CIDAI Centre of Innovation for Data tech and Artificial Intelligence

AGÈNCIA DE CIBERSEGURETAT DE CATALUNYA

Llibre blanc sobre la Intel·ligència Artificial aplicada a la Ciberseguretat

La IA per donar resposta als reptes del sector de la ciberseguretat a Catalunya

El contingut d'aquest document és titularitat de l'Agència de Ciberseguretat de Catalunya i resta subjecte a la llicència de Creative Commons BY-NC-ND 4.0.
Aquest informe es publica sense cap garantia específica sobre el contingut

Reconeixement: s'ha de reconèixer l'autoria de l'obra de la manera especificada per l'autor o el llicenciador (en tot cas, no de manera que suggereixi que gaudeix del suport o que dona suport a la seva obra)

No comercial: no es pot emprar aquesta obra per a finalitats comercials o promocionals

Sense obres derivades: no es pot alterar, transformar o generar una obra derivada a partir d'aquesta obra
En reutilitzar o distribuir l'obra cal que s'esmentin els termes de la llicència de la mateixa.
El text complert de la llicència es pot consultar a https://creativecommons.org/licenses/by-nc-nd/4.0/
Qualsevol mediació relacionada amb disputes derivades de la llicència es durà a terme d'acord amb les normes de mediació de la World Intellectual Property Organization.

Llista d'abreviacions

ACC	Agència de Ciberseguretat de Catalunya
ACIA	Associació Catalana d'Intel·ligència Artificial
AIRA	Artificial Intelligence Reearch Alliance
BEC	Business Email Compromise
CERT	Computer Emergency Response Team
CSIRT	Computer Security Incidents Response Team
CTI	Cyber Threat Intelligence
CSA	Cyber Security Act
DL	Deep Learning
DDoS	Distributed Denial of Service
DOGC	Diari Oficial de la Generalitat de Catalunya
DPIA	Data Protection Impact Assessments
ECCC	European Cybersecurity Competence Centre
ECSO	European Cyber Security Organisation
EDT	Emerging Disruptive Technology
EEE	Espai Econòmic Europeu
ENISA	Agència de la Unió Europea per a la Ciberseguretat
ETL	Extraction, Transformation and Load
GDPR	General Data Protection Regulation
GPU	Graphical Processing Unit
IA	Intel·ligència Artificial
IoT	Internet of Things
ISAC	Information Sharing and Analysis Center
ML	Machine Learning
NLP	Natural Language Processing
OSINT	Open-Source Intelligence
OTP	One Time Password
RIS3CAT	Estratègia per a l'especialització intel·ligent de Catalunya
SDN	Software Defined Network
SVM	Support Vector Machine
SOAR	Security Orchestration, Automation and Response
SOC	Security Operations Center
TTP	Tàctiques, Tècniques, Procediments
UE	Unió Europea

Resum executiu

Els darrers anys els avanços a la ciència de la Intel·ligència Artificial han estat remarcables, tenint gran impacte en diferents sectors industrials i socials. Existeixen en la actualitat una gran quantitat d'exemples que il·lustren com la IA s'aplica a sectors tan diversos com salut, agroalimentari, turístic, industrial, mobilitat i altres, i està present en equipaments, dispositius i aplicacions quotidianes i d'ús massiu, com mòbils, assistents virtuals i altres. El sector de la ciberseguretat no està al marge d'aquesta tendència, on la introducció de la Intel·ligència Artificial, tant des del vessant científic i de recerca, com des de la perspectiva empresarial es va iniciar fa dècades i compta amb un llarg recorregut. L'automatització, l'autonomia o la robustesa de les solucions i processos per a la gestió de la ciberseguretat sempre han estat objectius prioritaris pels professionals de la seguretat. El gran desenvolupament de les tecnologies de la informació, la connectivitat, la universalització de la digitalització, conjuntament amb l'expansió d'aplicacions en el seu moment considerades com disruptives, com per exemple la computació al núvol o Internet de les coses fan encara més necessària la integració de solucions basades en IA per a la gestió de la ciberseguretat, tant de les infraestructures com de la pròpia informació.

Es constata també, que el món de la ciberdelinqüència també adopta les possibilitats i avenços que ofereix la IA. Els ciberdelinqüents d'avui, gràcies a l'adopció de la IA, tenen el coneixement i la capacitat de millorar i automatitzar els seus atacs, així com la capacitat d'adaptar les seves reaccions de forma dinàmica segons la resposta del sistemes de protecció i defensa a què s'enfronten. Un cas extrem que il·lustra aquest escenari és l'existència de serveis que la ciberdelinqüència ofereix sota un model de *crime-as-a-service*.

Per tant és important comprendre de quina manera la intersecció entre la IA i el món de la ciberseguretat d'una banda pot redundar en una major capacitat de reacció davant d'un context amb més ciberamenaces i, per altra banda, pot ajudar a fer front a una major complexitat operativa deguda a la variada tipologia dels atacs i de les eines i processos implicats en la resposta. Aquesta tasca, però, no és senzilla, doncs no sempre l'aplicació de la IA en suport de la ciberseguretat s'ha mostrat tan eficient com hom esperava, ja sigui per un excés d'expectatives creades, perquè no se n'ha fet un desplegament correcte o per manca de prou coneixement de les solucions basades en IA per part dels professionals de la ciberseguretat. Aquests, i altres factors, poden reduir la confiança en les eines que aporta la IA aplicades al domini de la ciberseguretat.

Davant d'aquest escenari, el document fa una anàlisi per separat dels sectors de la ciberseguretat i de la IA per després abordar la confluència dels dos àmbits, desenvolupant pot esdevenir un factor de transformació per a la ciberseguretat i posant el focus d'aquest punt en el cas de Catalunya.

L'estudi de la ciberseguretat s'aborda a partir de les cinc accions relacionades amb un ciberatac segons el framework del NIST, emprat en la taxonomia radar de l'ECSO: identificar, protegir, detectar, respondre i recuperar, així com també els grans grups d'amenaces caracteritzats per ENISA, des del *malware* fins el ciberatacs específics dirigits contra la indústria del videojoc. Al mateix temps les diferents tècniques i metodologies de la ciberseguretat hauran d'evolucionar per afrontar reptes que sorgeixen degut a l'evolució de la tecnologia, com per exemple els que planteja la irrupció del metavers, la proliferació de la IoT o la ciberseguretat industrial en el marc de la Indústria 4.0/5.0.

Per la seva banda la IA és una disciplina que ha evolucionat enormement en els darrers anys gràcies a la disponibilitat de dades massives per a entrenar models i la millora i generació de nous algorismes, conjuntament amb l'increment de la potència de càlcul gràcies al processament d'altes prestacions i la irrupció de serveis i recursos disponibles al núvol.

La IA té diverses aplicacions en el món de la ciberseguretat que es poden categoritzar i avaluar segons el model denominat *Cyber Kill Chain*, que defineix set fases d'un ciberatac dirigit (reconeixement, armamentització, lliurament, explotació, instal·lació i comandament i control), i que és útil per categoritzar tant les estratègies d'atac emprades pels ciberdelinqüents com per les estratègies defensives dels equips de resposta contra incidents de ciberseguretat. Amb tot, l'aplicació de solucions de seguretat basades en IA presenta un clar efecte transformador, millorant les diferents etapes de la ciberseguretat i amb un impacte directe sobre les persones i equips de seguretat, sobre les tecnologies i infraestructures que es volen gestionar i, finalment, sobre els processos que governen la gestió de la ciberseguretat.

Tanmateix, les pròpies solucions o aplicacions de la IA presenten debilitats davant d'un ciberatac, ja que introdueix algunes febleses inherents a la naturalesa de la IA i que no es poden eliminar, com per exemple el robatori i/o manipulació d'algorismes, la introducció de mostres falses en l'entrenament de models, l'enverinament de l'aprenentatge i tècniques esteganogràfiques. A la vegada tenen certes limitacions, com per exemple la necessitat de disposar d'important volums de dades no sempre disponibles i mantenir la confidencialitat de les mateixes, els costos en el desenvolupament de models eficients, la participació de l'especialista humà que encara ha de verificar els resultats de la IA i, finalment, la manca de professionals que tinguin suficient coneixement i experiència en els dos àmbits, IA i ciberseguretat a la vegada.

Aquestes disciplines tenen la potencialitat d'evolucionar i créixer en el context català gràcies a diversos factors, com un sistema de coneixement consolidat amb universitats i centres de recerca i tecnològics capdavanters, la capacitat d'atreure talent essent un pol per a les *startups* tecnològiques, el lideratge d'iniciatives com les estratègies CATALONIA.AI i el Pla Nacional de Ciberseguretat, un important teixit empresarial i associatiu i la disposició d'infraestructures adequades, entre altres. Aquests factors fan que els sectors de la IA i de la ciberseguretat a Catalunya presentin les següents xifres clau:

- IA: engloba aproximadament un total de **179 empreses**, amb una facturació de **1.358 milions d'euros** anuals i ocupen un total d**e 8.483 professionals** (dades de 2019)

- Ciberseguretat: engloba aproximadament **495 empreses,** amb una facturació de **1.071 milions d'euros** i ocupant a **9.414 professionals** (dades de 2023)

Per altra banda també s'identifiquen algunes barreres significatives per al desplegament de solucions de ciberseguretat basades en IA:

- manca de formació creuada entre les dues disciplines que alenteix una adopció efectiva de la IA per a la ciberseguretat

- la dificultat de retenció dels professionals davant d'oportunitats laborals externes i les facilitats del teletreball

- expectatives sobredimensionades sobre la capacitat de la IA en la resolució de problemes.

- resistència al canvi, habitual en les organitzacions, davant l'adopció de canvis disruptius i on s'hi ha de sumar la reticència del professional e la ciberseguretat davant la poca transparència i explicabilitat en la majoria de les soluciona IA.

- manca de disponibilitat de dades de qualitat, homogeneïtat de les mateixes i de l'explicabilitat de les decisions preses per la IA

- manca d'entorns de prova realistes

- escenari dinàmic i canviant degut a l'activitat incessant dels ciberdelinqüents

- condicions organitzatives i normatives i ètiques

Com a resultat de l'anàlisi d'aquestes barreres sorgeixen una sèrie de recomanacions adreçades tant a l'administració com al sector empresarial per tal d'afavorir l'adopció de la IA en el món de la ciberseguretat, entre les quals destaquen:

- La creació d'itineraris de formació especialitzats on hi participin formadors professionals (locals i internacionals) d'ambdós sectors.

- La implementació d'un entorn col·laboratiu per poder compartir dades de forma segura i donant compliment a les normatives de privadesa existents.

- L'adopció d'un marc legal que asseguri la privadesa de les dades (sobretot personals) però que alhora fomenti i agilitzi la compartició d'aquestes dades.

- L'adopció d'un marc ètic per aconseguir un equilibri just entre els avenços tecnològics, els interessos empresarials i una sèrie de garanties per a la ciutadania.

1 Introducció

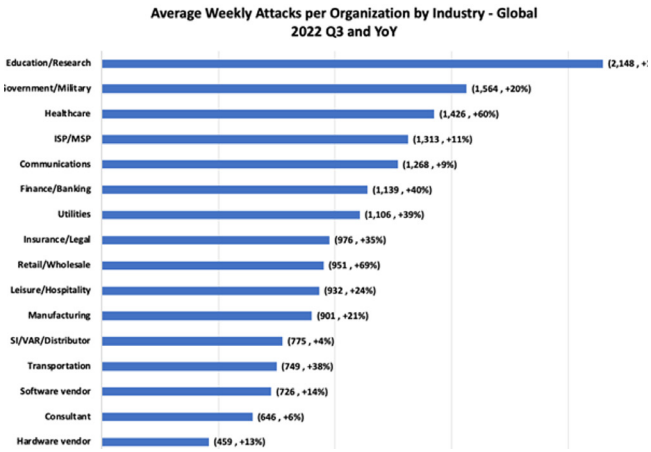
El procés de digitalització en què la societat actual es troba immersa està transformant molts aspectes de l'economia i de les relacions socials. En aquests moment s'estan consolidant les bases dels canvis que aquesta transformació està aportant i, a la vegada, es comença a visualitzar l'impacte de transformacions tecnològiques futures. Les grans possibilitats d'innovació que permeten les noves tecnologies d'informació combinades amb unes xarxes de comunicacions cada vegada més potents i ubiques, estan demostrant una enorme influència sobre la societat en conjunt. Un exemple ben conegut, i que posa de manifest aquesta influència, s'ha viscut a nivell global durant l'etapa de la pandèmia del Covid-19, en què la utilització massiva de les eines digitals, i en especial el teletreball, van ajudar a què les dinàmiques socials i l'economia no s'aturessin completament. Una conseqüència directa d'aquest procés de transformació és l'increment del risc i de les amenaces de ciberseguretat dels sistemes d'informació i dispositius connectats, ja siguin personals o corporatius, degut a dos factors. Per una banda, la connectivitat extrema augmenta el nombre de superfícies d'atac a traves de les quals es poden materialitzar els ciberatacs. Per altra banda, els ciberatacants disposen de manera continuada de més coneixement, recursos i capacitat per desenvolupar les eines necessàries per planificar i executar atacs cada cop més sofisticats i amb una major capacitat de de dany.

Aquesta situació es reflecteix clarament a les estadístiques que recullen l'activitat dels ciberatacants. Segons informes sobre intel·ligència de ciberamenaces[1], els atacs s'han incrementat en un 28% a nivell global durant el darrer quadrimestre de 2022 comparat al mateix període del 2021, i la mitjana d'atacs setmanals per organització a tot el món va superar els 1.130. I els incidents de ciberseguretat amenacen pràcticament tots els sectors d'activitat com es pot veure en la Figura 1, en què Educació, Governs/Exèrcits i Salut han estat els més atacats a nivell mundial durant els darrers mesos de 2022.

Per descomptat, les empreses, organismes i administracions públiques catalanes no queden al marge d'aquesta activitat ciberdelictiva. Només durant el 2020, Catalunya va patir 800 milions de ciberatacs. I en els darrers mesos s'han produït intrusions amb greus conseqüències a institucions educatives i sanitàries del país. Aquests incidents són només la punta de l'iceberg de la seriosa problemàtica associada a la ciberseguretat.

1 https://blog.checkpoint.com/2022/10/26/third-quarter-of-2022-reveals-increase-in-cyberattacks/

Figura 1: Mitjana setmanal global d'atacs per sector durant Q3 2022 i increment respecte 2021 (font: Check Point Research)



En aquest context, i per aquesta raó, la ciberseguretat emergeix com una disciplina fonamental i necessària, sota la qual s'agrupen totes aquelles tecnologies i metodologies destinades a combatre els ciberatacs, amb l'objectiu de protegir els actius, ja siguin digitals o no, susceptibles de patir les conseqüències d'aquesta activitat maliciosa. Les perspectives de futur indiquen que la rellevància de la ciberseguretat i la necessitat de protecció enfront de la creixent activitat ciberdelictiva anirà en augment.

Davant dels avanços tecnològics i de la transformació digital que es dibuixa en els propers anys, cal veure com s'aborden els nous riscos cibernètics i és precisament en aquest punt on la Intel·ligència Artificial (IA) té una importància cabdal, ja que esdevé fonamental per afrontar l'evolució i sofisticació dels ciberatacs que proliferen gràcies, entre altres elements, a la irrupció d'un mercat negre que obre la possibilitat d'executar atacs cibernètics com a servei (*Crime-as-a-Service*).

D'aquesta manera, la integració de solucions basades en IA als sistemes de la gestió de la seguretat ha de permetre donar resposta als següents reptes principals:

- Capacitat de reacció: els ciberatacs creixen en nombre i en velocitat, per sota dels 4 minuts de mitjana en què s'estima la durada d'un ciberatac. Els ciberatacants fan ús d'atacs automatitzats i veloços, generalment amb poca interacció humana que a més poden adaptar-se segons la resposta de l'objectiu.

- Complexitat operativa: la proliferació de sistemes i la diversitat de tecnologies, arran de la integració de tecnologies per al teletreball i de l'ús de plataformes de computació al núvol, així com la millora de prestacions dels sistemes TI, demanden que els serveis de ciberseguretat hagin de respondre molt ràpidament, més enllà del que la intervenció humana és capaç d'oferir. Per tant cal disposar de la capacitat de generar coneixement valuós i interpretable, amb un ampli ventall d'eines i processos complexes com, per exemple, analítica de grans volums de dades amb l'objectiu de donar una resposta eficient i immediata [2].

- Manca de talent en matèria de ciberseguretat: segons l'organització *International Information System Security Certification Consortium* (ISC)2, en l'actualitat manquen al voltant de 3,5 milions de professionals de la ciberseguretat a tot el món, una xifra que dobla la de l'any 2021, mentre que a l'estat espanyol s'hauria incrementat un 57% [3]. Els professionals de la ciberseguretat són uns recursos cobejats i amb una alta càrrega de treball, cosa que afavoreix la rotació laboral i l'aparició del fenomen de l'esgotament (*burnout*). Aquesta mancança generalitzada d'especialistes i experts en ciberseguretat, reforça encara més la necessitat de noves eines i processos automatitzats, intel·ligents i eficients que puguin assumir part de la càrrega de treball.

Cal dir que l'ús i aplicació d'algorismes i tecnologies d'IA i aprenentatge automàtic en el domini de la ciberseguretat no és nou. Fins fa relativament poc temps (les primeres aplicacions són prèvies a 1990[4,5]), aquest tipus de solucions s'integraven majoritàriament a la fase de detecció com, per exemple, *spam*, intrusions o *malware*. Així, ja des d'aquells inicis, s'ha aprofitat la capacitat d'aprenentatge dels algorismes i eines basades en IA per construir un model d'activitat normal d'un recurs (dispositiu, compte, etc.), i detectar amb posterioritat si actua fora d'aquesta normalitat. Els processos automatitzats de detecció aporten gran valor, doncs permeten monitoritzar cada un dels segments de la xarxa, millorant el rendiment del sistema de ciberseguretat, i reduint el seu temps de resposta.

Però, en els darrers temps, s'ha ampliat l'aplicació de les tecnologies basades en la IA en el domini de la ciberseguretat. Segons l'informe *Global AI Software Forecast 2022* de Forrester Research [6], el mercat del programari d'IA passarà dels 33.000 milions de dòlars de 2021 als 64.000 milions de dòlars el 2025. Segons aquest informe, "la ciberseguretat és la categoria de software d'IA de més ràpid creixement, amb especial atenció en els processos de supervisió en temps real dels atacs i la resposta als mateixos", amb una taxa de creixement anual compost (CAGR) del 22,3%. Les dues següents categories, gestió de clients i capital humà (22%) i optimització de processos, coneixement i intel·ligència de dades (18,3%), també tenen elements de ciberseguretat, per tant, l'impacte en els fabricants de solucions de seguretat podria ser encara més significatiu.

Segons Bruce Schneider [7], la IA és la tecnologia més prometedora per a la ciberseguretat, però destaca que també els grups maliciosos de ciberatacants es poden beneficiar de les seves capacitats. Per poder donar resposta a les noves amenaces, cal comprendre completament les estratègies, tàctiques i tècniques dels adversaris, i solament es podrà aconseguir si s'analitzen els ciberatacs tradicionals i els basats en IA. Aquest procés d'anàlisi i aprenentatge del comportament dels adversaris i el seus atacs avançats requereix el compromís i suma de capacitats de màquines i persones, on tradicionalment, la superioritat de les màquines destaca pel seu abast, velocitat i escalabilitat davant de les persones, que són bons raonant i pensant.

2 https://www.darkreading.com/operations/tool-overload-attack-surface-expansion-plague-socs

3 https://techmonitor.ai/technology/cybersecurity/burnout-is-rising-in-cybersecurity-industry

https://www.zdnet.com/article/cybersecurity-burnout-is-real-and-its-going-to-be-a-problem-for-all-of-us/

4 J. Jachner and V. K. Agarwal, "Data Flow Anomaly Detection," in IEEE Transactions on Software Engineering, vol. SE-10, no. 4, pp. 432-437, July 1984, doi: 0.1109/TSE.1984.5010256.

5 Kantzavelou, Ioanna. (1994). An Attack Detection System for Secure Computer Systems.

6 https://www.forrester.com/report/global-ai-software-forecast-2022/RES178146

7 Bruce Schneider es un criptógraf, expert en seguretat informàtica y escriptor. Es autor de diversos llibres de seguretat informàtica i criptografia i és el fundador i CTO de Counterpane Internet Security (https://es.wikipedia.org/wiki/Bruce_Schneier)

1.1. Objectiu del Document

El document ofereix algunes claus pel foment de l'adopció i la integració de solucions i sistemes basats en algorismes i tecnologies d'IA a les diferents capes i processos de la ciberseguretat des d'una perspectiva catalana. Dona una visió pràctica i senzilla de l'impacte de la IA sobre la ciberseguretat per tal que qualsevol tipus d'empresa i agent afectat, operant al territori català, pugui iniciar o continuar el seu procés de transformació digital, integrant els mecanismes necessaris per fer front al nou panorama d'amenaces i riscos emergents.

Els objectius específics del document són:

- Oferir una visió de l'estat de l'art i de la maduresa de la IA aplicada en la ciberseguretat a Catalunya.
- Reflexionar sobre les condicions que faciliten o obstaculitzen el desenvolupament de solucions basades en IA per a la ciberseguretat i el seu desplegament.
- Recomanar àrees d'aplicació i desplegament de solucions de ciberseguretat basades en IA com punt de partida de conscienciació del potencial de la IA aplicada a la ciberseguretat.
- Destacar les capacitats en matèria d'IA a desenvolupar per fer front als escenaris actuals i futurs en l'àmbit de la seguretat cibernètica.
- Identificar les oportunitats existents en el mercat dels serveis i productes de ciberseguretat que poden ser cobertes amb solucions d'IA.

1.2. A qui va adreçat

En ocasions les empreses, i en particular les petites, tenen dificultats a l'hora d'iniciar el camí cap a l'adopció de tecnologies disruptives al seu negoci. Aquest document convida aquests agents a prendre decisions corporatives i organitzatives encaminades a una millor protecció i defensa de les seves infraestructures i informació, on la IA serà la base de les solucions desplegades, per tal de prevenir i, en el seu cas estalviar-se, problemes majors amb greu impacte potencial, si es veuen afectats per ciberatacs com per exemple, un *ransomware, malware*, intrusió/atac, sabotatge o altres.

Concretament el document va adreçat a:

- Les empreses que ofereixen productes o serveis de ciberseguretat coneixeran les oportunitats actuals i futures per al desenvolupament de noves solucions basades en IA.
- Les empreses que ofereixen productes o serveis d'IA entendran el mercat de la ciberseguretat com un nou sector de gran potencial i les oportunitats que presenta.
- Tots aquells consumidors de productes i serveis de ciberseguretat obtindran una visió de les amenaces cibernètiques del futur que requeriran el desplegament de tecnologies basades en la IA.
- Els emprenedors i inversors per tal que tinguin coneixement de les iniciatives que apliquen la IA en ciberseguretat que presenten millors perspectives de desenvolupament.
- Els professionals de la IA i de la ciberseguretat identificaran les sinergies existents entre ambdós sectors d'activitat i la corresponent demanda laboral potencial.

1.3. Metodologia emprada

L'elaboració i redacció d'aquest document s'ha realitzat en 3 fases, amb la finalitat d'assolir els objectius definits i adaptar-los a l'audiència objectiu.

La primera etapa ha consistit en una anàlisi de les necessitats i tendències existents en el domini de la ciberseguretat, les solucions de mercat que s'estan proposant i les iniciatives que s'estan treballant des de l'àmbit científic. Es van mantenir diverses sessions de treball de posada en comú, establint l'estructura del contingut, així com el seu abast.

A la segona etapa s'ha fet una anàlisi col·laborativa i participativa de més profunditat sobre les diferents tecnologies d'IA i ciberseguretat, les sinergies entre elles i avaluant diferents casos d'ús i projectes on han participat membres de l'equip de redacció del document.

A la tercera i darrera etapa s'han consensuat opinions, potencials limitacions, oportunitats de futur i recomanacions per a la implantació de la IA en ciberseguretat a l'àmbit català.

Per a la concepció i redacció del present document s'han tingut en compte estudis i tallers de naturalesa similar [8], així com l'experiència de l'equip de redacció i consultes puntuals a altres col·laboradors experts.

8 Workshop on the Implications of Artificial Intelligence and Machine Learning for Cybersecurity on 12-13, 2019, in Washington, D.C.

Ciberseguretat: conceptes, tendències i reptes

2

Índex

La ciberseguretat es pot definir com el conjunt de pràctiques, metodologies i tecnologies que s'utilitzen per a protegir d'atacs maliciosos els sistemes d'informació, xarxes d'ordinadors, dispositius mòbils, bases de dades i altres actius digitals, així com la informació que s'hi tracta.

Per determinar l'abast de la ciberseguretat és útil referir-se a la taxonomia definida per la *European Cyber Security Organizacion* (ECSO)[1] en el seu radar de mercat (Figura 1). S'hi representen els diferents proveïdors de productes i serveis agrupats en cinc àmbits d'actuació relacionats amb la seqüència d'esdeveniments d'un ciberatac: identificar, protegir, detectar, respondre i recuperar. Aquests àmbits coincideixen amb els que defineix el *Cybersecurity Framework 1.1 del NIST*[2] *(National Institute of Standards and Technology)*.

Figura 2: RADAR de mercat de la ciberseguretat (font: ECSO)

A continuació es defineixen els àmbits o conjunt d'accions que estableix el radar de mercat de l'ECSO i alguns exemples dels productes i serveis corresponents:

- Identificar: desenvolupament organitzatiu i estratègic de la infraestructura de TI de ciberseguretat per a gestionar els riscos de ciberseguretat en sistemes, persones, actius, dades i capacitats. Inclou productes i serveis de *Gestió del Risc, Governança o Compliment Normatiu*.

- Protegir: desenvolupament de solucions per reduir la superfície d'atac a sistemes de TI, i garantir la confidencialitat, integritat, disponibilitat i auditabilitat, així com el rendiment de serveis informàtics essencials. Inclou productes i serveis de *Gestió de Vulnerabilitats, Tests de Penetració o Formació en la Conscienciació (Awareness Training)*.

- Detectar: desenvolupament de les mesures adequades per identificar l'aparició de ciberatacs. Inclou productes i serveis de *Detecció d'Intrusions* o la implementació d'un *Security Operations Center (SOC)*.

- Respondre: disseny i implementació de mesures per a actuar adequadament davant dels incidents de ciberseguretat detectats. Inclou productes i serveis d'*Anàlisi Forense* o *Gestió d'Incidents*.

- Recuperar: desenvolupament i implementació de solucions destinades a mantenir els plans, els processos i els recursos per a la resiliència dels sistemes TI d'una organització i restaurar les capacitats afectades a causa d'un incident cibernètic. Inclou, per exemple, els *Sistemes de Recuperació*.

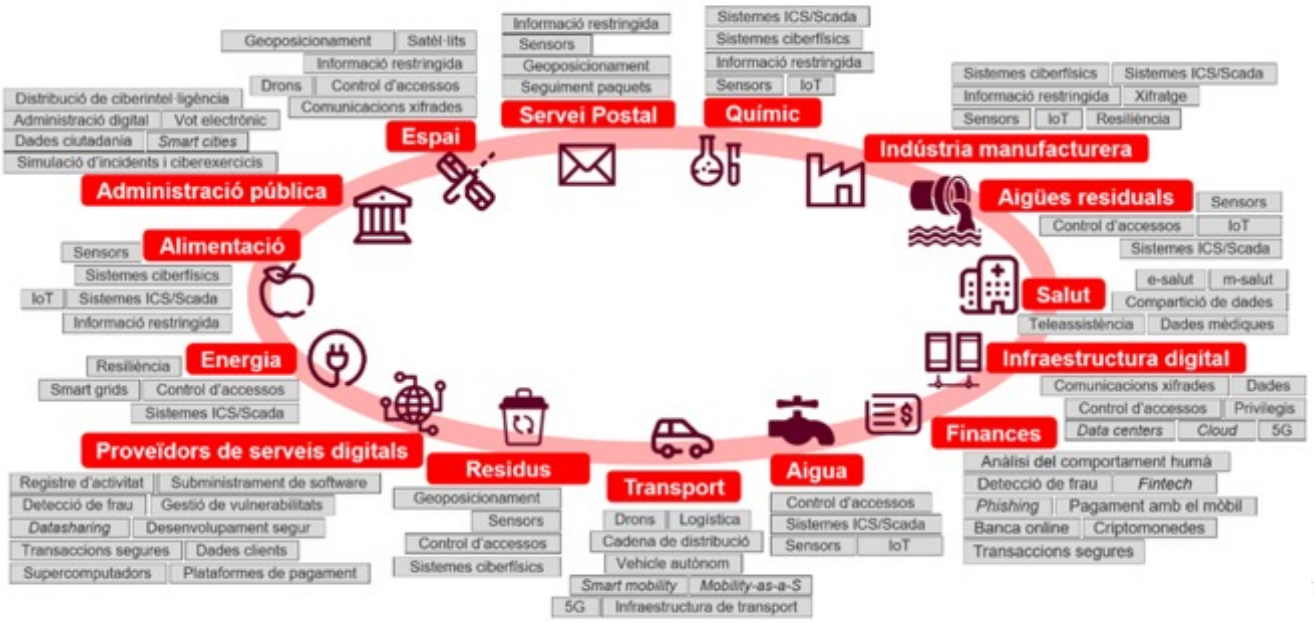
Degut a la transversalitat de la ciberseguretat, els serveis i productes descrits són demandats per tot tipus de sector d'activitat, amb èmfasi especial pels 15 àmbits essencials, a més de les indústries importants, que determina la recentment aprovada Directiva europea NIS 2 sobre seguretat de les xarxes i sistemes d'informació[3] (veure Figura 3).

