffic, recent encryption advances have made this task more difficult. However, it is possible to use AI/ML in order to better understand patterns that create risk within the network [11].

Not covered by this case study were even further steps that should be pursued, such as using dedicated 2-factor authentication keys for employees, which have practically eliminated the risk of phishing at large corporations such as Google [12].

Overall, though, even if not all the described measures are implemented by a water (or other) utility company, every additional layer has been proven to enhance security and improve IT governance and should be pursued if possible.

References

[1] He-Jun Lu și Yang Yu, „Research on WiFi Penetration Testing with Kali Linux," *Complexity*, vol. 2021, 2021.

[2] Niccolo Cascarano, Luigi Ciminiera și Fulvio Risso, „Optimizing Deep Packet Inspection for High-Speed Traffic Analysis," *Journal of Network and Systems Management*, pp. 7-31, 2011.

[3] Ying-Dar Lin, Po-Ching Lin, Viktor K. Prasanna, H. Jonathan Chao și John W. Lockwood, „Deep Packet Inspection: Algorithms, Hardware, and Applications," *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, vol. 32, nr. 10, pp. 1781-1783, 2014.

[4] Charlie King, „Cyber Security in the Power and Utilities Space," 2024. [Interactiv]. Available: https://cybermagazine.com/articles/cyber-security-in-the-power-and-utilities-space.

[5] Joshua M. Pearce, „Economic savings for scientific free and open source technology: A review," *Hardware X*, vol. 8, 2020.

[6] „America's Water Infrastructure Act," 2018. [Interactiv]. Available: https://www.epa.gov/ground-water-and-drinking-water/americas-water-infrastructure-act-2018-awia.

[7] „Raport de sustenabilitate Apa Nova Bucuresti," 2022. [Interactiv]. Available: https://www.apanovabucuresti.ro/assets/pdf/Raport-de-Sustenabilitate-2022-ANB.pdf.

[8] Associated Press, „Cyberattacks on water systems are increasing, EPA warns, urging utilities to take immediate action," 2024. [Interactiv]. Available: https://www.cbsnews.com/news/cyberattacks-on-water-systems-epa-utilities-take-action/.

[9] Jeff Pegues, „Feds tracking down hacker who tried to poison Florida town's water supply," 2021. [Interactiv]. Available: https://www.cbsnews.com/news/florida-water-hack-oldsmar-treatment-plant/.

[10] Tobias Roeder, „TLS 1.3, ESNI, ECH and QUIC: Taming new age cryptography with DPI and AI/ML-based encrypted traffic intelligence," 2023. [Interactiv]. Available: https://www.ipoque.com/blog/cryptography-with-dpi-and-eti.

[11] Ioan Florin Voicu și Daniel Constantin Diaconu, „Monitoring city water incidents via an Internet of Things-based sensor network," *Smart Cities International Conference (SCIC) Proceedings, Smart-EDU Hub*, vol. 10, pp. 207-214, 2022.

[12] Brian Krebs, „Google: Security Keys Neutralized Employee Phishing," 2018. [Interactiv]. Available: https://krebsonsecurity.com/2018/07/google-security-keys-neutralized-employee-phishing/.

Attacks against data security in smart cities: hypothetical scenarios or reality?

Irina-Ana DROBOT,
Technical University of Civil Engineering Bucharest, Faculty of Engineering in Foreign Languages, Department of Foreign Languages and Communication, Bucharest, Romania
anadrobot@yahoo.com

Abstract
The Objective is to show how changes in our lifestyle, which becomes more comfortable due to technological advancement, can lead to changes in which attacks, theft and various frauds can occur in smart cities. We need to be aware of risks. Prior work includes selected case studies and hypothetical risks scenarios, studied focusing not on the information, but on the way we relate to it. The 2018 Atlanta ransomware attack of the computer system of an entire smart city led to shutting down the municipal course, and to not allowing the citizens to pay their water bills and tickets for city traffic. In 2014, there was an attack on smart household appliances, worldwide, and home-networking routers were damaged. In Los Angeles, in 2016, an incident delayed work at MedStar hospital chain. The cyberattack caused data breaches in the healthcare industry, since medical equipment, such as pacemaker devices and MRI machines rely on information which expose them to risks. Hypothetical scenarios include cyberattacks damaging the way traffic functions. Approach: Data Security will be analysed from a psychological and cultural perspective, related to the anthropology of urban communities, and to the way in which individuals understand their right to personal space, data privacy, and the protective role of the cities. Case studies in previous research and news will be analysed. Results: Regarding the large-scale damage once cyberattacks occur, at the level of various areas of activity, affecting the entire city, what measures are taken by the European Union? Is digital democracy a simple utopia? Implications: Knowledge of previous experience means prevention for authorities and precaution for citizens. Value: Human beings need to cooperate with technology, but not passively. They need to be aware of its limits and not consider it a magical solution. Technology has shaped our urban culture and mindset.

Keywords: culture, values, mindset, practices.

1. Introduction
We can regard smart cities in the context of the phenomenon of the fast rate of urbanization [1]. Urbanization has been ongoing since the industrialization age, which was the first time when technological development reached a very high level, during the Victorian age. It continued further on, until today, until we reached the level of smart cities. Urbanization has meant a constant move of the population from rural to urban areas, to the point where the majority of the population will be living in cities. Since cities are the most frequent form of living and organization in today's world, and since urbanization is on the increase, we need to address the topic and to raise awareness regarding issues we are commonly confronting with today all over the world. We are, nowadays, organized in international, supranational and global communities, meaning that, next to our national specificities and forms of organizations, we can find urban culture which is present worldwide.

Cities are associated, especially, with "a better living environment" [1]. This means services and security. The comfort ensured by cities is related to the following services provided by cities: "water supplies and sewerage systems, residential and office buildings, education and health services and convenient transportation" ([2], mentioned in [1]). Cities, ever since ancient times, also meant protection for citizens, next to all the confort of daily living. We can be surprised at the high level of development of these services ever

since ancient times, yet the development is ongoing and adapting to the new possibilities of the cities, reaching the latest level of development, the smart city.

The smart city appeared as an answer to accommodate the needs of a continuously growing population [3], together with the organization and comfort it needs in case of a large population in the urban areas. These aspects lead to reliance on technology by local governments for ensuring the management of the city [3]. The management of the city in a smart city means using technology "to support a higher quality of urban spaces and a better offering of public services" [3].

Smart cities rely on technology, yet there are drawbacks to technology as much as there are advantages. One of the main issues can be, just as in the case of using computers and smartphones, and in the case of their software and applications, the security, theft and malfunctioning of data. It appears that thieves have adapted to the virtual environment, and to the possibilities offered by technology. Technology, through the use of various applications, has changed the way we relate to everyday life in the city, from using paid parking services to self-checkout, and to getting our public transport tickets, e.g. for bus and train, online, or using text messaging instead of relying on a seller. Even medical devices and smart household appliances, as well as traffic, can function in a smart city based on technology. Everything is programmed and the smallest problem with technology can ruin the functioning of a large segment of our lives in the city, if not the entire city.

We can notice how technology and culture, although at first sight unrelated, can be regarded as intertwined. Once we understand culture as an everyday life practice, we can see how technology has become one, in our daily lives. City life includes images of people of all ages with their smartphones, sending messages, checking social media, paying for various services, taking photographs and selfies. Contemporary art started to include selfies to the point where some art exhibitions, such as MoBU, recently organized as an art fair at ROMEXPO in Bucharest, Romania, during May 29-June 2, 2024, has considered telling visitors on a notice board to grab a doll and take a selfie. This is a form of interactive art, and selfies has also been included within various art pieces in the case of certain artists.

The way we use services in the smart city, which is a city relying on technology and which is developing all over the world continuously, can be considered part of our urban culture. We have come to perceive all this either as extremely comfortable, or as threatening and scary, once we do not know how to use it and once we have been confronted with various problems related to their use. For example, self pay in supermarkets has been, at least in some stores in Bucharest, Romania, quite recently a subject of debate in social media, as the self pay points were malfunctioning and did not help save time, but on the contrary, they made customers lose time. The problem was in a supermarket from the Auchan chain of stores, where all human cashiers were replaced by self checkout where customers had to scan their bought items themselves.

Still, these are just a few aspects of everyday life in a smart cities. We can also consider smart buildings, with plants and flowers on their surface, as well as with what look like gardens on top. Green spaces, and public parks are now the norm. We can consider how,

during the electoral campaign of 2024, before the elections in June in Bucharest, Romania, in district 6, Drumul Taberei neighbourhood there was an announcement about experimental green spaces with poppies where people were encouraged to take selfies.

Where do risks come in smart cities? Which data security issues can occur, even if, having in mind the comfort and utopia of everyday life in smart cities come in? Are citizens truly prepared for these data security attacks and why?

2. Materials and Methods

We can notice a contrast between the security of smart cities, which, through their use of high technology, can ensure "green environment and well-being for citizens" (R. P. Dameri, et al., 2013), social, "cultural needs" [4], entertainment, together with safety, promising a fantasy, utopic world, on the one hand and, on the other hand, the threat against the security of this very well-organized structure, revealing a distopic world, on the other hand. This contrast, however, regarding life in cities has existed since old times, when armed wars would threaten the stability of life in the city. We can speak of an adaptation to today's living conditions and technology when we consider attacks against data security.

Since a smart city includes optimization of "both tangible (e.g. transport infrastructures, energy distribution networks, and natural resources) and intangible assets (e.g. human capital, intellectual capital of companies and organizational capital in public administration bodies)" [5], the attacks against data security can target precisely these assets.

While life in a smart city can be associated, mostly, with entertainment and comfortable, peaceful, living conditions, once we look at cases that have occurred and at hypotheses regarding data security attacks, we notice how many risks are around and how many issues have actually occurred.

Examples of attacks against data security are presented in Fig. 1, taken over from a research paper [5]. As we can see, these attacks have targeted everyday life, housework appliances, such as smart refrigerators, tourists' hotels, and, thus, hotel businesses, in the case of a lock system, in most USA hotels, medical devices such as insulin pumps which would forget the right dosage after their batteries were changed and, therefore, could be dangerous to patients, the medical system, including surgery activity in hospitals, caused by devices that could be attacked since data was vulnerable due to its availability needed in order to purchase various devices, e.g. pacemakers and MRI machines, and, last but not least, passports, which could be copied using an RFID scanner, due to "unsecured wireless nodes" [5].

Fig. 1 below present each case, with its identifying details, in brief, including the year and place where they occurred, below the title line which helps identify the many situations that can become, from facilities and tools for a comfortable, highly-technologized life, sources of complete danger:

Fig. 1. Examples of Attacks on Data Security that have occurred in Smart Cities.
Source: Article shown in reference [5]

We can range the cases of attacks against data security, function of their consequences, according to the domains of activity and areas in life that they can affect, from mild to severe, while keeping in mind alternative solutions that may be found. If we look at the consequences related to the health and even life of patients in hospitals or people suffering from various conditions such as diabetes, we realize that technology has helped work in the medical field to become more and more efficient, and to help more patients in the course of a very short time by using high technology medical devices. The high technological development, through these high performance medical devices, can also help persons suffering from diabetes to have access to insulin pumps and be independent in their need for insulin through the very use of such devices. However, in Fig. 1 we notice that, once there is the slightest malfunctioning for various reasons, e.g. in the case of the battery exchange of insulin pumps, not only attacks against data security, then the lives of so many people dependent on and in need for treatment using these smart medical devices and needing them to be provided by the hospital will have their lives endangered. The health of the citizens is one important aspect of the security of any city. The malfunctioning of medical devices can be considered the most severe. Indeed, we notice how attacks against data security have led to an entire hospital chain to be "paralyzed" in its activity, failing to provide care, and how changing the batteries on insulin pumps caused them to reset and to no longer deliver the right dosage of insulin to the person suffering from diabetes, which could endanger their life [5].

High expenses, and high money loss can result for both businesses and individuals in case of attacks against the data security which can lead to the damaging of devices or stealing

of financial resources. We can see this in the cases related to the smart appliances attacks, the hotel lock system attack, and the stealing of data on passports, which can lead to stealing of goods and financial resources. New measures for protecting against stealing need to be taken, adapted to the current functioning of life in the city. While web cameras in the city and surveillance systems have been installed as part of security, in public places and institutions, and even in personal homes, and while these may discourage thieves to steal in the way they did in the past, new ways of stealing have been devised, using the very high technology that has helped discourage old forms of stealing. Even webcams are not present exactly everywhere, and pickpockets are still around, however. Security is never a solved issue, but an ongoing issue which is constantly in need for solutions.

The fact that a smart city's institutions function based on a computerized system can lead to issues which can cause an entire city to malfunction. As an example, in 2018, there was a ransomware attack of the computer system of Atlanta's City Hall [6, 7]. The citizens were affected, as they could not pay their water bills, and their tickets for city traffic. The computer system was compromised because of cybercriminals. According to Dean (2019), Atlanta is just one example of case of "ransomware against local governments". We can, therefore, notice several cases of this type, including the Baltimore ransomware attack, which occurred in 2019, once again targeting a large city in the USA, just like the Atlanta case. This could be seen as a sign of a frequently occurring problem, which needed attention. In the Baltimore ransomware attack, the following consequences were visible in the citizens' everyday life, which was functioning through a heavy use of high technology first of all, property transfer on the real estate market could not be operated, since the digital system had been shut donw, and the card payment system of the city was not functioning [8]. Neither was the debt checking application. City employees had to start creating Gmail accounts in order to use email sercices, as their email system was not working. However, the newly created Gmail accounts were blocked since so many new account had been created in such a short period of time. Therefore, they could not use Gmail either [9].

The activity in the smart city at public and personal level can, therefore, be hindered due to the data security attacks. Once we are so dependent on technology and computerized systems for government in the city and its used in institutions, we become all the more vulnerable in front of cyberattacks. The cyberattacks can, practically, stop all the normal course of activity and life in a smart city. Next to this immediate inconvenience, we can also expect personal data to be taken illegally and for the citizens to lose even more, especially money from their various cards and accounts.

Source [10] sums up the problem with attacks against data security in smart cities as being related to issues of trust of citizens and to the hindering of the efficiency of the entire technologically-based system, to the point where the smart cities are prevented from "the achievement of their full potential".

Source [11] devised a grid which can be used for analyzing cultures, which was called culture identity manifestations and which included the following categories of elements: traditions, rituals, practices, values, symbols, and personalities. We can fit in within this grid any culture, belonging to any country, as well as subcultures, and cultures which are

created based on supranational organizations and their instilled rules, norms, conventions, laws, and ideologically imposed lifestyle and values. From this point of view, once we understand culture as "patterns of thinking and doing" [11], then we can apply the culture identity manifestations grid to urban life and consider it a culture in its own right. The entire organization of the smart city can be seen as a reproduction of the culture identity manifestations grid.

Regarding values, we can include, as common concerns for people living in smart cities and for governments, together with the policies present at the level of supranational organizations such as the European Union, environmental care, transparent governance, citizens' involvement in political and social life, creativity, open-mindedness, social cohesion, etc. Security and trust regarding usage of data can also be considered a value. Among the symbols, we can include the recycling symbols, together with technology itself, symbolized through QR codes that can be scanned and by smartphones.

Fig. 2 shows an outline of categories based on which we can identify culture identity manifestations, regarding everyday life in smart cities, some of which were already mentioned:

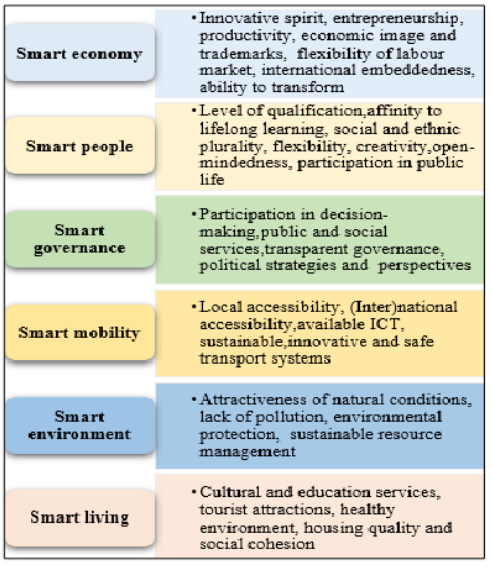

Fig. 2. Aspects of life in smart cities.
Source: [11], *shown in* [4]

We can look at the six axes which have been defined and used by [11] "to measure whether a Smart City is well-performing," and which consider the following dimensions: "Smart economy, Smart people, Smart governance, Smart mobility, Smart environment, and Smart living" [4], and identify culture identity manifestations in a smart city.

These axes include descriptions from where we can identify values which are further reinforced through rituals, traditions and practices, on special occasions, which may include pedestrian street events, where citizens are presented with cultural events samples,

under the form of street shows, e.g. acrobatics, circus elements, magic shows, theatre scenes, dancing scenes, presentations of contemporary artists showing their pantings in the street, street musicians performing live and offering their music CDs, as well as actitivities organized in parks during weekends where walking and using electric or usual scooters and bicycles are encouraged, fairs with handmade objects, natural and bio foods which are encouraged by European Union health policies, book fairs, etc. As far as environmental care is concerned, paying by credit card, and using less cash is encouraged, together with paying various taxes online and not receiving the invoice in print format by postal services. Less bureaucracy means less paper wasted, and more trees present, which can be made available through the digitalization of archives and through computerized systems. However, the security of all these practices and of all the infrastructure and organization of a smart city rely on a fundamental value, which should itself be reinforced through rituals, traditions, and practices, namely data security. Data security is ensured, at supranational level, through the General Data Protection Regulation (GDPR) established by the European Union, and which ensures privacy and individuals' image protection regarding sensitive data. Still, smart cities need to consider the security of personal data, as much as data that is related to the organization of the city by the government and various services, since they have proved in some cases to be vulnerable to cyberattacks. GDPR can be considered both a value and a practice, or a value that is reinforced through the practices represented by policies set up by the European Union.

Once data security is ensured, we can ensure trust of citizens in the authorities, which is always an important element for ensuring stability in the life of the city. As citizens, we project, psychologically, the role we ascribed to parents in our childhoods, namely that of a protector and of someone we trust. We expect protection and trust from leaders, and we are deceived and even angry if they do not comply with this role. Our childhood experiences shape our further relationship with the others, this time extended at the level of society. We can consider Freud's theory of infantile sexuality [12] and various fixations due to issues encountered during our psychological development. Once we have a more pronounced issue with authority, related to trust, it will resurface from childhood to maturity. In addition to psychological expectations, we have expectations created by ethics and conventions, as well as rules and laws, which ask of leaders to fulfill their duty towards the citizens.

Anthropology can explain the role of the leader as related to status, which can be understood as having "a position in a particular pattern," or as "a collection of rights and duties" ([13], mentioned in [14]). Expectations result from the rights, as well as duties, associated with the leader [14]. We deal, when it comes to a leader, with a role expected of him, therefore, and this role can be extended to the smart city itself. Ensuring data security can be associated by citizens with both government, with particular leaders and figures of authority, which can be included in the category of personalities from the culture identity manifestations grid, and with the smart city itself, which becomes itself a projection of the protective role of leaders and of stability of life, meaning ensuring data protection.

With the growing individualism [15] we expect to have ensured our personal space as individuals and to have data security ensured as well, as it can be considered part of our own, personal interest.

Today's mindset makes us willing to protect our personal data, as it is considered a part of our personal space. Everything is regulated through policies and rules, which legitimize these needs.

Additionally, as human beings we have the capacity to imagine and create preventive scenarios. One such hypothetical scenario with repsect to attacks against data security in smart cities is the one related to damaging the function of the traffic. The traffic lights are programmed and once cyberattacks occur, the entire traffic and life in the city could be disturbed, to the point where everyday life activity may no longer pe possible. This has far-reaching consequences, as institutions may not do their work properly once staff members are blocked in traffic, and public life may suffer. Public life has a strong influence on the way the city functions overall, which can disturb the routine and postpone certain activities and processes.

Smart transport system [16] may appear comfortable, and easy to use and operate, yet the danger comes when dealing with cyberattacks as well. We can consider the scenario in parallel with the way in which there is an accident in the city, or any deviation, traffic is delayed. Everything moves slower, and eventually all activity is going to be late.

3. Results

The following results can be identified as a consequence of the present research, and summed up as follows:

- Computerized systems can make some processes in public life in smart cities faster, yet they also expose the functioning of our everyday life to vulnerabilities related to data protection;
- Computerized systems can both ensure stability and instability in the city, function of whether they are not attacked or are attacked;
- Repeated cases of data security attack in similar situations may lessen the trust of citizens in government representatives and in leaders at the level of smart cities;
- Once citizens are no longer trusting authorities, then rebellion and protests can occur, which can destabilize life in the city even more than cyberattacks, to the point where usual activity can beak the natural flow of life activities in the smart city.

