Universitat Oberta de Catalunya

La IA, nou gestor de la Ciberseguretat.

Memòria del treball final de màster

TFM – Seguretat en sistemes operatius M1.724

Màster Interuniversitari de Seguretat de les TIC (UOC, URV, UAB) - MISTIC

Nom Estudiant : Jordi Fenés Castell
Nom Professors : Erik de Luis Gargallo
Jordi Serra Ruiz

Data Lliurament : 06/2022

Fitxa del treball Final

Títol del treball:	La IA, nou gestor de la Ciberseguretat
Nom de l'autor:	Jordi Fenés Castell
Nom del consultor:	Erik de Luis Gargallo
Data d'entrega (mm/aaaa):	06/2022
Àrea del treball Final:	Seguretat en sistemes operatius
Titulació:	Màster Interuniversitari de Seguretat de les TIC (UOC/URV/UAB) - MISTIC

Resum del treball (màxim 250 paraules):

Aquest treball pretén aportar una base sobre la IA (Intel·ligència Artificial) aplicada a la Ciberseguretat. Mostrarà el origen i fonaments de la IA, la seva evolució, algorismes emprats, aprenentatge automàtic, aprenentatge profund i les diferents orientacions de la IA. Afegirem en detall, la Ciberseguretat com una orientació o especialitat en que es pot integrar, aplicar i evolucionar la IA.

Actualment, les IA s'estan emprant, implementant i evolucionant, en molts camps i sectors, tant públics com privats, per donar solució a molts i variats problemes. Els punts forts de les IA son: la presa de decisions, treballar amb grans volums de dades, obtenir patrons o informació rellevant de les dades, el temps de resposta i pot combinar això amb sistemes experts. Enfocant això a la Ciberseguretat, la IA pot donar resposta primer a la detecció, mitigació i aturada d'atacs coneguts per a sistemes experts. I segon, pot donar suport o corregir mancances sobre el factor humà com:

- Falles humanes en la configuració
- Pèrdua d'eficiència humana en activitats repetitives
- Fatiga per excessos d'alarmes sobre amenaces
- Temps de resposta davant les amenaces
- Identificació i predicció de noves amenaces
- Reduir o mantenir la dotació de personal especialitzat
- Adaptació ràpida a l'entorn de treball

Com a tendències i visió final de tot això, tindrem la possible interacció de la IA amb els diferents elements o components (endpoints) de la xarxa. I també l'evolució de certes IA emprant altre "Hardware" per assolir reptes de treballar amb més volums de dades.

Abstract (in English, 250 words or less)

This paper aims to provide a basis for AI (Artificial Intelligence) applied to Cybersecurity. It will show the origin and foundations of AI, its evolution, algorithms used, machine learning, deep learning and the different orientations of AI. We will add in detail, Cybersecurity as an orientation or specialty in which AI can be integrated, applied and evolved.

AI is currently being used, implemented and evolved in many fields and sectors, both public and private, to solve many and varied problems. The strengths of AI are: decision making, working with large volumes of data, obtaining patterns or relevant information from the data, response time and you can combine this with expert systems. By focusing on Cybersecurity, the AI can respond first to detecting, mitigating, and stopping known attacks by expert systems. And second, it can support or correct deficiencies on the human factor such as:

- Human configuration failures
- Loss of human efficiency in repetitive activities
- Fatigue due to excessive alarms about threats
- Response time to threats
- Identification and prediction of new threats
- Reduce or maintain the staffing of specialized personnel
- Quick adaptation to the work environment

As trends and final vision of all this, we will have the possible interaction of the AI with the different elements or components (endpoints) of the network. And also the evolution of certain AI using other "Hardware" to meet the challenges of working with more volumes of data.