1 https://ecs-org.eu/

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

3 https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52020PC0823&from=EN

Figura 3: Sectors destacats en la directiva NIS 2 (font: La Ciberseguretat a Catalunya 2023)



2.1. Conceptes

La ciberseguretat engloba el conjunt de mesures físiques, lògiques i de governança que protegeixen les següents propietats de les dades i els sistemes d'informació:

- Confidencialitat: la informació ni es posa a disposició, ni es revela a individus, entitats o processos no autoritzats.
- Disponibilitat: les entitats o processos autoritzats tenen accés als mateixos quan ho requereixen.
- Integritat: l'actiu d'informació no ha estat alterat de manera no autoritzada.
- Autenticitat: una entitat és qui diu ser i garanteix ser la font de la qual procedeixen les dades.
- Traçabilitat: permet conèixer l'origen, l'ús, el recorregut i la localització de la informació.

Les mesures de ciberseguretat busquen reduir els riscos cibernètics fins a un nivell acceptable, ja que el risc nul és inviable. En aquest sentit, es destaquen els següents tipus de riscos:

- Operacionals: un incident cibernètic pot afectar l'operativa de les organitzacions i la presa de decisions, que depèn, cada vegada més, de l'ús de noves tecnologies. Les interrupcions també poden afectar clients i proveïdors de la cadena de subministrament i, en el cas dels serveis essencials, l'estabilitat social i econòmica.
- Econòmics: un incident cibernètic pot causar la pèrdua de dades o l'aturada de sistemes que impliquin una interrupció de la productivitat i els ingressos. A més, la recuperació posterior a l'incident també pot tenir un cost econòmic elevat: anàlisi forense, restauració de dades i sistemes, recuperació de la reputació, sancions, etc.
- Legals: un incident cibernètic pot revelar una negligència o que els sistemes d'informació no hagin estat degudament protegits, i això pot derivar en sancions. En el cas de les dades personals, que són un dels principals actius buscat pels cibercriminals, el tractament inadequat pot ser objecte de sancions econòmiques molt importants.
- Reputacionals: un incident cibernètic pot afectar l'opinió que els clients o la ciutadania tenen d'una organització, d'una marca, o bé d'un producte o servei, i això pot acabar impactant en el balanç econòmic. Recuperar la reputació després d'un incident pot representar un sobreesforç en termes econòmics i de temps.[4]

4 https://www.sindicom.gva.es/public/Attachment/2019/8/1567163056GPF-OCEX-5312-Glosario-de-Ciberseguridad_2.pdf

2.2. Tendències del sector de la ciberseguretat

En aquests últims anys, les tendències observades en l'àmbit de la ciberseguretat han evolucionat constantment degut a factors com l'aparició de noves tecnologies i la capacitat d'adaptació dels ciberatacants en cada context social, com per exemple la pandèmia o el conflicte bèl·lic entre Ucraïna i Rússia.

Entre aquests factors cal destacar l'adopció de la IA, tant per part pels ciberatacants com pels equips i sistemes de ciberseguretat. Els primers aprofiten la capacitat de la IA per fer més sofisticats els tipus d'atacs així com per evadir-se amb més facilitat dels sistemes de seguretat. Per als segons, la IA permet fer front a reptes específics com, per exemple, la manca de personal especialitzat o la necessitat d'automatització de processos per permetre la detecció del gran volum d'amenaces que es produeixen a dia d'avui, impossible de dur a terme de forma manual.

L'últim informe Threat Landscape 2022[5] de l'Agència de la Unió Europea per a la Ciberseguretat (ENISA) destaca les següents amenaces, en les quals la IA ha tingut un paper especialment rellevant:

- *Malware*: es tracta d'un programari maliciós que s'executa sense el coneixement ni autorització del propietari o usuari de l'equip infectat i realitza funcions en el sistema que són perjudicials per a l'usuari i/o per al sistema. El *malware* basat en IA s'oculta de tal forma que la càrrega viral s'activa en el moment adequat per no ser detectat pels sistemes de seguretat, per exemple, quan l'usuari realitza una acció concreta, des d'una ubicació geogràfica concreta, etc.

- Ransomware: aquest tipus de ciberatac, consistent en inutilitzar el sistema diana mitjançant un *malware* que encripta les seves dades i demanar un rescat econòmic per recuperar-lo, ha estat la principal amenaça per a les organitzacions de tot el món els darrers anys. Es tracta d'atacs cada vegada més automatitzats, fins al punt que al mateix temps que es xifren els sistemes d'informació de la víctima es filtren les dades i es publiquen al mercat negre per a la seva venda. La protecció enfront del segrest de sistemes s'ha convertit en l'objectiu número 1 de les empreses de seguretat, que ja estan aplicant la IA per detectar i bloquejar ràpidament aquest tipus d'amenaça[6].

- Enginyeria Social: és una tècnica que fan servir els ciberdelinqüents per guanyar-se la confiança de l'usuari i aconseguir enganyar-lo perquè faci alguna, com pot ser executar un programa maliciós, compartir dades privades o comprar en llocs web fraudulents. Dins d'aquesta categoria, els atacs de *phishing* s'han multiplicat exponencialment i estan impactant fortament a tots els sectors. Els atacants comencen a utilitzar la IA per executar campanyes de *phishing* de forma més intel·ligent, incorporant esquers més convincents adaptats a cada víctima, aprofitant les dades personals robades disponibles al mercat negre.

- Ciberatacs a la cadena de subministrament: es produeixen quan un atacant, a partir d'un ciberatac a una organització, aconsegueix expandir l'impacte sobre els seus clients, socis o proveïdors . L'atac experimentat per la plataforma SolarWinds [7] el 2020, que va acabar afectant centenars dels seus clients, va ser una de les primeres revelacions d'aquest tipus d'amenaça. La utilització de la IA ha de permetre conèixer el nivell de ciberseguretat d'entorns complexos, com les cadenes de subministrament, identificar les vulnerabilitats i preveure incidents de seguretat.

- Robatori de dades personals: existeix tot un conjunt d'amenaces contra les dades personals dels usuaris que permeten l'accés i divulgació no autoritzats a determinades fonts de dades, com les bases de dades mal configurades exposades a Internet, els sistemes al núvol i API vulnerables o l'ús de tècniques de *scraping* per a la captura massiva de dades de xarxes socials. L'immens volum de dades personals filtrades i, sovint, a la venda al mercat negre pot ser utilitzat per alimentar sistemes d'IA destinats a massificar i personalitzar de forma automatitzada els ciberatacs.

5 https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022

6 https://computerhoy.com/noticias/tecnologia/numero-ataques-ransomware-ha-reducido-2022-pero-no-te-emociones-hay-trampa-1146265

7 https://www.techtarget.com/whatis/feature/SolarWinds-hack-explained-Everything-you-need-to-know

- Denegació de Servei: interrompre la disponibilitat del servei de recursos connectats és l'objectiu dels atacs de DDoS (*Distributed Denial of Service*). Es tracta de ciberatacs amb voluntat eminentment destructiva, tot i que en alguns casos aquests ciberatacs tenen com a motivació l'extorsió econòmica de la víctima. En aquest tipus d'atacs la utilització d'IA és habitual per governar de forma automatitzada les *botnets* que inunden amb fluxos de dades les víctimes, les quals utilitzen tècniques cada vegada més evolucionades, com els atacs en ràfegues de polsos, canviant les IP objectiu o fent ús de paquets de dades especialment elaborats i difícils de detectar com a maliciosos.

- Desinformació: les campanyes de desinformació segueixen en augment, impulsades per l'ús creixent de plataformes de xarxes socials i mitjans en línia. De fet, s'estan emprant diverses tecnologies basades en IA per a la difusió de *fake news*, com l'ús de xarxes de bots immenses en les xarxes socials simulant ser usuaris legítims, o la generació de continguts audiovisuals fraudulents en què se suplanten celebritats per emetre opinions falses. En el vessant defensiu, la utilització de la IA esdevé l'única eina possible per fer front a un allau de continguts i notícies falses, per tal de bloquejar-les a temps i amb un alt grau d'efectivitat.

  Aquestes amenaces s'han vist magnificades i ampliades durant el 2022, en alguns casos a causa de la situació geopolítica global, fortament marcada pel conflicte bèl·lic entre Ucraïna i Rússia[8]:

- Ciberatacs amb motivacions geopolítiques: la tensió entre països de diferents bàndols ha propiciat la proliferació de ciberatacs de naturalesa diversa però amb una mateixa finalitat: desestabilitzar els serveis essencials i institucions de països rivals. Aquests ciberatacs es llancen per fer una demostració de força, exercir pressió, desprestigiar el rival, difondre desinformació o sabotejar infraestructures crítiques.

- *Hacktivisme*: ha estat un protagonista principal de la ciberguerra entre Ucraïna i Rússia. Dues organitzacions han estat particularment implicades en aquesta activitat, Anonymous i Killnet. Anonymous es va posicionar ràpidament a favor d'Ucraïna i va llençar atacs a diferents infraestructures crítiques russes, com per exemple TV, webs governamentals, etc. Com a resposta, Killnet també va fer pública la seva posició pro-russa i es va atribuir l'autoria dels atacs posteriors contra infraestructures ucraïneses i dels seus aliats.

- Programes espia: l'any 2022 han aflorat al domini públic diversos casos en què s'ha utilitzat el programari espia Pegasus, de l'empresa israeliana NSO Group, de forma abusiva per a l'espionatge dels telèfons mòbils de membres de la societat civil [9]. El fet és que, segons NSO Group, aquest programari només està disponible per a institucions estatals i per a la lluita contra el crim i el terrorisme.

- Estafes de correu professional (*Business Email Compromise*, BEC). Durant el 2022[10] les estafes de correu professional han augmentat en nombre i en les quantitats econòmiques defraudades, degut a la seva simplicitat tècnica i als beneficis que aporten. Tot i que són atacs que es poden considerar com una tipologia de *phishing*, la realitat és que es caracteritzen per estar encara més elaborats, ja que aconsegueixen suplantar convincentment empreses proveïdores o càrrecs de responsabilitat per sol·licitar algun pagament a favor de l'atacant sota un pretext oportú.

- Fraus i robatoris de criptomonedes: la indústria dels criptovalors ofereix múltiples oportunitats als inversors, però també als ciberdelinqüents per obtenir beneficis il·lícitament. Tot i que els fraus relacionats amb les criptomonedes s'han reduït respecte anys anteriors degut a la devaluació en el mercat d'aquest tipus de monedes digitals, els cibercriminals disposen d'altres formes per continuar obtenint beneficis d'aquests entorns tecnològics cada vegada més complexos, com per exemple l'explotació de vulnerabilitats en plataformes de canvi, *bridges* i protocols de *blockchain* per al robatori de fons o de la mineria furtiva (*criptojacking*) utilitzant servidors amb gran capacitat computacional del núvol.

8 https://ciberseguretat.gencat.cat/ca/Informes-de-Tendencies-en-Ciberseguretat/index.html

9 ENISA Threat Landscape 2022, https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022

10 Slashnext: The State of *Phishing* (2022)

- Ciberatacs a la indústria del videojoc: cada cop hi ha més jocs i més *gamers*, es creen nous actius digitals i també nous esdeveniments com els *e-games*, on hi participen milions de persones i s'hi mouen importants sumes de diners. Durant el 2022, diverses companyies del sector del videojoc han patit incidents de ciberseguretat, com el robatori i filtració d'informació de nous vídeojocs o el bloqueig de l'accés dels jugadors en línia amb atacs de DDoS. Els *gamers* són també objectiu d'atacs de *phishing* dirigit per robar-los les credencials dels comptes, fer-los caure en fraus o infectar-los amb programaris maliciosos.

2.3. Reptes del sector de la ciberseguretat

L'evolució de la tecnologia planteja nous escenaris que suposen autèntics reptes per al sector de la ciberseguretat, el qual haurà de demostrar una gran capacitat d'adaptació per donar resposta als punts que es detallen a continuació[11].

- Metavers: l'aparició de l'experiència del metavers i el desplegament de la web 3.0 accentuaran la participació dels usuaris en entorns virtuals complexos on es comunicaran persones, empreses i màquines. S'incrementarà l'ús i intercanvi de dades personals i es produiran interaccions digitals a nivells diferents dels actuals, amb la conseqüent aparició de nous riscos de ciberseguretat.

- Computació al núvol: les organitzacions esdevenen cada cop més *cloud-cèntriques* en utilitzar progressivament més recursos i serveis disponibles al núvol. Per tant, necessitaran la flexibilitat de la ciberseguretat oferta des del propi núvol per mitjà d'arquitectures de seguretat SASE (*Secure Access Service Edge*) i controls d'accessos amb sistemes CASB (*Cloud Access Security Broker*). Aquestes tecnologies existeixen des de fa alguns anys, però ara ha arribat el moment del seu ús generalitzat.

- 5G: les xarxes 5G permetran l'expansió de la utilització d'altres tecnologies més o menys consolidades o fins i tot emergents, com per exemple la IoT, les comunicacions hologràfiques, la salut digital o altres. Les comunicacions 5G presenten un nou paradigma amb diferents mitjans de tractament de dades, possibilitat d'interconnectar dispositius i sistemes, etc. Que augmentaran la complexitat i les superfícies d'atac amb moltes més parts interactives i heterogènies. I, mentre que el 5G no està encara plenament desplegat, cal tenir en compte que els protocols 6G ja estan sobre la taula.

- Ciutats Intel·ligents (*smart cities*): des de fa uns anys, i sota el concepte de "ciutat intel·ligent" cada vegada més ciutats a nivell global adopten noves tecnologies amb l'objectiu de millorar

11 https://ciberseguretat.gencat.cat/web/.content/PDF/InformeProspectives2022.pdf

la qualitat de vida i els serveis que ofereixen als seus habitants: sensorització per a la recopilació de dades, tecnologies de big data per optimitzar la presa de decisions en la gestió operativa i el govern de la ciutat, sistemes d'*smartmobility*, etc. Ara bé, al mateix temps, incorporen nous factors de risc cibernètic, motiu pel qual les mesures de ciberseguretat han d'adquirir un necessari protagonisme, especialment quan es pensa en infraestructures crítiques en l'entorn, o sota la responsabilitat de la ciutat.

- IoT: la proliferació imparable de dispositius i xarxes que constitueixen la IoT són la base de la interconnexió digital de qualsevol objecte per permetre millorar la vida quotidiana de la ciutadania i l'operació de les organitzacions. Ara bé, els dispositius IoT disposen de poca capacitat computacional i poques mesures de protecció, de manera que són especialment vulnerables davant de ciberatacs i infeccions amb *malware* de tipus *botnet*. Per tant, la introducció de la ciberseguretat en els sistemes per a l'administració de xarxes IoT és un element fonamental per a l'èxit del desplegament massiu d'aquesta tecnologia.

- Múltiples factors d'autenticació: s'estima que el 81% dels incidents de ciberseguretat es deuen a contrasenyes febles o robades. A dia d'avui, es recomana utilitzar sistemes d'autenticació addicionals a les contrasenyes tradicionals, com ara factors biomètrics com la retina o la veu, autenticació a través del telèfon mòbil personal, un *token*, o els *passwords* d'un sol ús (*One Time Password, OTP*).

- Sistemes ciberfísics: aquests elements constitueixen la base de la Indústria 4.0 (i per extensió també de la Indústria 5.0, una evolució de l'anterior que posa el factor humà al centre de la tecnologia en el context de la fàbrica del futur) traslladant el món físic de la manufactura al món digital mitjançant l'ús de sensors i actuadors. Al mateix temps els sistemes ciberfísics escurcen la distància entre la seguretat física i la ciberseguretat, amb el creuament dels riscos i dels impactes dels mons físic i cibernètic sobre les tecnologies operatives. Els sistemes ciberfísics, doncs, hauran d'estar preparats perquè un incident informàtic no impacti el món físic, o a l'inrevés.

- Computació quàntica: a mesura que la Computació Quàntica avanci durant els propers anys, augmentarà el risc que alguns mètodes de xifratge utilitzats per protegir les dades en repòs i en trànsit quedin obsolets. És per això que les empreses han de començar a establir plans de migració cap a algorismes de xifrat resistents a la computació quàntica.

- Bessons digitals (*digital twins*): esdevindran eines clau per analitzar els riscos i els impactes d'un incident digital sobre entorns físics complexos que incloguin sistemes ciberfísics, IoT, persones, cadenes de subministrament, processos, etc. Els bessons digitals permetran simular entorns físics i entrenar, en temps real, la capacitat de reacció d'una organització davant d'un ciberatac, la qual cosa no seria possible en entorns reals.

Breu introducció a la Intel·ligència Artificial

3

Índex

3.1. Definició i principals conceptes de la IA[1]

La IA és una disciplina tecnològica que s'enquadra en l'àmbit de les ciències de la computació i comparteix tècniques i dominis de coneixement amb altres disciplines com les matemàtiques, l'estadística o la neurociència. A causa de la creixent complexitat i la seva transversalitat, la IA esdevé cada vegada més interdisciplinària, amb sinergies amb la biologia, la filosofia, el món del dret i l'ètica, la psicologia, la sociologia i l'economia.

J. McCarthy, professor a Standford i reconegut com el pare de la Intel·ligència Artificial, el 1956 va definir la IA com "la ciència i enginyeria de fer màquines intel·ligents" i resulta dels treballs de l'escola d'estiu que ell mateix convocà al Dartmouth College per estudiar la conjectura que "tot aspecte de l'aprenentatge o qualsevol altra atribut de la intel·ligència es pot, en principi, descriure d'una forma tan precisa que una màquina ho podria simular[2]".

Posteriorment han sorgit altres definicions. Investigadors com S. Russel and P. Norvig defineixen la IA com "el disseny i construcció d'agents intel·ligents que reben percepcions de l'entorn i realitzen accions que afecten aquest entorn[3]". Altres com A. Kaplan i M. Haenlein proposen una definició alternativa com "la capacitat que té un sistema per interpretar dades externes correctament, aprendre d'aquestes dades, i fer servir els coneixements adquirits per completar tasques i assolir objectius específics mitjançant una adaptació flexible[4]". Finalment, La Comissió Europea, per la seva part, defineix la IA referint-se a "sistemes que mostren un comportament intel·ligent i amb capacitat de

realització d'accions –amb un cert grau d'autonomia– per assolir objectius específics[5]".

Malgrat aquestes, o altres definicions possibles, la naturalesa de la IA continua essent objecte de debat dins la comunitat científica. Una manera d'enfocar aquesta discussió és definint la IA en funció dels seus objectius, i) si l'objectiu és replicar el comportament o la biologia del cervell humà o ii) si l'objectiu és una racionalitat ideal i abstracta. Es tracta de considerar que l'objectiu de la IA consisteix en construir sistemes capaços de raonar o actuar. Per tant, aquesta aproximació en la definició de la IA es pot resumir com "la disciplina que cerca desenvolupar o construir sistemes que raonin i es comportin com els humans o pensin i actuïn racionalment[6]".

La Intel·ligència Artificial ha esdevingut una tecnologia d'ús estès degut a la confluència de diversos elements que ho han fet possible:

1 Llibre Blanc sobre la Intel·ligència Artificial aplicada a les indústries culturals i basades en l'experiència a Catalunya. CIDAI, Center of Innovation in Data technologies and Artificial Intelligence

2 McCarthy, J., Minsky, M., Rochester, N., Shannon, C.E., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1955

3 Stuart Russell and Peter Norvig. Artificial intelligence : a modern approach ISBN 0-13-103805-2 1, 1995

4 Kaplan, A., Haelnelin, M., Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons, Volume 62, Issue 1, Jan-Feb 2019, pg 15-25

5 A Definition of AI: Main Capabilities and Scientific Disciplines. High-Level Expert Group on Artificial Intelligence. Eureopan Commission, 2018.

6 https://plato.stanford.edu/entries/artificial-intelligence/

- La disponibilitat de dades massives: provinents de la digitalització i de la sensorització, amb un creixement exponencial en la generació de noves dades, elements fonamentals per la construcció i millora de sistemes basats en IA.

- Augment de la potència de càlcul i processament massiu: l'increment significatiu de la potència de càlcul dels processadors, i particularment l'adopció de les tecnologies GPU (*Graphical Processing Unit*) és el fonament de l'aprenentatge profund (*Deep Learning, DL*), una branca de la IA que està propiciant salts tecnològics disruptius en diversos camps d'aplicació.

- La millora en els algorismes d'IA: evolució de programes d'anàlisi de dades basats en regles, estadístiques, regressions i sistemes experts al desenvolupament de diverses noves famílies algorítmiques basades en xarxes neuronals profundes amb un gran potencial d'aplicació.

- La reducció del cost de la gestió de sistemes informàtics (processament i emmagatzematge) al núvol i la proliferació de serveis relacionats amb dades i IA: a mode d'exemple, el cost d'emmagatzematge per gigabyte ha passat de 500 mil dòlars el 1980 a menys de 2 cèntims el 2020.

L'àmbit d'actuació de la IA pot ser només del món virtual o pot tenir també incidència en el món físic. Si els sistemes de computació d'IA estan basats en programari la IA només serà virtual, com per exemple: assistents de veu, programari d'anàlisi d'imatges, sistemes de reconeixement de veu i discurs, cercadors, etc. Però els sistemes d'IA també poden trobar-se integrats dins d'algun tipus de dispositiu físic, dotant-lo de cert nivell d'intel·ligència i capacitat d'interacció amb l'entorn com és el cas de robots amb capacitats cognitives o drons i vehicles autònoms, entre altres.

En aquest punt és important diferenciar entre automatització i IA, dos conceptes que es poden confondre fàcilment. En general totes dues permeten realitzar de forma automàtica tasques sense la intervenció humana. La diferència rau en què un sistema automàtic utilitza un programari que segueix unes regles i uns passos preprogramats, mentre que un sistema basat en IA és capaç de prendre decisions o realitzar accions per a les quals no ha estat programat específicament degut a la seva capacitat d'aprenentatge a partir de dades o patrons previs i el corresponent entrenament. En altres paraules, un sistema automatitzat fa exactament allò per al que ha estat programat prèviament i res més, mentre què a un sistema d'IA no se li diu exactament i amb anterioritat què ha de fer, sinó que se li proporcionen les eines necessàries perquè, a partir d'un entrenament inicial, pugui prendre per si mateix la millor decisió en cada moment depenent de la situació en què es trobi.

3.2. Principals tecnologies d'IA[7]

A continuació es detallen algunes tecnologies com l'aprenentatge automàtic o *Machine Learning* i l'aprenentatge profund o *Deep learning* que s'engloben en el concepte genèric d'IA i, a la secció 3.3, s'avalua la seva aplicació en el sector de la ciberseguretat.

3.2.1. L'aprenentatge automàtic (Machine Learning, ML)

És la tècnica informàtica que permet que els sistemes d'IA aprenguin de forma automàtica i millorin a partir de l'experiència sense necessitat d'estar explícitament programats. El ML es classifica segons la tipologia d'algorismes en supervisats i no supervisats.

Els algoritmes supervisats es basen en deduir una funció capaç de predir el resultat corresponent a qualsevol valor d'entrada després d'haver-se exercitat amb una sèrie d'exemples d'entrenament (dades etiquetades). Els problemes característics que resolen els algorismes supervisats són els de regressió i classificació, i els mètodes per resoldre'ls són, per exemple, els de regressió lineal o xarxes neuronals pels de regressió, i mètodes com arbres de decisió o *Support Vector Machines* (SVM) pels de classificació. Un tipus d'aprenentatge supervisat especial són els mètodes d'aprenentatge de reforç (*reinforcement learning*), on el propi model aprèn a l'examinar els resultats de cadascuna de les seves decisions. Quan el resultat és positiu es recompensa i quan el resultat és negatiu no s'obté recompensa o aquesta és negativa. D'aquesta manera el model s'adapta el seu comportament d'acord a les recompenses obtingudes i acumulades. És tracta de l'aprenentatge més semblant al realitzat per un ésser humà.

En els algoritmes no supervisats el procés d'entrenament es basa en un conjunt de dades sense etiquetes o classes prèviament definides. El model ha de descobrir l'estructura subjacent que serveixi per descobrir els patrons que descriuen l'estructura de les dades. Els algorismes de *clustering* pertanyen a aquesta família algorítmica, així com els mètodes d'inducció de regles d'associació[8].

3.2.2. L'aprenentatge profund (Deep Learning, DL)

Es fonamenta en l'ús de xarxes neuronals artificials amb moltes capes ocultes que implementen processos progressius d'abstracció sobre les dades, aprenent atributs més i més generals a cada nivell. Els sistemes IA basats en DL poden no només aprendre conceptes, sinó també reconèixer contextos i entorns força complexos, com el reconeixement d'imatges i de la parla, entre altres.

La Figura 4 mostra les principals aplicacions del ML, reconeixement de patrons i automatització de processos robotitzats, i del DL, on s'hi integren les branques de processament de llenguatge natural, xarxes neuronals artificials i biometria.

Figura 4: Categories i aplicacions del ML (font: CIDAI, Llibre Blanc sobre la IA aplicada a les ICBE a Catalunya)

7 Llibre Blanc sobre la Intel·ligència Artificial aplicada a les indústries culturals i basades en l'experiència a Catalunya. CIDAI, Center of Innovation in Data technologies and Artificial Intelligence

8 Which method to use? An assessment of data mining methods in Environmental Data Science https://www.sciencedirect.com/science/article/abs/pii/S1364815218308715

3.3. Principals aplicacions de la IA per a la ciberseguretat

Des de la definició feta per J. McCarthy al 1955, les aplicacions d'IA aplicades a la ciberseguretat han anat evolucionant fins a l'actualitat. En aquesta secció es revisen les principals aplicacions des de dues perspectives: les tècniques defensives, és dir com es pot fer servir la IA per reforçar la ciberseguretat davant de les amenaces cibernètiques, i, de forma antagònica, les tècniques ofensives, doncs els actors maliciosos també utilitzen les capacitats de la IA per llençar atacs sofisticats.

Les tècniques presentades a continuació poden ser alineades en la *Cyber Kill Chain*, un model de set passos que descriu les fases d'un ciberatac dirigit. En aquesta cadena, la IA es una eina de valor tant des del punt de vista de l'atacant com pels serveis de resposta contra incidents de ciberseguretat (veure Figura 5).

Figura 5. Relació entre les etapes de la Cyber Kill Chain i els principals casos d'ús de la IA introduïts en aquest apart

	Reconeixement	Armamentització	Lliurament	Explotació	Instal·lació	Comandament i control
BAULA	**(Reconaissance)** Conjunt de tècniques que fan referència a conèixer informació sobre l'objectiu, com per exemple quins sistemes d'informació fa servir o bé informació extreta de xarxes socials.	**(Weaponization)** Després de conèixer les febleses de la víctima, l'atacant prepara els recursos per a l'atac, com per exemple els missatges de phishing per a una suplantació d'identitat.	**(Delivery)** En aquesta etapa es distribueix l'atac i es busca passar inadvertits per les capes de defensa dels grups de resposta d'incidents.	**(Exploitation)** En cas que l'etapa anterior hagi estat efectiva, en aquest punt es busca explotar una vulnerabilitat.	**(Installation)** L'atacant busca instal·lar programari maliciós amb el fi de poder operar de forma remota en la infrastructura.	**(Command and Control C&C and Actions)** L'atacant utilitza la màquina de la víctima pel seu interès personal, per exemple, robant informació o utilitzant-la com un bot per realitzar activitat maliciosa.
DEFENSIVA	• Detecció d'*exploit* de dia zero • Control biomètric • Identificació de punts febles	• Anàlisi d'amenaces per reconeixement de patrons de comportament d'usuaris • Anàlisi d'amenaces per detecció d'anomalies de xarxa • Detecció de *phishing / spam* • Automatització de les operacions de ciberseguretat				• Detecció avançada de programari maliciós
OFENSIVA	• Identificació del comportament humà • *Deep fakes* • Creació identitats falses o sintètiques	• Ciberatacs automatitzats • Evolució dels ciberatacs en funció del comportament de la xarxa • Atacs de força bruta amb dades personals robades				• Swarm *malware* (eixam de botnets) • Programari maliciós metamòrfic / polimòrfic

3.3.1. Tècniques defensives

3.3.1.1. Anàlisi d'amenaces per reconeixement de patrons de comportament dels usuaris

La premissa operativa bàsica de l'anàlisi del comportament dels usuaris és crear una imatge de l'activitat típica dels usuaris i com interactuen amb els recursos de l'organització mitjançant registres i altres fonts de dades. Les dades recollides podrien ser quan un usuari comença a treballar, com navega per llocs web, com mou el ratolí, l'accés als comptes, la seva geolocalització, l'ús d'aplicacions, l'activitat amb fitxers, la durada de la sessió, l'ús del xat i de missatgeria instantània i moltes més. L'objectiu és supervisar i analitzar totes aquestes activitats i ingerir i correlacionar les dades massives per avaluar si hi ha risc cibernètic gairebé en temps real.

Un dels punts crítics és reduir el nombre de falsos positius o falsos negatius, és a dir, garantir la precisió a l'hora de determinar si una activitat inesperada és realment malèvola. De fet, les solucions convencionals basades en anomalies sovint inunden amb falsos positius. En canvi, amb l'aplicació de la IA és possible analitzar el risc d'activitat amb més precisió: quan es produeixen nous comportaments d'usuari, s'analitzen mitjançant models d'aprenentatge automàtic per comprovar si s'ajusten al perfil d'aquest usuari i si es pot considera vàlid. Si hi ha una variació considerable, el sistema genera una alarma. Actualment al mercat hi ha moltes eines que proporcionen aquest tipus d'anàlisi, com IBM QRadar i Rapid7 InsightIDR. En un futur proper, s'espera que aquestes eines no només s'utilitzin per reduir el temps d'anàlisi de seguretat i identificar possibles anomalies (activador d'alarma), sinó també per oferir accions defensives o de mitigació automàtiques de manera que l'amenaça es pugui neutralitzar gairebé immediatament.