Smart cites can lead through the possibility actual actions to stable or unstable conditions. Once activity is delayed in smart cities, insability can can occur, as the natural course of everyday life may have included certain activities which, once postponed or cancelled, can disrupt the usual and expected routine, to the point where some activities can depend on others. In this latter case, the delays may depend on others delays, and entire segments of activities can be delayed. It all depends how urgent some actions are.

4. Discussion and Conclusions

The digital divide in the European Union is one of the frequently discussed topics. This refers to personal possibilities and access to technology, as well as to services that are publicly ensured. A smart city can ensure all these possibilities for its citizens, through providing public access to certain services. Once access to certain services is delayed, the entire city life and activity can be disturbed. It also depends on how soon certain actions can be retaken and on the extent to which activity can return back to normal. The frequence with which attacks occur can also establish the extent to which activities in the city can occur, namely their pace.

Smart cities bring along hopes for the better, as well as expected risks and threats to the stability of everyday life. Ensuring data security can mean a step further. Having citizens trust the leaders of the smart city also ensures harmony and efficiency in managing smart cities. Technology, as high as it can be, means that it is used by human beings, who need to be efficient in their jobs.

Examining previous real life cases and hypothetical scenarios previously discussed or imagined can create a basis for discussion and for reference regarding previous knowledge.

We can cooperate with technology, yet we can never consider it the final solution to our problems.

References

[1] C. Yin, et al., „A literature survey on smart cities," *Science China. Information Sciences,* vol. 58, nr. 10, pp. 1-18, 2015.

[2] K. Davis, „The urbanization of the human population," în *Menard S. W., Moen E. W., Perspectives on Population: an Introduction to Concepts and Issues*, Oxford University Press, 1987, pp. 322-330.

[3] R. P. Dameri, et al., „Searching for smart city definition: a comprehensive proposal," *International Journal of computers & technology,* vol. 11, nr. 5, pp. 2544-2551, 2013.

[4] S. Ouidad și T. Mazri, „Smart City Security Issues: The Main Attacks and Countermeasures," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 46, pp. 465-472, 2021.

[5] D. Popescul și L. D. Radu, „Data security in smart cities: challenges and solutions," *Informatica Economică,* vol. 20, nr. 1, 2016.

[6] G. Falco, et al, „A master attack methodology for an AI-based automated attack planner for smart cities," *IEEE Access,* vol. 6, pp. 48360-48373, 2018.

[7] C. Lamers, et al., „Ransomware: A Threat to Cyber Smart Cities," în *Cybersecurity for Smart Cities: Practices and Challenges*, Springer International Publishing, 2023, pp. 185-204.

[8] E. Stewart, „Hackers have been holding the city of Baltimore's computers hostage for 2 weeks, Vox," 2019. [Interactiv]. Available: https://www.vox.com/recode/2019/5/21/18634505/baltimore-ransom-robbinhood-mayor-jack-young-hackers.

[9] C. Lecher, „Google shut out Baltimore officials using Gmail after ransomware attack, The Verge," 2019. [Interactiv]. Available: https://www.theverge.com/2019/5/23/18637638/google-gmail-baltimore-ransomware-attacks.

[10] N. Neshenko, E. Bou-Harb și B. Furht, „Cyber Brittleness of Smart Cities," în *Smart Cities: Cyber Situational Awareness to Support Decision Making*, Springer International Publishing, 2022, pp. 19-40.

[11] S. Baciu, Culture: An Awareness-Raising Approach, Bucharest, Romania: Cavallioti Publishing House, 2012.

[12] S. Freud, Three essays on the theory of sexuality: The 1905 edition, Verso Books, 2017.

[13] R. Linton, The study of man: An introduction, 1936.

[14] G. Lang, „The Concepts of Status and Role in Anthropology: Their Definition and Use," *The American Catholic Sociological Review*, vol. 17, nr. 3, 1956.

[15] H. C. Santos, M. E. Varnum și I. Grossmann, „Global increases in individualism," *Psychological science*, vol. 28, nr. 9, pp. 1228-1239, 2017.

[16] R. I. Meneguette, R. De Grande și A. A. Loureiro, „Intelligent transport system in smart cities," *Springer International Publishing*, 2018.

A Romanian at the epicenter of controversy: the Guccifer case and its general impacts and destabilization of the American political scene

Mateo-Daniel DOGARU,
Police Academy "Alexandru Ioan Cuza", Bucharest, Romania
mateodogaru09@gmail.com

Abstract
In the digitalization era, the Guccifer case, involving Romanian hacker Marcel Lehel Lăzăreanu, has become emblematic of information sabotage. Operating under the pseudonym Guccifer, Lăzăreanu executed a series of high-profile cyber-attacks, targeting prominent organizations and political figures. His most infamous act involved hacking the email accounts of influential U.S. political figures and releasing compromising information on online platforms. These actions triggered widespread public confusion and unrest, significantly impacting the political landscape. By exposing confidential and sensitive information, Guccifer fueled speculation and controversy around various political candidates and organizations, thereby destabilizing the democratic process. The Guccifer case is a classic example of information sabotage, not only due to the exposure of secret information but also due to its strategic manipulation to influence public opinion and erode trust in democratic institutions. Guccifer's selective leaking and distribution of information allowed him to craft narratives that damaged the reputations of targeted politicians, fostering an atmosphere of uncertainty and distrust in the electoral system. This case has raised urgent questions about information security and integrity in the digital age. It underscored the fragility of modern political paradigms and demonstrated that threats to democracy can originate from the virtual realm. A single individual, equipped with technology and expertise, can cause significant damage, particularly in the realm of information sabotage. The Guccifer case thus serves as a stark reminder of the vulnerabilities inherent in our digital infrastructure and the critical importance of robust information security measures to protect democratic processes.

Keywords: information, democratic, digital, sabotage.

1. Introduction

Nowadays, everything is based on power, and often power comes from information. So the world is in a constant search for information, whatever the purpose or means. That's why some people try to get hold of sensitive information, which can affect a wide range of people, as in the case of information sabotage.

Information sabotage [1] refers to intentional and malicious activities aimed at disrupting, damaging, or manipulating information systems to achieve specific goals. This can include unauthorized data access, alteration or destruction, and dissemination of false information to undermine the credibility of organizations or individuals. These acts are a subset of cyber sabotage, encompassing any action designed to compromise the integrity and functionality of information and communication technologies.

Anyone can make use of this information sabotage, a relevant example being the Guccifer case that we will study throughout this paper from several perspectives. Marcel Lehel Lazar, known by his pseudonym "Guccifer" is a brilliant Romanian hacker responsible for serious attacks in the form of high-level computer security breaches in Romania and the United States of America.

Marcel Lehel Lazar, born on November 23, 1971, in the village of Sâmbăteni, Arad, became the most famous hacker without equipment and knowledge in the field. Before starting the so-called Guccifer "experiment", a pseudonym that comes from the iconic luxury brand Gucci and the angel kicked out of heaven Lucifer "Gucci's style and Lucifer's light", Marcel was a taxi driver without higher education and without high-performance equipment, but with exceptional intelligence. From behind a computer that at first glance looked like it wouldn't last much longer and a little Samsung smartphone (an old one), he managed to wreak havoc both at home in Romania and abroad, in the United States of America, which showed everyone how easy it is to be a "hacker" and how easy it is to be attacked yourself in our times, the times of digitization.

"He wasn't really a hacker, just a very smart, very patient, and very persistent guy," said Viorel Badea, the prosecutor who handled the case. Guccifer is also known for making public several self-portraits taken by former US President George W. Bush, for revealing to everyone the "flirtations" between Corina Cretu, a member of the European Parliament, and Colin Powel, and for obtaining numerous photos and private messages from national and international celebrities. "This is just a poor Romanian who wanted to be famous," Badea added.

Guccifer's list of attacks is long. In the first phase, in 2011, out of his desire to become famous in Romania, “The Little Smoke” as he was called back in 2011 launched attacks against well-known personalities in Romanian television such as Bianca Drăguşanu, Laura Cosoi, Corina Caragea, but also in the political arena in Romania, in 2013, personalities such as George Maior [2] (the head of the Romanian Intelligence Service between October 2006 - January 2015), Corina Cretu [3], Elena Udrea, Raed Arafat, Emil Boc, and many others.

After the success of his attacks in Romania, Guccifer wanted to expand his horizons to the realm of all possibilities, namely the United States. He launched attacks on major American political figures, managing to get his hands on sensitive information that destabilized elections and democracy in America. Some of the most famous personalities who have fallen into the hands of the hacker are Colin Powell [4], George W. Bush [5], Dorothy Bush Koch, Hillary Clinton [6], and many others. The attacker believes that everyone in power in America is corrupt, and everyone being controlled by Americans is controlled by corrupt people, that’s why Guccifer declared: "I hack the rich and powerful to show the world that they are not above scrutiny. Behind closed doors, they manipulate and deceive, but through my actions, I expose their corruption and bring transparency to the political elite."

2. Methods of attack leading to information sabotage

Guccifer has shown the world that you don't have to be an IT engineer or programmer to cause international damage or become a famous hacker. An old computer, motivation, and a lot of intelligence were behind his frauds that caused a lot of problems both locally in his home country and internationally, as mentioned above, in the United States.

He has managed to use hacking methods considered basic to create more problems than anyone expected. Among the methods used by Guccifer, we can mention: phishing, password guessing, social engineering, or account recovery processes.

Guccifer tried not to use these methods by their default meaning and that's why his hacking methods prominently featured social engineering tactics. While there is no specific documented instance of Guccifer using phishing techniques, his methods did involve elements of social engineering that are closely related to phishing or other methods above mentioned. For example, Guccifer used publicly available information to guess security answers and passwords, which is a common technique in phishing campaigns. He also sent emails that appeared to come from trusted sources to trick victims into revealing their credentials or allowing access to their accounts.

One notable method Guccifer used, which mirrors phishing techniques, involved gaining access to email accounts by exploiting weak security questions and publicly available personal information. For instance, he accessed the email account of Sidney Blumenthal, a close adviser to Hillary Clinton [7], by correctly answering security questions based on publicly available information about Blumenthal.

3. The beginning of Guccifer's highway to fame had its epicenter in Romania

The beginnings of the famous hacker were based in Romania. As we mentioned before, he attacked email accounts or tried to get sensitive data from several political personalities or personalities from Romanian show-biz. Next, we will analyze some important attacks that he initiated on some personalities in Romania.

"In this case there is a reasonable suspicion that, during 2013, on the basis of a single criminal resolution, the defendant L.M.L (Lazăr Marcel Lehel) repeatedly and unlawfully accessed, by violating security measures, e-mail accounts belonging to public persons in Romania, in order to gain possession of confidential data in the e-mail, after which he changed the authentication passwords, thus restricting the access of the right user to the computer data in the e-mail," a DIICOT spokesperson said in a press release about the case.

A big success for Guccifer was when he managed to hack the personal Yahoo account of the former director of the Romanian Intelligence Service (SRI), George Maior. During that time, in 2013, using the most basic forms of attack, Guccifer managed to break into the Yahoo account of the SRI director and reveal sensitive information that could have affected national security, so, the hacker created one of the most dangerous security incidents from Romania.

Guccifer conducted extensive research on George Maior, gathering publicly available information about his personal and professional life. By this, he could correctly answer the security questions that pop up when you try to access your, or in this case another email address or for password recovery. Many email accounts, including those on Yahoo, use security questions to facilitate password recovery. These questions often involve some basic personal information, such as the user's mother's maiden name, the name of a first pet, or the city of birth. Guccifer was able to find answers easily to such questions through

diligent online research. Guccifer could reset Maior's email password by correctly answering the security questions.

While the specific contents of George Maior's emails accessed by Guccifer are not publicly detailed, the hack itself is a significant example of the impact and risks of cyber-attacks on high-ranking officials. Guccifer's hack highlighted the weaknesses in the security measures used by public officials. His ability to breach the email account of a top intelligence official demonstrated the potential risks associated with insufficient security practices. These pieces of information offer a general understanding of Guccifer's methods and the broader implications of his hacks, even though specific revelations from George Maior's emails are not comprehensively detailed.

Guccifer did not only focus on attacking the high political or security pillars of the country. He also revealed sensitive information about personalities in the Romanian show biz, one of the victims being Corina Chiriac herself.

Guccifer's attack on Corina Chiriac was part of his broader hacking activities targeting various public figures. Corina Chiriac, a prominent figure in Romanian showbiz as a singer and actress, had her private emails and documents compromised by Guccifer. Using the same tricks as in George Maior's case, the leaked information about Corina Chiriac included personal communications, possibly revealing details about her personal life, career, and interactions with others, financial details, confidential correspondence, or any other data that was stored in Chiriac's compromised accounts. While the exact contents of the leaked information may vary and could include sensitive or private details, the specifics are not always widely disclosed or discussed publicly out of respect for individuals' privacy. Guccifer's hacking activities were illegal and unethical, and the leaked information could have potentially caused distress or embarrassment to those affected.

4. Guccifer's most impactful actions, ironically, took place in the realm of all possibilities, the United States of America

After the successive attacks against important personalities in Romania, Guccifer extended its horizons to big personalities in the USA, the attacks directed at them being also basic, with the same old equipment that at first glance seems useless but after being used properly it represented a powerful weapon against the American political plan, a weapon that managed to influence the subsequent elections.

The most significant information sabotage was directed at Colin Powell, the Former United States Secretary of State. The hacker on the morning of 11 March 2013 handiworked Colin Powells's Facebook page [8] and posted messages that disparaged, messages with vulgar contents like: 'You will burn in hell, Bush!' or 'Ass hole who would burn in hell for crimes purportedly committed along with Bush and Rockefeller family members'. In another post, Guccifer declared 'Kill the Illuminati! Tomorrow's world will be a world free of Illuminati or will be no more!'. After the former Secretary of State regained control of his Facebook page, he posted a message about the hack and tried to apologize to his followers for 'all the stupid, obscene posts that are popping up.' If that wasn't enough, instead of the messages, Guccifer uploaded to Powell's Facebook page screen grabs showing his prior access to e-

mail accounts of very important personalities like George W. Bush family members [8], including his siblings Neil and Dorothy.

Also, Guccifer's hack into Colin Powell's email uncovered a personal relationship between Powell and Romanian diplomat Corina Crețu. The leaked emails, spanning from 2010 to 2011, contained flirtatious and intimate messages between the two. While Powell denied having an affair, he admitted that their communication was personal. The release of these emails drew significant media attention and fueled speculation about the nature of their relationship.

The information sabotage started by Guccifer was about to get bigger and bigger. In his desire to find out sensitive information and destabilize the American political system he considered corrupt, he also breached the email account of Sidney Blumenthal, a former aide in the Clinton White House, and disseminated stolen memos sent to former Secretary of State Hillary Clinton regarding Benghazi. Guccifer uses basic but effective tactics, targeting the family and friends of his main objectives rather than going after them directly. By compromising their public email accounts, he exploits the relatively simple process of resetting passwords, which often requires answering a few personal questions. This method is not particularly challenging when the target is a celebrity. Guccifer's hack on Sidney Blumenthal, a close associate of Hillary Clinton, led to the exposure of a series of emails that Clinton had sent during her tenure as Secretary of State. While the hack was indirect (Blumenthal's email account was compromised, not Clinton's), it still revealed significant information like leaked emails including private discussions between Blumenthal and Clinton, providing insights into her thoughts and activities during her time as Secretary of State and also a conversation via email that contained sensitive information about the Benghazi attack, from Libya 2012. Also, we can talk about potential conflicts of interest because those emails highlighted some of the Blumenthal’s business interests in Libya.

These leaks contributed to the scrutiny and controversy surrounding Clinton's use of a private email server for official communications, which became a significant issue during her 2016 presidential campaign.

The brilliant hacker broke into the AOL accounts of Bush family pals Willard Heminway and CBS sportscaster Jim Nantz, as well as Dorothy Bush Koch, the sister of George W. Bush and the youngest child of George H.W. Bush. He obtained a wealth of private material about the Bush family, including correspondence between the two former presidents, through the penetration of other accounts. However, the hacker also gained access to private emails, documents, and images from Dorothy Bush Koch's account. For instance, the hacker obtained contact between Scott Pierce, the 82-year-old brother of Barbara Bush, and the 87-year-old former First Lady by breaking into his AOL account. Additionally, Patricia Legere, a friend of the Bush family and a former Miss Maine, and Josephine Bush, the 41st president's sister-in-law, and mother of Access Hollywood host Billy Bush, had their Comcast email accounts compromised by the hacker.

The hacker gained access to private information on the whereabouts, ailments, and travels of the Bush family thanks to the unauthorized incursions. Although the hacker managed to

gain access to AOL and Comcast accounts, it appears that they were not able to breach the personal email accounts of the two former presidents, who use distinct domains for their official post-presidential correspondence.

The hacker withheld information about the purpose of the attacks and the methods used to engineer them across several months of email discussions. However, it appears likely that the hacker's examination of previously hacked email accounts helped identify certain targets. The targeting of individuals of the Bush family's inner circle most likely makes sense given this daisy chain strategy. After falling in sequence, one of the hacker's victims thought their email account was a "domino."

5. Guccifer's payment for all his actions

Ironically said his payment, more concretely we will talk about what punishments Guccifer got for all his attacks so for all the private or classified information he stored or made public.

In Romania, in 2014, Guccifer was arrested, at the proposal of DIICOT Arad, and sent to trial for violating the correspondence of several public figures in Romania, including MEP Corina Cretu and even SRI Director George Maior. In the same year, the Bucharest Court sentenced him to 4 years in prison, plus the 3 years suspended from his previous sentence. The sentence of 7 years remained final and the man from Oradea was then imprisoned in Arad Penitentiary. Initially jailed in 2014 in the country after pleading guilty to charges of unauthorized access to a protected computer system and aggravated identity theft, Guccifer was eventually extradited to the US. The man spent four years in a Pennsylvania prison.

‘Marcel Lehel Lazar, 44, of Arad, Romania, a hacker who used the online moniker “Guccifer,” was sentenced today (September 1, 2016) to 52 months in prison for unauthorized access to a protected computer and aggravated identity theft [8].’ He pleaded guilty before U.S. District Judge James C. Cacheris of the Eastern District of Virginia on May 25, 2016. In exchange for a plea deal, Lazar admitted that from at least October 2012 to January 2014, he willfully obtained unauthorized access to about 100 Americans' [5] personal email and social media accounts to receive their personal information and correspondence. Lazar said that among his victims [9] were members of the immediate families of two previous US presidents, a former presidential advisor, a former member of the US Joint Chiefs of Staff, and a former member of the US Cabinet. According to the statement of facts submitted with his plea deal, Lazar frequently made his victims' private email communication, financial and medical records, and personal photos available to the public. The case was looked into by the Secret Service, DSS, and FBI. The case is being prosecuted by Assistant U.S. Attorneys Maya D. Song and Jay V. Prabhu of the Eastern District of Virginia, as well as Senior Counsel Ryan K. Dickey and Peter V. Roman of the Criminal Division's Computer Crime and Intellectual Property Section. The Office of International Affairs of the Criminal Division rendered noteworthy support. The Romanian government's help in this situation is much appreciated by the Justice Department.

After serving his sentence (August 2021), in an interview with The Intercept, the renowned Romanian hacker said several things about both the people he learned about and his actions.

He made a statement about Bush's unveiled self-portraits: ‘Thanks to Guccifer's infiltration of Dorothy Bush Koch's AOL account, the world now knows that her brother, George W. Bush, likes fine self-portraits in the bathroom,’ writes The Intercept.

Also, he declared: ‘I paid for it. People have to have privacy. But it's not like I want to know what my neighbors are talking about. I wanted to know what these guys in the United States were talking about and that's why (I resorted to this gesture - ed.). I was sure bad things were happening. That's the reason I did it, not for any other dubious reason. What I did is okay,’ Among other things, Guccifer said that he was disappointed that although he helped uncover Hillary Clinton's private email and then cooperated with federal officials, he was the one who ended up in jail, while she was not charged in connection with what he exposed. He also claimed that he came close to breaking into Trump's "inner circle" in October 2013. ‘I was about to break into Trump's boys, Ivanka, and others,’ he claimed. ‘But my computer broke...’

Guccifer declared related to The Intercept article that: ‘The article in The Intercept doesn't contain a lot of detail, it doesn't capture much of the turmoil I experienced. In the book, however, I went into a lot of things, much more deeply. The 300 pages I've written have already drained me of energy,’. ‘There's some incredible stuff in this book, because I'm splitting the difference,’ Guccifer said, predicting an ‘explosive’ return to the public arena.