Paraules clau (entre 4 i 8):

Intel·ligència Artificial (IA), Ciberseguretat, Machine Learning (ML), Deep Learning (DL), Algorismes, volums de dades, patrons, temps de resposta

Índex

Llista de Figures

1. Introducció

1.1. Context i justificació del treball

Històrica i generalment es posa el naixement i fonaments de base de la IA, a partir dels treballs en 1950 de Alan Turing, matemàtic britànic, que obren una nova disciplina de les ciències de la informació. De totes formes, hi ha indicis en dates anteriors, sobre emular el raonament humà de forma artificial i també idees, teories, invents o treballs que donen suport a tot això. I fins avui en dia, l'evolució de la IA, a patit diferents entrebancs o períodes de congelació en el seu desenvolupament, per varis motius. Encara que actualment, la IA es troba immersa dins una gran explosió d'aplicació amb èxit en molts sectors, entorns, especialitats i amb diferents finalitats.

Per entendre aquest canvi de visió i creixent ús de la IA, ens hem d'endinsar en com treballa la IA amb diferents algorismes[1] que donen resposta a molts i variats plantejaments de problemes. En base a l'anterior, també ens cal aprofundir en com aprèn la IA, en com ho fa, de quines formes pot fer-ho, que cal tenir en compte o quines son les directrius que cal considerar per tenir una "IA Fiable". I a partir de l'anterior, veure com es classifiquen les IA segons orientacions i els mètodes emprats per afrontar els reptes. Amb això, podrem veure quins son els punts forts de la IA, com els assoleix, i el perquè ens permet poder-la aplicar a moltes situacions complexes.

Plantejarem el cas de la Ciberseguretat, com una especificació complexa, en base a la seva gestió, manteniment, reptes, evolució i també el factor humà, a la que la IA pot donar una bona resposta. Aquí, desgranarem una mica l'estat real de la Ciberseguretat en la majoria de les empreses, les mancances o dèficits que si donen i els reptes de futur que es presenten. També, incidirem una mica en el factor humà a la Ciberseguretat en quant a capacitació, limitació i ineficiències que sorgeixen. A més, tindrem en compte l'evolució i reptes que es presenten a la Ciberseguretat, ja que actualment es planteja i es suposa l'ús de la IA per part dels Ciberdelinqüents.

Entrarem en com es pot combinar o implementar la IA a la Ciberseguretat, quines son les tendències, quins son els trets que hauria de tenir. Ens caldrà afrontar diferents situacions que es donen actualment a la Ciberseguretat, i

[1] Algorismes: (o algoritmes) conjunt finit d'instruccions o passos que serveixen per executar una tasca o resoldre un problema.

veure com les pot resoldre o mitigar la IA. Podrem veure com la IA ens aporta una millora, més control i ens ofereix una prevenció sobre possibles atacs. I per altra banda, podrem veure en detall i conèixer algunes les solucions actuals més punteres, sobre aquest plantejament de la IA a la Ciberseguretat. També, mostrarem els principals trets o característiques de cada una de les solucions revisades. Com a punt final, farem una especulació i donarem una visió de quin pot ser el futur de la IA a la Ciberseguretat.

Per tant, amb tot l'anterior, ens ha de quedar clar que més que un plantejament o una possibilitat, és una realitat i una necessitat poder disposar d'una IA a la Ciberseguretat. Un cop vista la realitat actual que es dona a la Ciberseguretat, per tal de poder tenir una gestió correcta i obtenir un valor afegit en el seu control, no hi ha més camí que afegir una IA. Per poder fer això amb èxit, ens cal valorar i entendre a fons els seus pros i contres, les necessitats i els tipus d'aprenentatge que es donen en una IA. Així, ens cal enfocar les possibles millores i els beneficis que aporti l'ús de la IA, al terreny de la Ciberseguretat i saber quines tendències hi ha. D'aquesta forma, amb tot l'anterior, podrem tenir una base i una visió critica del que interessa que faci una IA per la Ciberseguretat, per tal d'obtenir una millora en la productivitat i la resiliència[2].

1.2. Objectius del treball

En aquest TFM podem destacar els següents objectius o aspectes importants:

- Mostrar una breu cronologia de l'aparició de la IA, els seus fonaments i la seva evolució fins al dia d'avui.
- Donar alguns exemples generals en que es fa us de la IA i els beneficis que hi aporta a diferents sectors.
- Mostrar els diferents algorismes amb que treballa la IA, segons les tasques que hagi de resoldre.
- Veure en detall que es l'aprenentatge automàtic (ML-Machine Learning) i l'aprenentatge profund (DL-Deep Learning), i la seva correcta aplicació.
- Mostrar algunes classificacions de la IA segons els algorismes emprats, funcionalitat i ús o no de l'aprenentatge automàtic o profund.
- Entendre quins sons els punts forts de la IA, el perquè son aquestos, i de quina forma els assoleix.