3.3.1.2. Anàlisi d'amenaces per detecció d'anomalies de xarxa

Tant els antivirus com els sistemes de detecció d'intrusos usen mètodes per detectar possibles amenaces basats en signatures. Aquestes signatures serveixen per identificar patrons coneguts d'activitat maliciosa, ja sigui amb un fragment d'un codi maliciós, una llista blanca d'activitats legítimes conegudes o una llista negra d'activitats malicioses conegudes. El problema d'aquest mètode és que cal conèixer la signatura per detectar l'amenaça i, per tant, un nou atac passaria inadvertit ja que encara no es coneix la seva firma.

La IA ajuda a desenvolupar solucions basades en anomalies, que permeten la identificació d'amenaces emergents desconegudes a partir del coneixement adquirit prèviament d'amenaces anteriors i activitats històriques. Els valors atípics o comportaments anormals s'identifiquen com a amenaces potencials. Un dels principals avantatges d'aquesta metodologia és la capacitat de detectar els atacs a vulnerabilitats no conegudes pel sistema (*zero-days*) i dota el tècnic de capacitats preventives. Exemples d'aquestes eines són Dynatrace per a les xarxes internes d'empreses i Nokia NetGuard Cybersecurity Dome per a grans infraestructures, des de la xarxa d'accés de ràdio fins a les xarxes de transport.

3.3.1.3. Detecció avançada de programari maliciós

Avui, els adversaris generen milions de mostres noves de programari maliciós. Però, com que tot requereixen programació manual, es tracta d'un nombre finit i manejable de mostres, amb mètodes de detecció per mitjà de signatures. Amb la recent (i mediàtica) aparició del ChatGPT (un model d'IA conversacional que pot interactuar de forma similar a com ho faria un ésser humà i, fins i tot, és capaç generar codi font), un ciberatacant té al seu abast executar amb ChatGPT una comanda del tipus "Vull fer aquest atac i contra aquesta tecnologia objectiu" i automàticament una IA generarà diferents iteracions d'una peça de programari maliciós. Clarament, el ChatGPT i d'altres eines d'IA aproximen la ciberdelinqüència als script kiddies i a cibercriminals sense habilitats tècniques i, per tant, duran la producció de malware a nivells molt més elevats.

Fina ara, els mètodes de detecció més avançats es basen en l'anàlisi heurística. Aquesta tècnica determina la susceptibilitat d'un sistema davant d'una amenaça/risc particular mitjançant diverses regles de decisió o mètodes de ponderació. La idea és aïllar el contingut sospitós del sistema operatiu principal però alhora poder analitzar-ne el comportament. Per aquest motiu, el contingut sospitós s'executa en una màquina virtual especialitzada normalment anomenada sandbox. Això fa possible analitzar les ordres a mesura que s'executen, supervisant activitats ma-

licioses habituals com ara la replicació, la sobreescriptura de fitxers i els intents d'ocultar l'existència del fitxer sospitós. Si es detecten una o més accions semblants a un programari maliciós, el fitxer sospitós es marca com a virus potencial i s'avisa l'usuari. Un altre mètode comú d'anàlisi heurística consisteix en descompilar el programa sospitós i, a continuació, analitzar el codi màquina que hi ha. Aquest es compara amb el d'altres malwares coneguts i si un determinat percentatge coincideix llavors s'activa una alerta. En aquest sentit, la IA proporciona diferents avantatges. D'una banda, permet simplificar i analitzar grans volums d'activitat dels programaris per tenir coneixement de quin tipus d'atacs s'estan produint a cada moment o si sorgeixen noves amenaces desconegudes. D'altra banda, permet ajustar el nivell de seguretat que cal aplicar per a l'execució de cada procés, és a dir, millora les prestacions ja que aplicarà una anàlisi exhaustiva només quan sigui necessària. Per exemple, Kaspersky ja està implementant mètodes d'aprenentatge automàtic i aprenentatge profund per a la detecció de programari maliciós als seus productes.

Detectar el programari maliciós, doncs, requereix la IA per maximitzar la capacitat d'enginyeria inversa, i permet explicar als analistes amb llenguatge natural què creu que fa un determinat programari. La IA és capaç d'analitzar centenars de milers de mostres binàries i superar la presència de bucles enrevessats dissenyats per dificultar l'enginyeria inversa. Amb l'aparició del ChatGPT, cal tenir en compte que ja ha demostrat que és capaç de fer aquesta enginyeria inversa de manera molt més ràpida que un operador o analista humà.

3.3.1.4. Automatització de les operacions de seguretat

Una resposta ràpida i específica a un atac és crucial per protegir al màxim tots els components de la infraestructura. Tenint en compte que avui en dia els atacs utilitzen la IA per crear atacs massius i/o sofisticats i l'organització està ampliant la seva infraestructura digital (computació en núvol, empleats remots, models de negoci oberts, etc.), els equips de seguretat simplement no tenen la capacitat de supervisar, gestionar i actuar manualment sobre tots ells de manera ràpida i eficaç.

La IA i l'automatització milloren significativament la productivitat de la força de treball de ciberseguretat. Per exemple, pot automatitzar la recollida, integració i anàlisi de dades de centenars i fins i tot de milers de punts de control, sintetitzant registres del sistema, fluxos de xarxa i comportaments dels usuaris. Juntament amb la gestió d'amenaces i la priorització d'alertes, els equips d'operacions de seguretat poden entendre completament el context de les excepcions de seguretat, establir prioritats i dedicar recursos suficients a investigar les amenaces d'alt impacte.

3.3.1.5. Millora del control biomètric

La biometria estableix automàticament la identitat d'una persona mitjançant mesures fisiològiques com la veu, la cara, l'iris i les empremtes dactilars, i també característiques de comportament com la marxa, la signatura i la forma de teclejar. La biometria augmenta la seguretat quan s'utilitza en combinació amb altres tècniques d'identificació i simplifica l'accés a molts serveis eliminant la necessitat de recordar contrasenyes o portar coses, com ara targetes o claus.

Les tecnologies biomètriques han evolucionat enormement en els darrers anys, en quant a fiabilitat, cost i comoditat d'ús. Les seves aplicacions són molt diverses. Es pot utilitzar per controlar l'accés físic a edificis i altres instal·lacions, per controlar l'accés virtual a ordinadors, telèfons mòbils i comptes bancaris, i per personalitzar continguts audiovisuals. També, té aplicacions en videovigilància, per identificar persones a partir de fonts de vídeo o àudio, i investigacions forenses, per recollir empremtes dactilars o mostres d'ADN de les escenes del crim.

Les tècniques tradicionals de classificació han resultat exitoses per biometries fisiològiques amb poca variabilitat, com ara les empremtes dactilars o l'iris, però tenen problemes en casos de més variabilitat, com ara, l'afonia o el canvi d'idioma o d'estat d'ànim en el cas del reconeixement de veu, o l'envelliment, les mascaretes o el maquillatge en reconeixement facial. Actualment, les noves tecnologies d'IA, en particular l'aprenentatge profund amb grans xarxes neuronals, estan millorant de manera evident l'eficiència i la seguretat de la biometria en tots el casos, permeten inclús l'ús de comportaments d'alta variabilitat relacionats amb els nous dispositius, com ara l'activitat del ratolí, la manera de teclejar o de tocar les pantalles tàctils o els moviments del mòbil o la tauleta digital.

3.3.1.6. Anàlisi de riscos

Un aspecte important per a una organització és conèixer el seu propi sistema i tenir a disposició informes sobre els riscos als quals pot estar subjecta. En aquest sentit, la IA també pot ajudar. Com a primer pas, permet processar automàticament i tenir un inventari complet i actualitzat dels actius informàtics d'una empresa, així com classificar-los segons la seva criticitat. Aleshores, la IA pot analitzar grans quantitats de dades rellevants per obtenir informació sobre les amenaces emergents de ciberseguretat i identificar les vulnerabilitats que els atacants poden explotar. Així, la IA pot identificar els punts forts i febles d'una organització i de les solucions de ciberseguretat aplicades.

El resultat final serà una avaluació de la seguretat completa i gairebé en temps real de tots els paràmetres significatius, inclosa la pila tecnològica, els senyals de risc capturats a escala d'Internet, la topologia, el nivell d'amenaça, les prioritats empresarials, les obligacions reguladores, les observacions de sèries temporals dels actius orientats a Internet, visions històriques, etc. Aquesta avaluació es pot representar per un sistema de puntuació fàcil d'entendre que proporcioni un indicador prospectiu del risc de seguretat. OneTrust, per exemple, proporciona eines per identificar riscos i prioritzar les iniciatives de mitigació a tota l'organització.

3.3.1.7. Detecció de *phishing*/spam

La detecció de correu no desitjat és una de les aplicacions més prolífiques i conegudes en el domini de la ciberseguretat. En aquest sentit, és ben coneguda l'amenaça que suposen els correus de *phishing*, els quals cada vegada fan ús de paranys més elaborats i personalitzats, de manera que esdevé imprescindible dotar els usuaris d'eines de suport eficients que minimitzin el volum ingent de correus maliciosos que s'envia constantment.

La IA proveeix procediments automàtics que permeten als serveis de correu electrònic bloquejar missatges no desitjats tals com casos de correu brossa o el *phishing*. Les eines *anti-phishing* utilitzen tècniques de processament del llenguatge natural (NLP) per processar el contingut del correu electrònic, per exemple mirant paraules clau (còmput de les matrius TF-IDS) per la posterior classificació fent servir algoritmes supervisats tals com màquines de suport vectorial (SVM) i arbres de decisió, entre altres.

3.3.1.8. Compliment normatiu

La normativa en matèria de ciberseguretat està en contínua evolució i engloba diferents marcs reguladors que canvien segons la jurisdicció, un fet que afegeix més pressió a uns equips de ciberseguretat ja sobrecarregats. Aquesta càrrega creixent requereix l'adopció de noves eines tecnològiques que aju-

din a optimitzar el procés de compliment normatiu. La capacitat de la IA per analitzar volums massius de dades amb rapidesa i precisió té el potencial d'ajudar les entitats a comprendre les seves responsabilitats de compliment i prendre les mesures adequades en un menor període de temps. De fet, els sistemes d'IA poden efectuar auditories de ciberseguretat en què es validi el conjunt de normatives vigents i, fins i tot, s'implementin automàticament els controls necessaris per assegurar el seu compliment. Addicionalment, amb relació a la regulació de dades personals, les eines d'IA permeten tenir els sistemes d'informació sota control i, al mateix temps, donar una resposta automatitzada i veloç a les peticions dels usuaris sobre l'execució dels seus drets respecte les seves dades.

En l'actualitat ja es disposa de solucions d'IA per al compliment normatiu. De fet, diverses eines antivirus i de protecció dels sistemes d'informació ja inclouen formes d'aprenentatge automàtic, com ara serveis antivirus basats en núvol i sistemes de detecció d'intrusions, encaminats a resoldre nombrosos aspectes normatius.

3.3.2. Tècniques Ofensives

3.3.2.1. Ciberatacs automatitzats (sense intervenció humana)

El procés d'automatització proporciona als ciberatacants la possibilitat d'organitzar múltiples ciberatacs simultanis en paral·lel i corregir les tècniques d'atac a temps real sense la necessitat d'intervenir-hi directament. D'aquesta forma, els ciberatacs esdevenen escalables, requereixen menys esforç, augmenten la flexibilitat i serveixen per entrenar un model per aconseguir atacs cada vegada més eficients.

Un exemple el poden trobar en els atacs DDoS intel·ligents (*smart* DDoS), els quals coordinen milers de bots amb la capacitat de variar la tipologia dels paquets, elaborar peticions, emetre fluxos de dades en ràfegues i canviar les IP objectiu. Un altre àmbit d'aplicació interessant és la realització de campanyes de *phishing*, *smishing* o missatges fraudulents a xarxes socials. Precisament, una prova de concepte amb el ChatGPT ha demostrat que és capaç d'escriure correus electrònics maliciosos, convincents i gramaticalment correctes i a una gran velocitat[9]. La IA permet difondre aquestes campanyes automàticament, massivament i, si algú cau en l'engany i respon, es fa càrrec de la conversació, tal i com si fos un *chatbot*. Addicionalment, si la IA incorpora dades personals de les víctimes potencials, la IA podrà realitzar atacs dirigits i adaptats a cada perfil per tal d'adaptar l'esquer i maximitzar les probabilitats d'èxit.

3.3.2.2. Imitació del comportament humà

Una de les característiques que aporta més valor a un atac i que incrementa les seves possibilitats d'èxit és la imitació del comportament humà. Quan un ciberatacant aconsegueix tenir el seu *malware* desplegat a l'equip de la víctima, aquest programari no actua immediatament, sinó que dedica uns dies, setmanes o el que sigui necessari per poder aprendre el seu comportament habitual, per després poder emmascarar les seves accions i que no puguin ser detectables pels sistemes de detecció de defensa.

Un exemple d'imitació del comportament humà es pot donar quan es vol comprometre un sistema d'autenticació basat en el comportament d'un usuari. Aquest tipus d'autenticació, basat en l'ús i interacció amb les aplicacions o com es navega per la web, facilita la usabilitat d'accés als sistemes. En aquest cas, un programari maliciós al dispositiu de la víctima escoltarà i recopilarà dades del comportament de l'usuari. Gràcies a aquestes dades, l'atacant podrà construir un model de comportament de l'usuari que li permetrà suplantar-lo sigil·losament.
Es poden trobar més exemples de la imitació del comportament humà en les xarxes socials, on bots intel·ligents imiten el comportament d'una persona per interactuar amb altres usuaris, crear continguts, generar opinions, amb l'objectiu de manipular els usuaris de la xarxa, difondre fraus, robar identitats, generar desinformació, etc.

3.3.2.3. Evolució dels ciberatacs en funció del comportament de la xarxa

Com ja s'ha comentat, alguns sistemes de defensa integren algoritmes i tècniques d'IA per fer front als nous atacs a partir de l'ús de segments de xarxa per al seu anàlisi (*honeypots*). Per tal de fer front a aquests escenaris, els ciberatacants han de desenvolupat programaris intel·ligents, capaços de prendre

9 https://www.youtube.com/watch?v=_MVczQBk9yo

decisions de forma autònoma. Així, quan l'adversari aconsegueix introduir el seu programari maliciós en a la xarxa de la víctima, el programa inicia un procés d'anàlisi de comportament de la màquina on s'ha instal·lat. Posteriorment, el *malware* inicia les seves accions tenint en compte el comportament capturat, prenen les decisions adients per no ser detectat.

En aquest sentit, ja és habitual que els programaris maliciosos no despleguin les seves capacitats si detecten que estan executant-se en un entorn de simulació (*sandbox*) on es busca estudiar-lo. Aquesta capacitat de presa de decisions de forma autònoma elimina la necessitat de comunicació del *malware* amb l'exterior per rebre instruccions i, per tant, generant menys esdeveniments que podrien ser detectats.

3.3.2.4. Swarm *malware* (eixam de *botnets*)

Una altra capacitat de la IA explotada pels ciberatacants és la presa de decisions de forma col·laborativa i coordinada entre diferents elements cibernètics, proporcionant cert grau d'autonomia. Històricament, l'activitat del *malware* s'ha basat en la utilització de la comunicació C2 (*Command & Control*), com els habituals *malwares* de tipus *botnet*. El seu desavantatge prové precisament de la seva estructura de control centralitzada, ja que una vegada que el centre de C2 es bloquejat, el programari maliciós queda inactiu. Un pas endavant i de millora d'aquests tipus d'atac és la utilització d'una solució basada en el marc dels algoritmes amb intel·ligència d'eixam (*swarm intelligence*), similar al comportament dels sistemes d'eixam biològics, sense un punt de comunicació central, i que es basa en la comunicació i coordinació entre els diferents membres o bots.

La comunicació entre els membres de l'eixam pot realitzar-se de diferents maneres. Cada individu pot tenir una memòria col·lectiva, compartir coneixements i «aprendre d'ells», establint un marc de cooperació entre els diferents membres de la *botnet*, sense cap unitat centralitzada C2. Aquest tipus d'intel·ligència aporta capacitat d'auto-organització i auto-emergència i, per tant, és l'evolució natural dels *malwares* de C2. Probablement, els atacs de *botnets* amb intel·ligència basada en eixam seran la pròxima gran amenaça i repte de seguretat.

3.3.2.5. Programari maliciós metamòrfic / polimòrfic

Un dels objectius dels ciberatacants és aconseguir programaris maliciosos que tinguin la capacitat d'evadir els mètodes de detecció, per exemple, amb *malwares* capaços d'introduir mutacions al codi, renovar-se i esdevenir indetectable. Aquest procés de mutació té com objectiu modificar l'estructura del codi sense alterar necessàriament la seva funcionalitat, canviant de forma o reescrivint-se cada vegada que s'executa. Els algoritmes evolutius poden ser utilitzats com a mètodes d'aprenentatge adversarial, codificant el *malware* en una seqüència de ADN i fent-lo evolucionar fins que els classificadors d'aprenentatge de *malware* no siguin capaços de detectar-lo. Ara mateix, són tècniques amb molt potencial però encara estan lluny de ser utilitzades com un mitjà pràctic pels adversaris. Tanmateix, ja es va detectar el programari maliciós Zellome9, que utilitzava algoritmes genètics com mecanisme de força bruta per generar un encriptador polimòrfic, que no va aportar gran valor però sí va reclamar l'atenció sobre els algoritmes evolutius. Fent una analogia a la teoria darwiniana de l'evolució, es pot estimar que aquest tipus de *malware* continuarà adaptant-se als nous entorns i seguirà evolucionant fins convertir-se en una amenaça molt real pels actuals mètodes de detecció, principalment pel gran nombre de funcions que es podrien aplicar i que no poden ser filtrades. L'evolució del *malware* constaria de diversos passos: (1) se selecciona el *malware* entre diferents variants; (2) s'aplica l'algorisme evolutiu per simular la seva evolució, com el encreuament de gens (canvi de característiques de les mostres) i les mutacions (l'elecció de característiques) per produir una nova generació; (2) el codi maliciós generat es prova contra productes anti-*malware*. L'amenaça és real i alguns investigadors destaquen que el ChatGPT evidencia la capacitat de generar codi polimòrfic que canvia de xifratge criptogràfic per esquivar els mecanismes de seguretat tradicionals dissenyats detectar signatures concretes. Val a dir que el ChatGPT disposa de filtres per impedir el seu ús per crear programari maliciós, però l'ús de diferents subterfugis imaginatius permet aconseguir codis complexos per evitar les defenses[10].

Ara mateix, són tècniques amb molt potencial però encara estan lluny de ser utilitzades com un mitjà pràctic pels adversaris. Tanmateix, ja es va detectar

10 https://gizmodo.com/chatgpt-ai-polymorphic-malware-computer-virus-cyber-1850012195

el programari maliciós Zellome[11], que utilitzava algoritmes genètics com mecanisme de força bruta per generar un encriptador polimòrfic, que no va aportar gran valor però sí va reclamar l'atenció sobre els algoritmes evolutius. Fent una analogia a la teoria darwiniana de l'evolució, es pot estimar que aquest tipus de *malware* continuarà adaptant-se als nous entorns i seguirà evolucionant fins convertir-se en una amenaça molt real pels actuals mètodes de detecció, principalment pel gran nombre de funcions que es podrien aplicar i que no poden ser filtrades. L'evolució del *malware* constaria de diversos passos: (1) se selecciona el *malware* entre diferents variants; (2) s'aplica l'algorisme evolutiu per simular la seva evolució, com el encreuament de gens (canvi de característiques de les mostres) i les mutacions (l'elecció de característiques) per produir una nova generació; (2) el codi maliciós generat es prova contra productes anti-*malware*.

11 http://pferrie.epizy.com/papers/zellome.pdf?i=1

3.3.2.6. *Deep fakes*

Els *deep fakes* corresponen a una tecnologia en desenvolupament que utilitza l'aprenentatge profund per a la creació de muntatges audiovisuals, com imatges, vídeos, textos o missatges de veu, molt realistes que contenen situacions fictícies sobre persones reals. Aquesta tecnologia és de gran potencial pels ciberatacants, perquè obre un gran nombre de possibilitats per suplantar algú en atacs d'enginyeria social, superar sistemes de control d'accés biomètric, desinformar la ciutadania o difamar.

Hi ha dos mètodes principals per generar *deep fakes*. El primer es basa en situar la cara d'una persona sobre la cara d'una altra, essent necessàries milers de fotos de cares de les dues persones per un codificador basat en algoritmes IA. Un segon mètode per generar *deep fakes* és amb l'aplicació de *Xarxes Generatives Adversarials* (GAN), on s'enfronten dos algoritmes d'IA per la creació de noves imatges. Amb relació a la veu, existeixen diverses aproximacions, des de la clonació de veu a partir d'un nombre reduït de mostres de la persona objectiu (prèviament s'ha construït un model general amb un nombre elevat de mostres d'un conjunt de

persones), i que pot ser clonada a partir d'un text, fins la clonació a partir d'un nombre determinat de mostres de la víctima (si es tracta d'una persona pública, Internet pot estar ple de continguts a utilitzar). Un exemple d'èxit d'aquest tipus d'atacs és un cas de frau del 2019, on es va utilitzar un software basat en IA per a suplantar la veu d'un executiu i aconseguir desviar un pagament de 220k€.

3.3.2.7. Atacs de força bruta amb dades personals robades

Un exemple rellevant que il·lustra l'evolució de les ciberamenaces gràcies a l'ús d'IA són els atacs de nova generació per comprometre les contrasenyes, més eficients i intel·ligents que mai.

Les tècniques tradicionals per endevinar contrasenyes es basen en força bruta, un diccionari o en regles, on solament es poden capturar subconjunts de l'espai de contrasenyes que coincideixen únicament amb les regles disponibles generades per les persones i basades en la intuïció de com els usuaris selecciones les seves. Aquest procés es pot millorar amb la introducció de la IA.

Per exemple, un model d'IA pot generar un diccionari de millor qualitat. Per aconseguir-ho, es realitza un entrenament d'un model de xarxa neuronal recurrent (RNN) que aprèn a construir noves potencials contrasenyes mitjançant l'autoaprenentatge i la construcció del diccionari d'atac, basat en patrons derivats de contrasenyes anteriors. Si a aquest aprenentatge s'hi incorporen noves dades personals, ja sigui obtingudes de forma lícita o il·lícita, el procés permet incrementar la probabilitat d'endevinar la contrasenya. Anàlogament, es pot entrenar una GAN per aprendre la distribució de contrasenyes a partir de les dades recollides per filtració de contrasenyes reals. Com a resultat es pot obtenir un diccionari de contrasenyes que utilitza regles de contrasenyes generades per aprenentatge per determinar de forma autònoma les propietats de les mateixes.

3.3.2.8. Creació d'identitats falses o sintètiques

Aquest tipus de frau es defineix com l'ús d'una combinació d'informació personal d'un usuari per construir una nova persona o entitat amb la finalitat de realitzar una acció il·lícita. La combinació de les dades robades o comprades en el mercat negre, més l'ús d'algoritmes i tècniques d'IA per a la generació de dades versemblants d'una persona, o fins i tot la seva imatge, la veu o l'estil d'escriure, tenen gran potencial pels adversaris, ja que els permet la creació d'identitats sintètiques (identitats falses i no una suplantació d'una identitat) per a perpetrar diferents tipus de fraus, com la creació de comptes bancaris, sol·licitud de préstecs, obtenció de documents oficials, obrir cripotomoneders, subscriure's a serveis, etc.

3.4. Riscos de seguretat de la IA i limitacions

3.4.1. Ciberatacs contra la IA

Un dels riscos de la seguretat de la IA és la seva debilitat davant d'un ciberatac, ja que introdueix algunes febleses inherents a la naturalesa de la IA i que no es poden eliminar. A continuació, s'analitzen quatre casos diferents en què es pot atacar una IA.

3.4.1.1. Robatori d'algoritmes

Una manera d'atacar una defensa basada en IA és conèixer l'algoritme utilitzat. Recentment, s'ha demostrat que es pot fer enginyeria inversa a partir d'un algorisme d'IA i fins i tot reproduir-lo completament, és a dir, "robar-lo". El mètode consisteix a entrenar una IA amb la sortida de la IA objectiu enviant-li consultes i analitzant les respostes. Després d'uns quants milers o fins i tot centenars de consultes, la IA clonada és capaç de predir amb una precisió gairebé del 100 per cent les respostes de la IA original. Com que per desenvolupar i entrenar un algorisme es necessiten moltes dades i moltes hores de treball i execució, robar un algorisme ja entrenat o part d'ell proporciona un avantatge competitiu important. D'una banda s'estalvia temps i diners i de l'altra usant aquest mateix algorisme robat, es poden predir els resultats d'un adversari.

3.4.1.2. Introducció de mostres falses / Manipulació d'algoritmes

La manipulació de les dades que s'introdueixen en un sistema d'IA pot permetre alterar la seva respos-

ta sortida per a servir els interessos d'un ciberatacant. Com que, en el seu nucli, cada sistema d'IA és una màquina simple (pren una entrada, realitza alguns càlculs i retorna una sortida), la manipulació de l'entrada permet als atacants afectar la sortida del sistema. Per tant, sabent com funciona la IA, un atacant podria provocar que els sistemes d'IA cometin errors manipulant les entrades. Per exemple, els emissors de correu *spam* poden intentar evadir els sistemes de detecció ofuscant el contingut dels correus electrònics per tal de ser classificats com a legítims per la IA.

3.4.1.3. Enverinament de l'aprenentatge (Training-data poisoning)

L'enverinament de dades és un ciberatac que consisteix en manipular el conjunt de dades d'entrenament de la IA per controlar el comportament de predicció d'un model entrenat. A mesura que l'algoritme aprèn d'aquestes dades corruptes, estarà entrenat per a treure conclusions no desitjades i fins i tot perjudicials (p. ex. etiquetant un codi maliciós com a segur). El conjunt de dades enverinat es pot construir de manera que el model d'IA resultant funcioni bé en un conjunt de proves de servei, però es comporti malament amb activadors específics. L'inconvenient més important de la IA és que la seva eficàcia és gairebé directament proporcional a la qualitat de les seves dades. Per molt avançat que sigui el model, una informació de mala qualitat produirà resultats inferiors i, a més, un atacant pot amagar fàcilment un atac d'enverinament de dades d'entrenament. Els atacs d'enverinament de dades poden causar danys substancials amb un esforç mínim.

3.4.1.4. Enganys amb imatges ocultes per l'ull humà (esteganografia)

La criptografia consisteix en l'enviament d'informació xifrada per a què no resulti intel·ligible a una tercera part no autoritzada. En canvi, l'esteganografia pretén amagar que la comunicació s'està produint. Esteganografia ve del grec *steganos* (cobert o ocult) i *graphos* (escriptura). La idea és que el missatge ocult vagi incrustat en elements aparentment innocus, de manera que el missatge pugui passar completament desapercebut per a qui no en conegui l'existència, inclosa una AI.

L'esteganografia, per tant, es pot fer servir per amagar que s'està fent un atac, per exemple, dels tres atacs anteriors. Habitualment, el disparador és sovint visible i fàcilment reconeixible en cas d'una inspecció visual humana i una vegada els administradors comproven les entrades es trobaran les sospitoses. En canvi, l'esteganografia amaga el fet que hi ha una entrada que pot manipular o robar informació de forma invisible per als programes informàtics de detecció, com els antivirus. ObliqueRAT és un exemple de *malware* que utilitza tècniques d'estenografia per amagar codi, arxius, imatges i vídeos dintre d'arxius d'altres formats (en aquest cas, arxius .BMP), on els bytes executables activen la descàrrega d'un arxiu .ZIP que conté ObliqueRAT[12].

3.4.2. Limitacions de la IA

Els avantatges de la utilització de la IA en el camp de la ciberseguretat són clars i realment útils. No obstant això, l'ús generalitzat de solucions de seguretat basades en la IA està posant de manifest problemes en diversos fronts que podrien, en alguna mesura, limitar-ne el seu ús.

- Augment de la necessitat de dades. Els algoritmes d'IA requereixen dades per fer l'entrenament correcte i després detectar correctament l'atac. El nombre de dades necessàries per respondre ràpidament a un atac probablement augmentarà a causa de l'entorn més dinàmic creat per les amenaces que apareixen i canvien constantment. Aquest enfocament de la ciberseguretat té molts inconvenients, com ara la incapacitat de la IA per mantenir-se al dia amb l'expansió exponencial de les dades i la manca de dades disponibles perquè l'algoritme d'IA proporcioni resultats. No és fàcil o possible determinar el volum de dades necessari, existeixen diferents factors que influeixen, com el model IA utilitzat, la representativitat de les dades, la variabilitat dels diferents paràmetres, etc.

- Necessitat de control humà. Sens dubte, la IA pot ajudar amb el processament de dades, aprendre'n i generar informació. No obstant això, els mateixos especialistes en ciberseguretat són els

12 https://blog.underc0de.org/ciberdelincuentes-ocultan-carga-util-de-obliquerat-en-imagenes-para-evadir-la-deteccion/

veritables decisors en un camp tan delicat on els errors poden tenir greus conseqüències. Encara queda camí per recórrer fins que la IA estigui prou evolucionada per pensar i actuar com els humans abans que pugui substituir un expert, per a que el grau d'automatització i autonomia a la pressa de decisions sigui prou confiable.