6. Final conclusions

Guccifer conducted his hacking for a variety of political and personal motives. His primary goal was to reveal wrongdoing, call attention to abuses of authority, and highlight the weaknesses of powerful individuals.

Guccifer has been able to gather and divulge to the public sensitive and compromising information about a number of prominent political and public figures in the United States through his attacks. These include Sidney Blumenthal, Colin Powell, and other members of Hillary Clinton's inner circle's private correspondence. The public perception of the victims has been harmed by these revelations, which have sparked media scandals.

The political climate has been significantly impacted by Guccifer's acts, which have also brought attention to the significance of cyber security. His hacking operations have unintentionally exposed security flaws and breaches, which has led to informatical sabotage. This has highlighted the need for more stringent data protection measures and, consequently, raised awareness of the risks associated with cyberspace in the digital age.

During the 2016 presidential election, several people perceived Guccifer's actions as having the ability to sway public opinion and undermine the Democratic Party's campaign, which was perceived as a side benefit for Donald Trump's campaign. Even though Guccifer didn't have a direct relationship with Trump, the revelations stoked uncertainty and fueled political controversy.

References

[1] "Britannica ‘sabotaje subversive tactic‘," [Online]. Available: https://www.britannica.com/topic/sabotage-subversive-tactic . [Accessed 2024].

[2] "HotNews.ro, Vlad Barza, ‘Hackerul "Guccifer", care i-a spart contul sefului SRI, George Maior, a fost prins la Arad’," [Online]. Available: https://www.hotnews.ro/stiri-esential-16457219-hackerul-guccifer-fost-prins-arad.htm . [Accessed 2024].

[3] Adevărul.ro, "Octavian Palade '„Micul Fum“ şi marele noroc. Cum a reuşit Guccifer să spargă contul Corinei Creţu şi să bage spaima în familia Bush’," [Online]. Available: https://adevarul.ro/stil-de-viata/tehnologie/micul-fum-si-marele-noroc-cum-a-reusit-guccifer-1577962.html. [Accessed 2024].

[4] The Smoking Gun, "Colin Powell Facebook Page Was Hacked By Same Perp Who Broke Into Bush Family E-Mail Accounts," [Online]. Available: https://www.thesmokinggun.com/buster/colin-powell-guccifer-facebook-hack-467842 . [Accessed 2024].

[5] Independent, "Feliks Garcia, ‘Notorious hacker 'Guccifer' pleads guilty to hacking George W Bush and 100 others’," [Online]. Available: https://www.independent.co.uk/news/world/americas/hacker-guccifer-pleads-guilty-george-w-bush-hillary-clinton-emails-a7049001.html . [Accessed 2024].

[6] The Intercept, "Sam Biddle, ‘SORRY, NOT SORRY Guccifer, the Hacker Who Launched Clinton Email Flap, Speaks Out After Nearly a Decade Behind Bars’," [Online]. Available: https://theintercept.com/2023/01/15/guccifer-interview-hacked-clinton-emails/. [Accessed 2024].

[7] Wikipedia, "Guccifer," [Online]. Available: https://en.wikipedia.org/wiki/Guccifer#Arrests_and_convictions_in_Romania. [Accessed 2024].

[8] Office of Public Affairs U.S Department of Justice, "Romanian Hacker “Guccifer” Sentenced to 52 Months in Prison for Computer Hacking Crimes," [Online]. Available: https://www.justice.gov/opa/pr/romanian-hacker-guccifer-sentenced-52-months-prison-computer-hacking crimes.

[9] The Smoking Gun, "Bush Hacker's Victims Include U.S. Senator," [Online]. Available: https://www.thesmokinggun.com/documents/internet/bush-hackers-other-victims-637098. [Accessed 2024].

Cybersecurity: information and defence against data phishing

Cristiana SÎRBU,
University of Agricultural Sciences and Veterinary Medicine Bucharest, Faculty of Land Improvement and Environmental Engineering, "Gheorghe Ionescu Șișești" Academy of Agricultural and Forestry Sciences, Soil Science, Land Improvement and Environmental Protection Section, The Ecological Initiative and Sustainable Development Group Foundation
cris_sirbu@yahoo.com

Abstract
The paper presents the knowledge and information gained from participating in various events and workshops dedicated to cyber security. This highlights the need for information ownership in order to protect oneself in the jungle of data phishing, in a world that is constantly changing and where digital technology is the main means of conducting business in a modern society. Threats are evolving day by day and are difficult to contain, and this has highlighted the need for cyber security and cyber defence regulations. In the current context of digitization and technology, cybersecurity is a priority for every state.

Keywords: technology, digitization, security, strategy.

1. Introduction
Almost everyone and everywhere is talking about cyber defence, but it is a huge amount of information and thousands of years of work by programming and programming teams.

Threats in this vast field are evolving every day. It's an unmanageable avalanche that can swallow up the data protection work of super-professional teams like a giant in a fairy tale in seconds.

Cyber-attacks and cybercrime are growing in number and sophistication across Europe. The future is uncertain, with the trend expected to continue to grow.

2. Results and discussions
A number of critical sectors such as transport, energy, healthcare and finance have become increasingly dependent on digital technologies to run their core businesses.

Digitalization offers enormous opportunities and provides solutions to many of the challenges facing Europe.

Data phishing is the most common method of theft. Data phishing is a technique whereby an attempt is made to obtain sensitive data, such as bank account numbers, through a fraudulent request by email or on a website, where the perpetrator poses as a legitimate business or trusted person. This method of obtaining information can seriously damage both financial and reputational damage to corporations, institutions, business, academia and the many sectors that use digital technology [1].

Cyberspace is characterized by the absence of borders, creating opportunities for the development of the knowledge-based information society and risks to its functioning.

The more computerized a society is, the more vulnerable it is, and ensuring the security of cyberspace must be a major concern for all actors especially at the institutional level, where the responsibility for developing the security and implementation of coherent policies [2].

Cyber security is the application of technologies, processes and controls to protect systems, networks, software, devices and data from cyber-attacks. It aims to reduce the risk of cyber-attacks and protect against unauthorized exploitation of systems, networks and technologies.

The European Union Directive on cyber security measures, known as the NIS Directive2 (by Law No 362/2018) on ensuring a high common level of security of networks and information systems, in conjunction with the provisions of Article 72 of the Treaty on the Functioning of the European Union have become obvious national responsibility and require regulations in the field of cyber security and cyber defence.

The EU cyber security strategy aims to strengthen the Union's resilience to cyber threats and to ensure that all citizens can benefit from trusted digital services.

In October 2020, at the Extraordinary Meeting of the European Council, EU leaders called for strengthening the European Union's ability to: protect itself against cyber threats, ensure a secure communication environment and ensure access to data for law enforcement and judicial purposes.

On 22 March 2021, the Council adopted conclusions on the Cybersecurity Strategy, which underline that cyber security is essential for building a resilient, green and digital Europe [3].

As a dimension of national security, the issue of cyber security and defence has become a priority.

Romania's Cyber Security Strategy for the period 2022 - 2027 and the Action Plan for the implementation of Romania's Cyber Security Strategy for 2022 - 2027 must ensure complementarity with the provisions of the European Union but also with a number of domestic ordinances (e.g. Emergency Ordinance 104/2021 establishing the National Cyber Security Directorate and at the same time with measures on cyber defence with specific aspects manifested in cyberspace). So, we are on the right track.

Romania's National Recovery and Resilience Programme (NRRP) has taken on the implementation of the measure "Ensuring the cyber security of public and private entities that own critical infrastructures.

Cooperation between the relevant institutions and civil society, academia and the private sector is the basis for the development of effective cybersecurity partnerships and the legal and institutional framework for organizing and carrying out cybersecurity and freedom defence activities.

Public international law puts conceptual uncertainties and differences in interpretation under scrutiny, particularly with regard to attribution of malicious cyber activities.

Concepts, solutions from the international environment that are related to the realities and specificities of legislation and institutions in the field, the creation of networks and systems accompanied by the adoption and development of a regulatory and institutional framework strengthen confidence in our common cyber future, both in the European Union and in Romania.

By ensuring resilience through a proactive approach and deterrence, Romania becomes a relevant player in the international cyber security cooperation architecture.

3. Conclusions

As a result of actively participating in workshops and various events, I have learned some of the essentials of modern cybernetics:

- Collective intelligence, which is not all one with the IT-ist, is about perfecting systems where the citizen is not harassed, is informed and protected;
- Access to technical information must be restricted;
- Malicious' operators see it as a weapon and the cybersecurity system sees it as an art;
- Threats evolve daily.

Security is not the pinnacle of some companies it is the prerogative of us all.

Cybersecurity is considered an extremely important part of national security. There is therefore a need to develop a cyber security culture among users of information and communication systems, who are often insufficiently informed about potential risks and solutions to counter them [4].

Widespread knowledge of the risks and threats to which cyberspace activities are exposed and how to prevent and counter them requires effective communication and cooperation between the specific actors in this field.

Acknowledgements

The study was conducted by The "Ecological Initiative and Sustainable Development Group" Foundation with the main purpose to be a connection between institutions, academia and civil society in order to inform and facilitate the access to information.

References

[1] "Digital Innovation Summit Bucharest," 16-18 April 2024.

[2] *European Union Agency for Cybersecurity, Romanian Cybersecurity Strategy.*

[3] in *DigitALL 2023 Conference*, 2023.

[4] "ZF Cybersecurity Trends 2023: How many cyber defence solutions does a company need and who should "run" them?," 29 May 2023.

Navigating face recognition technology: A comparative study of regulatory and ethical challenges in China and the European Union

Ina VIRTOSU,
PhD in EU Law, University of Macau, SAR Macau, China
yb67199@connect.um.edu.mo, ivirtosu3@gmail.com

Chen LI,
PhD in Law, Southwest University of Science and Technology, Centre for Latin American and Caribbean Studies, Mianyang, China
yb77204@um.edu.mo

Abstract
Face recognition technology, while advancing rapidly, presents unique challenges in both China and the European Union (EU). This comparative study explores the distinct regulatory, ethical, and social obstacles each jurisdiction faces. In China, the widespread implementation of face recognition is facilitated by a supportive regulatory environment and a societal emphasis on security and surveillance. However, this has raised significant concerns regarding privacy, data security, and the potential for misuse by the authorities or private entities. In contrast, the EU's stringent data protection laws, particularly the General Data Protection Regulation (GDPR), impose rigorous constraints on the deployment of face recognition technologies. These regulations aim to safeguard individual privacy but also create hurdles for technological advancement and implementation. Furthermore, public skepticism and ethical considerations in the EU limit the adoption of face recognition. This paper highlights the dichotomy between China's rapid technological adoption with lesser regulatory constraints and the EU's cautious, privacy-centric approach, highlighting the need for a balanced framework that can navigate the ethical implications and privacy concerns while fostering technological innovation and addressing societal security needs in both regions.

Keywords: valid consent, GDPR, biometric data, bias issues, PIPL.

1. Introduction

Facial recognition technology (FRT) stands at the intersection of rapid technological advancement and profound ethical debate, particularly in its varied application across different global regions. This technology, which enables the identification and verification of individuals based on their facial features, offers potential benefits for security, efficiency, and convenience. However, it also raises significant concerns regarding privacy, data security, and civil liberties. The regulatory, ethical, and social challenges associated with FRT are pronounced, and they manifest differently in diverse geopolitical contexts. This article explores these challenges by comparing the approaches of China and the European Union (EU).

This article delves into the dichotomy between China's rapid, regulation-light deployment of FRT and the EU's stringent, privacy-focused regulatory environment. By examining the regulatory frameworks, societal attitudes, and ethical considerations in each region, this study aims to highlight the broader implications of these differing approaches. The comparison illuminates the need for a balanced framework that can navigate the ethical implications and privacy concerns of FRT, while also fostering technological innovation and addressing societal security needs.

Ultimately, this article seeks to contribute to the global dialogue on FRT by offering insights into how diverse regulatory landscapes shape the deployment and societal impact of this technology. It underscores the importance of developing balanced policies that harmonize the benefits of FRT with the imperatives of privacy, ethical standards, and social trust.

2. Definition of FRT in EU laws

According to Guidelines 05/2022, adopted by the European Data Protection Board (EDPB), facial recognition is considered a probabilistic technology that can automatically recognise and authenticate persons based on their facial features [1]. FRT belongs to the wider area of biometric technologies. Under Article 4(14) GDPR, biometric data is defined as "personal data resulting from specific technical processing relating to the physical, physiological, or behavioral characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data". The fact that facial images also constitute personal data was confirmed by both, the ECtHR [2] and the TCJEU [3]. The ECtHR has also stated that a person's facial image constitutes one of the key attributes of his/her personality, as it reveals the person's unique characteristics and distinguishes one person from another. The right to the protection of person's facial image is the essential components of personal development [4].

Using FRTs implies collecting, comparing or storing facial images for identification and authentication purposes, for border control, searching for people on police watch lists or tracking someone's activities in public places. FRT verifies a person's identity by examining the specific qualities and features of their face, in other words, biometric data to identify and/or verify a person's identification against previously recorded information [5].The use of AI-powered FRTs deploy more elaborate technologies and algorithms, involving the collection, storage, and processing of biometric data, which is considered highly sensitive under the GDPR (Article 4(13), (14) and (15) and Article 9) [6], but also under Law Enforcement Directive (LED) (Article 3(13) and Article 10) [7].However, LED is a more specialized regulation compared to the GDPR, so-called lex specialis, and applies specifically when public authorities handle personal data for the purposes of preventing, investigating, detecting, or prosecuting criminal offenses (Recitals 11 and 12 LED, and Recital 19 GDPR).

FRT is also regulated by the recent approved Artificial Intelligence Act (AI Act) [8]. The AI Act is the first of its kind in the world and it applies to the development, deployment, and use of AI in the EU or when it will affect people in the EU. AI Act covers all types of AI across a broad range of sectors, with exceptions for AI systems used solely for military, national security, research and non-professional purposes [8]. The AI Act categorizes AI applications not exempted from its regulations based on the potential harm they may cause, which range from unacceptable to high, limited, and minimal risk, with an additional classification for general-purpose AI [8]. Any applications posing unacceptable risks are prohibited, except in cases with specified exemptions. The EU AI Act forbids specific applications that influence individuals' choices or take advantage of their weaknesses, systems that assess or categorize individuals based on their social conduct or personal characteristics, and systems that forecast an individual's likelihood of engaging in criminal

activity [8]. Additionally, Article 5 (2) includes banning the use of "real-time remote biometric identification systems", i.e. AI systems from harvesting facial images from the internet or surveillance footage, deducing emotions within workplace or educational settings, and classifying individuals based on their biometric information [8]. This appears to encompass many algorithmic video surveillance applications. However, the restriction can be circumvented if the use of such systems is not conducted in real-time. Nonetheless, certain exemptions are granted for law enforcement activities, such as searching for missing persons, the localisation or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences or preventing terrorist attacks. The use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement shall be deployed only to confirm the identity of the specifically targeted individual, and it shall take into account the following elements: (a) the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm that would be caused if the system were not used; (b) the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those consequences [8]. "Real-time" biometric identification systems, including FRSs, can only be deployed if strict safeguards are met, e.g. its use is limited in time and geographic scope and subject to specific prior judicial or administrative authorisation. Using such systems post-facto is considered a high-risk use case, requiring judicial authorisation being linked to a criminal offence.

High-risk applications are those anticipated to present substantial risks to health, safety, or the fundamental rights of individuals. This notably includes AI systems employed in healthcare, education, recruitment, critical infrastructure management, law enforcement, or the justice sector [8]. Such applications are obligated to adhere to standards regarding quality, transparency, human oversight, and safety. In certain instances, they may necessitate a "Fundamental Rights Impact Assessment" prior to deployment. Evaluation is required both before market placement and throughout the lifespan of these applications. Additionally, the roster of high-risk applications can be expanded progressively over time, without requiring amendments to the AI Act itself.

3. Data protection and privacy concerns related to FRT in the EU

Using FRTs raise serious issues related to the right to personal data protection guaranteed in Article 8 of the Charter of Fundamental Rights of the EU (CFR), as well as the right to private life under Article 7 of the Charter [9]. Particularly, the initial video recording, continuing storage of the material, and the comparison of footage with database information for identification (matching) all interfere with or limit this right. Any limitation on these basic rights must be clearly justified and proportionate according to Article 52(1) CFR. To protect these rights, data controllers (and indirectly manufacturers) should design their intended data processing activities in full compliance with data protection principles, adhering to "data protection by design and by default" as stipulated in Article 25 GDPR and Article 20 LED [10]. Following the main legal principles of data protection (Article 5 GDPR and Article 4 LED), the processing of facial images must be based on lawful basis, valid consent, transparency, purpose limitation, privacy impact assessment, data minimisation, data accuracy, storage limitation, accountability and security measures.

3.1. Lawful basis

According to Article 52(1) CFR, any restriction on fundamental rights and freedoms must be established by law and must respect the core of those rights and freedoms [9]. Such restrictions must adhere to the principle of proportionality, meaning they can only be imposed if they are necessary and genuinely serve objectives of general interest recognized by the EU, or if they protect the rights and freedoms of others. For the processing of data to be lawful, it must comply with specific legal bases outlined in Recital 35 LED and Recital 40 GDPR. Video surveillance can be legally justified under Article 6 GDPR or under national laws implementing Article 8 LED. However, if it involves processing special categories of data, the processor must also meet the stringent requirements of Article 9 GDPR or Article 10 LED.

3.2. Valid consent

Processing personal data is generally prohibited, unless it is expressly allowed by law, or the data subject has consented to the processing. While being one of the more well-known legal bases for processing personal data, consent is only one of six bases mentioned in Article 6(1) GDPR, among others such as contract, legal obligations, vital interests of the data subject, public interest and legitimate interest [6]. Valid consent is one of the most problematic aspects when it comes to the deployment of FRT.

Consent is defined in Article 4(11) GDPR as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her" [6]. The basic requirements for a valid legal consent are defined in Article 7 and specified in recital 32 GDPR:

a) Voluntariness: Consent must be given voluntarily, without any form of coercion or undue pressure. Individuals should have a genuine choice and be able to refuse or withdraw consent without facing negative consequences. In situations where FRT is used by authorities, individuals may feel pressured to consent due to perceived or real power imbalances.

b) Explicitness: Under the GDPR, explicit (unambiguous) consent is required from individuals before their biometric data can be collected and processed for facial recognition purposes, which means it requires either a statement or a clear affirmative act. This ensures that individuals are fully informed about how their data will be used and have actively agreed to it. Consent cannot be implied and must always be given through an opt-in, a declaration or an active motion, so that there is no misunderstanding that the data subject has consented to the processing.

c) Informed and specific consent: For consent to be informed and specific, the data subject must be provided with clear and comprehensive information about the controller's identity, the purpose, scope, and potential implications of the data collection and use. This includes details on data storage, sharing, and security measures. Also, shall be notified about his or her right to withdraw consent anytime. Many individuals may not fully understand how FRT works, the data it collects, or the implications of its use. Without a clear understanding, consent cannot be truly informed [11].

Obtaining consent for facial recognition in public or semi-public spaces (like streets,

airports, or shopping centers) is particularly challenging. It is often impractical to inform every individual and obtain their explicit consent, raising significant privacy concerns. A case related to privacy concerns regarding FRT occurred in Germany. In 2019, a German court ruled against the use of FRT by a major property management company, Deutsche Wohnen, in a residential complex in Berlin [12]. The court found that the company's use of facial recognition violated the GDPR and the residents' right to privacy. The case stemmed from complaints filed by residents and privacy advocates who argued that the technology was being used without their consent and raised concerns about surveillance and data protection. The ruling set a precedent for the use of FRT in residential settings in Germany and emphasized the importance of respecting individuals' privacy rights when deploying such technologies.

3.3. Transparency

According to the transparency principle outlined in Article 5(1)(a) of the GDPR, it must be clear to individuals that their personal data is being collected, used, consulted, or otherwise processed, and to what extent this processing occurs (Recital 39 GDPR) [6]. Data subjects must be properly informed about the processing of their data, including through FRT. This information should be provided either when the personal data is collected or before consent is given. This principle does not, however, prevent competent authorities from conducting activities such as covert investigations or video surveillance (Recital 26 LED). Article 13(3) LED allows Member States to introduce exceptions to avoid hindering ongoing investigations or to protect public and national security [7]. Such exemptions can be crucial for law enforcement, as informing a suspect about the use of FRT might compromise their efforts. Given that these exceptions limit data subjects' ability to exercise their rights, they must be strongly justified.