- Costos. Les empreses han d'invertir molt de temps i diners en recursos com la potència de càlcul, la memòria i les dades per crear i mantenir sistemes d'IA si es pretén desenvolupar solucions propietàries, ja que s'han de realitzar fases d'entrenament periòdic i recurrent. En canvi, cada vegada més, el mercat ofereix solucions d'IA compartides i al núvol, amb un nivell de costos adequat, solament requerits a les fases d'entrenament periòdics per mantenir els sistemes i el seu rendiment.
- La manca de talent en IA i ciberseguretat. Si bé hi ha una coneguda falta de talent en ciberseguretat respecte a la demanda del mercat, aquest fet encara esdevé més rellevant si es busquen perfils de ciberseguretat amb formació o experiència en IA. Aquest fet implica que, habitualment, serà necessari utilitzar equips mixtes i coordinats que incloguin personal amb cadascuna de les capacitats, amb els habituals inconvenients de parlar un llenguatge comú, d'equilibri entre tecnologies, d'increment en recursos, etc. suposant major complexitat a l'hora de realitzar i desenvolupar projectes.

La IA com a factor de transformació de la Ciberseguretat

Com ja s'ha indicat a seccions prèvies, la irrupció de sistemes i solucions basades en IA a l'àmbit de la ciberseguretat està generant una veritable revolució. Segons un estudi realitzat al 2019 per Capgemini[1], el ritme d'adopció de solucions de ciberseguretat basades en IA està experimentant un fort increment. Les aplicacions més habituals són les de seguretat de xarxa, seguretat de les dades, i la seguretat dels punts finals, on els nivells de ciberseguretat més afectats són els de detecció, predicció o resposta.

4.1. Millora a les diferents etapes de la ciberseguretat

La necessitat d'integrar eines i solucions basades en IA per a la gestió de la seguretat és real, definint un nou model d'organització dels equips, delegant en determinants nivells i etapes la presa de decisions i accions a la IA, o l'avaluació de l'impacte del risc. En aquest sentit, la IA té el potencial d'ajudar encara més a millora la robustesa, la resposta i resiliència dels seus sistemes, tal i com s'entenen aquests conceptes:

- Robustesa: és la capacitat d'un sistema de mantenir la seva configuració inicial de forma estable davant a entrades no esperades gràcies a processos d'auto-comprovació i auto-reparació.
- Resposta: és la capacitat d'un sistema de respondre de forma autònoma a diferents atacs, identificar vulnerabilitats en altres dispositius, equips o sistemes amb la finalitat d'operar de forma estratègica.
- Resiliència: és la capacitat d'un sistema per a resistir i tolerar un atac mitjançant la detecció d'amenaces, intrusions i anomalies, i prendre les decisions més adients.

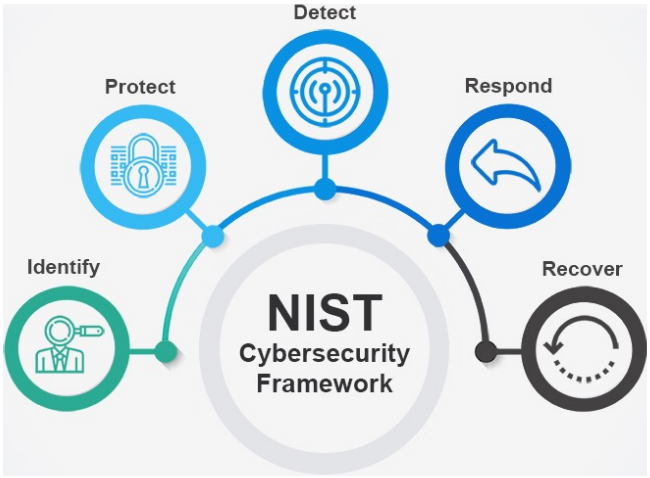
A continuació es descriu com integrar la IA en les cinc etapes definides al *NIST Cybersecurity Framework*[2],[3] (Figura 6), com pot ajudar la IA a la gestió de la seguretat, i com pot millorar la ciberseguretat davant dels nous i pròxims reptes, amenaces i atacs que aniran sorgint.

1 CAP Gemini Research Institute (2019), "Reinventing Cyber security with Artificial Intelligence. The new frontier in digital security

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

3 Inspiració i origen del ECSO Radar.

Figura 6. NIST Cybersecurity Framework



4.1.1. Identificació

A l'etapa d'identificació, la IA ajudarà en el procés de coneixement de la situació real del sistema, facilitant la presa de decisions a la resta d'etapes. Una primera acció a realitzar és la revisió del programari de les diferents aplicacions, on es pot fer de forma manual per revisors experts, o es pot automatitzar el procés mitjançant la introducció de sistemes IA, reduint el temps de revisió, i amb la capacitat de descobrir un nombre d'errors majors que els trobats manualment. Un exemple d'eina basada en IA és la CodeGuru, que va posar a disposició pública Amazon Web Services al juny de 2020[4]. CodeGuru utilitza algorismes d'aprenentatge i raonament automatitzat per ajudar a depurar el programari.

La cerca i avaluació automatitzada de vulnerabilitats és una altra àrea d'interès a l'etapa d'identificació, ja que aprofita les capacitats tècniques de la IA per a la detecció de vulnerabilitats en un temps més reduït, amb els beneficis econòmics i de seguretat que es deriven, així com la reducció de temps i cost en el procés de correcció de les vulnerabilitats detectades. En aquest sentit, cal destacar un article acadèmic recent titulat *"Do users write more insecure code with AI assistants?"*[5], el qual estableix que el codi és més insegur quan els programadors confien en un assistent d'IA i li permeten crear codi, ja que aquest fet introdueix més vulnerabilitats. Per descomptat, això només és l'inici de l'ús d'assistents d'IA per programar, perquè en un futur programar sense assistents d'IA serà una temeritat. De fet, el ChatGPT ja evidencia capacitats de cerca i correcció de vulnerabilitats, com SQL Injections o en el cas dels smart contracts, útils per avaluar i desenvolupar codi segur[6].

Els dos exemples comentats permeten avaluar l'estat de robustesa de les aplicacions i sistemes com a un pas previ a la gestió de la seguretat. Es podrà millorar aquesta etapa d'identificació aprofitant les capacitats dels algorismes d'analítica predictiva basats en IA, processant dades en temps real procedents de diverses fonts de dades, permetent identificar nous vectors d'atac, o processant un volum massiu de logs, alertes i informació contextual per a la gestió avançada de riscos.

Per altra banda, s'apliquen tècniques i algorismes d'IA, concretament NLP, per a extraure coneixement d'amenaces (*Cyber Threat Intelligence*, CTI) a partir de fonts heterogènies, com són dades provinent d'eines d'intel·ligència oberta (*Open-Source Intelligence*, OSINT), informes d'incidents, fòrums de ciberdelinqüents, mitjans de comunicació, etc. L'objectiu és poder treure informació sobre vulnerabilitats i la seva explotació, o detectar noves -tàctiques-tècniques-procediments (TTP) d'atacs, o trobar indicadors que permetin realitzar el procés d'atribució del atacs. En aquest cas, la IA s'aplica no solament en l'etapa d'extreure informació de les diferents fonts de text mitjançant NLP, sinó en les següents etapes de detecció, classificació i atribució. L'ús d'ontologies i tecnologies semàntiques són d'especial interès, ja que permetrà tenir un llenguatge comú entre els diferents actors, compartint el coneixement d'amenaces de forma intel·ligible i fàcilment processable per algorismes d'IA.

Un altre cas il·lustratiu, i que cada vegada desperta més interès, és la identificació de comptes falsos d'usuari en xarxes socials, que poden servir per comprometre els sistemes de control d'accés de les organitzacions. L'aplicació d'algorismes IA per a l'anàlisi de comportament dels comptes d'usuari, la detecció de canvis de comportament, així com les similituds en el comportament de grups de comptes d'usuari permet tenir un coneixement més precís de la situació. L'anàlisi de les relacions socials d'un compte d'usuari, així com la utilització d'algorismes NLP per a detectar diferents models lingüís-

4 https://aws.amazon.com/codeguru/?nc1=h_ls

5 https://arxiv.org/pdf/2211.03622.pdf

6 https://www.elladodelmal.com/2023/01/como-buscar-vulnerabilidades-en.html

tics d'usuaris legítims i adversaris facilitarà la detecció d'intents d'atacs als sistemes de control d'accés.

4.1.2. Protecció

Com en l'etapa d'identificació, la inclusió d'eines, algorismes i sistemes basats en IA en tasques de protecció milloraran la robustesa i resiliència dels sistemes d'informació i control de les organitzacions. Una de les primeres aplicacions IA integrades són les basades en la seguretat contextual, concretament aplicades als sistemes de control d'accés. En aquest cas, per exemple, es validen els usuaris en funció del patró de pulsacions o moviments del cursor, o permetent l'accés d'un usuari a un recurs sempre que es trobi en el mateix context (com seria el cas d'accés a l'*streaming* en vídeo a les persones que estan a la mateix sala de conferències o l'accés a les notes d'una reunió per part dels usuaris que comparteixen la mateixa sala).

Un altre àmbit d'ús de la IA per a protecció el podem trobar en la defensa contra els atacs DDoS, on es realitza una anàlisi profunda del tràfic de la xarxa per a la detecció d'un patró anòmal degut a la presència d'un atac DDoS, i descartant els paquets maliciosos o generant una alerta que després serà tractada a l'etapa de resposta i recuperació. Continuant amb exemples de protecció a partir de l'anàlisi de xarxa i models de comportament, es pot trobar l'aplicació de la IA per a la implementació de mecanismes de contradefensa contra l'*spam*, el *phishing*, per a la detecció de *malware*, o la protecció contra pagaments fraudulents i ordinadors o xarxes compromeses.

En aquesta etapa del *framework* de ciberseguretat també es poden desplegar solucions basades en SDN (*Software Defined Network*), on la IA gestiona i orquestra de forma automàtica el desplegament de recursos, així com la seva administració. Després, a la fase de resposta, també serà necessari canviar de configuració de connexions i recursos davant un incident amb l'objectiu de protegir la infraestructura, de fer-la resilient a l'atac.

4.1.3. Detecció

En l'etapa de detecció és on es poden trobar més aplicacions i solucions basades en algorismes i tecnologies d'IA, tot i que inicialment no s'obtenien resultats significatius degut a la falta de coneixement de les capacitats reals de la IA per part dels experts i especialistes de ciberseguretat. Addicionalment, arran d'unes expectatives molt altes, va arribar un punt en què es va deixar de confiar en l'aplicació de la IA a l'àmbit de la ciberseguretat. Tanmateix, en els darrers anys aquesta percepció ha canviat radicalment i la IA s'utilitza àmpliament en les tasques de detecció de ciberamenaces.

En aquest sentit, una aplicació molt interessant i de gran valor de la IA és la detecció d'elements maliciosos a la infraestructura d'una organització, on l'anàlisi dinàmic d'aplicacions permet la detecció d'activitats anòmales o malicioses durant la seva execució, basant-se en la modelització de comportament mitjançant aprenentatge. S'utilitzen processos similars per al descobriment de dispositius i sistemes a l'hora de realitzar l'inventari d'un infraestructura, i per a la detecció de dispositius aliens a la infraestructura, susceptibles d'estar controlats per un ciberatacant amb l'objectiu de crear un punt d'entrada per iniciar un atac o per interrompre el servei. Aquesta aproximació també és vàlida per descobrir servidors o dispositius que simplement no s'han retirat adequadament de la infraestructura, però que ni es gestionen, ni s'actualitzen, ni tenen una funcionalitat assignada al servei, i que poden esdevenir una porta d'entrada per a iniciar una intrusió o atac.

Una altra aplicació amb molta repercussió és la detecció d'intrusions o anomalies, aprofitant el fet que els algorismes d'IA poden detectar qualsevol canvi que no sigui l'habitual/normal, sense la necessitat de definir què és l'anormal. Pel que fa a la detecció d'anomalies, el sistemes d'aprenentatge poden ser útils per detectar-ne de noves mitjançant models de comportament construïts a partir de grans volums de dades històriques i la identificació de patrons de comportament anòmals. Pel que fa a la detecció d'intrusions, els algorismes d'aprenentatge poden interaccionar amb l'usuari facilitant que aquest pugui entendre el context local i, a la vegada, centrar-se en les dades que són més rellevants durant el procés de monitorització i detecció d'intrusions. I, donat que els models entrenats poden identificar i classificar en temps real les dades que processen, permeten que l'usuari doni una resposta més ràpida en les situacions més crítiques. S'han investigat i implementat diverses aproximacions IA en aquest àmbit, arribant a la conclusió que un sistema de detecció efectiu es caracteritza principalment per i) combinar solucions basades en IA amb aproximacions basades en signatures/IoC (Indicadors de Compromís), ii) necessitar la interacció dels especialistes/experts en el domini

de la ciberseguretat per analitzar les incidències i actualitzar el sistema de detecció, iii) re-entrenar de manera continuada els models d'IA per adaptar-se a les noves amenaces i atacs.

Per últim, l'ús de *honeypots* (sistemes trampa o esquer) que simulen una infraestructura objectiu, exposant-los intencionadament a atacs dels adversaris, té la finalitat de recopilar el major nombre de dades possibles sobre les característiques d'aquests atacs. A partir del processament d'aquestes dades i aplicant algorismes, tècniques i metodologies de sistemes de detecció d'intrusions i anomalies, es poden determinar nous indicadors de compromís per noves amenaces. Per altra banda, els sistemes d'IA són capaços de generar *honeypots* adaptatius, més complexos que els tradicionals, ja que canvien el seu comportament en funció de la interacció amb els atacants, de forma que les dades recopilades són més riques en informació i aporten més valor al procés posterior d'extracció de coneixement.

Malgrat les reticències comentades a l'inici d'aquesta secció, una vegada els experts en ciberseguretat han pogut constatar les capacitats reals de les soluciones basades en IA, han vist com la capacitat d'anàlisi predictiu d'un gran volum de dades en temps real ajuda a reduir els errors humans, la càrrega de treball dels analistes de seguretat i millorar la base de coneixement dels sistemes de generació i gestió d'alertes. La suma d'aquests avenços es manifesta en una reducció del nombre de falsos positius. Per altra banda, la col·laboració i interacció entre els sistemes basats en IA i els especialistes/experts en ciberseguretat permet la detecció de noves amenaces que no han estat modelitzades prèviament millorant, per tant, el comportament dels processos de detecció.

4.1.4. Resposta i recuperació

En les etapes de resposta i recuperació l'aplicació de la IA és també de gran interès, on es poden explotar les capacitats dels algorismes i tecnologies d'IA millorant particularment el grau d'autonomia i temps de resposta automàtica. El disseny i desenvolupament de sistemes de suport a la presa de decisions i recomanadors de resposta davant un incident per a SOCs, CERTs, InfoSec i altres equips de seguretat incrementa l'automatització del servei i redueix el temps de resposta. Un exemple d'aquesta aplicació és l'ús de recomanadors tenint en compte el risc de l'incident, generant diversos escenaris amb l'objectiu d'optimitzar la resposta segons el risc exposat i minimitzar el cost associat.

Una vegada presa la decisió i les accions que se'n deriven, aquestes s'han de gestionar adequadament mitjançant diferents motors d'orquestració (SOAR). Això implica accions com, entre altres, prioritzar les actualitzacions en funció de la vulnerabilitat i el risc, comunicar als diferents actors implicats davant una incidència o seqüenciar les accions necessàries per protegir el sistema atacat (per exemple segregar les xarxes de forma dinàmica per aïllar els actius en zones controlades de la xarxa o redirigir l'atac cap un sector de la xarxa on no hi han dades valuoses). Com s'ha comentat a l'apartat de protecció, les arquitectures basades en SDN hauran de re-configurar-se i gestionar els recursos de forma adequada per fer resilient a la infraestructura, on la IA proporciona el grau d'autonomia i automatització necessari.

Si es puja la presa de decisions a nivell estratègic i de direcció, la IA ajudarà a prioritzar les inversions en ciberseguretat tenint en compte un volum d'informació considerable, com incidències, riscos, costos, personal, panorama d'amenaces, sector de negoci de l'organització, etc. Les eines o solucions que es poden trobar en aquesta fase de la gestió de la seguretat es basen en sistemes experts, models semàntics amb un motor de raonament o les habituals aproximacions d'algorismes d'aprenentatge, tant supervisat, no-supervisat com per reforç.

Després de l'incident, la IA pot aportar més valor tant com eina pel procés d'anàlisi forense, com a eina per a la generació de coneixement d'amenaces (*threat intelligence*). Si hom es centra en l'anàlisi forense digital i la investigació retrospectiva dels incidents es podem aplicar mètodes i tècniques utilitzades per a la detecció d'intrusions i identificació d'objectes maliciosos, tenint com dades les recollides durant la incidència que es vol analitzar, i aplicant els algorismes de classificació comentats prèviament. També s'han aplicat algorismes i tècniques d'IA per a la priorització dels objectes a investigar, utilitzant un càlcul de puntuació d'anomalia de les imatges de disc recopilades durant les primeres fases d'una anàlisi forense.

4.2. Impacte de la IA a la ciberseguretat

Per tal que la integració d'algorismes i tecnologies IA en la gestió de la seguretat dels sistemes i infraestructures produeixi el màxim impacte positiu, és necessari establir una relació directa entre els sistemes i els equips de seguretat. És una relació sinèrgica en què la IA té un efecte directe sobre les persones i equips de seguretat, sobre les tecnologies i infraestructures que es volen gestionar i, finalment, sobre els processos que governen la gestió de la ciberseguretat.

4.2.1. Impacte sobre les persones

La integració de la IA a l'àmbit de la seguretat té un impacte directe sobre les persones que integren els equips de ciberseguretat d'una organització, i per extensió en els equips directius de la mateixa. Es remarquen 4 aspectes de l'impacte en aquest àmbit:

- Talent: la manca de professionals experts de domini és uns dels principals problemes a què s'enfronta el sector de la ciberseguretat, com ja s'ha indicat i quantificat anteriorment. En certa manera, l'adopció de la IA pot ajudar a pal·liar aquesta mancança gràcies a la capacitat d'aprenentatge, processament massiu de dades, modelització de coneixement i altres característiques que fan molt atractives les solucions basades en IA.

- Fatiga: un problema rellevant en la gestió de la ciberseguretat en una organització, és la fatiga que experimenten els operadors que han d'interpretar i donar resposta a les alertes d'incidències que es generen, molt sovint en temps real. Aquests professionals estan exposats a una pressió constant i la fatiga i tensió que acumulen poden provocar errors en el servei (per exemple determinació de falsos positius o també falsos negatius) o en el moment de decidir i realitzar les accions necessàries davant una incidència. La introducció de la IA en les fases de detecció (generació i interpretació d'alertes) i de resposta i recuperació (orquestració d'accions entre els diferents sistemes i persones) aconsegueixen reduir algunes de les tasques assignades als operadors, les més automàtiques, disminuint la seva pressió, i per tant reduint els errors deguts a la fatiga del servei.

- Interpretació: una de les capacitats que caracteritza les solucions basades en IA és el processament massiu de dades, generant coneixement i relacions entre dades heterogènies de diferents fonts. Aquest coneixement generat de manera automàtica ajuda tots els actors involucrats en el cicle de gestió de la ciberseguretat doncs permet que es concentrin de manera eficient en la interpretació de la informació i presa de decisions, i no en el processament de dades.

- Presa de decisions: lligat amb la possibilitat de fer una bona interpretació del coneixement generat mitjançant eines i solucions basades en IA es pot esperar una millor qualitat en la presa de decisions tant a nivell operatiu com a nivell estratègic. Així, el coneixement de l'estat de la situació, l'anàlisi de riscos, més la informació referent al context d'amenaces i atacs, entre d'altres serveis basats en IA, són de gran valor per la presa de decisions per part dels professionals dedicats a la gestió de la ciberseguretat.

4.2.2. Impacte sobre els sistemes

L'adopció de la IA a l'àmbit de la seguretat també té un impacte directe sobre els sistemes d'informació i electrònics d'una infraestructura, que s'analitza a continuació a partir de cinc aspectes:

- Fiabilitat: l'automatització de processos i la capacitat de gestionar volums importants de dades diverses inherents a les solucions basa-

des en IA incrementa la fiabilitat global dins del cicle de gestió de la ciberseguretat.

- Automatització: el temps de resposta davant un incident, sigui de la naturalesa que sigui (atac, intrusió, sabotatge, anomalia, etc.) ha de ser el mínim possible, tant a l'hora de la detecció, com a l'hora d'orquestrar les accions necessàries de resposta i recuperació. Les capacitats de les eines i solucions basades en IA incrementen el nivell d'automatització en determinades fases, reduint el temps de resposta, reaccionant davant una incidència de forma més ràpida i, per tant, reduint la finestra d'exposició i explotació de l'adversari.

- Autonomia: l'adopció de la IA proporciona una capacitat d'autonomia a l'hora de realitzar accions de defensa o d'engany. La IA permet que els diferents sistemes de seguretat s'adaptin de forma dinàmica al comportament de l'adversari durant un atac o intrusió, realitzant accions d'engany dirigint l'adversari a un *honeypot*. També es poden prendre les decisions de resposta o recuperació de forma dinàmica en funció de com evolucioni l'atac i l'impacte que pot tenir al servei o negoci de l'organització.

- Pro-activitat: la capacitat predictiva que aporten algunes solucions d'IA permet que les eines i sistemes de seguretat reaccionin de forma pro-activa davant un incident. Detectar una potencial intrusió, atac, o incidència (p.e. el manteniment predictiu d'un recurs per evitar la indisponibilitat del servei) abans que pugui començar a comprometre la infraestructura, és vital a l'hora de prendre decisions, d'orquestrar accions i de reduir-ne el seu impacte.

- Visualització i interpretabilitat: la introducció de solucions IA impacta directament en les tecnologies de visualització de dades i coneixement, existint la necessitat d'interpretar el coneixement generat, i poder-lo transmetre a les persones adequades, tant a nivell operatiu com de direcció. Existeix la necessitat de desenvolupar nous interfases persona-sistema, on les aproximacions semàntiques (ontologies i motors de raonament), de grafs (*graph neural networks*), o noves tecnologies de visualització milloraran aquest procés d'interpretació de resultats.

4.2.3. Impacte sobre els processos per a la gestió de la ciberseguretat

Els processos per a la gestió de la ciberseguretat de sistemes i infraestructures també es veuen afectats per la introducció de la IA, en particular en 4 àmbits d'impacte:

- Desenvolupament i desplegament: la introducció d'aproximacions i algorismes basats en IA requereixen d'un conjunt de processos de preparació, anàlisi, disseny i construcció del model més adient per realitzar una determinada funció, impactant directament en els processos de gestió i administració. En el cas d'iniciatives basades en aprenentatge automàtic, cal considerar la definició i la integració de les fases de DevML en els processos d'administració del sistema per a la gestió de la ciberseguretat.

- Interpretació i gestió d'alertes: altre aportació de gran valor de les solucions basades en IA és la seva capacitat d'analítica predictiva, tal i com s'ha indicat prèviament. Els algorismes i tècniques d'IA per a la interpretació i gestió de les alertes té un gran i directe impacte en el procés de detecció, així com en les etapes de presa de decisions que es transmetran a diferents processos d'orquestració i administració.

- Orquestració d'accions: una vegada s'ha detectat una incidència, s'ha de donar resposta amb un cert grau d'autonomia i de forma més o menys automatitzada. Les aproximacions basades en IA impacten sobre les etapes de planificació i coordinació de les accions que cal realitzar, i sobre els processos de resposta i recuperació. Aquesta capacitat no solament impacta sobre els processos, sinó també sobre les persones i sistemes.

- Definició d'estratègies: Com s'ha comentat durant aquesta secció, les eines, algorismes i tècniques IA permeten la generació de coneixement de la situació de seguretat de la infraestructura a partir de totes les dades que aporten els diferents sensors. Per tant, la IA impacta directament en els diferents processos avançats i intel·ligents que ajuden a la definició de l'estratègia a seguir a nivell operatiu, estratègic, de direcció o d'inversió (financera).

- Anàlisi forense i d'incidències: una de les aplicacions de la IA en la gestió de la seguretat és l'anàlisi posterior de les incidències, amb l'objectiu final de generar coneixement d'amenaces i atacs. La inclusió d'aquests algorismes i tècniques impacten directament sobre els diferents processos d'actualització dels mecanismes de defensa i protecció, incrementant el seu grau d'automatització i autonomia. Per tant, aquests processos de millora i actualització hauran de tenir en compte les noves aproximacions IA, integrant els processos necessaris.

4.3. Oportunitats de futur

La irrupció de la IA al domini de la ciberseguretat obre moltes oportunitats de millora de les eines i sistemes en la lluita contra les ciberamenaces i ciberatacs, sobre tot per abordar totes aquelles amenaces que estan emergint pel canvi tecnològic i social que s'està produint, que com ja s'ha vist, comporten un increment de les superfícies d'atac i més possibilitats per als adversaris.

Les aportacions més rellevants de la IA aplicada al domini de la ciberseguretat s'han desenvolupat en apartats anteriors. Però, amb tot, encara hi ha alguns àmbits que presenten oportunitats d'expansió en l'aprofitament de les sinergies entre IA i ciberseguretat i que mereixen atenció, tant des del món empresarial i científic com de l'administració pública. En aquesta línia doncs, es destaquen tres àmbits: la ciberseguretat pels sistemes IoT, la ciberseguretat dels propis sistemes IA, i el marc regulador.

4.3.1. Ciberseguretat pels sistemes IoT

El canvi tecnològic i social en l'ús de les comunicacions i internet ens ha portat a un món més connectat, més distribuït, menys centralitzat i en el que es produeix una convergència entre el món físic i el món cibernètic. Connectant milers de milions d'objectes i dispositius intel·ligents, la IoT té com objectiu construir un món intel·ligent que ens permeti percebre, analitzar, controlar i optimitzar els sistemes tradicionals utilitzant les tecnologies disruptives que contínuament estan apareixent. Encara que les aplicacions IoT esperen disposar d'una forta protecció de seguretat, no és gens senzill. Les greus limitacions de recursos i un disseny de seguretat insuficient són les dues principals raons dels problemes de seguretat actuals del sistemes IoT.

Molts del mecanismes de seguretat existent, inclosos els algorismes de seguretat avançats, com el control d'accés basat en atributs, l'autenticació basada en la signatura de grup, el xifrat homomòrfic i les solucions de clau pública, requereixen d'un dispositiu amb un alt nivell de computació, energia i espai de memòria per executar-lo, que en la majoria de casos no és aplicable. Una solució a aquest problema es el núvol (*cloud*), on hi ha recursos quasi il·limitats, però que no és factible en sistemes de control crític o en absència de connectivitat per una acció intencionada o no-intencionada. Davant d'aquest problema, el millor recurs és donar pas a la computació a l'extrem (*Edge Computing*) on es disposarà dels recursos necessaris prop dels punts finals, dels dispositius IoT o sistemes d'informació. Aquest nou paradigma de computació no només dona resposta a les limitacions de recursos als dispositius finals, sinó que també ajuda a optimitzar el rendiment dels sistemes, i a oferir un entorn per desplegar eines i solucions de seguretat per les infraestructures IoT i infraestructures deslocalitzades.

S'han fet algunes implementacions per dissenyar solucions de seguretat en la IoT, des de sistemes de detecció d'intrusos fins a protocols d'autenticació i autorització, tallafocs o mecanismes per preservar la privacitat. I és aquí on la IA té un paper a jugar, doncs la seguretat passa a ser distribuïda, descentralitzada i col·laborativa, i cal dotar de certa intel·ligència les eines i components de seguretat. Les eines de seguretat han de tenir capacitat de detecció local d'intents d'intrusió i d'adaptació de forma dinàmica a les noves amenaces. A la vegada, han de comunicar-se amb la resta de dispositius i sistemes de la mateixa xarxa per organitzar-se, prendre decisions i orquestrar les accions necessàries a dur a terme, etc. A nivell local, el dispositiu ha de tenir la capacitat de realitzar accions d'auto-protecció davant un intent d'intrusió o atac, d'auto-resposta i auto-recuperació, amb l'objectiu de tenir el major de resiliència possible. Ja a un nivell superior, s'hauran de realitzar tasques de recerca per aplicar solucions basades en *swarm intelligence* per coordinar tots els dispositius (per exemple aïllament de sectors de xarxes compromeses), mantenint la seva independència, però treballant de forma col·laborativa per lluitar contra les amenaces i atacs que poden realitzar-se de forma globalitzada, com un atac ransomware dirigit a cer-

tes infraestructures crítiques, o determinats grups d'usuaris. Podem comprovar doncs, que la IA impacta sobre l'automatització i autonomia a l'hora de prendre decisions relatives a la seguretat.

4.3.2. Ciberseguretat per a la IA

Com ja s'ha descrit a la secció 3, la integració de solucions basades en IA fan aparèixer noves superfícies d'atac, on les vulnerabilitats que presenta la IA no es poden actualitzar ni apedaçar, sinó que s'han de gestionat adequadament. Encara que aquest fet és una barrera en l'adopció i desplegament d'aquest tipus de solucions, també es pot considerar com una oportunitat científica-tècnica i de negoci de futur, on la protecció i gestió de la seguretat de la IA és una línia de recerca i desenvolupament de gran potencial.

Per exemple, en el cas de les *deep fakes*, si hom es centra en els atacs contra els sistemes de reconeixement de veu, tant pel control d'accés o assistents, on la majoria es basen en models IA, es poden classificar en: (1) clonació de veu, generant veu a partir d'un text o aprenent a partir de mostres de veu d'un objectiu concret; (2) conversió veu-text, transformant clips de veu en text, generant nous significats o ordres no esperades; (3) enverinament (*poisoning*), modificant les dades d'entrenament degut a vulnerabilitats presents, i generant models defectuosos. La resposta a aquest tipus d'atacs passa per l'anàlisi, recerca i disseny d'eines de protecció i detecció basades en IA, on la capacitat d'aprenentatge és la que proporcionarà els models necessaris per abordar aquest tipus d'amenaces.

En el cas que l'adversari vulgui manipular els sistemes IA que controlen sistemes crítics, o sistemes IA que formen part d'una solució per a la seguretat, pot aprofitar les vulnerabilitats innates de la IA. Així, en els sistemes basats en aprenentatge totes les etapes, des del desenvolupament del model fins el seu desplegament i el seu mode operatiu, presenten una sèrie de reptes a tenir en compte. Per exemple, un adversari pot dirigir el seu atac sobre les dades, tant en la fase d'adquisició com en el tractament previ. En un sistema d'aprenentatge basat en l'adquisició de dades, la qualitat de les mateixes (integritat) és el factor més crític, doncs les dades poden ser enverinades de forma intencionada, o de manera no intencionada. Per tant, ja des de l'etapa de tractament o preparació de les dades s'ha de garantir que els processos no tenen risc d'afectar-ne la qualitat i integritat. El procés d'etiquetat ha de ser confiable, i el procés d'augment de les dades ha de garantir que no afecta la integritat de les mateixes.

Durant l'etapa de construcció dels models, la confidencialitat del conjunt de dades d'entrenament pot ser compromesa per un adversari que tingui cert coneixement de la implementació de l'algorisme. Una vulnerabilitat d'integritat pot introduir-se a vegades mitjançant l'ús d'aprenentatge per transferència. Un altre tipus de vulnerabilitat que es pot introduir és una porta oculta (*backdoor*), incloent un patró especial que a l'etapa d'inferència pot generar una sortida determinada, controlada per l'adversari. Existeixen més vulnerabilitats que es poden introduir al model durant la fase de construcció, que són de difícil solució posteriorment. Per tant, s'hauran d'integrar i implementar les contramesures avançades per contrarestar les accions malicioses dels adversaris, obrint una altra línia d'aplicació d'IA per la seguretat dels sistemes IA.

Finalment, poden aparèixer vulnerabilitats introduïdes de forma intencionada o no, que poden ser explotades per un adversari per comprometre la confidencialitat, integritat o disponibilitat dels models IA durant la seva fase operativa, o durant la fase d'actualització. Per tant, és necessari investigar i treballar en el disseny, implementació i integració d'eines i mecanismes de seguretat durant tot el cicle de vida d'una solució basada en IA, i no cal dir que aquestes solucions han de tenir capacitats d'aprenentatge i adaptació dinàmica, així com aportar un grau important d'autonomia i automatització en tots els processos, és a dir han d'estar basades i integrar algorismes i tecnologies IA.

4.3.3. Escenaris derivats dels nous marcs reguladors

Finalment, l'adopció i les oportunitats reals d'aplicació i integració d'eines i tecnologies IA per a la seguretat de les infraestructures, sistemes d'informació i societat en general, han d'anar acompanyades d'un marc normatiu que estableixi mecanismes legals i de cooperació que incrementin les oportunitats reals d'ús de la IA en el domini de la ciberseguretat, i resolgui els aspectes ètics, de confidencialitat i privadesa.

Per tant, com a resultat de la definició dels marcs reguladors necessaris i recomanables, molt probablement es generaran un conjunt d'oportunitats de

millora per l'adopció de la IA en el domini de la ciberseguretat:

- Millora de la col·laboració entre els responsables polítics, la comunitat científico-tècnica i els principals representants de la indústria per fer recerca, prevenir i mitigar millor els possibles usos maliciosos de la IA en ciberseguretat.
- Aplicació i comprovació dels requisits de seguretat dels sistemes IA en les polítiques de contractació públiques, incloent els principis ètics.
- Intercanvi d'informació sobre les dades rellevants per a la ciberseguretat, com per exemple, dades per entrenar els models segons les millors pràctiques establertes.
- Suport a la fiabilitat de la IA, més que a la seva confiabilitat, en els estàndards i mètodes de certificació. Promoure els esforços proactius de certificació de la ciberseguretat de la IA.
- Control de l'apertura total pels resultats de recerca, com els algorismes o paràmetres dels models.
- Tractament de les normes de protecció de les dades personals que presenten els conjunts de dades mixtes personals i no personals.
- Millora de la cooperació entre les diferents entitats en temes de recerca i desenvolupament en IA, definint una arquitectura de referència per la ciberseguretat específica per a les aplicacions de la IA.
- Generació de competències i la distribució del talent i professionals entre els diferents agents del mercat.

Anàlisi de la IA en l'àmbit de la ciberseguretat a Catalunya

5

Per fer aquesta anàlisi es parteix d'una revisió sobre l'ecosistema global de la IA a Catalunya, identificant-ne a continuació les principal fortaleses. Per a il·lustrar aquestes capacitats s'identifiquen un conjunt d'iniciatives i projectes que presenten una clara intersecció entre la IA i la ciberseguretat, però només des del punt de vista dels equips de defensa, i no des del punt de vista de l'adversari o atacant.

Les característiques i problemàtiques a l'àmbit de la ciberseguretat compliquen el desplegament de solucions basades en IA, per tant, s'enumeraren també les principals barreres que dificulten l'adopció de la IA en el sector de la ciberseguretat. Aquests darrers punts, la identificació de barreres i la proposta de recomanacions (secció següent), han estat especialment treballats amb els diferents participants i col·laboradors en l'elaboració del document , així com a les sessions de discussió conjuntes.

5.1. La IA a Catalunya

L'informe *La Intel·ligència Artificial a Catalunya*[1] publicat per ACCIO el juliol de 2019 identifica l'ecosistema de companyies, universitats, centres de recerca i innovació, així com les institucions que conformen el sector de la IA a Catalunya. Amb dades publicades el 2019, el sector de la IA a Catalunya engloba aproximadament un total de 179 empreses, amb una facturació de 1.358 milions d'euros anuals i ocupen un total de 8.483 professionals.

Del total d'empreses, el 92% són petites i mitjanes, d'entre les quals un 63% són *startups*. Un altra dada interesant a destacar és que el 27% de les empreses són exportadores de productes, serveis o consultoria a l'exterior.

Per últim, el 71% d'aquestes empreses es va crear fa menys de 10 anys, la qual cosa indica que la creixent demanda de solucions basades en intel·ligència artificial en els darrers anys, estimulant el naixement d'un important nombre d'empreses del sector.

Si fem una divisió pels segments tipus de serveis (de forma anàloga RADAR de mercat de l'ECSO pel sector de la ciberseguretat), i tenint en compte que una empresa pot estar classificada en més d'un segment, podem observar que 134 de les empreses es dediquen al desenvolupament de software, 19 a la construcció d'algorismes, 19 a la consultoria tecnològica, i 7 a oferir serveis relacionats amb la IA.

Aquesta diagnosi del sector de la IA a Catalunya cal complementar-la amb l'escenari del sector de la ciberseguretat. Les sinergies entre ambdós àmbits són paleses, doncs algunes empreses catalanes del sector de la ciberseguretat incorporen la IA en els seus desenvolupaments, solucions i productes. Per tant, Catalunya disposa d'una base empresarial i de coneixement que afavoriran les sinergies i col·laboracions en la recerca i desenvolupament de solucions de ciberseguretat basades en IA, aportant valor i generant riquesa, així com millorant la seguretat de les infraestructures i la societat.

5.2. La indústria de la ciberseguretat a Catalunya

Amb dades publicades el 2023, el sector de la ciberseguretat a Catalunya engloba aproximadament 495 empreses que donen cobertura de consultoria, serveis i productes, amb un volum de facturació global propera a 1.071 milions d'euros i una ocupació de 9.414 llocs de treball.

Del total d'empreses, el 85% són petites i mitjanes, d'entre les quals un 9,5% són *startups*. Un altra dada interesant a destacar és que el 28,7% de les empreses són exportadores de productes, serveis o consultoria a l'exterior. Quant al volum, el 53,6% de les empreses facturen més d'un milió d'euros i el 21,7% més de 10. Pel que fa a la paritat home-dona només un 13,8% tenen dones en posicions directives.

Per últim, el 29,1% d'aquestes empreses es va crear fa menys de 10 anys, la qual cosa indica que la creixent demanda de ciberseguretat en els darrers temps ha estimulat el naixement d'un important nombre d'empreses del sector.

Si fem una divisió pels segments corresponents a les 5 categories vistes anteriorment amb l'estructura del RADAR de mercat de l'ECSO, i tenint en compte que una empresa pot estar classificada en més d'un segment, podem observar que el 89,7% de les empreses es dediquen a la protecció, el 58,7% a la

1 https://www.accio.gencat.cat/web/.content/bancconeixement/documents/informes_sectorials/informe-tecnologic-inteligencia-artificial.pdf

Identificació, el 37% a la detecció, el 33,6% a la resposta i el 20,6% a la recuperació.

Podem afirmar que en aquests darrers anys el sector de la ciberseguretat a Catalunya ha mantingut una tendència de creixement constant, amb poc impacte de factors rellevats com l'aturada generalitzada de l'economia per la pandèmia i la incertesa deguda al conflicte bèl·lic entre Ucraïna i Rússia, sinó tot al contrari, aquests dos factors han incrementat la necessitat de professionals i solucions a l'àmbit de la ciberseguretat. S'espera que en els propers anys es mantingui aquesta tendència de creixement, o inclús més exagerada, degut als reptes que hem vist anteriorment condicionada, però, a l'evolució global de l'economia[2].

5.3. Fortaleses i acceleradors a Catalunya

Fins l'elaboració d'aquest informe no s'ha realitzat cap estudi que abordi l'impacte sectorial a Catalunya (per exemple en xifres de facturació o llocs de treball generats) de la intersecció dels àmbits de la IA i de la ciberseguretat. Però, en base als resultats recollits en aquest document, les consultes a experts i la pròpia activitat de les diverses entitats que han participat en l'elaboració de l'informe, es pot afirmar que l'ús de la IA en el camp de la ciberseguretat pot esdevenir una domini d'aplicació i negoci amb projecció a Catalunya. Existeixen diversos factors en suport a aquest posicionament, que s'examinen a continuació.

5.3.1. Un sistema de coneixement consolidat

Les activitats de formació al voltant de la IA són cada vegada més importants per als departaments tecnològics de les universitats. Catalunya disposa d'un ecosistema ric en universitats de referència. La Universitat Politècnica de Catalunya (UPC) va establir el 2005 el primer Màster oficial en IA de Catalunya, implementat des de la seva creació en coordinació amb la Universitat de Barcelona (UB) i la Universitat Rovira i Virgili (URV) i en el que més del 50% dels estudiants, de manera consistent en els anys, són internacionals.

La UAB per la seva part, coordina el Màster oficial en Visió per Computador, implementat en col·laboració amb la UOC, la UPC i la UPF, i on el Centre de Visió per Computador (CVC) juga un paper molt important.

Per altra banda, la UAB (amb la participació del Centre de Visió per Computador, CVC, i l'Institut d'Investigació en IA, IIIA-CSIC) i la UPC al curs 2021-22 van posar en marxa respectivament els dos primers graus universitaris en IA de Catalunya. La UPC es converteix, per tant, en la primera universitat espanyola que ofereix simultàniament grau, màster i doctorat en aquesta especialitat científica.

Altres universitats catalanes també estan fent una forta aposta per la incorporació de la IA en el seu currículum educatiu, com els màsters d'investigació de la UPF enfocats a tecnologies relacionades com visió per computadors, sistemes intel·ligents interactius, etc. Es també destacable la incorporació d'assignatures d'IA en màsters de Ciència de Dades i altres cursos de curta durada (La Salle-URL, UoC, UdL, UdG, etc).

Però, malgrat que el nombre de propostes formatives relacionades amb la IA és rellevant, i que a Catalunya existeixen 11 graus i màsters impartits per diferents universitats i centres formatius, l'escletxa existent actualment pel que fa a formació creuada d'IA i ciberseguretat és encara significativa, ja que no existeixen propostes formatives que englobin les dues àrees de coneixement.

Catalunya disposa d'un potent model de recerca i innovació amb centres de recerca i centres tecnològics capdavanters. En IA destaca el Centre de Visió per Computador (CVC), l'Institut de Robòtica i Informàtica Industrial (IRI), l'Institut d'Investigació en IA (IIIA-CSIC), el Barcelona Super Computing Center (BSC), el Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (TALP), i el recentment creat Centre de Recerca en Ciència de Dades Intel·ligent i IA (IDEAI-UPC). En l'àmbit de la innovació i de la transferència cal mencionar centres tecnològics com Eurecat, i2CAT o Leitat, que es caracteritzen per l'aplicació de la IA als principals sectors d'activitat a Catalunya (indústria, salut, recursos, etc.).

La Figura 7 mostra els agents existents a Catalunya

2 Informe Agència de Ciberseguretat de Catalunya i ACCIO – La Ciberseguretat a Catalunya de 2023, https://ciberseguretat.gencat.cat/web/.content/PDF/Ciberseguretat_Informe-tecnologic-2022.pdf

Figura 7: Agents de l'ecosistema de la Ciberseguretat a Catalunya (font: La Ciberseguretat a Catalunya, AC-CIO, març 2022)

amb relació a l'ecosistema de la ciberseguretat. Si bé en el cas de la formació s'ha indicat l'existència d'un buit d'estudis combinats d'IA més ciberseguretat, la situació és diferent pel que fa a l'àmbit de la innovació. On alguns centres de recerca i tecnològics presenten grups o unitats de recerca, desenvolupament i transferència tecnològica consolidats pel que fa a la ciberseguretat i solucions basades en IA.

5.3.2. Atracció de talent

Catalunya, i en especial Barcelona, han esdevingut pol d'atracció i demanda de talent per diversos factors com per exemple l'increment notable d'iniciatives empresarials i *startups* o la implantació de nous centres operatius d'empreses multinacionals, però també per aspectes relacionats amb la qualitat de vida, culturals o altres.

L'informe Digital Talent Overview 2022[3] elaborat pel Barcelona Digital Talent de la Mobile World Capital Barcelona, fa una anàlisi detallada de les tendències relatives al talent digital a nivell global, català i amb un focus específic per a la regió de Barcelona. L'informe indica que els perfils professionals més demandats a Catalunya són en els àmbits de les Telecomunicacions, la Ciberseguretat i la IA i també el de programador, que és un perfil que continua amb un gran auge. Això es plasma en la demanda de talent existent, on el 2021 ha tingut un increment respecte el 2020 del 33% en el domini de la ciberseguretat, i del 59.4% a l'àmbit de la IA.

Aquest informe també indica que de tot el talent que prové de fora de Barcelona i la seva àrea d'influència durant el 2021, el 40,55% correspon a perfils de ciberseguretat.

5.3.3. Lideratge d'iniciatives

La declaració de Barcelona sobre la IA (2017), liderada per la comunitat científica catalana de la IA, és una de les iniciatives pioneres a Europa en debat ètic, jurídic, socioeconòmic i cultural sobre els usos futurs d'una IA segura i confiable. Els científics en IA catalans són molt actius en múltiples iniciatives sobre l'ètica en la IA, inclòs el desenvolupament de directrius ètiques de la Comissió Europea per a la IA, i que afecten directament al desplegament de solucions basades en IA al sector de

3 https://barcelonadigitaltalent.com/en/report/digital-talent-overview-2022/

la ciberseguretat. Les administracions públiques tenen voluntat de donar impuls a iniciatives d'IA. La Generalitat de Catalunya, assessorada per un equip d'experts, va publicar el seu Pla Estratègic en Intel·ligència Artificial, denominat CATALONIA.AI[4] que es va fer públic el juliol de 2019. Amb el desplegament d'aquesta estratègia s'han posat en marxa els quatre pilars que la fonamenten: el Centre of Innovation in Data and Artificial Intelligence (CIDAI) amb focus en la innovació tecnològica, l'Artificial Intelligence Research Alliance (AIRA) per promoure la recerca especialitzada i la formació, l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya (OEIAC) per vetllar sobre els aspectes ètics de la IA i la Digital Catalonia Alliance-IA (DCA-IA) amb focus en l'emprenedoria. L'estratègia CATALONIA.AI és pionera i ha esdevingut un model de referència per a altres regions europees.

Per altra banda, el conjunt d'Administracions públiques catalanes van instar la Generalitat de Catalunya a definir un servei públic de ciberseguretat i desenvolupar una Estratègia i Pla Nacional de Ciberseguretat[5], liderada per l'Agència de Ciberseguretat de Catalunya (ACC), amb la responsabilitat de definir una planificació, gestió o control en matèria de seguretat de les TIC de la Generalitat de Catalunya, per prevenir i donar resposta als incidents que puguin posar en perill els sistemes d'informació de la Generalitat de Catalunya i el seu sector públic.

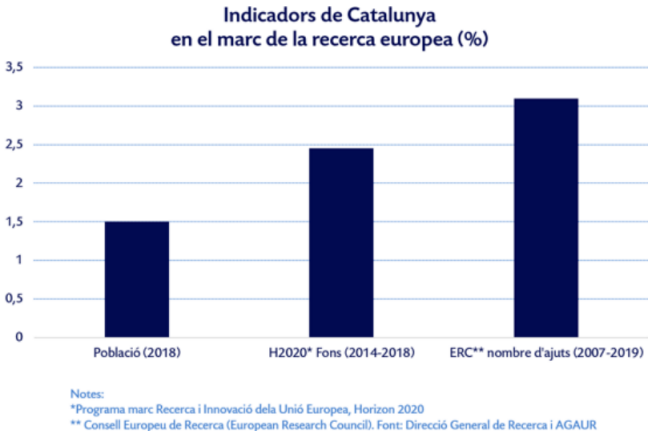
A nivell europeu existeixen deferents iniciatives relacionades amb la IA i ciberseguretat, com ECCC, ECSO i ENISA, on totes elles disposen de grups de treballs concrets amb relació a la sinèrgies entre els dos dominis de coneixement, on hi participen membres d'empreses i grups de recerca.

5.3.4. Pol d'innovació

El Centre Europeu de Competència en Ciberseguretat (ECCC), junt amb la Xarxa de Centres Nacionals (NCC), formen el nou marc d'Europa per recolzar la innovació i la política industrial en ciberseguretat. Són l'estructura de coordinació de la inversió pública europea en ciberseguretat que, aprofitant el que s'ha denominat Comunitat tecnològica de ciberseguretat, establirà una agenda comuna pel desenvolupament tecnològic i pel seu ampli desplegament, que es mourà a través de la inversió en projectes estratègics de ciberseguretat. És en aquest marc, i per fer aquesta visió una realitat, l'Agència de Ciberseguretat de Catalunya, dins del seu àmbit competencial i territorial, crea el CIC4Cyber (Centre d'Innovació i Competència en Ciberseguretat). El CIC4Cyber és una estructura organitzativa pública que té un clar objectiu de coordinació, cohesió i capacitat de l'ecosistema de ciberseguretat de Catalunya, es recolza en el coneixement i la innovació com a palanca de transformació i creixement del sector, i en el treball en xarxa com a via per maximitzar l'impacte dels seus programes i projectes, fomentant la captació de fons i la internacionalització.

Catalunya presenta uns bons indicadors en el marc de la recerca europea (Figura 8) ja que entre el 2014 i 2018 ha rebut 1.070 milions d'euros del programa Horizon2020. Això equival al 2,4% dels fons H2020 distribuïts a Europa, per sobre de la població de Catalunya respecte el total de la Unió Europea (1,5%). Entre el 2007 i el 2019 Catalunya ha rebut un 3,2% dels fons europeus de recerca (ERC). Aquests fons estan destinats a recerca capdavantera i basats en l'excel·lència científica.

Figura 8: Indicadors catalans en el marc de la recerca europea



Encara, en el marc europeu, i segons s'acredita a l'estudi Anàlisi de l'especialització en Intel·ligència Artificial[6], publicat el gener de 2021 i realitzat en el marc del monitoratge de la RIS3CAT, el 8,4% dels

4 https://politiquesdigitals.gencat.cat/web/.content/00-arbre/economia/catalonia-ai/Estrategia_IA_Catalunya_VFinal_CAT.pdf

5 https://ciberseguretat.gencat.cat/web/.content/PDF/EstrategiaCiberseguretat-de-Catalunya-2019.2022.pdf

6 http://catalunya2020.gencat.cat/web/.content/00_catalunya2020/Documents/estrategies/fitxers/analisi-especialitzacio-intelligencia-artificial.pdf

projectes H2020 amb participació catalana versaven sobre IA (la mitjana europea és 6,6%) ocupant la 3a posició en nombre de projectes (228) i la 4a posició en finançament atorgat (130 milions d'euros). Per altra banda, i segons es recull al ja referit informe La Ciberseguretat a Catalunya, el programa H2020 ha finançat 22 projectes a l'àmbit de la ciberseguretat on participen 22 entitats catalanes, amb una inversió de 7.7 milions d'euros i amb una relació amb 357 socis de l'exterior, oferint noves oportunitats per projectes de recerca.

5.3.5. Fort teixit empresarial

Catalunya s'ha posicionat com un atractiu pol d'innovació empresarial en el sector de la ciberseguretat i, com ja s'ha destacat a la secció 5.2, compta amb 432 empreses i 8.188 llocs de treball, on el 18,1% són *startups*. Aquesta proporció d'empreses *startups* demostra l'aposta real per l'emprenedoria tecnològica, posicionant Barcelona com un important pol d'innovació internacional amb derivades com, per exemple, la celebració d'esdeveniments internacionals en IA i ciberseguretat. Gran part de les *startups* o empreses de nova creació inclouen serveis o productes amb solucions basades en IA, tant per l'etapa d'identificació, protecció, detecció o en la gestió d'alertes, resposta i recuperació.

Barcelona ha esdevingut el tercer hub més atractiu i popular[7] per a *startups* a Europa. En particular, l'ecosistema de *startups* catalanes en software i IA és molt potent, amb diversos i destacables exemples. També s'han atret acceleradors d'empreses emergents, *venture builders* i de capital de risc que ofereixen una oportunitat única per desenvolupar nous negocis al voltant de la majoria de tecnologies disruptives. Destaquen organitzacions vinculades a l'emprenedoria tecnològica com el Barcelona Tech City.

5.3.6. Actiu teixit associatiu

Catalunya destaca per la forta activitat associativa, també en els àmbits tecnològics com la IA i la ciberseguretat. L'any 1994 es va constituir l'Associació Catalana d'Intel·ligència Artificial, (ACIA) i en l'actualitat té més de 200 membres, reunint la major part de la comunitat científica catalana d'IA, així com antics alumnes, professionals del sector i alguns associats institucionals. Tot i ser l'associació d'un petit territori, l'ACIA és, des de l'any 1995, un capítol de l'Associació Europea d'IA (EurAI) i organitza una conferència anual internacional des de l'any 1998. Per altra banda, Catalunya disposa del Col·legi Oficial d'Enginyeria Informàtica de Catalunya, el COEINF, creat pel Parlament de Catalunya l'any 2001 amb la missió de vetllar per la professió d'enginyeria informàtica. Aquesta organització professional és un suport fonamental per a la IA a Catalunya. Des de l'any 2016, el COEINF inclou una posició específica en el seu equip directiu dedicada a la IA: el vicedeganat de Big Data, Ciència de les Dades i IA. Aquesta posició promou el desenvolupament fructífer del sector empresarial en aquests camps, inclosa la necessitat de reduir l'escletxa de gènere en el sector, amb una comissió dedicada a aquest tema: donesCOEINF. La bretxa de gènere també es va convertir en objecte de consideració per a l'ACIA i el març del 2019 va fundar un grup de treball de dones en IA a Catalunya anomenat donesIAcat.

Una altra associació d'interès, ja identificada en el marc de la CATALONIA.AI (secció 5.3.3) és la Digital Catalonia Alliance (DCA), iniciativa que agrupa els principals sectors tecnològics emergents del territori català en una aliança de comunitats tecnològiques innovadora, visionària, disruptiva i col·laborativa, on la ciberseguretat, IA, IoT NewSpace i Blockhain tenen un paper destacat, facilitant les sinergies entre aquestes diferents tecnologies. Per altra banda, les diferents universitats catalanes amb departaments que treballen a l'àmbit de la ciberseguretat s'han agrupat en l'associació CYBERCAT, amb l'objectiu d'establir un espai comú de col·laboració i compartició de coneixement.

5.3.7. Disposició d'infraestructures adequades

Infraestructures tècniques d'alt nivell com el Supercomputador MareNostrum, instal·lat a Barcelona l'any 2005, al Centre Nacional de Supercomputació (BSC-CNS), una referència internacional crucial per al processament de dades massives, o el Centre de Visió per Computador (CVC), que disposa de més de 200 unitats de processament gràfic (GPU) connectades en xarxa, el que el converteix en un dels centres amb major nombre de GPUs treballant simultàniament del món. També s'ha de tenir en compte la xarxa de CERs i CSIRTs de titularitat pública i privada existent, que poden facilitar la re-

7 https://startupheatmap.eu/

alització i integració de solucions IA a l'àmbit de la ciberseguretat. Quant a altres infraestructures, Barcelona ha creat recentment el Labs.5G Barcelona, en connexió amb la Mobile World Capital, per donar suport a la innovació en tecnologia 5G.

No hem de deixar de banda les diferents infraestructures de transport destacables i obertes a nivell global com són el port o l'aeroport de Barcelona, que faciliten l'arriba del talent necessari. L'existència de *hubs* industrials com la Zona Franca o els parcs automobilístics, ajuden d'igual forma a incrementar l'atractiu de Catalunya per atraure talent i empreses que puguin implementar solucions IA i ciberseguretat.

5.3.8. Seu de grans esdeveniments

Barcelona acull alguns esdeveniments de més renom internacional en l'àmbit de la tecnologia i la digitalització de l'economia. Des del 2011 Barcelona és la Mobile World Capital, i per aquesta raó ha esdevingut una ciutat de referència en la tecnologia mòbil. I, a més, és seu del Mobile World Congress, de l'Smart City Expo World Congress, Smart Mobility World Congress I l'IoT Solutions World Congress. També de congressos sectorials com l'Advanced Factories o l'AI & BDg Data Congres. En l'àmbit de la ciberseguretat Barcelona acull el Barcelona Cybersecurity Congress. Aquesta intensa activitat congressual fomenta el coneixement internacional de la ciutat i del seu ecosistema tecnològic i l'intercanvi i creació d'oportunitats a tots nivells.

5.3.9. Impuls de tecnologies facilitadores

La ràpida evolució i accés per a empreses i organismes catalans en els darrers anys de tecnologies al núvol, 5G, tecnologies de processament de dades com Big Data, tecnologies de captació de dades com sensòrica avançada i IoT, etc. contribueixen a ampliar el rang d'aplicació de la IA al sector de la ciberseguretat, ja que el model de seguretat basat només en la protecció del perímetre d'atac és insuficient per respondre als nous reptes de seguretat que comporten aquestes noves tecnologies facilitadores. Destaquen tecnologies i disciplines acceleradores com:

- Sensorització i IoT per a la comunicació màquina a màquina on unes infraestructures segures per a la intercomunicació dels dispositius connectats és cabdal.
- *Big data* o gran volum de dades de fonts heterogènies i anàlisi de dades. La capacitat d'operar de manera fluïda en aquest àmbit permet apostar per iniciatives que seran clau pel sector de la seguretat per millorar el coneixement de les noves amenaces i el comportament dels atacants.
- *Edge computing,* paradigma de computació emergent que requerirà d'una presa de decisions de seguretat distribuïda per una ràpida resposta davant una incidència i reduir el seu impacte.
- Robòtica, sector que està experimentant un creixent desenvolupament i implantació de de màquines autònomes cognitives.
- 5G i compartició de dades generant Infraestructures intel·ligents i serveis compartits convergits.

5.4. Iniciatives d'IA i ciberseguretat a Catalunya

Per tal d'il·lustrar amb exemples reals l'aplicació de solucions basades en IA en el sector de la ciberseguretat, s'identifiquen a la Taula 1 diversos projectes i casos d'ús desenvolupats a Catalunya. Consisteix en una selecció de 9 iniciatives considerades rellevants i inspiradores per a l'ecosistema de la ciberseguretat a Catalunya, i que han estat impulsades o compten amb una important participació d'integrants d'aquest ecosistema: petites i grans empreses, *startups*, administració pública, centres de recerca, tecnològics i universitaris. A la vegada, adrecen els principals reptes de la ciberseguretat identificats a l'inici del document. El detall i característiques d'aquests casos il·lustratius es troba a l'ANNEX.

Aquesta breu relació d'iniciatives i projectes de ciberseguretat on la IA hi juga un paper clau. palesa la capacitat dels actors catalans pel que fa al dinamisme, iniciativa i visió per explorar el potencial de la IA al domini de la ciberseguretat a diferents sectors (salut, financera i assegurances, industrial, energia, automoció, etc.). S'ha de puntualitzar que alguns dels

Taula 1. Taula resum d'iniciatives de ciberseguretat amb solucions integrades d'IA.