For video surveillance/FRT driven by AI under the GDPR, the EDPB recommends a two-layered approach to meet transparency requirements. Key information should be provided through a warning sign so that individuals can recognize the surveillance before entering the monitored area. Additional details can be made available through other accessible means, such as posters or websites, clearly referenced on the initial warning [1].

A notable case related to privacy concerns and legal challenges regarding facial recognition technology occurred in France. In 2020, the French data protection authority, Commission Nationale de l'Informatique et des Libertés (CNIL), fined a major retailer, Carrefour, for violating the GDPR due to its use of FRT in some of its stores [13].The CNIL found that Carrefour had failed to obtain proper consent from customers and did not provide sufficient transparency regarding the use of facial recognition technology. This case highlighted the importance of complying with GDPR regulations and ensuring transparency and consent when implementing FRT in commercial settings in France and the wider EU.

3.5. Fairness

The EDBP stated in its guidelines that "fairness is an overarching principle which requires that personal data should not be processed in a way that is unjustifiably detrimental, unlawfully discriminatory, unexpected, or misleading to the data subject" [1]. However, some scholars consider this principle is somewhat ambiguous that can be applied in

situations where data processing might be legally permissible but still seems unfair in the specific context [14].

3.6. Purpose limitation

The principle of purpose limitation dictates that personal data may only be processed for a specifically defined, explicit, and legitimate purpose and it is reflected in Article 8(2) CFR, Article 5(1)(b) GDPR, and Article 4(1)(b) LED [6, 7, 9]. This principle mandates that personal data must be processed solely for specified purposes, which must be explicitly defined by law, allowing individuals to foresee the intended use of their data. These principles also apply to the processing of data via facial recognition technologies, prohibiting the unlimited retention of such data. In this context, the intended purpose must be clearly articulated so that the individual concerned can understand how their data will be used and adhere to a high threshold, primarily focused on combating terrorism, serious crimes, identify missing persons and victims of crime, including children, which is the established purpose limitation under EU law for law enforcement access to various large-scale EU databases. However, designing IT systems, including facial recognition systems, for purposes such as combating serious crimes, terrorism, improving public safety, and curbing irregular migration carries the risk of function creep, where personal data (facial images) may be used for unintended purposes [5]. Given the significant risk of "function creep" associated with FRT, related systems and processes should incorporate safeguards, such as a compartmentalised architecture, to prevent unauthorized use. Even if access falls within the scope of the legitimate purpose, the principles of proportionality and data security may further limit access conditions [15]. To prevent this, safeguards must be implemented to ensure that facial recognition technology is not unlawfully used to access large-scale EU databases, particularly when considering interoperability of these databases.

3.7. Privacy impact assessment

Article 35 GDPR requires from controllers a Data Protection Impact Assessments (DPIA) prior to the processing of personal data, when these activities is likely to result in a high risk to the rights and freedoms of natural persons [6]. For instance, a DPIA is required when data processing activities includes systematic and extensive profiling with significant effects, processing of special categories of data or criminal offense data on a large scale, large-scale monitoring of publicly accessible areas (CCTV).

A DPIA shall include several key elements: a) description of processing activities; b) assessment of necessity and proportionality; c) risk assessment (identify the risks to individuals' rights and freedoms, considering both the likelihood and severity of these risks); d) mitigation measures; e) consultation with stakeholders [6]. Specifically, while considering the deployment of FRTs in uncontrolled environments, law enforcement authorities will have to assess and explain in their assessment the strict necessity and proportionality of the deployment of these technologies; address the risk to different fundamental rights, including data protection, privacy freedom of expression, freedom of assembly, freedom of movement or antidiscrimination, depending on the potential uses in different places [6]. The impact assessment could be carried out either by entities themselves or by an independent monitoring body or by an auditor having relevant expertise to help find out, measure or map out impacts and risks over time.

3.8. Data minimisation, data accuracy and storage limitation

The principle of data minimisation, as outlined in Article 5(1)(c) GDPR and Article 4(1)(c) LED outlines that the amount of data collected should be limited to what is necessary for the intended purpose and should not be excessive. EDBP suggest that this principle also involves anonymising data where feasible [16]. Thus any video material not relevant to the purpose of the processing should always be removed or anonymized, for instance by blurring with no retroactive ability to recover the data before deployment [7]. The EDPS has observed that FRT systems may not always comply with the principle of data minimisation [1].

The principle of data accuracy, stipulated in Article 5(1)(d) GDPR and Article 4(1)(d) LED, requires that personal data be factually and temporally accurate, meaning that certain data must be kept up to date [6, 7]. Accuracy is assessed based on the purpose for which the data was collected. Minor errors may not affect overall accuracy, such as a single faulty data point in a large dataset. The EU Agency for Fundamental Rights notes that accuracy typically means correctness for each individual, though it can be interpreted more broadly [17]. The Council of Europe's guidelines on facial recognition stress the need to avoid mislabeling and to test systems to eliminate demographic disparities, thereby preventing unintended discrimination [10]. Data controllers must check the quality of images and biometric templates in watch-lists to prevent false matches. The Article 29 of Guidelines on Automated individual decision-making suggests that even inaccurate inferences from accurate data could violate the accuracy principle, implying that algorithms must be trained on representative datasets with minimal hidden biases [18].This aspect of the principle remains debatable and unresolved.

The principle of data retention (storage limitation) mandates in Article 5(1)(e) GDPR and Article 4(1)(e) LED that data should not be retained in an identifiable form longer than necessary for its intended purposes. Typically, 72 hours is sufficient to determine whether data needs to be retained longer, allowing for the deletion of unnecessary footage. If storage exceeds 72 hours, substantial justification for the purpose and necessity of the extended storage must be provided. Data may be kept longer for specific surveillance purposes. The EPDB advises that data extracted from digital images to create templates should not be excessive and should only contain necessary information, thus preventing further unnecessary processing [19, 20, 21]. Additionally, depending on the purpose, the raw data used to generate facial templates should be deleted once the template is created.

3.9. Data security and accountability

The principle of data security requires that data be processed securely, protecting personal data against unauthorized or unlawful processing, as well as accidental loss, destruction, or damage, through appropriate technical and organizational measures (Article 5(1)(f) GDPR and Article 4(1)(f) LED) [6, 7]. Articles 32 GDPR and 29 LED (indirectly) mandate that controllers and processors implement measures to prevent unauthorized disclosure or access to personal data. The EDPB advises that controllers must protect the system and data during storage, transmission, and processing [22]. Measures should include compartmentalizing data during transmission and storage, storing biometric templates and raw data on separate databases, encrypting biometric data, especially templates,

establishing a policy for encryption and key management, implementing fraud detection measures, associating an integrity code with the data, and prohibiting external access to biometric data [22]. These measures should adapt as technology advances. The Council of Europe also emphasizes the need to prevent technology-specific attacks, such as presentation and morphing attacks [10].

4. Concerns about violating fundamental rights and freedoms through indiscriminate use of FRT in public spaces

Mass surveillance and concerns for fundamental rights have been highlighted by many authors in relation to the widespread adoption of FRT [19, 20, 23]. The use of technology to process biometric data on a mass scale, whether for law enforcement, public authority, or commercial purposes, poses unique and serious threats to privacy and security. The Council of Europe defines mass surveillance as any monitoring that is not directed in a "targeted" manner at a specific individual [24]. Extending the use of these systems beyond their initially authorized and controlled purposes introduces potential risks over time. Such extensions might include using data from social networks or databases initially intended for different purposes, repurposing a database beyond its allowed scope, or adding new functionalities to an existing system. Critics argue that this gradual extension may be part of a deliberate strategy by proponents to first implement facial recognition in seemingly legitimate contexts and then progressively broaden its application [19, 20, 23]. The use of technology to process biometric data on a mass scale, whether for law enforce [19, 25]. This type of surveillance lacks sufficient transparency, leaving people unaware of what is happening, unable to provide informed consent, and without a genuine, free choice to opt in or out.

European Commission investigations indicate that wherever such a system operates, the movements of individuals in the reference database can be tracked [26]. Investigations by the European Commission indicate that the deployment of such systems allows for tracking the movements of individuals within the reference database, significantly impacting personal data, privacy, autonomy, and dignity.[26] This practice raises new social concerns, such as the inability to move anonymously in public spaces and the pressure to conform, which could undermine free will. The Commission highlighted the necessity of an ex-ante mechanism to ensure compliance with requirements and obligations, ensuring that providers of AI systems, including FRT, implement measures to minimize risks to fundamental rights by design [27]. Without such measures, AI systems will not be allowed on the Union market. Additionally, ex post market surveillance and supervision by competent authorities are essential to investigate and sanction any violations of fundamental rights in a proportionate, effective, and dissuasive manner [6].

According to Article 52(1) of the Charter, any restrictions on fundamental rights and freedoms must be legally established and must not violate the core of those rights and freedoms. These restrictions must adhere to the following criteria:

1) *Provided by law*: This requirement ensures that any restriction on rights and freedoms has a clear legal basis and is subject to the rule of law. This legal foundation must be clear enough to inform citizens about the conditions and circumstances under which authorities can collect data and conduct secret surveillance. It must clearly outline the scope and manner in which public

authorities can exercise their discretion to ensure that individuals receive the minimum level of protection required by the rule of law in a democratic society. Since biometric data falls under the special categories of data listed in Article 10 of the LED, most FRT applications would require a dedicated law that clearly defines the application and conditions of its use, including specifying the types of crimes and, where applicable, the appropriate severity threshold.

2) *Respect the essence of rights and freedoms:* The essence of a fundamental right refers to its very core, which must always be respected, even when the right is restricted [28]. Human dignity must also be upheld in all circumstances. This means that even if a limitation is justified, it cannot be so extensive that it destroys the fundamental nature of the right or freedom [9]. Potential indicators of an infringement on this inviolable core include a) provisions that impose limitations regardless of an individual's conduct or specific circumstances; b) barriers that prevent or hinder access to the courts [29]; c) situations where the individual's circumstances are not considered before imposing a severe limitation [30].
3) *A legitimate aim* is a fundamental requirement for justifying any limitation on fundamental rights and freedoms. In the context of limitations under Article 52(1) CFR, legitimate aims typically include: a) public safety and security measures taken to protect national security, prevent crime, and maintain public order; b) efforts to protect public health and uphold societal moral standards; actions necessary to safeguard the rights and freedoms of other individuals; c) policies aimed at supporting the economic stability and well-being of the state; d) ensuring the proper functioning of democratic institutions and processes [9].
4) *Necessity and general interest:* According to established case law of the CJEU, any derogations and limitations concerning the protection of personal data must be applied only to the extent that they are strictly necessary [30, 31]. This also means that no less intrusive means are available to achieve the intended purpose and objectives of general interest recognized by the EU or to protect the rights and freedoms of others. These objectives include those stated in Article 3 TEU and other interests protected by specific provisions of the Treaties, such as establishing an area of freedom, security, and justice and preventing and combating crime. However, the deployment must be accompanied by strict safeguards to prevent abuse and ensure that it is used only for its intended purpose. Differential treatment can be justified if it aims to achieve a legitimate objective and the means used are necessary and proportionate [32].
5) *Principle of proportionality* is crucial when considering the deployment of FRT and shall correspond to the following criteria: a) appropriateness (the use of FRT must be suitable to achieve a legitimate aim, such as enhancing public security or preventing crime); b) necessity (there should be no less intrusive means available to achieve the same objective); c) balancing interests (the benefits of using FRT must outweigh the potential negative impact on individuals' rights and freedoms. This requires a careful and case-by-case assessment).

According to Amnesty International, the widespread and invasive nature of mass surveillance imposes constraints on everyone's engagement in social, public, and political activities [33]. According to a United Nations Human Rights Council report, using FRT to

identify individuals in the context of assemblies significantly undermines not only privacy, but also freedom of expression, and peaceful assembly [34]. It affects individuals' capacity to lead autonomous lives without altering their behaviors out of fear of constant surveillance and this situation hinders people from fully exercising their political and civil rights [34].

Religious freedoms are also at stake with the deployment of FRTs. Individuals practicing certain religions may be subject to heightened surveillance and discrimination based on their appearance or attire. Such surveillance can lead to a chilling effect, where people may feel compelled to alter their behavior or conceal their religious practices to avoid being targeted. This undermines the fundamental right to freely practice one's religion without fear of state interference or social discrimination.

The freedom of assembly and association is similarly jeopardized by the use of FRTs. Surveillance of public gatherings and protests can have a deterrent effect, discouraging individuals from participating in these activities due to fears of being identified and possibly facing repercussions [35].This is particularly concerning in contexts where people are advocating for political or social change. The deployment of biometric surveillance systems establishes a dynamic wherein the powerful observe while the powerless are subjected to observation [36]. This dynamic empowers disproportionately influential groups to reinforce their control over socially marginalized communities, including individuals living in poverty, experiencing social exclusion, people of color, and human rights activists [9].

5. Accuracy and bias issues

Article 21 CFR prohibits discrimination on various grounds, including sex, race, color, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership in a national minority, property, birth, disability, age, or sexual orientation. Additionally, Article 20 of the Charter states that everyone is equal before the law. Discrimination in data-supported algorithmic decision-making can arise for various reasons. Demographic bias in FRT refers to the tendency of these systems to perform differently across various demographic groups. This bias can manifest as varying levels of accuracy and error rates for different genders, ages, ethnicities, and other demographic categories. Biases, whether intentional or not, can be embedded during the design, testing, and implementation of facial recognition algorithms. Additionally, discrimination can occur based on how officers respond to matches produced by these algorithms. If an algorithm performs inconsistently across different groups, removing such bias through mathematical or programmatic means is often very challenging, and sometimes impossible.

One significant cause of discrimination is the quality of the data used to develop these algorithms and software [37]. If the datasets used to train facial recognition algorithms are not diverse, the resulting models may not perform well for underrepresented groups. For facial recognition software to be effective and accurate, it needs a large volume of facial images. More images generally lead to more accurate predictions. However, accuracy also depends on the quality of the images and having a representative set of faces from diverse groups. The design of the algorithm itself can introduce bias if it does not account for

demographic diversity. The way FRT is deployed and used can also contribute to bias, especially if it is not regularly monitored and adjusted for fairness [26]. Data accuracy is crucial for ensuring reliable identification, both factually and temporally. This accuracy is particularly vital in FR systems, where any discrepancies can lead to significant errors. Inaccuracies in data can lead to false positives, where individuals are wrongly identified [38]. This can have serious consequences, potentially leading to mistaken identity and causing harm or inconvenience to innocent individuals. Similarly, inaccuracies in data can result in false negatives, where individuals are not recognized when they should be. This poses a threat to security and can undermine the effectiveness of facial recognition systems [38]. Ensuring data accuracy is key to minimizing these errors and maintaining the integrity of such systems.

The EU Fundamental Rights Agency's 2019 report indicates that certain demographic groups are more susceptible to misidentification by FRT [39]. These groups typically include:

a) *Ethnic minorities*: Studies have shown that facial recognition systems often have higher error rates for individuals from ethnic minority groups due to biases in the training data and algorithms.
b) *Women*: Research has indicated that facial recognition systems tend to have higher error rates for women compared to men.
c) *Elderly people*: Age can impact the accuracy of facial recognition, with elderly individuals often facing higher misidentification rates.
d) *Children:* As vulnerable individuals deserving of heightened protection, children are particularly at risk when these technologies are employed in law enforcement and border management. The primary issue stems from the lower accuracy rates of FRTs in detecting and recognizing the rapidly changing facial features of young people. This inaccuracy can lead to higher rates of misidentification, resulting in potential harm and undue scrutiny of children.

As result there are several types of demographic bias:

a) Ethnic and racial bias, when FRT systems often show higher error rates for people of color. For example, currently, facial images used to develop algorithms in the Western world often over-represent white men and under-represent women and individuals from other ethnic backgrounds. Consequently, facial recognition systems tend to perform well for white men but poorly for black women [39].
b) Gender bias: Many FRT systems have been found to perform better on male faces compared to female faces.
c) Age bias, when there are differences in accuracy based on age, with systems often performing less accurately on younger and older individuals compared to middle-aged individuals. Given the vulnerability of children, processing their biometric data, including facial images, must undergo a stricter necessity and proportionality test compared to adults. This ensures that the use of such data is not only justified but also carefully limited to protect children's rights and well-being.
d) Bias can also arise based on factors like facial hair, glasses, or other accessories, which may be more common in some demographic groups than others.

Demographic bias in facial recognition technology is a significant concern that can lead to

unequal treatment and discrimination. By taking proactive steps to ensure diverse training data, detect and mitigate biases, and maintain transparency and accountability, developers and users of FRT can work towards more equitable and accurate systems.

6. Ethical and social implications of implementing FRTs in the EU

The public perception of FRT within the EU is multifaceted and complex. While some individuals acknowledge the potential benefits of facial recognition for enhancing security and streamlining various processes, there is a growing unease about its widespread deployment. This unease stems from concerns about privacy violations, data security, and the potential for misuse. People are increasingly aware of the implications of having their faces scanned and stored in databases without explicit consent, leading to fears of constant surveillance and loss of anonymity. Such concerns are particularly pronounced when the technology is used in public spaces, where individuals feel they have little control over their personal data.

A survey conducted across various EU countries provides insight into public perception regarding FRT [17].The data highlights levels of support, opposition, and neutrality towards the technology, as well as concerns related to privacy and discrimination.

Table 1. Public perception regarding implementation of FRTs

Country	Support FRT(%)	Oppose FRT(%)	Neutral FTR(%)	Concern Privacy(%)	Concern Discrimination (%)
France	45	35	20	70	55
Germany	50	30	20	68	60
Italy	48	33	19	72	58
Spain	47	34	19	69	57
Netherlands	52	28	20	65	53
Poland	49	32	19	67	56
Sweden	55	27	18	64	54

Source: EU FRA, Your rights matter: Data protection and privacy - Fundamental Rights Survey, 2020

Support for FRT varies across the EU, with Sweden showing the highest level of support at 55%, and highest opposition in France at 35%. Privacy concerns are significant, with Italy having the highest at 72%. Discrimination concerns are prominent, with Germany expressing the highest concern at 60%. The survey data indicates a complex and cautious public attitude towards facial recognition technology in the EU. While there is notable support for its potential benefits, significant opposition and neutrality reflect ongoing public debates about its implementation. Privacy and discrimination concerns are particularly prevalent, highlighting the need for robust regulatory frameworks and transparency measures to address these issues.

The ethical concerns surrounding the use of facial recognition technology primarily revolve around issues of surveillance and privacy. The idea that one's movements and activities can be continuously monitored and recorded raises significant ethical questions. There is a fear that this level of surveillance could lead to a society where people alter their behavior out of fear of being watched, thereby undermining personal freedoms and autonomy.

One of the critical ethical issues is the potential for abuse by those in power [17]. Facial recognition technology can be used to target and discriminate against specific groups, whether based on race, religion, or political beliefs [40]. The ability of authorities or private entities to track individuals without their knowledge or consent infringes on fundamental human rights, such as the right to privacy and freedom of expression. This concern is particularly relevant in the context of political protests or social movements, where surveillance could be used to intimidate or suppress dissent [41].

The ethical and privacy concerns significantly impact public trust and acceptance of facial recognition technology. Trust is eroded when people feel that their privacy is being invaded without adequate justification or oversight. The lack of transparency in how data is collected, stored, and used further exacerbates these concerns. For facial recognition technology to gain public acceptance, there must be clear and robust legal frameworks that regulate its use. These frameworks should ensure that the technology is used in a manner that is transparent, accountable, and respects individuals' rights.

For the technology to gain widespread acceptance, several measures need to be in place to address these concerns: a) implementing robust and comprehensive laws that clearly define when and how FRT can be used, ensuring that the use of such technology is always necessary, proportionate, and in line with human rights standards; b) transparency and accountability; c) data security; d) independent oversight; e) public engagement and education; f) ethical design and implementation, which includes addressing potential biases in the technology and ensuring that it does not disproportionately affect vulnerable groups.

7. Definition of FRT in Chinese laws

In China, FRT is the most extensively adopted form of AI, utilized across various sectors for diverse purposes such as identification and enhancing efficiency. The Chinese government acknowledges the efficiency benefits that facial recognition brings to both public and private sectors, and has prioritized its research, development, deployment, and commercialization [42]. Recognizing the role of FRT in enhancing public security, the Chinese government has widely implemented it as part of a broader national security framework, which also includes mechanisms like the social credit system [43]. Increasingly, state-owned enterprises in sectors such as telecommunications, banking, and transportation are recording citizens' facial data for their FRT systems. This technology is also prevalent in the private sector, where it is used for online payments, residential security, and hospital check-ins. The rapid advancement and extensive use of FRT have positioned China as a global leader in this field. Consequently, facial recognition has permeated nearly every aspect of daily life in China; for instance, it has been instrumental in managing the COVID-19 pandemic by enabling identity verification without physical contact.