CAPA D'ACTUACIÓ	REPTES	CASOS IL·LUSTRATIUS
Identificació	Modelització de comportament de dispositius electrònics	SECUTIL
Protecció	Autenticació biomètrica evolutiva	NEWIDENTITY
	Proporcionar un marc de resiliència cibernètica per a PYMES	PALANTIR
	Sistema de prevenció d'intrusions anti-pirateria per a la indústria de l'automoció	CARAMEL
Detecció	Detecció d'atacs combinats en infraestructures crítiques	STOP-IT
	Seguretat distribuïda	SECUTIL
	Detecció d'adversaris interns	PRODUCTIO
	Sistema de detecció d'intrusions anti-pirateria per a la indústria de l'automoció	CARAMEL
Resposta	Resposta col·laborativa davant un incident	INTSEG
	Marc de resiliència cibernètica per a la continuïtat i la recuperació del negoci.	PHOENI2X
Recuperació	Marc de recuperació d'incidències explotant les tecnologies NFV i SDN	PALANTIR
	Marc de resiliència cibernètica per la resposta a incidents i l'intercanvi d'informació.	PHOENI2X
Anàlisi d'incidències	Ajudar als equips i especialistes de contra-intel·ligència.	SPARTA
	Modelització d'amenaces i atacs mitjançant *honeypots*	INTSEG
	Modelització d'usuaris susceptibles de ser atacats per analitzar el risc associat	TDA

projectes presentats estan encara en fase de prova pilot, o bé són projectes demostratius o proves de concepte. Per tant, serà clau apostar per la continuïtat d'aquestes iniciatives i, a la vegada, inspirar-se dels projectes o explorar sinergies que es puguin derivar dels agents que ja han començat a fer els primers passos en l'adopció d'IA a la ciberseguretat.

5.5. Barreres per l'adopció de la IA en ciberseguretat a Catalunya

Fruit de l'experiència acumulada en la interacció de la IA en la cadena de la gestió de la ciberseguretat de les infraestructures i la informació, i conjuntament amb discussions i aportacions dels diferents professionals especialistes en IA i ciberseguretat que han contribuït a aquest estudi, és possible detectar un conjunt de barreres o limitacions que dificulten una adopció òptima de la IA per aplicacions de ciberseguretat.

Cal destacar que aquestes barreres no són només de naturalesa tècnica degut, per exemple, a la manca de maduresa de la tecnologia en algunes aplicacions o la complexitat de processament de grans volums de dades per a la generació d'alertes o presa de decisions en temps real, o la manca de qualitat de les mateixes, sinó que també són degudes a factors no tecnològics com pot ser la necessitat d'una adaptació permanent a noves amenaces i models d'atacs que varien constantment, obligant a què els models existents s'hagin de redefinir i re-entrenar per tal que les solucions basades en IA no esdevinguin obsoletes. I tot, sense deixar de banda l'exigent normativa i marc jurídic en aspectes de privadesa i ètica, o limitacions de tipus pressupostari que dificulten la disponibilitat d'infraestructures, entre altres.

Les barreres analitzades en aquest estudi s'agrupen en tres blocs. El primer aborda les limitacions associades al factor humà, el segon analitza barreres relacionades amb la tecnologia i el tercer tracta de limitacions d'origen organitzatiu i normatiu.

5.5.1. Factor humà

Aquest bloc inclou algunes barreres degudes al factor humà implicat en el procés d'adopció d'aproximacions IA al domini de la ciberseguretat:

- Manca de formació creuada entre disciplines IA i ciberseguretat. La manca de currículums especialitzats, suficient professorat expert o sufi-

cients instal·lacions i recursos formatius dedicats (del tipus Cyber Range, simuladors, CSIRT o similars) provoca la manca de professionals especialitzats en aquesta disciplina, circumstància global però que en el cas de Catalunya és especialment crítica degut a la impossibilitat de satisfer la demanda de les empreses. Addicionalment, els professionals que es volen dedicar a aquesta activitat en la majoria dels casos aprenen sobre el terreny en base a l'experiència i això resta eficiència a les empreses i organitzacions en veure's sotmeses a l'efecte de les necessàries corbes d'aprenentatge.

- Dificultat de retenció dels professionals. Actualment s'observa una elevada rotació en els perfils tecnològics i, en especial, en els experts en ciberseguretat. Donades les circumstàncies econòmiques actuals és freqüent que alguns professionals optin per millors oportunitats laborals que arriben de fora de Catalunya i que, en el cas específic de la IA, de la ciberseguretat i de la seva intersecció, aquesta rotació s'incrementi degut al factor teletreball.

- Expectatives sobre la capacitat de la IA en la resolució de problemes. Existeix un risc en la interpretació antropomòrfica de la IA, és a dir, atribuir-li unes capacitats similars a les humanes que, òbviament, no té. Això pot generar expectatives desmesurades sobre la capacitat de les màquines i sistemes que implementen IA en produir resultats que després no s'obtenen, o només parcialment. I, com a conseqüència, pot comportar certa desmotivació i pèrdua de credibilitat de les solucions IA, que només es poden combatre amb una formació rigorosa i una vigilància contínua de l'estat de l'art en aquest àmbit.

    - Continuant amb l'antropomorfisme, aquest pot portar a situacions justament contràries. Un excés de confiança en les expectatives de l'impacte positiu de la IA en la resolució dels problemes de ciberseguretat pot provocar una falsa sensació de seguretat i crear noves vulnerabilitats o riscos de ciberseguretat.

- Resistència al canvi de les persones i organitzacions vers l'impacte que pot suposar la incorporació de tecnologies com la IA en els processos, operacions, serveis i models competencials vigents, ja que la integració de solucions basades en IA pot derivar a canvis en responsabilitats, fluxos d'informació, formació, aparició d'altres funcions, etc. Això és una barrera que el sector de la ciberseguretat sempre ha tingut present, però amb l'arribada de la IA es veu amplificada, i s'hi ha de sumar la reticència del professional/especialista de la ciberseguretat a incloure eines basades en IA, amb poca transparència i explicabilitat en la majoria dels casos.

5.5.2. Tecnològiques

A continuació es destaquen algunes barreres de caire tecnològic:

- Manca de disponibilitat de dades de qualitat. Hi ha diverses casuístiques relacionades amb la qualitat de les dades. Les dades poden mancar perquè no tenen la mateixa freqüència de recol·lecció, poden incloure valors espuris que cal descartar, etc. Aquesta situació es pot veure agreujada si es tracta de dades confidencials relacionades amb infraestructures crítiques, amb la reticència de compartir les dades per part de les organitzacions, o la falta de garantia en l'anonimat del proveïdor de les dades. Per tant, cal extremar la curació de les dades i realitzar processos ETL adients per tal de poder alimentar a un algoritme o sistema IA de manera eficient.

- Manca d'homogeneïtat de les dades. El conjunt de dades de partida pot incloure dades qualitatives, quantitatives, categòriques, i on el procés de processament d'adequació no millora la seva qualitat, tot el contrari, simplifica la informació i en la majoria casos generen resultats no esperats, i a vegades pitjors que les clàssiques tècniques de ciberseguretat. A més, existeix el problema de la privadesa de les dades que es processen pels algorismes IA, que obliga a incloure etapes d'anonimització o xifrat, més encara en espais de compartició de dades *(Data Spaces)*.

- Explicabilitat de la IA. Una de les barreres més rellevants en l'adopció generalitzada de la IA és la poca o limitada confiança en les decisions preses degut a la poca transparència (caixa negra) i explicació dels resultats obtinguts. Saber perquè es prenen certes decisions i no altres ajuda a lluitar contra les amenaces i atacs i per tant ajuda també a millorar les contramesures que s'han d'implementar. L'explicabilitat de la IA és una línia de recerca molt important actualment.

- Manca d'entorns de prova realistes. Molt sovint la verificació dels sistemes i solucions IA es realitza en un entorn de laboratori amb dades sintètiques degut a la manca de prou volum de dades reals. En la majoria dels casos aquest nivell de verificació no és suficient en no ser equivalent a la infraestructura real on han d'operar.

- Detecció i mitigació d'entrades adversàries. Les entrades adversàries comprometen els sistemes de gestió de la ciberseguretat basats en IA. Les vulnerabilitats innates d'aquests sistemes, basats en l'aprenentatge i que no es poden actualitzar, estan sempre presents i s'han de tenir en compte (secció 3). Això incrementa el cost d'adopció d'aquest tipus de solucions, ja que és necessari un sistema de filtratge de les entrades adversàries.

- Escenari dinàmic i canviant. La naturalesa del sector de la ciberseguretat és canviant perquè els ciberdelinqüents i el seu entorn evolucionen constantment des de tot els punts de vista: millorant les seves infraestructures, adquirint amb agilitat noves tècniques o tàctiques avançades basades en IA, implementant noves estratègies, etc. Per altra banda el concepte de *Crime-as-a-Service*, i que cada vegada és més senzill realitzar accions malicioses (campanyes de *phishing* mitjançant ChatGPT), incrementa el nombre d'atacs. Tot això comporta que els sistemes de defensa implementats amb IA estan obligats a evolucionar dinàmicament per tal de poder respondre en tot moment al comportament canviant de l'atacant. Això implica un esforç important per als equips de defensa que no sempre es veu recompensat per resultats exitosos i pot conduir a un desencantament sobre l'efectivitat de la IA aplicada a la ciberseguretat.

5.5.3. Organitzatives i normatives

Com darrera classe de barreres o limitacions per a l'adopció de solucions i sistemes basats en IA, hi ha les que són de caire intern o normatiu i/o legal que una organització, ja sigui pública o privada, hauria de considerar.

- Dificultat per desplegar i validar solucions, ja que aquestes dues fases són difícils d'executar en un entorn operatiu, tenint en compte que s'han de validar davant a diferents atacs i proves, i això no és gens fàcil i assumint grans riscos. L'aparició del concepte bessó digital (*digital twin*) pot ajudar a salvar aquesta barrera.

- Dificultat i poca agilitat per la compartició de dades. Per tal de lluitar eficientment contra les amenaces de ciberseguretat és crucial la màxima compartició entre organismes i empreses de dades relacionades amb potencials atacs rebuts. Això permet analitzar les estratègies, tàctiques i tècniques emprades pels adversaris i facilitar el disseny i implantació de contramesures més efectives. Existeixen iniciatives sectorials en aquest sentit, sota la denominació genèrica d'ISAC (*Information Sharing and Analysis Center*), però encara no són prou madures i tenen un abast limitat. En canvi, la compartició de dades en altres fases del cicle d'atac no és tan crítica. Per exemple, l'etapa de detecció depèn molt de les característiques específiques de la infraestructura atacada, i com que cada infraestructura és diferent, les dades compartides no serien rellevants.

- Informació restringida del proveïdor i dades d'incidències. Aquesta barrera està clarament lligada a l'anterior amb una relació causa-efecte. Per qüestions reputacionals moltes organitzacions que reben atacs es resisteixen a compartir la informació sobre aquests incidents si no se'n pot garantir la confidencialitat. Mentre això no sigui possible de manera efectiva serà difícil disposar d'aquesta valuosa informació perquè la majoria d'organitzacions no són proactives a l'hora de posar de manifest les seves vulnerabilitats enfront dels atacs rebuts.

- Manca d'una estratègia clara per part de les diferents organitzacions respecte del paper positiu que la IA pot aportar per a millorar la seva ciberseguretat. Aquesta limitació està estretament relacionada amb la resistència al canvi, l'asimetria cognitiva[8] (és a dir diferents nivells de comprensió global i de pensament crític respecte la IA dins d'una organització) i per tant el desconeixement real de què pot aportar la IA, i què no pot resoldre. Com a conseqüència, si la presa de decisions amb relació a aquesta qüestió és sota la responsabilitat de persones

8 CEPS Task Force Report. Artificial Intelligence and Cybersecurity. https://www.ceps.eu/wp-content/uploads/2021/05/CEPS-TFR-Artificial-Intelligence-and-Cybersecurity.pdf

amb aquestes carències, pot derivar en a una adopció de la IA deficient o, senzillament, que s'ignori per complet.

- Manca de transparència i dificultat per garantir el comportament ètic dels models IA. La falta de transparència i explicabilitat dels sistemes IA, més concretament aquells basats en aprenentatge, és un hàndicap en el procés d'adopció i desplegament de solucions IA en seguretat. L'adopció de tecnologies IA té impacte en la robustesa i capacitat de resposta dels sistemes tecnològics implicats. Sorgeixen interrogants de caire ètic que poden frenar l'adopció de la IA. Respecte la robustesa, com es pot assegurar que el sistema es comportarà com s'espera?. Respecte la capacitat de resposta, on recau la responsabilitat de les decisions preses i resultats obtinguts quan s'utilitzen sistemes autònoms?. La manca de respostes adequades a aquests interrogants pot frenar l'adopció de la IA en el camp de la ciberseguretat.

- Incertesa respecte al risc legal i la responsabilitat civil associada als resultats fruit de decisions atribuïbles als sistemes i models d'IA incorporats als sistemes i processos de gestió de la seguretat d'infraestructures i sistemes d'informació, amb més incertesa i preocupació en aquelles infraestructures catalogades com a crítiques. A hores d'ara existeix una proposta de Llei Europea d'IA (*Artificial Intelligence Act*), on assigna a les aplicacions de la IA tres categories de risc, amb uns requeriments legals per a cada una de les categories.

Recomanacions per fomentar l'adopció de la IA en l'ecosistema de ciberseguretat

6

6.1. Accions per superar les barreres per a la implantació de la IA en la ciberseguretat

En aquest apartat s'identifiquen algunes accions que poden ajudar a superar les barreres identificades anteriorment. Cal tenir en compte que només es proporcionen idees ja que és un tema actualment encara en fase de recerca i en contínua evolució.

Pel que fa al factor humà, la manca de coneixement i experts en ambdós àmbits de la IA i ciberseguretat es pot superar creant itineraris de formació especialitzats on hi participin formadors professionals (locals i internacionals) d'ambdós sectors. Actualment, s'estan creant a Catalunya molts nous graus a l'àmbit de la IA que abasta un ampli espectre de disciplines (algorítmica, governança, negocis, administració d'empreses, etc.). En aquest sentit, caldria crear itineraris especialitzats en IA/ciberseguretat en graus universitaris o inclòs graus i màsters específics. Amb relació a la retenció de talent, aquesta només es pot aconseguir millorant tot l'ecosistema de l'entorn de treball. És a dir, no només és una qüestió de salari (que és clarament un factor motivant), si no que també cal oferir mesures complementàries tant per retenir el talent existent com atraure'n de nou. Això implica aconseguir un ambient de treball altament dinàmic, col·laboratiu, que fomenti la creativitat i el pensament lateral (*outside the box*), amb gran participació de tots els sectors, sobretot amb suport i empenta per part de l'administració pública. Finalment, aquest entorn ha de poder fomentar el coneixement i la participació ciutadana perquè s'eliminin les barreres sobre les pors de la IA i els recels al canvi de paradigma. Aquesta participació ha de poder ser transversal per a tots els sectors i segments, des d'iniciatives pensades des de i per a les famílies fins a mesures d'impacte empresarial.

A nivell tecnològic, la manca de dades homogènies i de qualitat és un factor altament limitant a l'hora de competir amb altres països, en especial amb relació a les dades personals. Aquesta barrera es pot superar creant un entorn col·laboratiu per poder compartir dades de forma segura i donant compliment a les normatives de privadesa existents. La creació d'aquest tipus d'entorn és, precisament, cap a on s'encamina l'estratègia europea, amb l'aposta per al desenvolupament d'espais de dades (*Data Spaces*). En aquest sentit, el repte és aconseguir contribuïdors a aquest entorn que comparteixin les seves dades per a un benefici col·lectiu, això sí, amb les degudes restriccions. Cal que hi hagi entorns col·laboratius en lloc de competitius, doncs només fomentant la col·laboració i la compartició de recursos es podran aconseguir resultats significatius davant d'un cibercrim que no para d'evolucionar i no coneix fronteres. En aquest sentit, caldria una iniciativa de coordinació a gran escala, entre sectors públic i privat, per a la compartició de dades i la realització de proves que permetin reforçar el desenvolupament de solucions d'IA per a la ciberseguretat.

Finalment, les barreres a nivell normatiu ja s'estan debatent a nivell europeu. Se'n parlarà en més detall a les seccions a continuació. D'una banda (Secció 6.3), s'està fomentant l'adopció d'un marc legal que assegura la privadesa de les dades (sobretot personals) però que alhora fomenti i agilitzi la compartició d'aquestes dades. Alhora, s'està discutint sobre la creació d'un marc de certificació de la IA en línia amb allò que ja s'ha fet per a la ciberseguretat. D'altra banda (Secció 6.4), s'està definint el marc ètic per aconseguir un equilibri just entre els avenços tecnològics, els interessos empresarials i una sèrie de garanties per a la ciutadania.

6.2. Accions de control/ mitigació pels riscos de seguretat de la IA

La IA es pot utilitzar per millorar significativament les mesures de ciberseguretat i defensa, i fomentar l'estabilitat del ciberespai. No obstant això, l'adopció de la IA va acompanyada d'importants reptes ètics i legals com s'ha indicat en la secció 5.5.3. En aquesta direcció, hi ha un consens generalitzat sobre la necessitat de discutir límits fiables en el desenvolupament de sistemes d'IA, perquè el seu ús pot tenir conseqüències negatives molt importants per a la vida de les persones o reproduir models socials que es consideren moralment reprovables. Els límits, però, no estan clars, i és difícil establir o acordar un marc ètic, polític o normatiu que pugui regular el desenvolupament de formes d'IA que després puguin tenir un gran impacte en les decisions socials. Una de les dificultats que sorgeix en aquest sentit és la tensió entre una sèrie de garanties per a la ciutadania i, alhora, la competitivitat en recerca i innovació.

La implementació d'una governança adequada de la IA esdevé fonamental per oferir aquest marc normatiu. La definició més completa diu que *"la governança de la IA és un sistema de regles, pràctiques, processos i eines tecnològiques que s'utilitzen per garantir que l'ús de les tecnologies d'IA s'alinea amb les estratègies, objectius i valors de l'organització, compleix els requisits legals i els principis d'IA ètica seguits per l'organització"*[1]. En poques paraules, la governança de la IA hauria de tancar l'escletxa que hi ha entre la responsabilitat i l'ètica en l'avenç tecnològic[2] i assegurar-se que s'estableixin límits fiables dins de la tecnologia, de manera que no perjudiqui i agreugi encara més les desigualtats mentre funciona. El marc legal i ètic es discuteix a les seccions 6.3 i 6.4, respectivament.

Un cop establerta una governança de la IA, tres dominis tenen una influència significativa a l'hora d'analitzar la funció de la IA en la ciberseguretat a nivell de sistemes: robustesa del sistema, resiliència del sistema i capacitat de resposta del sistema[3]. Cal abordar els tres dominis per tenir una solució global d'IA millorada i estable. Les regulacions haurien de vetllar per tal que la seguretat dels usuaris i del sistema es tinguin en compte durant tota la vida útil d'un producte d'IA[4]. Això demana l'ampliació dels processos de ciberseguretat actuals per incloure tècniques de desenvolupament de programari segur, certificacions de seguretat, auditories de seguretat i controls. El desenvolupament de nous estàndards i certificacions d'IA farà ús d'eines legals suggerides recentment o ja existents.

A part del riscos de caire ètic i de desenvolupament de programari, la introducció de solucions de ciberseguretat amb IA generen nous riscos de seguretat deguts a la pròpia naturalesa de la IA, tal i com s'ha indicat en seccions prèvies, tant a l'etapa de desenvolupament, com a l'etapa de desplegament i operativa. Llavors es poden establir un conjunt de recomanacions en dos grups, les més rellevant es mostren a continuació:

Recomanacions de mitigació en l'etapa d'entrenament (desenvolupament):

- Millorar la qualitat de les dades és una millora del model a la cadena de subministrament de les dades. S'ha de protegir la cadena de subministrament de les dades contra les manipulacions per frustrar atacs que alterin o degradin la qualitat de les dades. El risc vers la qualitat de les dades disminueix si la recollida es realitza en un entorn controlat.

- El sanejament de les dades té com objectiu detectar les dades que tenen un impacte negatiu en el rendiment del model i extreure'ls del conjunt d'entrenament, d'aquesta forma reduirem falsos positius i negatius.

- Amb l'objectiu d'evitar l'existència de *backdoors* i *triggerings* al model, s'introduiran mètodes de detecció basats en diferències estadístiques entre les entrades benignes i dels *triggers*.

1 Matti Mäntymäki, Matti Minkkinen, Teemu Birkstedt, Mika Viljanen, "Defining organizational AI governance", AI and Ethics, February 2022

2 KOSA AI, The importance of AI governance and 5 key principles for its guidance, https://kosa-ai.medium.com/the-importance-of-ai-governance-and-5-key-principles-for-its-guidance-219798c8f407, accessed on October 2022

3 Mariarosaria Taddeo, "Three Ethical Challenges of Applications of Artificial Intelligence in Cybersecurity", Minds & Machines, vol. 29, pp. 187–191, June 2019

4 Ronan Hamon, Henrik Junklewitz, Ignacio Sanchez, "Robustness and Explainability of Artificial Intelligence", JRC Technical Report, 2020, https://publications.jrc.ec.europa.eu/repository/handle/JRC119336 , accessed on November 2022

- Si un *trigger* és detectat a partir d'una mostra de dades d'inferència determinada, la mostra de dades pot tractar-se de forma aïllada, on el *trigger* es pot descartar o purificar amb la introducció de soroll, o directament l'eliminació del *trigger*.

Recomanacions de mitigació en l'etapa d'inferència (desplegament i operativa):

- Els atacs d'evasió impliquen dues fases, manipular les entrades i presentar l'entrada manipulada al model. Les mitigacions existents contra aquests atacs d'evasió en incrementar el nombre d'exemples adversos a la fase de desenvolupament i en detectar si una entrada es un exemple advers a l'etapa de desplegament.

- El robatori de models es pot materialitzar mitjançant dos tipus d'atac: obtenint directament els paràmetres del model gràcies a les vulnerabilitats del sistema, o mitjançant atacs de canal lateral. Les mitigacions per aquest tipus d'atac es focalitzen en assegurar els paràmetres e hiperparàmetres del model, detectar consultes no-normals al model en mode operatiu, l'ofuscació de la sortida de la inferència i la gestió de la propietat intel·lectual.

- La tercera recomanació va dirigida a mitigar els atacs d'extracció de dades (atac d'inversió del model, i atacs d'inferència). Les mitigacions necessàries per a aquest tipus d'atac es centren en la integració de la privacitat de les dades, l'entrenament amb privacitat i la limitació de la informació filtrada.

Les recomanacions descrites se centren en mitigar aquells riscos tècnics que aporta la integració d'IA al domini de la ciberseguretat. Altres riscos que es poden trobar per la seguretat de la IA poden estar relacionats amb factors humans, tal i com s'ha comentat a la secció 5.5.1, deguts a falta de coneixement i capacitats de ciberseguretat dels experts en disseny i desenvolupament de sistemes IA, i la falta de coneixement i capacitats dels especialistes de ciberseguretat amb relació a la IA. Aquesta mancança de capacitats pot derivar en eines i sistemes de ciberseguretat basats amb IA de baix rendiment, o que resultin més un problema o inconvenient que una ajuda. Encara que ja s'ha comentat, per reduir aquest risc, s'haurien de formar adequadament als equips.

6.3. Marc regulatori

Els sistemes d'IA entren en l'àmbit d'aplicació de la Llei de Ciberseguretat (*Cybersecurity Act*, CSA), proposada per la Comissió Europea el 2017 i acceptada el 2019. En breu, la Llei de Ciberseguretat estableix el primer marc de certificació de ciberseguretat a tota la UE per garantir un enfocament comú de certificació de ciberseguretat al mercat interior europeu i, en definitiva, millorar la ciberseguretat en una àmplia gamma de productes, processos i serveis digitals (per exemple, Internet de les coses)[5].

Aquesta llei CSA implica que les empreses que fan negocis a la UE es beneficiaran d'haver de certificar els seus productes, processos i serveis TIC només una vegada i veure els seus certificats reconeguts a tota la Unió Europea. De fet, un dels principals objectius és incrementar la competitivitat i el creixement de les empreses europees, per passar de ser importadors de ciberseguretat a exportadors de ciberseguretat[6]. Actualment, ENISA ha publicat el primer esquema europeu de ciberseguretat per a productes TIC segons les directrius del CSA anomenat EUCC 1.1.1 (*European Candidate Cybersecurity Certification Scheme* – Esquema de certificació ciberseguretat candidat europeu basat en criteris comuns)[7].

El següent pas de la UE ha estat proposar la *Cyber Resilience Act*[8]. En aquest cas l'interès és regular els requeriments de seguretat de productes maquinari i programari. Els objectius en concret són dos: i) crear les condicions per al desenvolupament de productes segurs garantint que es comercialitzin amb menys vulnerabilitats i que els fabricants es prenguin de debò la seguretat al llarg del cicle de vida d'un producte; ii) crear les condicions que permetin als usuaris tenir en compte la ciberseguretat en seleccionar i utilitzar productes amb elements digitals

5 European Commission, "Proposal for a regulation on ENISA, the EU Cybersecurity Agency, and repealing Regulation (EU) 526/2013, and on Information and Communication Technology cybersecurity certification", COM(2017) 477 final, September 2017, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2017:477:FIN

6 https://www.jtsec.es/es/entrada-blog/90/la-certificacion-de-ciberseguridad-en-europa-2-anos-del-cybersecurity-act

7 https://www.enisa.europa.eu/publications/cybersecurity-certification-eucc-candidate-scheme-v1-1.1

8 https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act

La Comissió Europea també construeix un enfocament europeu sòlid de la IA, basat en l'estratègia de 2018[9] i reforçat pel Llibre Blanc de 2020 sobre la IA[10]. En aquest sentit, és probable que la UE sigui la primera a promulgar la legislació reguladora d'IA[11]. És clar que qualsevol llei que regula la IA afecta l'ús de la IA a la ciberseguretat. L'*Algorithmic Accountability Act*[12] dels EUA exigeix les grans empreses amb accés a grans quantitats de dades auditar els sistemes basats en IA per a la justícia, la privadesa, la precisió i els riscos de seguretat. Una iniciativa destacada és el Marc de Governança de la IA de Singapur[13]. És el primer model desenvolupat a Àsia i la seva força és que tradueix els principis en un marc pràctic i operatiu per a una acció immediata, disminuint les barreres d'entrada a l'adopció de la IA. Aquest marc es basa en dos factors: i) les solucions d'IA han de centrar-se en l'ésser humà, i ii) les decisions preses o assistides per la IA han de ser transparents, explicables i justes.

Recentment, la UE ha reconegut que existeixen nombroses maneres possibles en què els usos relacionats amb la cibernètica de les tecnologies disruptives emergents (EDT) podrien afectar profundament la imatge de seguretat a Europa. La *Cybersecurity Strategy for the Digital Decade* (Estratègia de Ciberseguretat de la UE per la Dècada Digital) de la UE, ja ha destacat des de 2020 que tecnologies importants com la IA, l'encriptació, la computació quàntica i les xarxes de nova generació són crítiques per a la ciberseguretat. El pla destaca la importància d'incorporar fonaments de ciberseguretat a tots els dispositius digitals connectats als dominis tecnològics esmentats i la integritat i convergència de les seves cadenes de subministrament. Les conclusions del consell sobre el desenvolupament de la postura cibernètica de la Unió Europea a partir del maig de 2022 subratllen "la importància de fer un ús intensiu i d'integrar les noves tecnologies, en particular la informàtica quàntica, la IA i el *Big Data*, per aconseguir avantatges comparatius en termes de ciberseguretat".

En línia amb el que s'ha fet amb la CSA, la UE pot establir un marc de certificació sobre l'ús i l'aplicabilitat de la IA a Europa. En aquesta situació, un programa de certificació de la IA es podria construir sobre dos pilars[14]: en primer lloc, una anàlisi exhaustiva dels riscos dels efectes de la IA en el seu entorn, així com dels perills que representen els possibles adversaris i les debilitats actuals. En segon lloc, un examen exhaustiu de la transparència dels sistemes d'IA, una avaluació del seu bon funcionament en escenaris de punta i l'explicabilitat.

D'altra banda, el Reglament General de Protecció de Dades[15] (GDPR en les seves sigles en anglès) és un reglament de la legislació de la UE sobre protecció de dades i privadesa a la UE i l'Espai Econòmic Europeu (EEE). L'objectiu principal del GDPR és millorar el control i els drets de les persones sobre les seves dades personals i simplificar l'entorn regulador per als negocis internacionals i, per tant, la transferència de dades personals fora de les àrees de la UE i l'EEE. El GDPR és important en aquest context perquè introdueix les *Data Protection Impact Assessments* (DPIA), que inclouen un conjunt d'eines per avaluar els riscos associats a l'ús de processos automatitzats de presa de decisions o de generació de perfils[16]. La DPIA es pot utilitzar per determinar i posar en marxa les salvaguardes necessàries per gestionar eficaçment els riscos associats i proporcionar l'explicació i la transparència necessàries dels algorismes i dades utilitzats. El GDPR és una barrera per a l'ús de Big Data i l'avaluació de riscs i protecció per part de la IA ja que, como s'ha comentat anteriorment, aquesta necessita dades com ara les ac-

9 European Commission, "Artificial intelligence for Europe", COM(2018) 237 final, Brussels, April 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN

10 European Commission, White paper on artificial intelligence - A European approach to excellence and trust, COM(2020) 65 final, Brussels, February 2020, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

11 European Commission, Proposal for a Regulation laying down harmonised rules on artificial intelligence, 2021/0106 (COD) and Annexes (2021). https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

12 US Congress, Algorithmic Accountability Act of 2019, H.R.2231, 116th Congress, April 2019, https://www.congress.gov/bill/116th-congress/house-bill/2231, accessed on August 2022.

13 https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework

14 Ronan Hamon, Henrik Junklewitz, Ignacio Sanchez, "Robustness and Explainability of Artificial Intelligence", JRC Technical Report, 2020, https://publications.jrc.ec.europa.eu/repository/handle/JRC119336 , accessed on November 2022.