FRT as defined under Chinese law generally refers to a biometric identification method that involves the automated recognition and analysis of individuals' facial features to verify identity. While there is no single comprehensive legal definition explicitly codified in a specific law, several regulations and guidelines provide context for how FRT is understood and regulated in China. Personal Information Protection Law of China (PIPL) The PIPL

categorizes facial data as sensitive personal information, which includes biometric characteristics. The law requires that processing such information must have a specific purpose and necessitate stringent protection measures [44]. Article 28 of PIPL stipulated personal information processors can only handle sensitive personal information if they have a specific purpose and sufficient necessity under protection by strict measures. Article 29 stressed that the processing of sensitive personal information needs the separate consent of the individual. Cybersecurity Law mandates the protection of personal data, including biometric information, emphasizing the need for consent and the secure handling of data to prevent misuse and breaches [45].

Thus, FRT in the context of Chinese law can be defined as "a biometric identification technology that uses automated processes to capture, analyse, and verify individuals' facial features for the purpose of identity verification, subject to regulations governing the processing of sensitive personal information, consent requirements, and data protection measures as stipulated by the PIPL and related cybersecurity regulations. This definition encompasses the core principles of FRT as regulated in China, emphasizing both its technical function and the legal framework governing its use.

8. Mandatory use of FRT and the issues of consent

FRT has raised significant privacy issues globally, and China is no exception. While some observers and the survey presented above suggest that Chinese culture may be more accepting of privacy infringements compared to Western cultures, and many Chinese people support FRT due to enhanced security and convenience, there have been extensive discussions about the rationale and appropriate extent of FRT deployment in the country [46]. China has been actively developing a regulatory framework for FRT since 2020. Despite aiming to significantly improve personal data protection, this framework faces growing risks and challenges in safeguarding citizens' data within the FRT landscape.

Undoubtably, FRT brings convenience to Chinese citizens in various scenarios, including cashless payments and bypassing security queues at metros, libraries, train stations, and airports. However, this convenience comes with challenges related to privacy and personal data protection, raising public concerns about the potential misuse of sensitive personal data [47]. The proliferation of FRT in numerous sectors has sparked growing concerns. Numerous media reports indicate that its application in the private sector is susceptible to issues such as lack of transparency and cybersecurity vulnerabilities, including data leaks. Regulatory concerns have also been raised, since a multi-agency task force report highlighted widespread privacy issues, noting that mobile applications using facial recognition often force users to provide facial data, lack clear rules for data collection, and fail to offer mechanisms for users to withdraw consent for the collection and use of their facial information [48].

On August 20, 2021, the National People's Congress passed the PIPL, marking the country's first comprehensive legislation on personal information protection set to take effect on November 1, 2021. Article 26 of the PIPL imposes restrictions on the use of FRT [44], stating that installation of devices for image collection and personal identity recognition in public places is permissible only if necessary to safeguard public security,

comply with relevant state regulations, and prominently display notices. The article also specifies that personal images and identification information collected can only be used to protect public security and must not be disclosed to others, except with explicit consent from individuals or as stipulated by laws and administrative regulations. Under the PIPL, biometric characteristics are categorized as sensitive information, requiring personal information processors to obtain consent from the data subjects and explain the necessity and potential impact of collecting such information.

FRT has become prominently utilized in the public sector, especially for law enforcement purposes such as identifying and tracking criminal suspects. Additionally, the government has designated facial recognition as a primary technology for identity verification in various regulations. It is strongly encouraged and often mandated for administrative tasks such as notarization, obtaining driver's licenses, and delivering social benefits to residents [49]. In April 2019, the General Administration of Customs authorized the use of facial recognition technology at Customs registration counters. Since September 2017, the Ministry of Justice has required parties seeking notarization to undergo identity verification using methods like facial recognition, cross-checking against the Ministry of Public Security's databases [50]. From January 2020, the Ministry of Public Security mandated that online traffic schools under the Traffic Management Department verify user identities using technical methods such as facial recognition [51]. Furthermore, during in-person traffic law education sessions organized by the Traffic Management Department, drivers' identities must also be verified through facial recognition technology [51]. Moreover, in February 2020, in response to monitoring and controlling COVID-19, Ant Financial introduced a QR code system that assigns users a color code indicating their health status [52]. Users obtain these codes by providing their name, national identity number, and registering with facial recognition. As it can be noticed these cases refers to public security and public health and all these regulations do not specify usage parameters or provide specific guidelines on how facial recognition technology should be deployed in public setting. Additionally, none of the rules issued address security measures aimed at protecting facial information.

With strong governmental backing, state-owned enterprises across various sectors have begun adopting facial recognition technology for identity verification purposes. For instance, the People's Bank of China, which has issued rules mandating FRT for verifying bank account identities since 2016. Banks are encouraged to utilize this technology to assist in reading, collecting, and verifying client information during account opening processes. The National Health Commission also promoted the use of FRT in pilot medical institutions starting from February 13, 2019, to strengthen the management. China Railway, where users are notified in the privacy policy that facial scans are required for logging into accounts using facial recognition. The Beijing Municipal Commission, which mandated the incorporation of FRT in public housing projects starting from January 2019 [53]. This is primarily aimed at enhancing security at entryways to prevent unauthorized access. And many other cases deployed by central and local authorities. These initiatives illustrate the extensive use of facial recognition technology by state-owned enterprises, primarily for streamlining identity verification processes across various administrative and service sectors.

Benefiting from government support and sometimes even mandates, numerous private companies are increasingly integrating facial recognition technology to improve operational efficiency. Across diverse industries, these companies are employing facial recognition primarily for managing user authentication processes. For instance, since December 1, 2019, mobile phone users in China are required to undergo facial recognition scans when registering new SIM cards. The Ministry of Industry and Information Technology mandated telecom companies to implement technical measures that compare the facial features of users with their identification cards [54]. Network access is only granted when the facial comparison matches the identification card information.

Using facial recognition in strictly personal settings can enhance efficiency, but it also poses challenges when private rights are disregarded. An example of this is the compulsory use of facial recognition without offering alternative solutions. In 2021, a property management company PMC "Wuye" was sued that it does not provide alternative methods for neighbourhood entrance verification [55]. The plantiff claimed that this PMC forces residents to use facial recognition and does not allow person to enter the neighbourhood if they refuse such technology [55]. The defendant argued that FRT is "the symbol of updated and reconstruction of old verification system" and it got consent with most residents only except the plantiff. From one side, facial recognition in Chinese society is considered to be a fashionable solution and become a key element evaluating the level of digitalization or "smartness" of neighbourhood management. Refusing FRT would be regarded as a "conservative or outdated" lifestyle. From the other side, the PMC do asked for consents from residents. However, according to the Judicial Interpretation of the Supreme Court [56] property owners shall be provided with alternative verification methods if a property management company insists on using FRT as the sole method for entry [56]. The parties reached an agreement that the PMC will provide entry method of using card key. From the Judicial Interpretation and litigation result, it can be found that the consent from the majority is not enough. Even if there is only single person who refuse to use FRT, the company must provide alternative solutions of the entrance verification.

The litigation dispute in question in the opinion of Court does not directly involve privacy violations or misuse of personal data, but rather concerns the compulsory use of personal information. Unlike cases where personal data is unlawfully used by third parties, mandatory facial recognition obtains user consent but may not be voluntary. While facial information collectors do not disclose any privacy, this still constitutes a breach of civil law because user consent may not be freely given. Article 1024 of Civil Code of China involves the protection on facial information from the perspective of civil law [57]. The Judicial Interpretation explained that the use of facial recognition without consent is a violation on right of personality [56], but kept silent on whether it is illegal in administrative cases and in public places. Article 4 of Judicial Interpretation rules that courts shall not support information processors' defenses of obtaining consents if: (1) the information processor refuses to provide products or services unless the natural person consents to the processing of facial information, except when the processing of facial information is necessary for the provision of products or services; (2) the information processor requires that the natural person should consent to the processing of facial information by means of

tie-in authorization; (3) the information processor forces directly or in a disguised form the natural person to consent to the processing of his/her facial information [56].

On August 2023, the Cyberspace Administration of China first disclosed the draft of "Provisions on the application of safety management of Face Recognition Technology (Trial)" (FRT Provisions) [58]. The FRT Provisions sets out that consent is compulsory requirement if information processors need to collect face images from users of applications. Article 5 FRT Provisions rules that the use of FRT to process face information shall obtain individual consent or written consent according to law, except for those who do not need to obtain personal consent according to laws and administrative regulations [58]. For the same arrangement, Article 13 rules that the separate or written consent of the parents or other guardians of the minors should be obtained if face information of minors under the age of 14 is processed [58]. Administrative regulation rules differently from judicial interpretation of civil law because it arranges exceptions for consent. If a civil litigation is triggered, the parties have to be both private ones and it is no doubt that there is no possibility for a private party to have right to use other's biometric information without consent. The PMC's behavior is also prohibited by the FRT Provisions: PMC shall not use face recognition technology to verify personal identity as the only way to enter and exit the property management area [58]. If individuals do not agree to use FR system verification, PMC shall provide other reasonable and convenient verification methods [58].

The Judicial Interpretation treat mandatory use "without consent" but it still makes an exception when FRT is necessary for realization of product or service functions [58]. The rules of FRT Provisions are stricter than those in Judicial Interpretation because it prohibits any use of FRT without consent for private purpose. It could be argued that continuing to use FRT in practice instead of discontinuing its use could serve as evidence of consent obtained, as supported by exceptions outlined in Judicial Interpretations, but Article 5 FRT Provisions even emphasized that written consent is necessary in some scenarios [58]. Therefore, consent should have a valid form.

Besides requirement on consent in facial information collection, the FRT Provisions also set other scenarios where consent is also mandatory. Article 7 rules that the installation of FRT in public places areas should satisfy the requirement of necessity for the maintaining public safety [58]. Entities operating such facilities and collecting facial images have the obligation to keep confidentiality of the obtained facial images and personal information, which does not allow relevant information to be illegally disclosed to the public or provided to third parties [58]. Even though consent to collect facial information is not required for public security purposes, the use of relevant information should be limited to such purposes. If relevant entities want to use collected facial information in other scenarios, they must obtain the consent of each individual [58], even though such entities may be bodies of the government. These provisions apply also in scenario when analysing other sensitive personal information via FRT, including race, ethnicity, religious belief, health status and social class, etc [58]. The exceptions to non-consent involve the maintenance of national or public security, and the protection of individuals' life, health, or property in emergencies. These two conditions differ in their requirements. While the FRT Provisions do not mandate an emergency element for national and public security scenarios, they do

require it for the protection of relevant rights. Therefore, consent from individuals is necessary if the situation is not urgent [58]. The FRT Provisions restrict not only the collection of facial information but also the handling of such information after collection. The provisions outlined in Article 12 emphasize the importance of balancing public safety with the protection of individual privacy rights in the context of using image acquisition and personal identification equipment in public places [58]. Here are some key points:

1. *Necessity and compliance*: The law mandates that the installation of such equipment should only be done when necessary for public safety. This is a reasonable measure to ensure that surveillance is not overused or implemented without justification. Moreover, the requirement to comply with national regulations and provide prominent notifications is crucial for transparency and public awareness.
2. *Confidentiality obligations:* The duty imposed on units to maintain the confidentiality of collected data underscores the importance of protecting personal information. By prohibiting illegal disclosure and external provision of data, the law aims to prevent misuse and unauthorized access, thereby safeguarding individuals' privacy.
3. *Purpose limitation:* Restricting the use of collected data exclusively to preserving public safety is a significant measure to prevent the abuse of surveillance technologies. This provision ensures that personal data is not exploited for other purposes, such as commercial gain or unwarranted monitoring.
4. *Consent requirement:* Allowing the use of personal images and identification information for other purposes only with the individual's specific consent is a critical aspect of data protection. It empowers individuals to have control over their personal information and ensures that their rights are respected.

These regulations reflect a thoughtful approach to integrating surveillance technologies into public spaces. They aim to harness the benefits of such technologies for public safety while imposing strict controls to protect privacy and prevent potential abuses. This balanced approach is essential in fostering public trust and ensuring that technological advancements do not come at the expense of fundamental privacy rights. Given the significant trust placed in governments using FRT, the exemption from consent collection for security reasons aligns with public concerns. Consent primarily becomes necessary when FRT is utilized for private purposes or by private entities, which underscores people’s apprehensions about the subsequent use of their facial information post-collection.

9. Abuse of facial information and issues of violation of civil rights

Misuse of Facial Recognition Technology (FRT) is anticipated to cause harm to individual rights such as personal identity, privacy, and other civil liberties. Typically, the misuse of FRT involves the absence of consent from individuals. In other words, any action taken without consent can be considered misuse of FRT, especially when it is mandated. In China, the misuse of FRT violates various regulations found in judicial interpretations of civil litigation, administrative laws and regulations, and criminal law.

9.1. Infringement of civil rights

In 2012 China's top legislative authority, the Standing Committee of the 11th People's Congress, expressed its commitment to safeguarding digital privacy. Plans were made to introduce legislation that included principles for data protection, such as restrictions on personal information collection and measures to ensure privacy protection [59]. The enactment of the 2020 PRC Civil Code marked a significant change in China's regulatory framework concerning the safeguarding of personal information, including biometric data. Prior to the Civil Code, regulations concerning personal data, including FRT, were fragmented, primarily addressed in laws related to cybercrime and cybersecurity breaches [57]. The Civil Code introduced a new chapter dedicated to privacy laws in China, recognizing personal information as a fundamental civil right. Article 1035 of the Civil Code sets forth general principles for data protection, including limitations on purposes and scope, as well as the requirement for informed consent from data subjects in the processing of personal information [57].

In the Judicial Interpretation of Supreme Court, abuse of FRT is consider to be "an action of infringing on the personality rights of a natural person [56]." It listed eight categories of abuse: (1) conducting facial verification, recognition, or analysis in business premises and public places by using FRT in violation of laws and regulations; (2) failing to disclose rules on the processing of facial information or failing to explicitly state the purposes, methods, and scope of such processing; (3) failing to obtain the separate consent; (4) not complying with the specified purposes, methods, and scope for processing facial information as stated by the information processor or agreed upon by all parties involved; (5) failing to take proper technical measures or other necessary measures for ensuring the security of facial information collected and stored, which results in leaks, distortion, or loss of facial information; (6) providing others with facial information in violation of the provisions of laws and administrative regulations or the agreement of both parties concerned; (7) processing facial information in violation of public order and good moral; (8) other circumstances where facial information is processed by violating the principles of lawfulness, legitimacy, and necessity [57].

9.2. Administrative law and regulations

According to the information disclosed by Institution of Judicial Case Study of the Supreme People's Court on August 18, 2021, the reporter found a total of 422 cases involving "face recognition" and "administrative penalty" in Weike Advanced Database (wkinfo) [60]. Among them, 29 cases are related to the protection of personal rights and interests in case of using FRTs and all occurred in the housing sales industry [60]. The report classifies into four categories:

Table 2 Types of abuse corresponding Judicial Interpretation

Type of Infringements:	Item in in the Judicial Interpretation
Not informing consumers of the collection of biometric data (face image)	Article 2(3)
Not clearly informing the purpose, method and scope of collection and use	Article 2(2)

Having informed the way to collect and use of biometric information (face image), but not specifying the true purpose and scope of their collection and use, nor having obtained the consent of consumers	Article 2(4)
Not specifying the purpose, method, and scope of collecting and using information to consumers with the consent of consumers	Article 2(2)

Source：Chu Xia, Analysis & interpretation on 400 administrative punishment cases of "face recognition"

Administrative authorities imposed fine on relevant parties in all 29 cases, according to Law on the Protection of Consumer Rights and Interests of China (LPCRI) [60]. Article 29 of LPCRI rules that operators shall follow the principles of legality, legitimacy, and necessity, specify the purpose, method and scope of collecting and using information, and obtain the consent of consumers [61]. When a business operator collects and uses consumer's personal information, it shall disclose its rules of collection and use. Such entity shall not collect and use information in violation of the provisions of laws and regulations and the agreements of both parties [61].

The analysis shows that different law rules concerning the use of FRT have similar norms. Even though Provisions on Facial Recognitions are still in the draft, other laws started to protect citizens from abusive use of FRT by commercial entities.

9.3. The application of criminal law for the abuse of FRTs

Besides civil liability and administrative fines, abuse or illegal obtaining of facial information may also receive criminal penalties. The Supreme Court of China disclosed Guiding Case no. 192 on Mr. Li Kaixiang's infringing citizens' personal information [62]. This case is a combination of criminal and civil litigation case for public interest purposes [62]. From June to September 2020, Li Kaixiang made a mobile phone "hacker software" with the function of illegally stealing the photos of the installer's album [62]. Through it, he stole a total of 1,751 photos from the installers' album, some of which containing 100 pieces of citizens' personal information including facial information [62]. On August 23, 2021, the People's Court of Fengxian District of Shanghai found that Li Kaixiang had committed the crime of infringing citizens' personal rights by stealing information and sentenced him to three years in prison but three years' probation, and a fine of 10,000 CNY [63]. The Fengxian court stated in its decision that "facial information" is recognized as citizens' personal information under the principle of law and order. Article 1034 of the Civil Code and the PIPL includes "facial information" in the category of sensitive information. Using hacker software to steal "facial information" is socially harmful and punishable by law. As sensitive information, "face information" is crucial for identifying individuals and has strong social attributes. It is easily misused or synthesized, potentially leading to privacy violations, reputation damage, theft, and fraud, posing significant social risks [62, 63].

The Supreme Court emphasized that face information generated or processed by FRT is highly recognizable [62]. It can be used to identify the identity of a specific natural person, or it can reflect the activities of a specific natural person alone or in combination with other information. Such information is regarded as the personal information under the criminal law. Article 5(4) of the Interpretation on Several Issues Concerning the Application of the Law in Criminal Cases of Infringement of Citizens' Personal Information may apply if a

person (1) collect or use facial information without the consent of the citizen himself, (2) does not have the legal reasons for the handling of personal information stipulated in the PIPL or the authorization of relevant departments; (3) steals or illegally obtains the above information by other means such as software programs [62].

Another similar case occurred when the first civil public interest litigation initiated by the Procuratorate regarding the protection of citizens' personal facial information was publicly adjudicated in Guangzhou [64]. The defendants collected high-definition ID card photos, ID card numbers and other personal sensitive information, then used the avatar in the photo to make AI videos and sell for money. The court ordered that the four defendants should immediately stop the infringement of citizens' personal information, pay compensation and damages, and apologize publicly [64].

10. Special regulation on FRT.

Article 6 FRT Provisions prohibits image acquisition and personal identification equipment to be installed in locations that might infringe on others' privacy, such as hotel rooms, public bathhouses, dressing rooms, and bathrooms [58]. The installation of FRT and personal identification equipment in public places must be done only when necessary to ensure public safety, adhere to applicable national regulations, and include clearly visible notifications [58]. Relevant entities have obligations to keep the obtained personal images and identification information confidentially and shall not be illegally disclosed or provided to the public [58]. Even though the use of FRT is for implementation of internal management, relevant entities should reasonably determine the image information collection area according to the actual needs, and take strict protection measures to prevent illegal access, copying, disclosure, external provision, dissemination of personal images, etc [58]. They should prevent the leakage, change, lost or illegally acquisition or use of personal information [58].

Security issues are also related to privacy. FRT Provisions give several requirements on security which aims to protect security:

Table 3. requirements on security in FRT Provisions

Article	Item	Content
17	Information preservation	Except under legal conditions or with individual consent, FRT users must not save original face images, pictures, or videos unless anonymized. Systems providing face recognition services must meet network security level protection above the third level and implement data encryption, security audits, access control, authorization management, and intrusion defenses. Critical information infrastructure must also comply with relevant security protection requirements.
18	Deletion or anonymization	The use of FRT to process face information shall try to avoid collecting face information that has nothing to do with the provision of services. If it cannot be avoided, it shall be deleted or anonymized.
19	Evaluation on security and risk	Users of face recognition technology must annually assess and mitigate security risks of image and identification equipment, adjust security strategies and confidence thresholds, and implement measures to protect against attacks, invasions, interference, and destruction.
20	Requirement on facilities	Image collection equipment and personal identification equipment listed in the catalogue of key network equipment and special products for network security

		in accordance with the relevant provisions of the State shall be sold or provided only after the qualified institutions have passed the certification or met the requirements in accordance with the mandatory requirements of relevant national standards.
21	Regular check	The network information department, together with the competent telecommunications department, the public security organ, the market supervision department and other relevant departments, shall strengthen the supervision and inspection of the use of face recognition technology according to their responsibilities, guide and urge users of face recognition technology to complete the filing procedures, find potential safety hazards in a timely manner and urge rectification within a timely limit.