15 https://gdpr-info.eu/

16 European Commission, "Regulation 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 (General Data Protection Regulation)", Official Journal of the European Union, L119, April 2016, https://eur-lex.europa.eu/eli/reg/2016/679/oj

tivitats dels usuaris, informació de dades personals (direcció IP, geolocalització, etc.).

Per agilitzar i regular aleshores l'accés a les dades, l'EU està treballant actualment en dues iniciatives noves, el *Data Act*[17] i el *Data Governance Act*[18]. La idea del *Data Act* és eliminar les barreres d'accés a les dades no personals, tant per als organismes del sector públic com privat, alhora que es preserva els incentius per invertir en la generació de dades assegurant un control equilibrat de les dades per als seus creadors. Per tant, la Llei de dades amplia el dret del GDPR a la portabilitat a les dades no personals generades per productes connectats i serveis relacionats i facilita l'intercanvi de dades i l'ús/reutilització de les dades generades per part dels usuaris i de tercers seleccionats, establint estàndards a tota la UE.

En canvi, la *Data Governance Act* vol fomentar l'àmplia reutilització de les dades en poder dels organismes del sector públic i preveu un marc legal quan els organismes del sector públic decideixen posar les dades (incloses les dades personals) disponibles per a la seva reutilització a tercers. Per incentivar les persones a compartir les seves dades, la UE proposa crear "intermediaris de dades" (com a alternativa europea a les principals plataformes tecnològiques existents), que s'encarregaran de l'intercanvi de dades per part de particulars, organismes públics i empreses privades. Ambdues, situen els Espais de Dades com un element clau en la compartició de grans volums de dades per alimentar la IA en ciberseguretat.

Finalment, com ja s'ha indicat a l'inici del document, el Parlament Europeu va aprovar el novembre de 2022 l'actualització de la Directiva sobre Ciberseguretat[19] (coneguda com a NIS2.0, per les sigles de *Network and Information Security*) que substitueix l'anterior Directiva NIS (Directiva 2016/1148/CE). NIS2.0 estableix les mesures de gestió del risc de ciberseguretat i les obligacions d'informació en tots els sectors que estan coberts per la directiva, com ara l'energia, el transport, la salut i la infraestructura digital. La directiva revisada té com a objectiu eliminar les divergències en els requisits de ciberseguretat i en la implementació de les mesures de ciberseguretat en els diferents estats membres, establint normes mínimes i establint mecanismes per a una cooperació efectiva entre les autoritats pertinents.

A nivell espanyol, el Consell de Ministres va aprovar el desembre de 2021 la nova *Estrategia de Seguridad Nacional* (ESN) 2021[20]. Aquesta estratègia pretén desenvolupar el sistema de seguretat espanyol per a la gestió de crisis de naturalesa polièdrica amb polítiques preventives. En aquest sentit, preveu utilitzar sistemes d'alerta primerenca basats en dades i indicadors, a través de l'aplicació de noves tecnologies, com ara l'IA. D'altra banda, el Consell de Ministre també va aprovar el novembre de 2020) l'Estratègia Nacional d'IA (ENIA)[21]. Un dels eixos principals és impulsar accions en l'àmbit de les plataformes de dades, models, algorismes, motors d'inferència i ciberseguretat per catalitzar la investigació i la innovació orientada a millorar la seguretat, la certificació, l'eficiència i la governança de les dades oberts i tancats i la IA. En aquest sentit, es vol fomentar la conscienciació de tots els agents implicats, administració pública, empresa i ciutadania, i una bona coordinació entre tots els agents que intervenen en la protecció de la xarxa, des dels centres de ciberseguretat de l'Estat i les comunitats autònomes fins a l'INCIBE, encarregat de promoure la ciberseguretat de les empreses, i els corresponents homòlegs europeus i a tercers països.

A nivell català, el 2019, es va publicar al DOGC, l'Estratègia de Ciberseguretat de la Generalitat de Catalunya 2019-2022[22]. Aquest document defineix l'estratègia a seguir pel Govern català pel que fa a la ciberseguretat i dins l'àmbit de les Administracions Públiques catalanes per als anys 2019-2022, el seu sector públic, la ciutadania i aquelles entitats que, en qüestions de relacions administratives i/o comercials es relacionen amb i/o tracten la informació titularitat de la Generalitat de Catalunya. Es destaca en aquest document el propòsit de fomentar la innovació en la ciberseguretat i que aquesta ha de donar suport i seguretat a la irrupció de noves TIC, com el *blockchain*, les comunicacions 5G, la IA

17 https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113

18 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52020PC0767

19 https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52020PC0823&from=EN

20 Estrategia de Seguridad Nacional 2021, https://www.dsn.gob.es/es/documento/estrategia-seguridad-nacional-2021

21 Estrategia Nacional de Inteligencia Artificial 2020, https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/201202_ENIA_V1_0_Resumen_ejecutivo.pdf

22 https://ciberseguretat.gencat.cat/web/.content/PDF/EstrategiaCiberseguretat-de-Catalunya-2019.2022.pdf

i la proliferació de la IoT. Per tant, el conjunt de les administracions públiques de Catalunya té el repte d'impulsar la innovació en matèria de ciberseguretat i fomentar el desenvolupament d'un sector de la ciberseguretat català que sostingui el creixement de les TIC i, al mateix temps, doni suport a l'activitat econòmica del teixit empresarial. UUna d'aquestes iniciatives ha estat la fundació de l'Agència de Ciberseguretat de Catalunya[23] que es va posar en marxa l'1 de gener del 2020. L'organisme s'encarrega de vetllar per la seguretat digital i de desplegar l'Estratègia de Ciberseguretat de la Generalitat de Catalunya, dins l'àmbit de les competències de la Generalitat. D'altra banda, com ja comentat en la secció 5.3.3, la Generalitat de Catalunya va impulsar l'estratègia CATALONIA.AI el 2019 amb l'objectiu de convertir Catalunya en un pol tecnològic reconegut en tecnologia d'IA. Aquesta estratègia destaca pel propòsit: i) de proporcionar directives ètiques per a una IA fiable, molt en línia amb el que es comenta a continuació a la secció 6.4; ii) fomentar la creació de talent a la IA, inclòs el cas específic de formació en IA aplicada a ciberseguretat; iii) potenciar el rol de la dona a la IA.

23 https://ciberseguretat.gencat.cat/ca/inici

6.4. Marc ètic

Les mesures de governança proactiva són cada cop més reconegudes com una característica diferenciadora per a les empreses que busquen establir una reputació de fiabilitat. Hi ha una sèrie de marcs mundials sobre la governança de l'IA i els conceptes d'ètica. Per exemple, el marc ètic de la UNESCO[24] proporciona en la seva resolució la base per fer que els sistemes d'IA funcionin pel bé de la humanitat, les persones, les societats, el medi ambient i els ecosistemes, i per prevenir els danys. El GDPR de la EU comentat en la secció anterior inclou un conjunt especial de normes que es relacionen amb el dret del consumidor a una explicació quan les empreses utilitzen algorismes per prendre decisions automatitzades. No obstant això, també va atreure algunes controvèrsies, ja que no ofereix dret a una explicació de la presa de decisions automatitzada[25].

En resum, per ser eficaç i oferir el correcte compromís entre les estratègies i els objectius de l'empresa i els requisits legals i l'ètica, molts actors treballen per identificar els principis fonamentals. Per exemple, la Universitat de Harvard[26] va crear un mapa de visualització de 32 conjunts de principis d'IA. Com a resum de totes aquestes aportacions al marc ètic, es consideren a continuació un conjunt de sis principis per la governança de la IA que són funcionalment independents de l'algoritme, de la tecnologia i del sector (i, per claredat, es manté la terminologia anglesa amb què de facto s'identifiquen habitualment):

- *Accountability*: la rendició de comptes requereix una identificació clara de qui té la responsabilitat de les decisions i les accions en dissenyar, desenvolupar, operar i/o desplegar el sistema d'IA. Donar la responsabilitat a una màquina és

24 UNESCO, Recommendation on the ethics of artificial intelligence, SHS/BIO/REC-AIETHICS, 2021, https://unesdoc.unesco.org/ark:/48223/pf0000380455, accessed on November 2022

25 Sandra Wachter, Brent Mittelstadt, Lciano Floridi, "Why a right to explanation of automated decision-making does not exist in the general data protection regulation", International Data Privacy Law, vol. 7, no. 2, pp. 76–99, May 2017.

26 Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, Madhulika Srikumar, Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI, Berkman Klein Center Research Publication No. 2020-1, February 2020.

innecessari (sempre hi haurà una persona física o corporació responsable dins de les lleis i els marcs legals existents), immoral (la responsabilitat és una propietat intrínsecament humana), poc pràctic (és impossible responsabilitzar les màquines per les violacions dels seus drets i obligacions) i oberta a l'abús (facilitaria que les persones dolentes es protegissin). Han de ser persones o organitzacions les que en última instància siguin responsables dels actes dels sistemes d'IA sota el seu disseny o control, per molt complex que sigui. En aquest sentit, es necessari tenir normes, estàndards, regles o lleis perquè la rendició de comptes s'apliquin de manera uniforme.

- *Transparency:* La transparència fa referència a la capacitat d'explicar per què un sistema d'IA es comporta d'una determinada manera per augmentar la confiança de les persones en l'exactitud i l'adequació de les seves prediccions. A l'hora de considerar quins graus d'explicació són adequats, cal tenir en compte els criteris que aplicaria una persona per prendre una decisió en la mateixa situació. És important considerar quin tipus d'explicació és més adequada en un context determinat: diferents públics requereixen necessitats diferents. Definitivament, com més els usuaris sentin que entenen el sistema d'IA en general, més inclinats i més equipats estaran per utilitzar-lo.

- *Fairness:* Hi ha diversos debats sobre què és justícia. Tanmateix, cal fer un judici global sobre els valors compartits i les normes morals comunes, i també s'han de transmetre als sistemes d'IA. El principi d'equitat també garanteix que es tingui en compte i es redueixi qualsevol possible biaix en les dades històriques o biaix induït per humans. En resum, l'equitat ha de garantir que els sistemes d'IA siguin ètics, lliures de biaix, lliures de prejudicis i que no s'utilitzen atributs protegits. No obstant això, hi ha moltes situacions conflictives en què és difícil complir amb l'equitat amb la satisfacció general, ja sigui que les decisions les prenguin persones o màquines. Aquest problema es pot mitigar parcialment si quan es desenvolupa una eina d'IA per ajudar a la presa de decisions, és vital decidir la tècnica exacta d'equitat que cal utilitzar per endavant i fer-les transparents. Com que hi ha tants punts de vista i tècniques diferents per definir l'equitat, algunes definicions poden contradir-se directament, mentre que altres poden fomentar la justícia al preu de la precisió o l'eficiència. Tanmateix, si es fa correctament, un mètode algorítmic pot ajudar a millorar la coherència en la presa de decisions, sobretot quan es contrasta amb l'alternativa dels humans que decideixen basant-se en les seves pròpies idees personals (i, per tant, probablement variables) d'equitat.

- *Safety:* Pel que fa a la seguretat, és fonamental prendre mesures contra l'abús inadvertit i intencionat de la IA que suposa una amenaça per a la seguretat humana. Tanmateix, això s'ha de fer d'una manera raonable, tenint en compte el potencial de dany i la practicitat de les mesures preventives suggerides pel que fa als factors tecnològics, legals, econòmics i culturals. Els problemes de seguretat també poden estar relacionats amb accidents o per una explotació de seguretat i un atac maliciós intencionat. Per exemple, els sistemes d'IA aprenen en temps real en un entorn del món real i què passa si les dades estan malmeses? La IA és capaç de reaccionar i mitigar els seus impactes? Un altre exemple si com podem assegurar-nos que si les dades d'entrenament són incompletes i ometen determinats detalls importants, o si les característiques importants del món han canviat després de l'adquisició de les dades d'entrenament? Cal abordar aquest tipus d'aspectes i provar i validar el mecanisme adequat.

- *Human control:* hi ha la necessitat de tenir un control humà de la tecnologia, és a dir, que les persones han d'estar en un o més punts del procés de presa de decisions d'un sistema automatitzat. Els humans i els sistemes d'IA tenen punts forts i defectes al final. L'elecció de la combinació més raonable requereix una avaluació integral de la millor manera de garantir que es faci una selecció acceptable en les condicions actuals. No obstant això, prendre aquesta determinació no és fàcil. Per exemple, mentre que la major part de l'atenció s'ha centrat en els perills dels sistemes d'IA mal dissenyats i implementats que tinguin un biaix integrat, els mateixos perills existeixen per a les persones. D'altra banda, podem tenir la síndrome de "ho ha dit l'ordinador" (*computer says yes syndrome*)[27], on els

27 Kevin Hoff and Masooda Bashir, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust", Human Factors - The Journal of the Human Factors and Ergonomics Society, vol. 57, no. 3, p. 407-4, May 2015.

empleats que han passat molt de temps tractant amb un sistema on els errors són poc freqüents (com hauria de ser el cas dels sistemes d'IA comercials) tenen menys probabilitats de desafiar la correcció del sistema amb el temps. A mesura que la tecnologia d'IA ha avançat, s'ha convertit en una eina sòlida per detectar informació problemàtica ràpidament i a escala. Tanmateix, els algorismes basats en IA continuen cometent nombrosos errors en treballs sensibles al context, per això és fonamental mantenir una persona al corrent mentre revisa la nova informació. Aquest aspecte humà manté la responsabilitat alhora que detecta errors i produeix millors dades d'entrenament, donant lloc a un model millor per a futures iteracions. Definitivament, independentment de la precisió d'un sistema d'IA o dels beneficis en temps/cost de l'automatització completa, gairebé sempre hi haurà situacions delicades en què la societat vol que un humà faci el judici final.

- *Universality:* El principi d'universalitat recomana la definició i aplicació d'estàndards tècnics, ètics i reglamentaris durant el desenvolupament, l'avaluació i el desplegament d'algorismes per tal de disposar d'interoperabilitat, cooperació i un determinat nivell de qualitat, seguretat i confiança. L'emissió d'una col·lecció de documents per evitar i minimitzar els riscos de fallades mitjançant l'estandardització és una tècnica potent per reduir els riscos associats a l'ús de la IA als sistemes. A més, els processos de certificació compleixen els requisits. Un exemple pertinent del que es pot aconseguir amb la IA són les bones pràctiques recentment publicades per ENISA per a la seguretat de diferents sistemes ciberfísics, com ara xarxes IoT o automòbils intel·ligents[28].

28 ENISA publication on emerging technology, https://www.enisa.europa.eu/topics/iot-and-smart-infrastructures/?tab=publications

Conclusions

7

L'entusiasme actual al voltant de la IA està recolzat aquesta vegada, com hem vist, per tres grans progressos tecnològics: 1) els ordinadors són molts més potents; 2) la disponibilitat de grans quantitats de dades que inclouen comerç electrònic, empreses, xarxes socials, ciència i govern; 3) i models intel·ligents basats en aprenentatge automàtic millorats.

Pel que fa a la ciberseguretat, la IA ha esdevingut una tecnologia que, sota certes circumstàncies, es mostra molt efectiva a l'hora de protegir les infraestructures i actius digitals de les empreses i organitzacions en front dels ciberatacs. Però els atacants també tenen accés a eines que utilitzen la IA, de manera que el joc del gat i el ratolí continua. De fet, segons diverses fonts, el cibercrim és en constant augment. Com més la nostra vida depengui del món digital (tant a nivell social, com de govern o empresarial), més complexes són les infraestructures, més dades es generen i clarament més complicat és protegir-les. Per tant, és prioritari donar algunes claus per al foment de l'adopció i la integració de solucions i sistemes basats en algoritmes i tecnologies d'IA a les diferents capes i processos relatius als sistemes de ciberseguretat.

La IA es pot fer servir des d'una estratègia defensiva. Per exemple una IA pot reconèixer patrons de comportaments tant de les persones com dels dispositius i detectar desviacions anòmales amb tal precisió que es poden minimitzar els falsos positius i negatius. Però la IA també pot utilitzar-se en ciberatacs per passar inadvertits, bé per maximitzar l'ofensiva amb atacs massius automàtics o bé generar *deep fakes*. Addicionalment, la IA comporta uns riscos de seguretat inherents, intrínsecs als algorismes d'aprenentatge, als models utilitzats i al seu procés de construcció. En aquest context, cal destacar l'augment de la complexitat per defensar-se dels ciberatacs, sobretot d'aquells que es basen en models d'IA, derivant en una necessitat de grans volums de dades per entrenar una IA, per obtenir un model que ens ajudi en les diferents etapes de defensa.

Els beneficis que ofereix l'ús d'eines i solucions basades en IA beneficien la identificació, la protecció, la detecció, la resposta i la recuperació, millorant la gestió de la seguretat en totes les seves fases. El seu impacte és avantatjós sobre persones, tecnologies i processos, encara que s'ha de continuar progressant en tots els aspectes de la tecnologia IA. Resumidament, la IA té la capacitat de simplificar qualsevol operació, de manera que pot facilitar abastament accions clau com la presa de decisions, la interpretació de les alarmes, l'automatització de tasques, la definició d'estratègies, l'anàlisi forense, etc.

Amb relació amb la situació de la IA a l'àmbit de la ciberseguretat a Catalunya, podem considerar que Catalunya ocupa una posició capdavantera a Europa degut a factors com disposar d'un ecosistema consolidat de generació de coneixement, amb generació de talent i desenvolupament tecnològics, també d'un ecosistema digital d'emprenedoria, amb capacitat d'atracció d'inversions, a part de ser la seu de grans esdeveniments com el MWC, comptar amb infraestructures tecnològiques capdavanteres i ser un centre de lideratge en iniciatives i en innovació. Tot i així, en contra, també es presenten diverses barreres. La més evident és la manca de professionals amb el coneixement conjunt de la IA i la ciberseguretat, degut principalment a que no hi ha una formació especialitzada en la intersecció dels dos dominis. Altra barrera important del procés d'integració de la IA en la gestió de seguretat és la falta de dades de qualitat, homogènies i en grans quantitats i la necessitat d'entorns d'entrenament (per exemple *bessons digitals*) per simular i validar models en condicions realistes. En aquesta línia, les limitacions a l'hora de tractar dades personals compliquen més aquesta barrera.

Per tal de donar resposta a aquestes barreres per a la implantació de la IA en la ciberseguretat, calen nous itineraris formatius que incloguin les dues disciplines, així com desenvolupar entorns per a l'accés a dades per a entrenament respectuosos amb la normativa en matèria de protecció de dades, com espais de dades (*Data Spaces)*. En aquest sentit, l'estratègia de la UE vol reduir o eliminar les barreres d'accés a les dades, definint normatives i regulacions, com la *Data Act* i el *Data Governance Act*.

Es pot concloure que Catalunya es disposa d'un potencial real per treballar i avançar en l'aplicació de la IA a la ciberseguretat, en particular les capacitats provinents del món acadèmic i de la recerca, un ecosistema empresarial i de *startups* que dissenyen productes i ofereixen serveis de valor, la tasca d'innovació en suport a les empreses que desenvolupen els centres tecnològics i les polítiques de suport institucionals. Caldrà, doncs, alinear tots aquests eixos per tal d'assolir aquest objectiu en un moment oportú com l'actual.

A tot això, com a reflexió final, cal fer una menció especial a la recent posada en marxa del ChatGPT, el passat 30 de novembre de 2022, esdeveniment que marcarà un abans i un després en l'ús de la IA en la societat i, també, en l'àmbit de la ciberseguretat. Tanmateix, el ChatGPT només és la punta de l'iceberg de l'autèntica revolució de la IA que està per venir.

Autoria i agraïments

Aquest document ha estat impulsat per l'Agència de Ciberseguretat de Catalunya (ACC) i realitzat en col·laboració amb el CIDAI, Center of Innovation for Data Tech and Artificial Intelligence, entitat definida a l'estratègia CATALONIA.AI i que té com a missió fomentar i accelerar l'adopció de tecnologies innovadores d'explotació de dades i IA a Catalunya.

El CIDAI és una associació coordinada per Eurecat, i integrada per la Generalitat de Catalunya, l'Ajuntament de Barcelona, BSC, CVC, Fundació i2CAT, IDEAI-UPC i les empreses Huawei, NTT Data, Microsoft, SAP i SDG Group.

El Llibre Blanc sobre l'aplicació de la IA en l'àmbit de la ciberseguretat s'ha elaborat amb la valuosa contribució de diversos experts del sector, la qual cosa ha permès copsar i analitzar la realitat del sector a Catalunya. CIDAI agraeix la seva dedicació i suport en la redacció d'aquest document. La participació ha estat la següent:

Redactors
- Fundació EURECAT: Mario Reyes de los Mozos.
- Fundació i2CAT: Abel Pozo, Albert Calvo, Nil Ortiz
- IDEAI-UPC: Davide Careglio, Francisco Javier Hernando Pericás, Karina Gibert

Contribuïdors
- Fundació Eurecat: Joan Mas
- Fundació i2CAT: David Company, Daniel Rodríguez, Xavi Galcerán

Revisors
- CIDAI: Joan Mas i Marco Orellana
- Fundació i2CAT: Jordi Guijarro, Àngel Martin
- Agència de Ciberseguretat de Catalunya: Santi Romeu, Isart Canyameres

Assessors
- Fundació Eurecat: Antonio Antón
- UPC-IDEAI: Víctor García Domínguez
- Agència de Ciberseguretat de Catalunya: Tomàs Roy

Annex. Presentació de casos d'ús il·lustratius

9

En el present estudi s'han identificat un conjunt de 10 casos il·lustratius de projectes destacables promoguts per agents catalans en l'aplicació d'IA a l'àmbit de la ciberseguretat. Per descriure els diferents casos d'ús hem de ser conscients del tipus de projecte que mostrar, i la confidencialitat associada segons la seva tipologia, per tant, per a cadascun dels casos d'aplicació s'identifiquen els següents aspectes (amb lleugeres modificacions segons la naturalesa del projecte):

1. REPTE: Quin és el repte que vol solucionar?
2. SOLUCIÓ: En què consisteix la innovació (projecte/solució)? Quin és l'objectiu del projecte?
3. PERFIL TECNOLÒGIC: Com s'aplica la IA en el cas? Com funciona la solució proposada?
4. DURACIÓ: Quina és la durada del projecte?, indicar si ha finalitzat o està en curs. No cal posar data de començament ni finalització (si es vol posar, cap problema)
5. TIPUS DE PROJECTE: Innovació/Recerca/...
6. TIPUS DE FINANÇAMENT: Públic o privat. En cas que sigui públic, àmbit del finançament (autonòmic, nacional, europeu), i si és considera interessant el número de projecte.
7. IMPACTE: etapa de la gestió de la seguretat on impacta el resultat (segons l'esquema ECSO), quin impacte real té el resultat assolit?
8. BENEFICIS DERIVATS: Quines millores aporta pel sector de la ciberseguretat? Quins aprenentatges es poden concloure del cas d'ús presentat?
9. GRAU DE MADURESA: S'indica el nivell de maduresa de la tecnologia en base al TRL (*Technology Readiness Level*)
10. LÍNIES DE FUTUR: Quin és el potencial de creixement i escalat del cas il·lustratiu? Quines són les futures línies estratègiques dels impulsors?
11. CONTACTE: Identificació del web, correu-ei telèfon de contacte

Cadascun dels casos presentats s'ha classificat amb relació a quins dels reptes principals de la ciberseguretat té el potencial a donar resposta. Degut a la criticitat on les solucions presentades han estat validades o desplegades, i que alguns dels casos d'ús estan sota contractes de confidencialitat, no es poden enumerar de forma explícita les empreses i organitzacions implicades, i la descripció es centra en la solució i aportació tecnològica d'algorismes i tècniques IA.

SPARTA, Strategic Program for Advanced Research and Technology in Europe	
REPTE	Ajudar als equips i especialistes de contra-intel·ligència en el procés d'atribució d'atacs i incidències, desenvolupant una eina que permeti automatitzar algunes de les etapes del procés d'atribució, com pot ser l'anàlisi de models d'atacs i la identificació de comportaments similars. Projecte centrat en desenvolupar coneixement i capacitats al voltant de la Intel·ligència Col·laborativa sobre Amenaces (CTI, *Collaborative Threat Intelligence*)
SOLUCIÓ	L'objectiu és explotar la informació que és emmagatzemada a la plataforma MISP amb relació a atacs i incidències, on aquesta informació s'expressa seguint el estàndard MITRE ATT&CK. Aplicant algorismes i tècniques d'IA, es pretén detectar comportament similars que puguin ajudar a detectar atacs o variants d'atacs que es poden atribuir a grups maliciosos. La solució s'ha validat a nivell de laboratori contra dades de grups maliciosos coneguts.
PERFIL TECNOLÒGIC	En aquest cas, els algorismes d'IA s'apliquen sobre les dades MITRE ATT&CK explotant aquesta informació, i integrant un procés de cerca de patrons de comportament similars, generant coneixement valuós pels equips d'especialistes. L'eina s'integra amb la plataforma MISP, per una banda utilitzant dades de la plataforma, i posteriorment, emmagatzemant els resultats obtinguts per ser consultats i utilitzats pels equips i especialistes en ciber-intel·ligència.
DURACIÓ	Projecte finalitzat el juny de 2022, amb una duració de 3 anys.
TIPUS DE PROJECTE	Recerca, amb un consorci d'empreses privades, i centres de recerca públics: Indra, Vicomtech, Tecnalia, Eurecat, etc. (https://sparta.eu/partners/).
TIPUS DE FINANÇAMENT	Públic, amb finançament d'EU per la convocatòria H2020.
IMPACTE	Aquest cas d'ús té un impacte directe a la darrera etapa de la gestió de la seguretat, en la de l'anàlisi d'incidències i atacs. L'explotació dels models d'atacs permet generar coneixement que permet accelerar el procés d'anàlisi i interpretació de les dades d'atacs i incidències. Aquest fet permet reduir el temps en el procés d'atribució, i el temps necessari en la pressa de decisions.
BENEFICIS DERIVATS	Com marca l'UE la resposta davant les ciber-amenaces ha de ser global, participant tots els actors, tant Estats com empreses, de forma col·laborativa i coordinada. La compartició d'informació d'incidències i atacs és necessària i vital. L'anàlisi i explotació de les dades per conèixer millor als adversaris afavoriran aquesta resposta global i la pressa de decisions. Però, la compartició d'informació d'incidències i atacs és complicat, ja que encara hi han aspectes a resoldre i que s'estan treballant, com l'anonimat i privadesa de qui aporta les dades a la plataforma de compartició d'informació, en aquest cas MISP.
GRAU DE MADURESA	TRL6
LÍNIES DE FUTUR	Encara queda camí per recórrer, ja no sol a nivell polític i organitzatiu, promovent la participació de tots els actors, i protegint a tots els participant. Sinó que també queda molt treball per realitzar a nivell tecnològic-científic, des de les dades utilitzades, com en els algorismes utilitzats degut a l'aparició de nous grups i tipus d'amenaces/atacs.
CONTACTE	https://www.sparta.eu/ juan.caubet@eurecat.org

STOP-IT, Strategic, Tactical, Operational Protection of water infrastrcture agains cyber-physical Threats (Detection of combinated attacks)	
REPTE	L'objectiu clau de la tasca feta en aquest projecte desenvolupar metodologies i eines per a la detecció primerenca d'atacs combinats (físics i lògics) sobre infraestructures crítiques (en particular relacionades amb la gestió de l'aigua), utilitzant dades recollides de sistemes de detecció físics (càmeres, sensors, etc.) i lògics (IDS, etc.). Es busca reduir el temps d'exposició i impacte d'un atac sobre la infraestructura.
SOLUCIÓ	L'objectiu es desplegar un sistema de detecció d'atacs o intrusions a infraestructures crítiques, concretament en el sector aigua, amb la finalitat de reduir d'exposició davant una incidència o atac. Com és conegut, molts dels atacs sobre infraestructures crítiques s'inicia amb accions físiques, ja que moltes vegades els sistemes estan aïllats com a mesures de seguretat. La innovació que aporta aquest cas d'ús és que té en compte dades i alertes que es generen pels diferents dispositius i sistemes desplegats en una infraestructura crítica, siguin relacionats amb el domini ciber, com el domini físic. La solució ha estat desenvolupada sobre una plataforma Big Data i s'ha validat desplegant-se en infraestructures crítiques d'aigua a diferents països.
PERFIL TECNOLÒGIC	En aquest cas, els algorismes d'IA s'apliquen a diferents capes. Per una part en els sensors o càmeres de videovigilància, per poder processar de forma avançada les dades i la generació d'alertes. I per altra part en les dades recollides pels diferents dispositius de la xarxa com tallafocs o IDS. Integrat de la plataforma Big Data està la solució basada en IA per a la detecció d'atacs combinats. Aquesta solució requereix d'una etapa d'entrenament per modelar el comportament habitual de la infraestructura, i posteriorment està l'etapa operativa que és capaç de detectar comportaments estranys que seran revisats pels equips de seguretat. Les conclusions i coneixement assolides per aquests equips s'incorporaran al sistema, de forma que a mesura que vagi passant el temps el seu rendiment millori. S'ha de dir que el rendiment de la solució va millorant segons va passant el temps, quan més temps estigui en l'etapa operativa.
DURACIÓ	Projecte finalitzat l'octubre de 2021, amb una duració de 4 anys.
TIPUS DE PROJECTE	Recerca i innovació, amb un consorci d'empreses privades, i centres de recerca públics: Aigües de Barcelona, Cetaqua, Oylo Trust Engineering, World Sensing, EMASAGRA, ATOS, Eurecat, etc. (https://stop-it-project.eu/about-stop-it/meet-the-team/).
TIPUS DE FINANÇAMENT	Públic, amb finançament d'EU per la convocatòria H2020.
IMPACTE	Aquest cas d'ús té un impacte directe sobre l'etapa de detecció. La problemàtica de la seguretat de les infraestructures critiques fins fa uns anys s'havia abordat de forma separada entre la part física i la part ciber, però des de fa uns anys, a partir de determinats atacs coneguts, la seguretat és un tot, i s'està abordant de forma conjunta, ja que no són mons aïllats i separats, com passava fins fa pocs anys. Generalment hi existien dos departaments separats, un per la seguretat física i un per la seguretat ciber, i era molt complicat la seva coordinació. Amb aquesta solució, s'aborda la seguretat global de la infraestructura, i es dona resposta a aquelles amenaces no conegudes, com atacs dirigits o que volen explotar vulnerabilitats dia-zero, reduint l'exposició de les infraestructures i reduint el temps de resposta davant una incidència.
BENEFICIS DERIVATS	Aquest tipus de solució permet incrementar la seguretat de les infraestructures crítiques, tenint en compte no solament les amenaces cibernètiques, sinó també aquelles combinades amb accions físiques, ja que moltes vegades un atac dirigit s'inicia amb una acció física. Aquest tipus de solució, amb aprenentatge continu, increment del coneixement, i processament d'un gran volum de dades de diferents fonts és una bona aproximació per l'etapa de detecció, sobretot per front a les noves amenaces i atacs que combinen accions del món real i del món lògic.
GRAU DE MADURESA	TRL7
LÍNIES DE FUTUR	Varies línies de futur estan obertes, des de reduir el temps necessari per construir els corresponents models intel·ligents, des de la capacitat de processament en mode operatiu, fins a l'adaptació d'aquest models davant dels canvis de la infraestructura o els sistemes de captació de l'estat de la infraestructura (càmeres, sensors, IDS, etc.). Aquest es un handicap d'aquest tipus de solucions, on qualsevol canvi significatiu en les dades d'entrada fa que el model s'hagi de re-entrenar, adaptar-se, i això afecta al rendiment i al temps necessari per l'adaptació.
CONTACTE	https://stop-it-project.eu/ juan.caubet@eurecat.org

SECUTIL – Inmonological Detectors	
REPTE	El procés de transformació digital dels sectors de distribució d'energia i serveis incrementa el nombre de superfícies i vectors d'atac, així com l'aparició atacs més complexos. Aquest cas d'ús té com a repte donar resposta a aquest nou panorama, amb una gestió de la seguretat més autònoma, automatitzada i col·laborativa.
SOLUCIÓ	La transformació digital dels sistemes de distribució ha derivat als sistemes que anomenem *Smart Systems*, com per exemple els *Smart Grids* en el sector elèctric. L'increment de superfícies i vectors d'atac s'ha de gestionar eficientment, sent necessari que la gestió de la seguretat estigui distribuïda per la infraestructura, i que sigui una solució amb cert grau d'autonomia i automatització per donar una resposta adequada en un temps acceptable davant una amenaça, atac o intrusió. L'objectiu és dotar de certa intel·ligència als diferents dispositius de la infraestructura, per poder-se auto-protegir, auto-defendre, auto-respondre i auto-recuperar. La gestió de la seguretat ha de ser col·laborativa i coordinada, per tant, també s'apliquen algorismes swarm per poder aconseguir l'objectiu de tenir una gestió de la seguretat distribuïda.
PERFIL TECNOLÒGIC	S'apliquen diferents algorismes d'IA segons la fase, sigui de preparació i desplegament, operatiu o de resposta. En l'etapa de preparació i desplegament es realitza una fase d'aprenentatge per construir el model de comportament del dispositiu que serà la base del detector d'intrusions i atacs, i posteriorment una fase d'optimització mitjançant l'ús d'algorismes genètics. Ja en l'etapa de desplegament es torna a aplicar una etapa d'aprenentatge i optimització per millorar el comportament del detector integrat al dispositiu. Posteriorment, durant l'etapa operativa i resposta es prenen les decisions necessàries tant a nivell de dispositiu com a nivell d'infraestructura. D'aquest forma la intel·ligència per detectar i prendre decisions vers un intent d'atac o intrusió està distribuïda en els diferents elements de la infraestructura, millorant la resiliència i temps de resposta.
DURACIÓ	Projecte finalitzat el març de 2021, amb una duració de 3 anys.
TIPUS DE PROJECTE	Recerca i innovació, amb un consorci d'empreses privades, i centres de recerca públics: Naturgy, Cellnex, Cetaqua, UdL, Eurecat.
TIPUS DE FINANÇAMENT	Públic amb finançament català mitjançant la convocatòria R3CAT.
IMPACTE	Realment, aquesta solució impacte en l'etapa de detecció, encara que també incorpori un aprenentatge per identificar el tipus de dispositiu i el seu comportament. De igual forma, impacta en l'etapa de resposta i recuperació, encara que durant aquest projecte el cas d'ús no l'ha treballat en profunditat. Ara mateix no té un impacte real, ja que el cas d'ús s'ha validat en un entorn de laboratori amb dispositius hardware (físics), per calcular les mètriques d'acceptació, com consum d'energia, temps de resposta, capacitat de processament i memòria, etc. Però els resultats obtinguts deriven a una major autonomia i automatització en la pressa de decisions, i una millor gestió de la seguretat.
BENEFICIS DERIVATS	Les millores s'han comentat a l'apartat anterior, a la descripció de l'impacte. Els anomenats *Smart Systems (Smart Grids)*, producte de la transformació digital dels sistemes de distribució d'electricitat, gas o aigua, són d'especial interès pels adversaris per ser infraestructures crítiques, i la seva seguretat és un gran repte per totes les organitzacions implicades. Encara que el cas d'ús pot millorar molts aspectes de la seva seguretat, com l'autonomia, automatització, temps de resposta, etc., un dels aprenentatges és que incorporar algorismes i solucions basades en intel·ligència requereix d'una inversió econòmica i de temps. Quan més es trigui en adaptar-se els sistemes de seguretat d'aquest tipus d'infraestructures, major temps s'està deixant a l'adversari a que actuï.
GRAU DE MADURESA	TRL6
LÍNIES DE FUTUR	La millora en els algorismes d'aprenentatge per a que la solució global evolucioni i s'adapti per respondre a les noves amenaces, i les estratègies, tàctiques i tècniques dels adversaris és la principal línia de recerca i treball, ja que és el gran repte d'empreses i organitzacions en el domini de la seguretat.
CONTACTE	mario.reyes@eurecat.org

INTSEG, Threat Intelligence	
REPTE	Ajudar a conèixer i modelar el comportament d'atacs mitjançant l'ús de *honeypots*, per poder determinar contramesures de prevenció i protecció de les infraestructures IoT.
SOLUCIÓ	L'ús de *honeypots* com a eina de protecció i anàlisi del comportament dels adversaris no és nou, però en aquest cas d'ús l'objectiu consisteix en incrementar el nivell d'automatització en les diferents etapes, des de l'etapa de desplegament, la recopilació d'informació, la classificació de dispositius que es sospita estan sent atacats, l'agrupació de dispositius on l'adversari es comporta de forma similar, així com la determinació d'indicadors de compromís relatius al comportament de l'adversari. Per aconseguir aquest objectiu s'utilitzen algorismes de classificació i optimització en les diferents fases que s'han indicat. De totes formes, la interacció i col·laboració entre el sistema desplegar i l'equip d'experts de ciberseguretat és essencial per millora la seguretat de la infraestructura que es vol assegurar.
PERFIL TECNOLÒGIC	Com s'ha comentat, a la fase de desplegament s'apliquen algorismes d'optimització en base a l'experiència acumulada de desplegaments anteriors, amb la finalitat de ser eficient amb relació als costos, amb relació al temps, i amb relació als servidors més efectius per recopilar informació. Una vegada recollida la informació, s'apliquen algorismes d'aprenentatge per modelar el comportament dels dispositius a desplegar, i posteriorment s'apliquen algorismes de classificació per agrupar aquells dispositius sospitosos d'estar sent atacats, i extraure aquells indicadors que poden caracteritzar el comportament de l'atac. Tota aquesta informació després es comparteix amb l'equip d'experts per analitzar i prendre decisions. Per tant, aquesta solució inclou tres tipus d'algorismes d'IA en diferents etapes.
DURACIÓ	Projecte finalitzat amb una duració de 3 anys.
TIPUS DE PROJECTE	Recerca: Eurecat
TIPUS DE FINANÇAMENT	Públic, amb finançament català i inversió interna.
IMPACTE	Aquesta solució impacta directament en el coneixement dels atacs, del comportament dels adversaris (estratègies i tàctiques), ajudant als equips d'especialistes de seguretat a reduir el seu temps de resposta i pressa de decisions. Hores d'ara, les proves realitzades s'han fet en el context de la fase de validació de la solució, desplegant un honeypot amb un nombre controlat de dispositius virtuals en un nivell internacional. Els resultats obtinguts són realment satisfactoris, però al no estar desplegat en un sistema real, amb la interacció amb l'equip de seguretat, encara no podem concloure quin és l'impacte real de la solució. Sí es pot afirmar que s'ha millorat el nivell d'automatització de tot el procés de generació de coneixement sobre amenaces *(threat intelligence)*.
BENEFICIS DERIVATS	Com s'ha indicat, les solucions per a la generació de coneixement de comportament d'adversaris mitjançant *honeypots* són de gran ajuda pels equips de seguretat, però la majoria de vegades és un procés massa manual, amb la introducció d'algorismes i tecnologies d'IA s'incrementa el grau d'automatització de tot el procés, reduint el temps necessari, així com la gestió de tot el procés. D'aquest cas d'ús es pot concloure que aquest tipus de sistemes són de gran valor pels equips de seguretat, però que encara s'ha d'incrementar el nivell d'automatització i pressa de decisions per poder combatre i compatir la capacitat d'evolució dels atacs o intrusions realitzats pels adversaris.
GRAU DE MADURESA	TRL6
LÍNIES DE FUTUR	La proposta de l'ús de *honeypots* amb IA té un gran potencial per la lluita contra els ciberatacs, en la generació de coneixement relatiu al comportament dels atacs i en suport als equips de seguretat. Aquest cas d'ús es podria completar amb l'aplicació de tecnologies de processament del llenguatge per poder extreure coneixement de les dades recopilades pels *honeypots*, i afegir informació addicional de fonts externes, de forma que podríem ser capaços de modelar els atacs sobre els dispositius IoT amb l'estàndard MITRE ATT&CK.
CONTACTE	mario.reyes@eurecat.org

PRODUCTIO, Insiders detection	
REPTE	Els serveis de tractament de dades en el cloud tenen el repte de detectar totes les incidències relatives als usuaris, com per exemple si els operadors introdueixen les dades correctament, com si algun usuari té un comportament maliciós. Aquest cas d'ús es centra en la detecció d'aquestes tipus d'incidència.
SOLUCIÓ	L'objectiu del cas d'ús és la detecció de comportaments no esperats per part dels usuaris d'un servei de processament de dades al cloud, i poder discernir d'aquelles que són degudes la fatiga dels usuaris, o el comportament és sospitós de ser maliciós. En el moment de realització del cas d'ús, aquesta innovació integrada al servei va suposar una millora en determinants processos del servei, com un millora en la seva seguretat.
PERFIL TECNOLÒGIC	En aquest cas es van aplicar algorismes d'IA per modelar el comportament habitual (normal) dels diferents usuaris, així com per modelar la qualitat de les dades que eren introduïdes de forma manual pels diferents operadors. Una vegada obtinguts aquests models, existeix una capa addicional per detectar comportament d'usuaris no-habituals, així com anomalies a les dades degudes a la variació del comportament de l'operador que introdueix les dades de forma manual.
DURACIÓ	Projecte finalitzat amb una duració de 3 anys.
TIPUS DE PROJECTE	Recerca i innovació amb una empresa com impulsor del projecte.
TIPUS DE FINANÇAMENT	Privat (Tecnomatrix, Eurecat)
IMPACTE	Aquest cas d'ús impacte directament a la fase de detecció de la gestió de la seguretat del servei. La detecció del comportament no adequat en la fase d'introducció de les dades de forma manual permet corregir aquests errors millorant la qualitat de les dades, i per tant el servei. Per altre banda, la detecció de comportaments sospitosos per part d'usuaris autoritzats millora la seguretat de la plataforma i la seva confiança en el sector. Aquestes dos propostes millorant el servei ofert per l'empresa impulsora del projecte.
BENEFICIS DERIVATS	Aquest cas d'ús es va realitzar fa uns anys, i va ajudar a detectar diferents línies de continuïtat en aquest àmbit, com el procés de modelatge de comportament dels usuaris amb les dades reduïdes que es poden recol·lectar, el nombre de fals positius i negatius que es generen, i el grau d'automatització i autonomia vers la detecció d'incidències sospitoses amb relació als usuaris.
GRAU DE MADURESA	TRL7
LÍNIES DE FUTUR	Com s'ha indicat, existeixen diferents línies de continuïtat per part de l'empresa impulsora, millorant el rendiment dels mòduls basats en IA, tant en la reducció de fals positius i negatius, així com en el grau d'automatització i autonomia de la solució.
CONTACTE	juan.caubet@eurecat.org

NEWIDENTITY, Biometric authentication	
REPTE	Un dels inconvenients d'utilitzar un sistema d'autenticació biomètrica basat en senyals biològiques és la seva variació en el temps i condicions de context, i les dades que es poden utilitzar per crear el model de l'usuari. El repte d'aquest cas d'ús és resoldre aquests inconvenients, mantenint el grau de privadesa segons el marc jurídic.
SOLUCIÓ	Els sistemes biomètrics d'autenticació continua és un sector amb gran potencial, on la majoria d'elles es basen en dades biològiques o comportament del usuari. Aquestes dades són difícils d'obtenir, i varien segons l'edat de l'usuari, el context, etc. L'objectiu del cas d'ús es obtenir un sistema d'autenticació biomètric amb poques dades, i capaç d'evolucionar en el temps.
PERFIL TECNOLÒGIC	Les característiques del problema que es tracta, com el reduït conjunt de dades per generar un model de l'usuari, així com la variabilitat en el temps i el context, deriven en l'ús de diferents algorismes en diferents etapes. En una primera etapa de generació d'un model general, s'utilitzen dades anonimitzades i una xarxa neuronal convolucional. A partir d'aquest model, es realitza un procés de personalització en el dispositiu on es farà el procés d'autenticació d'un usuari, utilitzant únicament les dades de l'usuari, i una xarxa neuronal siamesa. I, per últim és necessari un procés d'optimització d'alguns paràmetres mitjançant un procés d'optimització dinàmic. La privadesa es garanteix perquè les dades tractades en el servidor (requereix de gran capacitat de processament i dades) són anonimitzades, i quan es personalitza el model les dades no surten del dispositiu de l'usuari i no és necessari emmagatzemar-les.
DURACIÓ	Projecte finalitzat amb una durada de 2 anys
TIPUS DE PROJECTE	Recerca
TIPUS DE FINANÇAMENT	Per una primera fase va tenir finançament públic, i per una segona fase va tenir finançament privat.
IMPACTE	Aquest cas d'ús impacta directament en l'etapa de protecció, amb un sistema d'autenticació biomètrica continua, facilitant la usabilitat del procés del control accés. Com l'autenticació és continua mitjançant un dispositiu que porta l'usuari (mòbil o wearable), el procés de control d'accés és transparent a l'usuari.
BENEFICIS DERIVATS	En el procés de transformació digital que s'està produint en la societat i a l'àmbit productiu, la identitat i el control d'accés cada vegada té un paper més rellevant, ja que la deslocalització dels recursos, el desplegament d'elements IoT i la pèrdua de perímetre, fa necessari identificar en tot moment com, qui i què realitza cada acció en cada moment. Aquest tipus de solució, autenticació biomètrica basat en dades biològiques, requereixen d'algorismes intel·ligents per la generació dels models, així com la seva adaptació en el temps. Però, encara queden coses per resoldre, com la millora de la seva autonomia en l'etapa d'adaptació en el temps.
GRAU DE MADURESA	TRL6
LÍNIES DE FUTUR	Està clar que és necessari la integració i desplegament de solucions d'aquest tipus en el sistemes productius i d'informació. Com s'ha indicat, una de les principals línies de futur i recerca és treballar en els algorismes d'adaptació, mantenint la privadesa, i tenint en compte els recursos limitats dels potencials dispositius (wearables) que portarà un usuari.
CONTACTE	juan.caubet@eurecat.org

TDA en Ciberseguretat	
REPTE	L'exposició dels usuaris a una ciberamenaça és un gran repte a resoldre, segons l'Agència de Ciberseguretat de Catalunya. Aquest projecte pretén desenvolupar solucions que hi donin resposta. El projecte s'emmarca en el Programa de Recerca i Innovació en tecnologies digitals avançades (TDA) d'SmartCatalonia.
SOLUCIÓ	Es proposa una framework de codi obert pel càlcul dels usuaris a ser víctimes d'atacs en xarxa.
PERFIL TECNOLÒGIC	El projecte està dividit en dues etapes, en la primera etapa es modela el comportament dels usuaris a través de seqüències d'esdeveniments i s'extreuen els comportaments dels usuaris compromesos en campanyes. En la segona etapa, els patrons apresos es fan servir per calcular el risc dels usuaris envers campanyes.
DURACIÓ	Projecte en curs
TIPUS DE PROJECTE	Recerca i innovació, amb un consorci d'empreses privades, i centres de recerca públics: I2CAT Foundation, Agencia Ciberseguretat de Catalunya, CRM, Univ. Lleida, UOC, Open Cloud, Grupo Sirt, UPC, Across Legal, Aubay Spain, URV
TIPUS DE FINANÇAMENT	Autonòmic
IMPACTE	El projecte, encara en curs, i en procés de validació. Ha sigut presentat en diverses conferències a nivell internacionals tals com el tnc o el TF-CSIRT.
BENEFICIS DERIVATS	L'usuari és avui dia l'eslavó més feble i per aquest motiu és important conèixer el comportament d'aquests per preveure les amenaces actuals que poden esdevenir. Aquest cas d'ús pretén predir mitjançant el comportament dels usuaris quins poden ser susceptibles d'un ciberatac de forma automatitzada sense necessitat que els equips de seguretat hagin de revisar tots els logs de forma manual i determinar si és o no un atac. Com a aprenentatge, podem concloure que a causa de la gran quantitat de dades que generen les aplicacions i els usuaris, és inviable analitzar de forma manual aquestes dades i s'han de trobar mecanismes que automatitzin la detecció de forma adequada (com si fos un analista) de les possibles amenaces perquè els equips de seguretat únicament decideixin quines mesures s'han d'aplicar.
GRAU DE MADURESA	TRL7
LÍNIES DE FUTUR	El comportament dels usuaris, es dir com interactuen a la xarxa es el punt més debil en la kill-chain. Com a lines de futur es preveu avançar en l'eina per ser capaços de preveir altres tipus d'atacs com «insider threats» i també ser capaços de desplegar l'eina en entorns zero-trust i entorns federats.
CONTACTE	https://tdacyber.cat/

PALANTIR	
REPTE	Els ràpids avenços en la tecnologia digital requereixen trobar formes de garantir la seguretat digital ajudant a les petites i mitjanes empreses (PIMES) a ser resilients dels atacs cibernètics. El projecte PALANTIR, té com a objectiu implementar un marc que combina aspectes de privacitat, protecció de dades, detecció i recuperació d'incidències . El projecte també es centrarà en la resiliència cibernètica i garantirà el compliment de la part de les pimes de les normes pertinents de privacitat i protecció de dades.
SOLUCIÓ	PALANTIR es centra en el desenvolupament de les següent característiques a) aplicant i explotant les tecnologies de virtualització de funcions de xarxa (NFV) i xarxes definides per programari (SDN); b) considerar paradigmes emergents com ara l'aplicació d'IA escalable, estandardització i tècniques de compartició d'amenaces a l'anàlisi de riscos, l'operació de la xarxa, el seguiment i la gestió i c) assegurant el compliment de la PIME amb les normatives rellevants de privacitat i protecció de dades en l'era de la violació de dades. , implementant els principis de «Privacitat per defecte» i els principis de «Privacitat per disseny» sobre com es recullen, s'utilitzen, es transfereixen i s'emmagatzemen les dades personals entre empreses i entitats de tercers.
PERFIL TECNOLÒGIC	La Intel·ligència Artificial esdevé un punt clau per la detecció i caracterització de threats en les diferents PIME's
DURACIÓ	Projecte en curs
TIPUS DE PROJECTE	Recerca i innovació, amb un consorci d'empreses privades, i centres de recerca públics: I2CAT Foundation, Universitat de Murcia, Telefónica Investigación y desarrollo S.A.U. (TID), Universidad Politécnica de Madrid
TIPUS DE FINANÇAMENT	Europeu
IMPACTE	Els resultats del projecte proporcionaran a aquestes empreses eines de seguretat que augmentaran la seva resiliència a un cost raonable.
BENEFICIS DERIVATS	TRL6
CONTACTE	https://www.palantir-project.eu/

PHOENI2X, A European Cyber Resilience Framework with Artificial Intelligence-Assisted Orchestration & Automation for Business Continuity, Incident Response& Information Exchange	
REPTE	PHOENi²X té com a objectiu dissenyar, desenvolupar i oferir un marc de resiliència cibernètica que proporcioni IA (capacitats d'orquestració assistida, automatització i resposta) per a la continuïtat i la recuperació del negoci, la resposta a incidents i l'intercanvi d'informació, adaptat a les necessitats dels operadors de serveis essencials (OSE) i de les autoritats nacionals dels estats membres de la UE encarregades de la ciberseguretat
SOLUCIÓ	L'enfocament holístic de PHOENi2X integra la prevenció, la detecció i la resposta mitjançant un conjunt d'eines de referència amb totes les funcions. En aquesta direcció, el projecte: 1) proporcionarà capacitats fiables de consciència i predicció de la situació assistides per IA, amb avaluació de l'impacte del risc, facilitant la priorització, recomanació i adaptació de la resposta del sistema; 2) dissenyar i desenvolupar mecanismes d'orquestració, automatització i resposta de la resiliència, que inclouen tasques de continuïtat del negoci, recuperació i gestió d'incidències proactives i reactives; 3) oferir una preparació millorada mitjançant una gamma cibernètica de resiliència i jocs seriosos; 4) proporcionar mecanismes i processos d'alertes, informes i intercanvi d'informació que permetin la col·laboració entre els actors crítics del sector públic i privat a nivell nacional i europeu.
PERFIL TECNOLÒGIC	Consciència de la situació millorada amb capacitats de predicció, prevenció, detecció i resposta assistides per IA i priorització basada en l'avaluació de l'impacte del risc empresarial
DURACIÓ	Projecte en curs, Setembre 2022 – Agost 2025.
TIPUS DE PROJECTE	Recerca i innovació, amb un consorci d'empreses privades, centres de recerca públics i universitats
TIPUS DE FINANÇAMENT	Públic, amb finançament d'EU per la convocatòria H2020.
IMPACTE	L'objectiu és que PHOENi²X ajudi els operadors dels sectors crítics a assolir la resiliència cibernètica i també donarà suport a les autoritats dels Estats membres de la UE per millorar les capacitats nacionals de ciberseguretat, la col·laboració transfronterera i la supervisió nacional dels seus sectors crítics, d'acord amb els requisits de la Directiva NIS.
BENEFICIS DERIVATS	Tenint en compte que l'efecte de la globalització és més evident a Internet, la defensa contra els ciberatacs, com també indiquen les directives de la UE, cal que ara sigui un esforç coordinat entre diferents sectors i diferents estats. Aquest projecte pretén proporcionar mecanismes i processos d'alertes, informes i intercanvi d'informació que permetin la col·laboració entre els actors crítics del sector públic i privat a nivell nacional i europeu, així com les agències de seguretat nacional.
GRAU DE MADURESA	TRL7
LÍNIES DE FUTUR	El projecte acaba de començar.
CONTACTE	https://fishy-project.eu/

H2020 CARAMEL: Artificial Intelligence based cybersecurity for connected and automated vehicles	
REPTE	CARAMEL és un projecte que pretén introduir un innovador sistema de detecció/prevenció d'intrusions anti-pirateria per a la indústria de l'automoció europea. Els efectes perjudicials dels ciberatacs a una indústria com la Cooperative Connected and Automated Mobility (CCAM) poden ser enormes. Des dels menys importants fins als pitjors, es pot esmentar, per exemple, el dany a la reputació dels fabricants de vehicles, l'augment de la negativa dels clients a adoptar CCAM, la pèrdua d'hores de treball (que té impacte directe en el PIB europeu), danys materials, etc. l'augment de la contaminació ambiental a causa, per exemple, d'embussos de trànsit o modificacions malicioses en el firmware dels sensors i, en definitiva, el gran perill per a vides humanes, ja siguin conductors, passatgers o vianants.
SOLUCIÓ	L'objectiu de CARAMEL és abordar de manera proactiva els reptes de ciberseguretat dels vehicles moderns aplicant tècniques avançades d'intel·ligència artificial (IA) i d'aprenentatge automàtic (ML) i també buscar contínuament mètodes per mitigar els riscos de seguretat associats. Per abordar les consideracions de ciberseguretat dels vehicles autònoms i connectats que ja estan aquí, s'han adoptat metodologies consolidades procedents del sector TIC, que permetin avaluar les vulnerabilitats i els impactes potencials dels ciberatacs. Tot i que les iniciatives anteriors i els projectes de ciberseguretat relacionats amb la indústria de l'automòbil han arribat als marcs de garantia de seguretat per als vehicles en xarxa, diverses dimensions tecnològiques recentment introduïdes com el 5G, els pilots automàtics i la càrrega intel·ligent de vehicles elèctrics (EV) introdueixen llacunes de ciberseguretat, encara no abordades de manera satisfactòria. Tenint en compte tota la cadena de subministrament d'operacions d'automoció, CARAMEL ha tingut com a objectiu arribar a productes comercials IDS/IPS anti-pirateria per a la ciberseguretat de l'automoció europea i demostrar el seu valor mitjançant escenaris d'atac i penetració extensos.
PERFIL TECNOLÒGIC	Amb els vehicles "intel·ligents" connectats i autònoms que transporten cada cop més persones i càrrega, la seva seguretat i seguretat són cada cop més preocupants. CARAMEL proposa la ciberseguretat dels vehicles que és una solució integral conscient del risc, basada en processos, des del disseny fins a l'operació.
DURACIÓ	Projecte finalitzat amb una duració d'octubre 2019-Juny 2022
TIPUS DE PROJECTE	Recerca i innovació, amb un consorci d'empreses privades, centres de recerca públics i universitats: i2cat Foundation, atos, nextium by idneo
TIPUS DE FINANÇAMENT	Públic, amb finançament d'EU per la convocatòria H2020. (Grant Agreement No. 833611)
IMPACTE	El projecte H2020 CARAMEL, finançat per la UE, ha desenvolupat solucions de ciberseguretat per a la nova generació de cotxes sota quatre pilars: i) Mobilitat autònoma: Els ciberatacs no requereixen accés físic al vehicle ni manipulació del sistema de comunicació. ii) Mobilitat connectada 5G: Les aplicacions V2X interconnecten no només els vehicles, sinó també la infraestructura i els vianants, per tant, és fonamental protegir les funcions V2X d'un mal ús. iii) Electromobilitat : Detecció d'accessos no autoritzats de les estacions EVSE i les modificacions del firmware. iv) Vehicle de control remot (RCV) Algorisme d'estimació i detecció d'intrusions als controladors per evitar un mal ús. https://cordis.europa.eu/article/id/442445-making-smart-vehicles-more-cybersecure
BENEFICIS DERIVATS	Aquest projecte de recerca europeu ha permès la generació de més ecosistema i coneixement local, en aquest cas en el camp de la ciberseguretat basada en IA per a vehicles connectats i autònoms on a mode d'exemple la Fundació i2CAT ha liderat la implementació d'un sistema de protocols de comunicació de vehicle a vehicle i infraestructura (V2X) que té per objectiu principal detectar i mitigar els atacs que pot rebre un vehicle autònom i que poden comprometre la seva seguretat.
GRAU DE MADURESA	TRL7
LÍNIES DE FUTUR	Ciberseguretat de nova generació: Com que el transport forma part de la infraestructura crítica d'Europa, està destinat a una protecció especial en temps de crisi i es considera un pilar central de l'estratègia europea de ciberseguretat. Els resultats de CARAMEL són un trampolí cap a la ciberseguretat de nova generació per a vehicles autònoms, connectats i d'electromobilitat. Però perquè els actors europeus puguin aprofitar-se, cal abordar la fragmentació de la indústria i de determinades polítiques.
CONTACTE	https://www.h2020caramel.eu/ Project OpenBook: https://www.nowpublishers.com/Article/BookDetails/9781638280606 Coordinador i2CAT. Fundacio@i2cat.net

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

AGÈNCIA DE CIBERSEGURETAT DE CATALUNYA

info@cidai-catalonia-ai.eu
www.cidai.eu

Membres del CIDAI

Generalitat de Catalunya Departament d'Empresa i Treball

Ajuntament de Barcelona · Barcelona Activa

Barcelona Supercomputing Center Centro Nacional de Supercomputación

CVC Centre de Visió per Computador

eurecat Centre Tecnològic de Catalunya

i2cat

Microsoft

NTT DATA

SAP

sdg group

UPC Universitat Politècnica de Catalunya BarcelonaTech Intelligent Data Science and Artificial Intelligence Research Center

IDEAI