Source: Provisions on the application of safety management of face recognition technology (Trial)

Besides substantive rules and ex-post regulation, FRT Provisions requires ex-ante compliance. Article 15 rules that FRT processors should conduct an impact assessment of personal information protection in advance and record the processing [58]. The impact assessment of personal information protection mainly includes the following: (1) whether it meets the provisions of laws, administrative regulations and the mandatory requirements of national standards, and whether it conforms to ethics; (2) whether the processing of face information has a specific purpose and sufficient necessity; (3) whether it is limited to the accuracy, accuracy and distance requirements necessary to achieve the purpose; (4) whether the protective measures taken are legal, effective and compatible with the degree of risk; (5) the risk of leakage, loss, destruction or illegal acquisition or illegal use of face information and possible harm; (6) the damage and impact that may be caused to the rights and interests of individuals, and whether the measures to reduce the adverse effects are effective [58].

The personal information protection impact assessment report shall be kept for at least three years. If the purpose and method of processing face information change, or a major security incident occurs, the user of face recognition technology shall re-evaluate the impact of personal information protection.

As for large scales of using FRT, extra evaluation process should be implemented. Article 16 of FRT Provisions rules that FRT processor who use FTR in public places or store more than 10,000 face information shall file with Cyberspace Administration at or above the municipal level within 30 working days [58]. The following materials shall be submitted for filing: (1) the basic situation of users of face recognition technology and their person in charge of personal information protection; (2) explanation of the necessity of handling face information; (3) the purpose, processing method and security protection measures of face information; (4) rules and operating procedures for the handling of face information; (5) personal information protection impact assessment report; (6) other materials that the network information department deems need to be provided.

FRT Provisions only provide general requirements, and detail things rely on different standards. Some of these standards are shown below:

Table 4. Standards related to FRT in China

Number	Type	Title	Promulgation	Validation
DB31/T 1467-2024	Shanghai Standard	Application Guide for Face Recognition Classification in Public Places	2024-04-02	2024-07-01
GB/T 42981-2023	National Standard	Information technology – Biometrics - Test methods for face recognition system	2023-09-07	2024-04-01
GA/T 1093-2023	Industrial Standard	Security prevention, face recognition application, entrance and exit control face recognition technical requirements	2023-07-28	2023-12-01
GB/T 41987-2022	National Standard	Public security - Face recognition applications - Test methods for presentation attack detection with fake face	2022-10-12	2023-05-01
GB/T 41819-2022	National Standard	Information security technology - Security requirements of face recognition data	2022-10-12	2023-05-01
GB/T 41772-2022	National Standard	Information technology - Biometrics— Technical requirements for face recognition system	2022-10-12	2023-05-01
YD/T 4087-2022	Industrial Standard	Mobile Intelligent Terminal Face Recognition Security Technical Requirements and Test Evaluation	2022-09-30	2023-01-01
SF/T 0106-2021	Industrial Standard	Inspection specifications for face recognition technology in portrait identification	2021-11-17	2021-11-17
GB/T 38671-2020	National Standard	Information security technology - Technical requirements for remote face recognition system	2020-04-28	2020-11-01
GB/T 35678-2017	National Standard	Public security - Face recognition application - Technical requirements for face images	2017-12-29	2018-07-01
SJ/T 11608-2016	Industrial Standard	General Specification for Face Recognition Equipment	2016-01-15	2016-06-01
GB/T 31488-2015	National Standard	Technical requirements for face identification of video surveillance in security systems	2015-05-15	2015-12-01

Source: National public service platform for standard information.

11. Ethical and social implications of implementing FRTs in China

In October 2020 Artificial Intelligence Ethics Research Group and the App Special Governance Working Group of the Nandu released the "Report on Face Recognition Application and Investigation on the Public (2020) (Nandu Report) [65]. In this report, the research groups mainly discussed the scenarios of using FRT, questions on public's acceptance on FRT and potential public concerns on FRT's risks. The Nandu Report listed ten types of scenarios using FRT, such as money transfer, opening and canceling accounts, real-name registration, unlocking and decrypting, face-changing applications, government affairs, traffic security inspection, access control attendance in campus/online education, public safety supervision [66]. The investigation shows that 94.07% of interviewees admitted that they used FRT in daily life [66]. Contrasting with this high percentage, the proportion of giving consent or having an agreement on using face information collection or privacy protection is much lower, only reaching 61.81% [66]. 18.59% of interviewees show that they did not see relevant agreements or consent polices [66]. The research

indicated that 61.81% of participants felt that their willingness to give consent for the use of FRT would vary depending on the specific scenario in which it is used. This suggests that compliance with consent protocols may be better in some situations but significantly worse in others. The research revealed that the use of FRT is higher in specific scenarios such as money transfers (67.17%), unlocking and decrypting devices (54.09%), traffic security inspections (49.63%), and real-name registrations (47.68%) [66]. This higher usage correlates with increased rates of obtaining consent and adherence to privacy policies. However, the potential risks of using FRT without proper consent and privacy protections are still prevalent. Big companies, including financial institutions, may conduct good due diligence process according to relevant rules on consent collection and privacy protection, but small developers of narrowly-used applications often may be very weak in such works. Even though FRT appears in many daily scenarios, it is not the most popular way for verification [66]. With a percentage of 32.98%, FRT ranks no.5 in the most popular method of verification, lower than fingerprint (55.7%), verification code on smartphones (50.66%), password (48.57%), and ID card (39.87%).[69] The primary concern arises when FRT is extensively and involuntarily implemented in everyday situations, either due to undesirable circumstances or mandatory requirements.

People concern on security issues caused by the use of FRT, with 63.64% of interviewees worry about leakage of facial information, ranking on the top of FRT risks. Others are personal tracking being recorded (54.4%), money loss (53.72%), fake news with manipulation of faces (49.59%), impersonation (37%) and reputation (13.91%) [66]. These risks are not unique for using FRT. Such risks may also appear when using other verification methods.

In contrast, interviewees support the use of FRT in public areas even though they suspect that it may bring risks on privacy since FRT in China is considered to be the guardian on public security. Up to 67.64% of interviewees can accept using FRT on detection of infringement of traffic rules and surveillance on urban roads and public transportation receives 64.77%. In contrast, only 39.22% of interviewees can accept that vendors use FRT to collect and analyse consumers' behaviours and preferences [66]. From this perspective, people welcome FRT in public scenarios but warry to be used for commercial purposes. Another evidence is that governments (74.06%), schools and universities (66.63%), SOEs (62.16%), and financial institutions (51.37%) received very high trustiness in using FRT [66]. Meanwhile, private companies only received 31.79% of trustiness [66].

Taking into account mandatory use of FRT in public by the government, it can also be concluded that people react very negatively on using FRT by private companies. Governments implement FRT to guarantee public security even though some personal rights may be sacrificed and limited, but in Chinse mindset this is for common good to all members in the society. If such mandating is implemented by private enterprises, there would be sufficient reason to suspect that face information would be misused or even abused, particularly for illegal purposes, which bring risks on individual security, privacy, and reputation, even though such risks are not unique in scenarios of using FRT.

12. Conclusions

The comparative study of face recognition technology (FRT) implementation in China and the European Union (EU) reveals stark contrasts driven by differing regulatory frameworks, ethical considerations, and societal values.

12.1. Differences in implementation

a) Regulatory environment and adoption:

- The EU has stringent data protection laws, primarily governed by GDPR, which imposes strict requirements on the collection, storage, and use of biometric data, including explicit consent from individuals, DPIAs, and strong security measures. These regulations are designed to safeguard individual privacy and ensure transparency and accountability in data processing. The strict regulatory framework and high public skepticism result in more cautious and limited adoption of FRT. There are significant legal and ethical hurdles that organizations must navigate to implement FRT.
- China has a more permissive regulatory environment regarding the use of facial recognition technology. The regulations are less stringent compared to the GDPR, allowing for broader deployment of the technology in various sectors, including public surveillance, without the same level of consent and transparency requirements. There are few legal restrictions on the collection and use of biometric data, giving authorities broad leeway to deploy FRT without significant oversight or accountability.

b) Purpose and usage:

- Facial recognition technology in the EU is often used in law enforcement, controlled settings such as airports, banks, and retail for purposes like security, identity verification, and customer service. There are significant restrictions on its use in public surveillance and law enforcement due to privacy concerns.
- In China, facial recognition technology is widely used for law enforcement, public surveillance, not only in in controlled settings, but also social credit systems. The government employs this technology extensively for monitoring, ensuring security, and maintaining social order.

c) Public perception and acceptance:

- There is significant public concern and debate about the use of facial recognition technology in the EU, driven by privacy advocates and civil rights organizations. The general public tends to be wary of extensive surveillance and the potential for misuse of biometric data.
- Public acceptance of facial recognition technology is higher in China, partly due to the government's narrative on the benefits of enhanced security and social order. The population is more accustomed to state surveillance and the use of technology for monitoring purposes.

d) Technological development and innovation:

- The EU's cautious, privacy-centric approach highlights the challenges of balancing technological innovation with robust privacy protections and ethical considerations. Innovation in facial recognition technology in the EU is influenced by strict regulatory requirements, which can slow down rapid deployment but ensure privacy and ethical considerations. European companies often focus on developing privacy-preserving technologies.
- China is a global leader in the rapid development and deployment of facial recognition technology, but at the cost of significant privacy and ethical concerns. Chinese companies benefit from a supportive regulatory environment and significant government investment, allowing for faster innovation and widespread implementation.

12.2. Similarities in implementation

a) Security applications:

- Both the EU and China use facial recognition technology for security purposes, such as access control in secure facilities, identity verification at airports, and enhancing public safety in crowded places.

b) Commercial use:

- In both regions, businesses are leveraging facial recognition technology for customer service improvements, personalized marketing, and efficient transaction processing. Retail stores, banks, and hospitality sectors are common adopters.

c) Technological capabilities:

- Both the EU and China have advanced technological capabilities in facial recognition. Companies in both regions are developing sophisticated algorithms and hardware to improve accuracy, speed, and reliability of facial recognition systems.

d) Ethical and privacy debates:

- Despite regulatory differences, there are ongoing ethical and privacy debates in both the EU and China regarding the use of facial recognition technology. Concerns about data security, potential misuse, and impacts on civil liberties are prevalent in discussions in both regions.

e) Need for a balanced framework:

- *Ethical and privacy safeguards:* Both regions need to find a balanced framework that addresses ethical implications and privacy concerns while fostering technological innovation. Such a framework should ensure that FRT is deployed in a manner that respects individual rights and societal values.
- *Regulatory harmonization:* There is a need for regulatory harmonization that can provide clear guidelines for the ethical use of FRT. This includes establishing international standards and best practices that protect privacy and prevent misuse while enabling technological advancement.

In conclusion, in China, the implementation of FRT is characterized by largely supportive regulatory environment that facilitates its widespread use into various aspects of public life,

including law enforcement, public surveillance, and even everyday commercial transactions. This broad adoption reflects a regulatory framework that prioritizes public safety over individual privacy. While this approach has enabled rapid technological deployment and enhanced security measures, it has also sparked significant ethical concerns. Issues such as privacy concerns, data security risks, and the potential for misuse of FRT highlight the dark side of unchecked technological growth.

In stark contrast, the EU's approach to FRT is governed by stringent data protection laws, reflecting a strong commitment to protecting individual privacy and data security. This regulatory framework ensures that technological advancements do not compromise fundamental rights, resulting in a more cautious adoption of FRT. Public skepticism and ethical concerns further constrain the deployment of FRT in the EU, where there is significant public and governmental scrutiny over its potential impacts on privacy and civil liberties. This cautious approach underscores the EU's prioritization of privacy and ethical considerations over rapid technological adoption.

References

[1] European Data Protection Board, "Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement," 2023.

[2] ECtHR, "Szabó and Vissy v. Hungary," no. 37138/14, 2016.

[3] CJEU, "M. Schwarz v. Stadt Bochum," no. C-291/12, 2013.

[4] ECtH8R, "Guide on Article 8 of The European Convention on Human Rights, Right to Respect for Private and Family Life," *Home and Correspondence*, 2019.

[5] M. Tambiama and M. Hendrik, "Regulating Facial Recognition in the EU," 2021. [Online].

[6] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119".

[7] "Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities".

[8] Corrigendum, "Artificial Intelligence Act," *Interinstitutional file: 2021/0106 (COD)*, 2024.

[9] "Charter of Fundamental Rights of the European Union, 2000/C 364/01. Explanation on Article 52 - Scope and Interpretation of Rights and Principles," 2000. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT.

[10] Council of Europe, "Guidelines on facial recognition, Directorate General of Human Rights and Rule of Law, T-PD (2020)03rev4," 2021. [Online].

[11] European Data Protection Board, "Guidelines 05/2020 on Consent under Regulation 2016/679," 2020.

[12] Verwaltungsgericht Berlin , "Urteil des VG Berlin vom 27.06.2019 - VG 1 K 129.17," [Online]. Available: https://www.berlin.de/gerichte/verwaltungsgericht/.

[13] Commission Nationale de l'Informatique et des Libertés (CNIL), "Délibération SAN-2020-012," 2020. [Online]. Available: https://www.cnil.fr/fr/sanction-de-3-millions-deuros-pour-carrefour-france-et-800-000-euros-pour-carrefour-banque.

[14] P. Kramer, "'Artikel 5 DSGVO', in M. Eßer et al., Auernhammer DSGVO BDSG," 2020.

[15] European Union Agency for Fundamental Rights, "Under watchful eyes: biometrics, EU IT systems and fundamental rights, FRA," 2018.

[16] European Data Protection Board, "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default," 2020.

[17] EU Agency for Fundamental Rights, "Your rights matter: Data protection and privacy - Fundamental Rights Survey," 2020.

[18] European Commission, "Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679," 2018.

[19] S. Zuboff, "The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power," *Public Affairs,* 2019.

[20] B. Schneier, "Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World," *W.W. Norton & Company,* 2015.

[21] V. Eubanks, "Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor," *St. Martin's Press,* 2018.

[22] European Data Protection Board, "Guidelines 3/2019 on processing of personal data through video devices," 2020.

[23] B. Wagner, "Ethics of AI and Robotics," *Springer,* 2020.

[24] Council of Europe, "Declaration on Mass Surveillance," 2015.

[25] K. Crawford, "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence," *Yale University Press,* 2021.

[26] European Commission, "White Paper on Artificial Intelligence: A European approach to excellence and trust. COM (2020) 65 final," 2020.

[27] European Commission, "Impact assessment accompanying the Proposal for a Regulation of the European Parliament of the Council Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts. Commission Staff Working Document," 2021.

[28] CJEU C-279/09, "DEB Deutsche Energiehandels- und Beratungsgesellschaft mbH v Bundesrepublik Deutschland. Report of Case, 2010 I-13849," 2010.

[29] European Court of Justice, "Case C-362/14 Maximillian Schrems v Data Protection Commissioner," 2015.

[30] European Court of Justice, "Case C-293/12 and C-594/12 Digital Rights Ireland Ltd v Minister for Communications," Marine and Natural Resources and Others, 2014.

[31] "Case C-473/12, Institut professionnel des agents immobiliers (IPI) v Geoffrey Englebert and Others," 2013.

[32] CJEU, "C-356/12, Wolfgang Glatzel v. Freistaat Bayern," 2014.

[33] Amnesty International, Russia, "Intrusive facial recognition technology must not be used to crackdown on protests," 2020.

[34] Office of the High Commissioner for Human Rights , "The impact of new technologies on the promotion and protection of human rights in the context of assemblies, including peaceful protests. A/HRC/44/24," 2020.

[35] D. Harwell and C. Timberg, "As Protests Spread Across the U.S., Facial Recognition and Surveillance Technology Comes Under Scrutiny," *The Washington Post,* 2020.

[36] E. Jakubowska and N. Naranjo, "Ban biometric mass surveillance," *EDRi,* 2020.

[37] EU Fundamental Rights Agency, "Data quality and artificial intelligence- mitigating bias and error to protect fundamental rights," *Publications Office,* 2019.

[38] Commission Nationale de l'Informatique et des Libertés (CNIL), "Facial Recognition: For a Debate Framed by the Law," 2019.

[39] EU Fundamental Rights Agency, "Fundamental Rights Report 2019," Luxembourg: Publications Office of the EU, 2019.

[40] J. Buolamwini and G. Timnit, "Gender shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in *Proceedings of Machine Learning Research*, 2018.

[41] A. M. Bedoya, "The Color of Surveillance," *Georgetown Law Technology Review,* no. 4(2), pp. 109-142, 2020.

[42] Ministry of Industry and Information Technology of China, "Three-Year Action Plan to Develop a New Generation of the Artificial Intelligence Industry," 2017.

[43] J. A. Lee and P. Zhou. , "FRT Regulation in China," in *The Cambridge Handbook of Facial Recognition in the Modern State. Cambridge Law Handbooks,* Cambridge University Press, 2024.

[44] "Personal Information Protection Law of the People's Republic of China," 2021.

[45] "Cyber Security Law of the People's Republic of China," 2016.

[46] D. Ren, "AI, Machine Learning Tech Promises US$6000 Billion Annually for China Economy as It Pervades Industries, Says Mckinsey," 2022. [Online]. Available: www.scmp.com/business/banking-finance/article/3186409/ai-machine-learning-tech-promises-us600-billion-annually.

[47] T. G. Brown and a. et, "Public Debate on Facial Recognition Technologies in China," 2021. [Online]. Available: https://mit-serc.pubpub.org/pub/public-debate-on-facial-recognition-technologies-in-china/release/1.

[48] Cyberspace Administration of China, "Without the right of choice or the right to be informed, can facial recognition be trusted?," 2020.

[49] Y. Luo and R. Guo, "Facial recognition in China: Current Status, Comparative Approach and The Road Ahead, Penn Carey Law: Legal Scholarship Repository," 2022.

[50] Ministry of Justice of China, "Notice of the Office of the Ministry of Justice on Practical Guidance for Notarization," 2017.

[51] Ministry of Public Security of China, "Rules on Accepting Traffic Safety Education to Reduce Illegal Traffic Behavior (Trial)," 2020.

[52] M. H. Hu, "Beijing Rolls Out Colour-Coded QR System for Coronavirus Tracking Despite Concerns Over Privacy, Inaccurate Ratings," *South China Morning Post,* 2020.

[53] Beijing Commission of Housing and Urban-Rural Construction, "Notice on Further Strengthening of Supervision and Administration of Subletting and Leasing Public Rental Housing, Beijing Construction Regulation," no. 23, 2018.

[54] SOHU TECH, "China's Ministry of Industry and Information Technology's new implementing regulation. Beginning from December. 1, Application for Cards Requires 'Facial Recognition," 2019.

[55] R. S. Du and C. Y. Wan, "Suzhou court closed the first case of the latest judicial interpretation applicable to face recognition: Don't you want to enter the community without brushing your face? Court: Don't force it!," 2021.

[56] Supreme People's Court of China, "Provisions of the Supreme People's Court on Several Issues concerning the Application of Law in the Trial of Civil Cases Relating to Processing of Personal Information by Using the Facial Recognition Technology, Interpretation No 15 [2021]," 2021.

[57] "Civil Code of the People's Republic of China," 2020.

[58] Cyberspace Administration of China, "Notice on the Provisions on the Safety Management of the Application of Face Recognition Technology (Trial) (Draft for Comments)," 2023.

[59] "Decision of the Standing Committee of the National People's Congress on Strengthening Information Protection on Networks, issued by the Standing Committee of the National People's Congress," 2012.

[60] X. Chu, "Analysis & interpretation on 400+ administrative punishment cases of "face recognition"," *Institution of Judicial Case Study of the Supreme People's Court,* 2021.

[61] "Law of the People's Republic of China on the Protection of Consumer Rights and Interests," 2013.

[62] The Supreme Court of China, "Guiding Case No. 192: Li Kaixiang's Criminal Incidental Civil Public Interest Litigation Case for Infringing Citizens' Personal Information," 2022.

[63] "Criminal Judgment No. 828 (2021)," Shanghai, 2021.

[64] Y. Y. Zhong and et al, "Guangdong's first civil public interest lawsuit involving face recognition and personal information protection was sentenced", Jiancha Daily, Supreme People's Procuratorate (SPP) of the People's Republic of China," 2022.

[65] L. L. Fu, "Face Recognition Application Public Research Report (2020), Science and Technology Daily," 2020.

[66] "Artificial Intelligence Ethics Research Group and the App Special Governance Working Group of the Nandu (Southern Metropolis Daily), Report on Face Recognition Application and Investigation on the Public (2020)," 2020.

[67] EU Fundamental Rights Agency, "Data quality and artificial intelligence- mitigating bias and error to protect fundamental rights," *Publications Office*, 2019.

Bridging the AI divide: The evolving arms race between AI-driven cyber attacks and AI-powered cybersecurity defenses

Guy WAIZEL,
"Alexandru Ioan Cuza" University of Iasi, Romania
guy.waizel@gmail.com

Abstract
The rapid advancement of artificial intelligence (AI) has significantly transformed both offensive and defensive dimensions of cybersecurity. This article explores the burgeoning landscape of AI-driven cyber-attacks and the corresponding AI-powered cybersecurity defenses. Through an extensive literature review, we establish a foundational understanding of current AI techniques used in cyber-attacks, such as machine learning-based malware and AI-generated phishing schemes. Concurrently, we examine state-of-the-art AI-driven defense mechanisms, including anomaly detection systems and automated response strategies. To provide concrete examples, we conduct detailed case studies of high-profile cyber incidents where AI played a pivotal role. These case studies illustrate the sophistication and effectiveness of AI-driven attacks and highlight the defensive measures deployed to counteract them. By juxtaposing the capabilities of offensive AI with defensive AI, we reveal a significant gap between the two, underscoring the challenges faced by cybersecurity professionals in keeping pace with rapidly evolving threats. The findings from the research underscore the need for continuous innovation and collaboration in the cybersecurity field to enhance AI-powered defenses. By synthesizing insights from academic research, industry practices, and real-world case studies, this article offers a comprehensive view of the current state of the AI cybersecurity arms race. The analysis not only illuminates the existing disparity between AI-driven attacks and defenses but also suggests strategic pathways for narrowing this gap, ultimately aiming to bolster global cyber resilience.

Keywords: Ransomware, Stealth techniques, AI techniques, APT, Malware, Cybersecurity

1. Introduction: AI in Cybersecurity: Confronting Evolving Threats with Innovative Defenses

The rapid advancement of artificial intelligence (AI) has significantly reshaped the landscape of cybersecurity, influencing both offensive and defensive strategies. This article delves into the emerging realm of AI-driven cyber-attacks and the corresponding AI-powered defenses through an extensive literature review. It establishes a foundational understanding of contemporary AI techniques employed in cyber-attacks, such as machine learning-based malware and AI-generated phishing schemes, and examines state-of-the-art AI-driven defense mechanisms, including anomaly detection systems and automated response strategies. Detailed case studies of high-profile cyber incidents illustrate the sophistication and efficacy of AI-driven attacks and the defensive measures deployed to counteract them. The juxtaposition of offensive and defensive AI capabilities reveals a notable disparity, highlighting the challenges faced by cybersecurity professionals in keeping up with rapidly evolving threats. The findings emphasize the necessity for ongoing innovation and collaboration in the cybersecurity field to enhance AI-powered defenses, providing a comprehensive view of the current state of the AI cybersecurity arms race and suggesting strategic pathways to bolster global cyber resilience.

Following this introduction, Chapter 1.1 outlines the various AI-driven attack types and methods, while Chapter 1.2 explores the cybersecurity defenses that leverage AI to combat these and other sophisticated threats.

1.1. AI-Driven Cyber Attack Types and Methods

Artificial Intelligence (AI) has become a double-edged sword in the realm of cybersecurity. While it offers advanced tools for defending against threats, it also equips cybercriminals with sophisticated methods to launch attacks.

This section outlines the various AI-driven attack types and methods reported to date [1, 2, 3, 4, 5, 6, 7, 8].

1.1.1 Phishing and Spear Phishing

AI-Powered Phishing: Attackers use AI to craft highly convincing phishing emails by mimicking writing styles and creating personalized content.

Spear Phishing: Leveraging AI to gather information from social media and other sources, attackers create targeted phishing campaigns aimed at specific individuals or organizations [9, 10, 11, 12, 13, 14, 15, 16].

1.1.2. Malware and Ransomware

AI-Enhanced Malware: AI helps in developing malware that can adapt and evade detection by traditional antivirus software. Polymorphic Malware: Uses AI to continuously change its code, making it hard to detect with signature-based systems.

AI-Driven Ransomware: Employs AI to identify valuable data and encrypt it, optimizing the attack's impact and ransom demands [17, 18, 19].

1.1.3. Social Engineering

Deepfake Technology: AI-generated audio and video content used to impersonate individuals, tricking victims into disclosing sensitive information or transferring funds.

Automated Social Engineering: AI algorithms analyze and exploit human psychology, enhancing the effectiveness of social engineering tactics [20, 21, 22].

1.1.4. Adversarial Attacks

Adversarial Examples: Slight modifications to input data that cause AI systems to make incorrect decisions, used in attacks on image recognition and other AI models.

Model Poisoning: Injecting malicious data into training datasets to corrupt AI models, leading to erroneous outputs [23, 24, 25, 26].

1.1.5. Automated Attacks

Credential Stuffing: Using AI to automate the process of testing stolen username-password pairs on multiple websites. Botnets: AI-powered bots that can carry out distributed denial-of-service (DDoS) attacks, spreading malware or conducting large-scale spam campaigns [27, 28, 29, 30, 31].

1.1.6. Supply Chain Attacks

Compromised Updates: AI is used to infiltrate and manipulate software updates in the supply chain, distributing malware to multiple targets [32, 33].

1.1.7. Additional Findings from Recent Studies

Voice Control Systems Vulnerabilities: In their survey, Wang et al. (2023) revealed that AI-driven audio attacks expose new security vulnerabilities in voice control systems, highlighting that current defense strategies are not completely effective against advanced attacks. This emphasizes the need for enhanced defense mechanisms in voice control systems (VCS) [34].

Microgrid Cyber-Attacks: Beg et al. (2023) found that the attack surface in microgrids has increased, making them vulnerable to various AI cyber-attacks. They propose using AI-based cyber-attack mitigation in distributed cooperative control-based AC microgrids and suggest future research directions including transfer learning and explainable AI to improve trust in these systems [35].

Cybersecurity Threats and AI as Both Threat and Solution: Murphy (2024) identified the increasing sophistication of cybersecurity threats. The findings suggest that while AI can introduce new vulnerabilities when manipulated by attackers, it also offers powerful tools for enhancing threat detection, automating security processes, and improving overall defense strategies [11].

Adoption of Machine Learning in Banking: Gonaygunta (2023) highlighted that despite the effectiveness of machine learning algorithms in detecting cyber threats, financial institutions have low adoption rates. The study using the Unified Theory of Acceptance and Use of Technology (UTAUT) model found that performance expectancy and facilitating conditions positively influence the intention to use machine learning, while effort expectancy and social influence have less impact. This underscores the need for better integration of AI in the banking sector [36].

Cyber-Physical Systems in Manufacturing: Mamun (2023) noted that cyber-physical systems (CPS) in manufacturing are vulnerable to significant cyber threats due to the integration of AI, IoTs, Cloud Computing, ICSs, and Big Data analytics. The research proposes frameworks to enhance security, including detecting unauthorized changes, addressing high-volume data issues, and recovering sensor data with missing entries using advanced data reduction methods [17].

Smart Home Technology: Benedict (2023) demonstrated that machine learning techniques are effective in detecting SSH brute force and botnet attacks on smart home technology networks by analyzing representative network data, addressing limitations of outdated benchmark datasets [37].

AI in Cybersecurity Decision Making: Gusman (2023) explored how AI and ML technologies influence cybersecurity decision-making, revealing that while AI is expected to become more prevalent, human professionals will remain crucial due to technology

limitations. This highlights the need for a learning curve and adjustment period for employees [10] [38].

FinTech Cyber Development Challenges: Boonyapredee (2023) revealed that FinTech companies in Southeast Asia prioritize rapid deployment over cybersecurity due to consumer demands, leading to vulnerabilities. The study stresses the need for a balance between cybersecurity measures and rapid technological advancements [39].

Explainable IDS: Ables (2023) found that while black box intrusion detection systems (IDS) are accurate, they lack transparency, prompting the need for eXplainable IDS (X-IDS) using techniques like Competitive Learning (CL) and Rule Extraction (RE) to achieve both accuracy and trustworthiness [40].

Distributed AI Defense: Gonzalo (2021) emphasized the need for specialized edge-level systems using deep learning to counter escalating cyber threats, particularly in IoT environments lacking control. The proposed model includes synthetic data frameworks and distributed neural networks to address these challenges [41].

Agricultural Cyber-Physical Systems: Zhou (2023) discussed enhancements in IoT and machine learning architectures that improve prediction accuracy and energy efficiency in intelligent frost protection systems for agriculture, providing practical solutions for changing risk patterns and rising costs [42].

Countermeasures Against AI Malware: Jalaluddin (2020) investigated countermeasures against AI-powered malware targeting facial recognition, identifying that recommended technical and non-technical measures, combined with security awareness programs, can effectively reduce threats posed by AI-generated malware [12].

1.2. Cybersecurity Defenses Against AI-Driven Attacks

As AI-driven cyber threats evolve, so must the defenses against them. This section explores current cybersecurity measures that leverage AI to combat these sophisticated attacks, providing a robust defense framework.

1.2.1. AI-Based Threat Detection

Behavioral Analysis: AI models analyze normal behavior patterns and detect anomalies that may indicate a cyber threat.

Intrusion Detection Systems (IDS): AI enhances IDS by integrating advanced machine learning algorithms that improve the accuracy and efficiency of detecting malicious activities within a network. These systems can process large volumes of data in real-time, identifying suspicious patterns and behaviors that traditional methods might miss. AI-driven IDS can also adapt to new threats by learning from past incidents and evolving their detection capabilities accordingly.

Explainable Intrusion Detection Systems (X-IDS): Traditional IDS often use black-box AI models, which, despite their high accuracy, lack transparency. This can hinder trust and

understanding among security professionals. Explainable IDS (X-IDS) address these issues by incorporating white-box techniques that provide clear insights into the AI decision-making process. Methods such as Competitive Learning (CL) and Rule Extraction (RE) help elucidate how decisions are made by the AI, enhancing transparency and accountability.

Regulatory Compliance: By making AI decisions interpretable, X-IDS facilitate compliance with data protection laws and industry standards, as they provide auditable insights into security decisions [43] [40] [11].

1.2.2. AI-Augmented Threat Intelligence

Threat Intelligence Platforms (TIPs): These platforms aggregate, analyze, and act on threat data from multiple sources using AI. They provide real-time insights and predictive analytics to preemptively identify and mitigate potential threats.

Automated Threat Hunting: AI-powered tools continuously scan for threats, enabling proactive threat hunting instead of reactive measures [44] [39].

1.2.3. Adaptive Security Architectures

Zero Trust Architecture: This principle requires strict verification for every device, user, and application, regardless of their location within or outside the network. AI enhances the zero-trust model by continuously assessing risk and adapting security policies in real-time.

Dynamic Defense Mechanisms: AI systems dynamically adjust defenses based on real-time threat assessments and intelligence. This includes automated patching and configuration adjustments to mitigate vulnerabilities immediately as they are detected [45] [18] [41].

1.2.4. Behavioral Biometrics

User and Entity Behavior Analytics (UEBA): AI-driven UEBA systems monitor the behavior of users and entities to identify unusual patterns that may indicate compromised accounts or insider threats. These systems leverage machine learning to understand normal behavior and flag deviations. Continuous Authentication: Instead of relying on a one-time authentication, AI systems continuously monitor user behavior to ensure ongoing verification throughout a session [46, 47].

1.2.5. Machine Learning for Threat Intelligence

Automated Threat Hunting: AI and machine learning (ML) models analyze vast datasets to identify patterns and anomalies that signify potential threats. This enables proactive threat hunting and faster detection of sophisticated attacks.

Predictive Analytics: By analyzing historical data, ML models predict potential future attacks, allowing organizations to implement preemptive measures to mitigate risks [48, 49].

1.2.6. Enhanced Authentication Methods:

AI-Powered Biometrics: Utilizing AI to improve biometric authentication methods such as facial recognition, fingerprint scanning, and voice recognition ensures that access controls are robust against spoofing and other types of attacks.

Behavioral Biometrics: AI models analyze user behavior, such as typing patterns and mouse movements, to continuously authenticate users, adding an additional layer of security [50, 51].

1.2.7. Adaptive Security Measures:

Dynamic Risk Assessment: AI continuously assesses the risk level of users and systems, adapting security measures in real-time based on detected threats and vulnerabilities.

Automated Incident Response: AI-driven systems can automatically respond to detected threats by isolating affected systems, blocking malicious traffic, and initiating recovery protocols [52].

1.2.8. Secure Software Development

AI for Code Analysis: AI tools analyze source code to detect and mitigate vulnerabilities during the development phase, reducing the risk of exploitable flaws in deployed software.

Automated Patch Management: AI systems manage and deploy patches automatically, ensuring that systems remain up-to-date and protected against known vulnerabilities [53].

1.2.9. Network Defense

AI-Enhanced Network Monitoring: AI systems monitor network traffic in real-time to detect unusual patterns indicative of cyber threats, enabling rapid identification and mitigation of attacks such as DDoS and data exfiltration.

Anomaly Detection: Machine learning algorithms detect deviations from normal network behavior, identifying potential intrusions and suspicious activities [54].

1.2.10. Data Protection and Privacy

AI for Data Encryption: AI optimizes encryption algorithms to enhance data protection without compromising system performance, ensuring that sensitive information remains secure.

Data Anonymization: AI techniques anonymize data to protect individual privacy while maintaining data utility for analysis and decision-making [55].

1.2.11 Integration with Existing Security Frameworks

Security Information and Event Management (SIEM): Integrating AI with SIEM systems enhances their ability to process and analyze security logs, providing more accurate threat detection and incident response.

Endpoint Detection and Response (EDR): AI augments EDR solutions by providing deeper insights into endpoint activities and improving detection of sophisticated malware and exploits [56].

1.2.12 Explainable AI (XAI) in Cybersecurity:

Transparency in AI Models: Implementing explainable AI techniques helps in understanding and interpreting the decision-making process of AI systems. This enhances trust and reliability in AI-driven cybersecurity measures by providing clear explanations for detected threats and recommended actions.

Regulatory Compliance: Explainable AI aids in meeting regulatory requirements by providing auditable insights into how security decisions are made, ensuring accountability and transparency [57].

1.2.13. Federated Learning:

Collaborative Defense: Federated learning enables multiple organizations to collaborate on training AI models without sharing sensitive data. This improves the collective ability to detect and respond to cyber threats while maintaining data privacy [58].

1.2.14. AI-Driven Threat Intelligence Platforms

Crowdsourced Threat Intelligence: AI platforms can aggregate threat intelligence from multiple sources, including open-source data, to provide comprehensive insights into emerging threats and vulnerabilities [59].

2. Method

2.1. Literature Review

Objective: To establish a foundational understanding of the current state of AI-driven attacks and AI-powered cybersecurity defenses.

Approach: A comprehensive literature review was conducted to systematically examine existing academic papers, industry reports, white papers, and relevant publications. This review covered key areas, including AI techniques used in cyber-attacks, such as machine learning-based malware and AI-generated phishing schemes, and AI-driven defense mechanisms, including anomaly detection systems and automated response strategies.

The selection criteria for the literature included relevance to AI in cybersecurity, recent publication dates to ensure up-to-date information, and the credibility of sources. Databases such as IEEE Xplore, Google Scholar, ProQuest, ResearchGate and more repositories were utilized to gather the necessary materials.

Outcome: The literature review synthesized existing knowledge, identified gaps in current research, and outlined the evolution and sophistication of AI in both offensive and defensive cybersecurity applications. This synthesis provided a comprehensive understanding of how AI is currently utilized in cyber threats and defenses, setting the stage for further analysis.

2.2. Case Studies Analysis

Objective: To provide concrete examples of AI-driven cyber-attacks and the corresponding AI-based defensive measures.

Approach: Detailed case studies of high-profile cyber incidents where AI played a pivotal role were selected and analyzed. The selection criteria for these case studies included incidents where AI techniques were notably used by attackers for evasion, automation, or enhancement of traditional cyber-attack methods, and incidents where AI-driven defense mechanisms were employed to detect, mitigate, or respond to the cyber-attacks.

Sources for case studies included publicly available incident reports and cybersecurity threat analysis publications. Each case study was analyzed to illustrate the sophistication and effectiveness of AI-driven attacks and the defensive measures deployed.

Outcome: The case studies highlighted specific instances of AI use in cyber-attacks and defenses, demonstrating the capabilities and limitations of both. By juxtaposing the capabilities of offensive AI with defensive AI, these case studies revealed the significant gap between the two, underscoring the challenges faced by cybersecurity professionals in keeping pace with rapidly evolving threats.

2.3. Integration and Analysis

The integration of findings from the literature review and case studies offered a comprehensive view of the current state of AI in cybersecurity. The analysis illuminated the existing disparity between AI-driven attacks and defenses and suggested strategic pathways for narrowing this gap. This integrated approach aimed to bolster global cyber resilience by providing actionable insights and recommendations for continuous innovation and collaboration in the cybersecurity field.

3. Results

3.1 AI-Driven Cyber Attack Case Studies

The landscape of cyber threats has been significantly transformed with the advent of artificial intelligence (AI). AI has empowered attackers to conduct more sophisticated, precise, and impactful cyberattacks across various industries. Table 3.1 provides an overview of real-world case studies where AI-driven tactics were employed to compromise organizations. These examples illustrate the wide-ranging applications of AI in cybercrime, highlighting the urgent need for advanced cybersecurity measures to counter these evolving threats.

Table 3.1. Summary of AI-Driven Cyber Attacks

Industry	Description	Entry Point	Type of Attack	Impact of Attack
Automotive	Attackers used AI to exploit supply chain vulnerabilities, leading to data theft.	Phishing Emails	Data Breach	Significant data theft and operational disruption
Financial Services	Breach exploited cloud misconfigurations using	Cloud Infrastructure	Data Breach	Exposure of personal information of

	AI, exposing sensitive customer data.			over 100 million customers
Healthcare	AI-enhanced ransomware targeted critical systems, optimizing disruption.	Network Analysis	Ransomware	Major operational disruption, ransom demands
Entertainment	AI-crafted phishing emails led to data exfiltration and destruction of IT infrastructure.	Phishing Emails	Data Breach, Destruction	Data exfiltration, major IT infrastructure damage
Hospitality	Attackers used AI-powered social engineering to breach systems, disrupting digital services.	Social Engineering	Social Engineering	Disruption of digital room keys, slot machines; over $100 million loss
Telecommunications	AI-driven supply chain attack, compromising software update to distribute malware.	Compromised Employee Credentials	Supply Chain Attack	Malware distribution, significant security breach
Energy	Deepfake technology used to impersonate executives, leading to fraudulent financial transfers.	Deepfake Impersonation	Social Engineering, Financial Fraud	Fraudulent transfer of €220,000
Technology	AI-driven malware designed to evade detection, demonstrating advanced capabilities.	AI-generated Malware	Polymorphic Malware	Highlighted vulnerabilities in traditional cybersecurity measures

Sources: [60, 61, 62, 63, 64, 65, 66, 67, 68, 69]

3.3 Summary of AI-defenses methods and Potential Gaps

In the face of increasing AI-driven cyber threats, various AI-enhanced defense mechanisms have been developed to bolster cybersecurity. These methods range from sophisticated threat detection systems to adaptive security architectures and enhanced authentication techniques. While these defenses leverage advanced machine learning and artificial intelligence to provide robust protection against cyber-attacks, they are not without their limitations. Table 3.3 summarizes the current AI-based defense methods, providing a brief description of each and highlighting potential gaps that attackers might exploit. These insights are crucial for understanding both the capabilities and vulnerabilities of modern AI-driven cybersecurity strategies.

Table 3.3. Summary of AI-defenses methods and Potential Gaps

AI Defense Method	**Description**	**Potential Gaps that can be Exploited by AI-Driven Cyber Attack**
AI-Based Threat Detection	Detects anomalies in behavior patterns	Sophisticated evasion techniques

Intrusion Detection Systems (IDS)	ML algorithms enhance IDS accuracy	Adaptation to evade detection
Explainable Intrusion Detection Systems (X-IDS)	Transparent AI decision-making process	Complexity and lack of comprehensive rules
Regulatory Compliance	Facilitates compliance with data laws	Exploiting non-compliance or loopholes
Threat Intelligence Platforms (TIPs)	Aggregates and analyzes threat data	Overwhelming with false data
Automated Threat Hunting	Proactive scanning for threats	New, undetected threat patterns
Zero Trust Architecture	Strict verification for all entities	Exploiting verification gaps
Dynamic Defense Mechanisms	Real-time defense adjustments	Rapidly changing attack methods
User and Entity Behavior Analytics (UEBA)	Monitors user/entity behavior for anomalies	Mimicking normal behavior patterns
Continuous Authentication	Ongoing user behavior verification	Subtle changes in behavior to avoid detection
Machine Learning for Threat Intelligence	Analyzes datasets for threat patterns	Evasion through novel attack vectors
Predictive Analytics	Predicts future attacks from historical data	Unpredictable or novel attack methods
AI-Powered Biometrics	Enhances biometric authentication	High-quality spoofing techniques
Behavioral Biometrics	Analyzes behavior for authentication	Mimicking genuine behavior patterns
Dynamic Risk Assessment	Real-time risk level assessment	Gradual, undetected risk escalation
Automated Incident Response	Automatic threat response	Automated system manipulation
AI for Code Analysis	Detects vulnerabilities in code	Introducing subtle, hard-to-detect flaws
Automated Patch Management	Manages and deploys patches automatically	Exploiting patch deployment delays
AI-Enhanced Network Monitoring	Real-time network traffic monitoring	Generating benign-looking malicious traffic

Anomaly Detection	Detects deviations in network behavior	Blending malicious activity with normal traffic
AI for Data Encryption	Optimizes encryption algorithms	Breaking encryption with advanced techniques
Data Anonymization	Anonymizes data to protect privacy	Re-identification attacks
Security Information and Event Management (SIEM)	Enhances SIEM with AI for better log analysis	Flooding logs to hide malicious activity
Endpoint Detection and Response (EDR)	Provides insights into endpoint activities	Exploiting endpoint vulnerabilities
Transparency in AI Models	Makes AI decision-making transparent	Misinterpreting AI outputs
Collaborative Defense (Federated Learning)	Trains models without sharing sensitive data	Poisoning shared learning processes
Crowdsourced Threat Intelligence	Aggregates threat intelligence from multiple sources	Information overload and data manipulation

4. Discussion

The research presented in this article highlights the profound impact of artificial intelligence (AI) on both the offensive and defensive aspects of cybersecurity. Through a comprehensive examination of current AI techniques employed in cyber-attacks and corresponding AI-enhanced defense mechanisms, several critical insights have emerged.

4.1. The Evolving Threat Landscape

The advent of AI has revolutionized the threat landscape, enabling attackers to execute more sophisticated, targeted, and effective cyberattacks. The case studies detailed in Table 3.1 underscore the diverse applications of AI in cybercrime, from machine learning-based malware to AI-generated phishing schemes. These examples reveal a clear trend: AI is not merely an incremental improvement over traditional cyberattack methods but represents a transformative leap that significantly enhances the capability of malicious actors. This transformation necessitates an urgent response in the form of equally advanced and adaptive cybersecurity measures.

4.2. AI-Driven Defenses: Progress and Challenges

On the defensive front, the development of AI-enhanced cybersecurity mechanisms has shown promising advancements. As detailed in the results section, contemporary defenses utilize sophisticated machine learning algorithms for threat detection, adaptive security architectures, and improved authentication techniques. However, despite these advancements, there remain notable gaps and limitations. Table 3.3 highlights these

vulnerabilities, which attackers might exploit to circumvent current defenses. This disparity between the capabilities of offensive and defensive AI underscores a critical challenge in the cybersecurity arms race: the need for continuous and rapid innovation in defense technologies to keep pace with evolving threats.

4.3. Comparative Analysis and Strategic Implications

By juxtaposing AI-driven offensive tactics with defensive mechanisms, the research reveals a significant gap in effectiveness. Offensive AI technologies, being inherently innovative and aggressive, often outpace the defensive strategies currently in place. This imbalance is particularly evident in the sophistication and adaptability of AI-driven attacks compared to the relatively static nature of many defense systems. The findings suggest that while current AI-based defenses are robust, they are not sufficiently adaptive or anticipatory to counteract the most advanced AI-driven threats effectively.

The strategic implications of these findings are profound. There is an evident need for a paradigm shift in cybersecurity strategy, moving from a reactive to a proactive approach. This shift involves not only the development of more advanced AI technologies but also fostering greater collaboration between academia, industry, and government to share knowledge, resources, and strategies. By leveraging a multidisciplinary approach, the cybersecurity community can enhance the resilience of defense mechanisms against AI-driven threats.

4.4. Future Directions and Recommendations

To bridge the gap between offensive and defensive AI in cybersecurity, several pathways can be pursued. First, ongoing research and development must focus on creating more adaptive and intelligent defense systems capable of anticipating and responding to new attack vectors in real-time. Second, there must be an emphasis on integrating AI with human expertise, leveraging the strengths of both to develop more nuanced and effective cybersecurity strategies. Finally, fostering a culture of continuous learning and innovation within the cybersecurity community is crucial. This includes regular updates to defense protocols, continuous monitoring of the threat landscape, and incorporating feedback from real-world incidents to refine AI-driven defense mechanisms.

In conclusion, while AI has dramatically transformed the cybersecurity landscape, presenting both new opportunities and challenges, it is clear that continuous innovation and strategic collaboration are essential to enhancing global cyber resilience. By understanding the current capabilities and limitations of AI in both offensive and defensive contexts, cybersecurity professionals can better prepare for the future, developing more sophisticated, adaptive, and effective defense strategies to safeguard against the ever-evolving threat of AI-driven cyberattacks.

References

[1] "4 Types of AI Cyberattacks Identified by NIST. (n.d.)," [Online]. Available: https://www.lumenova.ai/blog/4-types-of-ai-cyberattacks-identified-nist/.

[2] R. Gurzeev, "Seven AI attack threats and what to do about them," 2024. [Online]. Available: https://www.scmagazine.com/perspective/seven-ai-attack-threats-and-what-to-do-about-them.

[3] "Hacking AI? Here are 4 common attacks on AI, according to Google's red team. (n.d.)," [Online]. Available: https://www.zdnet.com/article/hacking-ai-how-googles-ai-red-team-is-fighting-security-attacks/.

[4] C. W. Elizabeth Montalbano, "Google Categorizes 6 Real-World AI Attacks to Prepare for Now," 2023. [Online]. Available: https://www.darkreading.com/cyberattacks-data-breaches/google-red-team-provides-insight-on-real-world-ai-attacks.

[5] A. Lundqvist, "Backdoor Attacks on AI Models," 2024. [Online]. Available: https://www.cobalt.io/blog/backdoor-attacks-on-ai-models.

[6] "Critical Scalability: Trend Micro Security Predictions for 2024. (n.d.)," [Online]. Available: https://www.trendmicro.com/vinfo/us/security/research-and-analysis/predictions/critical-scalability-trend-micro-security-predictions-for-2024.

[7] P. Uy, "AI Cyber-Attacks: The Growing Threat to Cybersecurity and Countermeasures," 2023. [Online]. Available: https://ipvnetwork.com/ai-cyber-attacks-the-growing-threat-to-cybersecurity-and-countermeasures/.

[8] "Proliferation of AI-driven Attacks Anticipated in 2024. (n.d.)," 2024. [Online]. Available: https://newsroom.trendmicro.com/2023-12-05-Proliferation-of-AI-driven-Attacks-Anticipated-in-2024.

[9] "The Need For AI-Powered Cybersecurity to Tackle AI-Driven Cyberattacks. (n.d.)," [Online]. Available: https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2024/the-need-for-ai-powered-cybersecurity-to-tackle-ai-driven-cyberattacks.

[10] "AI-powered cyber-crime: Barclays Private Bank. (n.d.)," [Online]. Available: https://privatebank.barclays.com/insights/2023/september/the-rise-of-ai-powered-cyber-crime/.

[11] H. Murphy, "Is artificial intelligence the solution to cyber security threats? FT.Com,," 2024. [Online]. Available: https://www.proquest.com/trade-journals/is-artificial-intelligence-solution-cyber/docview/2915062394/se-2.

[12] A. Z. Jalaluddin, "An Exploration of Countermeasures to Defend Against Weaponized AI Malware Exploiting Facial Recognition (Order No. 28094887)," 2020. [Online]. Available: https://www.proquest.com/dissertations-theses/exploration-countermeasures-defend-against/docview/2446975948/se-2.

[13] C. Barry, "5 Ways cybercriminals are using AI: Phishing," 2024. [Online]. Available: https://blog.barracuda.com/2024/03/28/-5-ways-cybercriminals-are-using-ai--phishing.

[14] "OneLogin (n.d.). Watch Out for AI-Powered Spear Phishing," 2024. [Online]. Available: https://www.onelogin.com/resource-center/infographics/cybersecurity-ai-spear-phishing.

[15] "How AI Will Supercharge Spear Phishing Attacks: Darktrace: Darktrace Blog. (n.d.)," [Online]. Available: https://darktrace.com/blog/ai-will-supercharge-spear-phishing.

[16] "We're All the Target: Generative AI and the Automation of Spear Phishing. (n.d.)," [Online]. Available: https://www.f5.com/company/blog/generative-ai-automation-of-spear-phishing.

[17] A. A. Mamun, "AI-Enabled Modeling and Monitoring of Data-Rich Advanced Manufacturing Systems (Order No. 30567656)," 2023. [Online]. Available: https://www.proquest.com/dissertations-theses/ai-enabled-modeling-monitoring-data-rich-advanced/docview/2852486421/se-2.

[18] "AI-Generated Malware and How It's Changing Cybersecurity. (n.d.)," [Online]. Available: https://www.impactmybiz.com/blog/how-ai-generated-malware-is-changing-cybersecurity/.

[19] "Global ransomware threat expected to rise with AI, NCSC warns. (n.d.)," [Online]. Available: https://www.ncsc.gov.uk/news/global-ransomware-threat-expected-to-rise-with-ai.

[20] "What Is Deepfake: AI Endangering Your Cybersecurity? (n.d.)," [Online]. Available: https://www.fortinet.com/resources/cyberglossary/deepfake.

[21] "What is a Deepfake Attack?: CrowdStrike," 2024. [Online]. Available: https://www.crowdstrike.com/cybersecurity-101/social-engineering/deepfake-attack/.

[22] S. Sjouwerman, "Council Post: Deepfake Phishing: The Dangerous New Face Of Cybercrime," 2024. [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2024/01/23/deepfake-phishing-the-dangerous-new-face-of-cybercrime/?sh=20c8b5484aed.

[23] "What Is Data Poisoning? - CrowdStrike," 2024. [Online]. Available: https://www.crowdstrike.com/cybersecurity-101/cyberattacks/data-poisoning/.

[24] "What Is Adversarial AI in Machine Learning? (n.d.)," [Online]. Available: https://www.paloaltonetworks.com/cyberpedia/what-are-adversarial-attacks-on-AI-Machine-Learning.

[25] "What Is Adversarial AI in Machine Learning? (n.d.)," [Online]. Available: https://www.paloaltonetworks.com/cyberpedia/what-are-adversarial-attacks-on-AI-Machine-Learning.

[26] N. Hassan, "Adversarial machine learning: Threats and countermeasures: TechTarget," 2023. [Online]. Available: https://www.techtarget.com/searchenterpriseai/tip/Adversarial-machine-learning-Threats-and-countermeasures.

[27] "OneLogin (n.d.). What Is Credential Stuffing? Akami," 2024. [Online]. Available: https://www.akamai.com/glossary/what-is-credential-stuffing.

[28] "Credential Stuffing Attacks: Examples and Prevention: Wiz," 2024. [Online]. Available: https://www.wiz.io/academy/credential-stuffing.

[29] "The Growing Threat of Credential Stuffing and 6 Ways to Defend Your Organization. (n.d.)," [Online]. Available: https://www.hackerone.com/knowledge-center/growing-threat-credential-stuffing-and-6-ways-defend-your-organization.

[30] P. Sullivan, "How does credential stuffing enable account takeover attacks?: TechTarget," 2018. [Online]. Available: https://www.techtarget.com/searchsecurity/answer/How-does-credential-stuffing-enable-account-takeover-attacks.

[31] DataDome, "Credential Stuffing Attacks & Methods for Prevention," 2022. [Online]. Available: https://securityboulevard.com/2022/09/credential-stuffing-attacks-methods-for-prevention/.

[32] "What is a supply chain attack?," 2024. [Online]. Available: https://www.sailpoint.com/identity-library/supply-chain-attack/.

[33] "(N.d.)," [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2022/04/11/supply-chain-attacks-on-ai/?sh=6ed6d377edc7.

[34] Y. Wang, Q. Yan, N. Ivanov and X. Chen, "A Practical Survey on Emerging Threats from AI-driven Voice Attacks: How Vulnerable are Commercial Voice Control Systems?," 2023.

[35] O. A. Beg, A. K. Asad, W. U. Rehman and A. Hassan, "A Review of AI-Based Cyber-Attack Detection and Mitigation in Microgrids. Energies, 16(22), 7644.," 2023. [Online]. Available: https://doi.org/10.3390/en16227644.

[36] H. Gonaygunta, "Factors Influencing the Adoption of Machine Learning Algorithms to Detect Cyber Threats in the Banking Industry (Order No. 30811800)," 2023. [Online]. Available: https://www.proquest.com/dissertations-theses/factors-influencing-adoption-machine-learning/docview/2915921368/se-2.

[37] C. O. Benedict, "Detecting Security Anomalies Using Machine Learning for Smart Homes (Order No. 30571460)," 2023. [Online]. Available: https://www.proquest.com/dissertations-theses/detecting-security-anomalies-using-machine/docview/2838610558/se-2.

[38] J. Gusman, "The Deployment of Artificial Intelligence and Machine Learning Within the Field of Cybersecurity for Intelligent Decision Making: A Qualitative Study (Order No. 30639374)," 2023. [Online].

[39] K. Boonyapredee, "Southeast Asia Cyber Development Challenges in the FinTech Industry (Order No. 30525741)," 2023. [Online]. Available: https://www.proquest.com/dissertations-theses/southeast-asia-cyber-development-challenges/docview/2822572703/se-2.

[40] J. Ables, "Explainable Intrusion Detection Systems Using White Box Techniques (Order No. 30812542)," 2023. [Online]. Available: https://www.proquest.com/dissertations-theses/explainable-intrusion-detection-systems-using/docview/2903798888/se-2.

[41] D. L. T. P. Gonzalo, "Distributed AI-Defense for Cyber Threats on Edge Computing Systems (Order No. 28715667)," 2021. [Online]. Available: https://www.proquest.com/dissertations-theses/distributed-ai-defense-cyber-threats-on-edge/docview/2572563940/se-2.

[42] I. Zhou, "Intelligent Frost Prediction and Active Protection Cyber-Physical Systems in the Agricultural Sector (Order No. 30757869)," 2023. [Online]. Available: https://www.proquest.com/dissertations-theses/intelligent-frost-prediction-active-protection/docview/2901818834/se-2.

[43] "The Next Paradigm Shift: AI-Driven Cyber-Attacks," 2023. [Online]. Available: https://bridgeforum.io/ideas/the-next-paradigm-shift-ai-driven-cyber-attacks/.

[44] Akitra, "Automated Threat Hunting: Leveraging AI and Machine Learning for Proactive Security Measures," 2024. [Online]. Available: https://medium.com/@akitrablog/automated-threat-hunting-leveraging-ai-and-machine-learning-for-proactive-security-measures-cddca54d6517.

[45] SPGLOBAL , "How AI is changing defense technology. Spglobal," 2024. [Online]. Available: https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/how-ai-is-changing-defense-technology-81571291.

[46] "What is User Entity and Behavior Analytics (UEBA)? (n.d.)," [Online]. Available: https://www.fortinet.com/resources/cyberglossary/what-is-ueba.

[47] "What is User and Entity Behavior Analytics (UEBA)?: CrowdStrike," 2024. [Online]. Available: https://www.crowdstrike.com/cybersecurity-101/identity-protection/user-and-entity-behavior-analytics-ueba/.

[48] J. C. Haass, "Cyber Threat Intelligence and Machine Learning," in *Fourth International Conference on Transdisciplinary AI (TransAI)*, Laguna Hills, CA, USA, 2022.

[49] A. Sidhu, "AI-Driven Threat Intelligence: Leveraging Machine Learning to Empower Cybersecurity Applications for Enhanced Threat Detection and Response," 2023. [Online]. Available: https://zenodo.org/records/8050866.

[50] A. Turgeman, "Council Post: Machine Learning And Behavioral Biometrics: A Match Made In Heaven," 2018. [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2018/01/18/machine-learning-and-behavioral-biometrics-a-match-made-in-heaven/?sh=7ffa592C3306.

[51] Y. B. W. Piugie, J. D. Manno, C. Rosenberger and C. Charrier, "How Artificial Intelligence can be used for Behavioral Identification?," in *International Conference on Cyberworlds (CW)*, Caen, France, 2021.

[52] A. Chan, "Can AI Be Used for Risk Assessments?," 2023. [Online]. Available: https://www.isaca.org/resources/news-and-trends/industry-news/2023/can-ai-be-used-for-risk-assessments. [Accessed 2024].

[53] J. Chiappetta, "How Automated AI Code Analysis Can Scale Application Security," 2023. [Online]. Available: https://betterappsec.com/how-automated-ai-code-analysis-can-scale-application-security-667002ad63c4.

[54] SeventhQueen, "AI-Driven Networks Anomaly Detection: Best Guide 2024: Infraon," 2024. [Online]. Available: https://infraon.io/blog/a-guide-on-ai-driven-networks-anomaly-detection/.

[55] S. Ambassadors, "AI Cryptography: Enhancing Security and Privacy in the Digital Age," 2023. [Online]. Available: https://medium.com/@singularitynetambassadors/ai-cryptography-enhancing-security-and-privacy-in-the-digital-age-db5c1bbf5fdb.

[56] D. Pissanidis and K. Demertzis, "Integrating AI/ML in Cybersecurity: An Analysis of Open XDR Technology and its Application in Intrusion Detection and System Log Management," 2023.

[57] D. Praveenraj, M. Victor, C. Vennila, A. Alawadi, P. Diyora, N. Vasudevan and T. Avudaiappan, "Exploring Explainable Artificial Intelligence for Transparent Decision Making. E3S Web of Conferences," 2023.

[58] C. Hacks , "Federated Learning: A Paradigm Shift in Data Privacy and Model Training," 2024. [Online]. Available: https://medium.com/@cloudhacks_/federated-learning-a-paradigm-shift-in-data-privacy-and-model-training-a41519c5fd7e.

[59] "Introducing Google Threat Intelligence: Actionable threat intelligence at Google scale | Google Cloud Blog. (n.d.)," [Online]. Available: https://cloud.google.com/blog/products/identity-security/introducing-google-threat-intelligence-actionable-threat-intelligence-at-google-scale-at-rsa.

[60] B. N, "Volkswagen Hacked - Hackers Stolen 19,000 Documents From VW Server," 2024. [Online]. Available: https://cybersecuritynews.com/volkswagen-hacked/.

[61] M. E. Tara Seals, "Capital One Attacker Exploited Misconfigured AWS Databases," 2023. [Online]. Available: https://www.darkreading.com/cyberattacks-data-breaches/capital-one-attacker-exploited-misconfigured-aws-databases.

[62] NHS, "Data breach: trusts shared patient details with Facebook without consent," 2023. [Online]. Available: https://www.theguardian.com/society/2023/may/27/nhs-data-breach-trusts-shared-patient-details-with-facebook-meta-without-consent.

[63] A. Thompson, "The MGM Resorts Attack: Initial Analysis," 2023. [Online]. Available: https://www.cyberark.com/resources/blog/the-mgm-resorts-attack-initial-analysis.

[64] C. W. Jai Vijayan, "3CX Supply Chain Attack Tied to Financial Trading App Breach," 2023. [Online]. Available: https://www.darkreading.com/cyberattacks-data-breaches/3cx-supply-chain-attack-originated-from-breach-at-another-software-company.

[65] H. Chen and K. Magramo, "Finance worker pays out $25 million after video call with deepfake “chief financial officer," 2024. [Online]. Available: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html.

[66] B. Toulas, "Activision confirms data breach exposing employee and game info," 2023. [Online]. Available: https://www.bleepingcomputer.com/news/security/activision-confirms-data-breach-exposing-employee-and-game-info/.

[67] "Chinese hackers are using AI to inflame social tensions in US, Microsoft says," 2024. [Online]. Available: https://therecord.media/china-ai-influence-operations.

[68] S. Bocetta, "Has an AI Cyber Attack Happened Yet?," 2020. [Online]. Available: https://www.infoq.com/articles/ai-cyber-attacks/.

[69] K. White, "Real-Life Examples of How AI Was Used to Breach Businesses," 2024. [Online]. Available: https://oxen.tech/blog/real-life-examples-of-how-ai-was-used-to-breach-businesses-omaha-ne/.