[2] Resiliència: capacitat de recuperar-se d'una falla i conservar la confiança del servei, ha de garantir la protecció de les operacions de forma que una amenaça o incompliment no afecti a la resta de negoci.

- Entendre els reptes a que s'enfronta actualment la Ciberseguretat, les necessitats i el seu panorama laboral, dins de moltes empreses.
- Veure com esmerçar el factor humà, millorar productivitat i resiliència, mitigar reptes i mancances donats a la Ciberseguretat actual, emprant la implementació d'una IA que treballi amb ella.
- Comentar quines son les tendències actuals en el tema i revisar algunes IA per Ciberseguretat, que ens poden ajudar a millorar productivitat i resiliència, i mostrarem els seus trets principals.
- Orientar als nouvinguts en el que és una IA per Ciberseguretat, que cal valorar segons plantejaments funcionals, organitzatius, capacitació de personal i nous reptes a superar.
- Entreveure el futur amb millores de base que es poden fer a les IA per poder tractar més volums de dades, en menys temps i més eficiència.

1.3. Enfocament i mètode a desenvolupar

Existeixen moltes publicacions relacionades amb el mon de les IA, si bé, moltes son teòriques o matemàtiques, encaminades a resoldre o visionar les funcions primordials d'una IA amb aprenentatge i raonament humà. Per altra banda, també creixen amb força les publicacions més practiques sobre les IA, que s'enfoquen a resoldre problemes més concrets en diferents sectors, aplicant una sèrie de principis ètics, i guiades per uns algorismes especialitzats. Aquest altres casos, son els que ens serveixen per veure com ha de ser una IA, que es vol sols per resoldre un tema en concret, aplicant raonament en les decisions i que aprengui de la informació de que disposa.

També hi ha moltes publicacions relacionades amb la Ciberseguretat, de tota mena: gestió, manteniment, organització, estratègia, eines, automatismes, etc. De totes elles, pel cas donat, ens interessa veure les referents als punts febles de la Ciberseguretat. És a dir, on hi ha els principals problemes, que és el que costa més en la seva gestió, veure els seus riscos, el cost de la prevenció, els temps de resposta, la baixa productivitat i mala resiliència, etc. Així podrem contrastar les troballes efectuades, veure que en majoria de casos no hi ha una situació ideal de funcionament, i que el futur no es presenta massa prometedor si no es dona un canvi, és a dir, un altre enfocament de treball.

Per tant, en base al creixent èxit de l'aplicació de les IA en diferents temàtiques, creiem necessari poder-la aplicar a la Ciberseguretat. D'aquesta forma es poden mitigar o alleugerir els principals punts febles, que s'hagin detectat dins de la Ciberseguretat, cosa que fa millorar la productivitat i la resiliència, i a més,

aprofitem els punts forts de les IA. El fet, es que ja hi ha alguns productes punters que apliquen la IA a la Ciberseguretat, dels que ens cal veure una mica els seus principals trets o característiques, i també quines son les tendències més actuals. A més, hi ha alguns plantejaments visionaris, que persegueixen incrementar el rendiment de procés a la IA per la Ciberseguretat, sense afegir més endpoints o hardware superior.

En resum, en aquest treball es vol mostrar una visió vertical, en part de forma una mica cronològica, del per què tot plegat. És a dir, origen, fonaments, implementacions, classificació, evolució i expansió d'ús real de les IA. I per altra banda, les necessitats, mancances i reptes actuals de la Ciberseguretat, i les possibles solucions que pot aportar una IA. Es vol donar una visió i fonaments de base, que permetin entendre millor el perquè cal emprar una IA en la Ciberseguretat, veure que pot fer, quines solucions ens aporta, quins beneficis podem obtindre i com ens pot millorar la gestió de la Ciberseguretat.

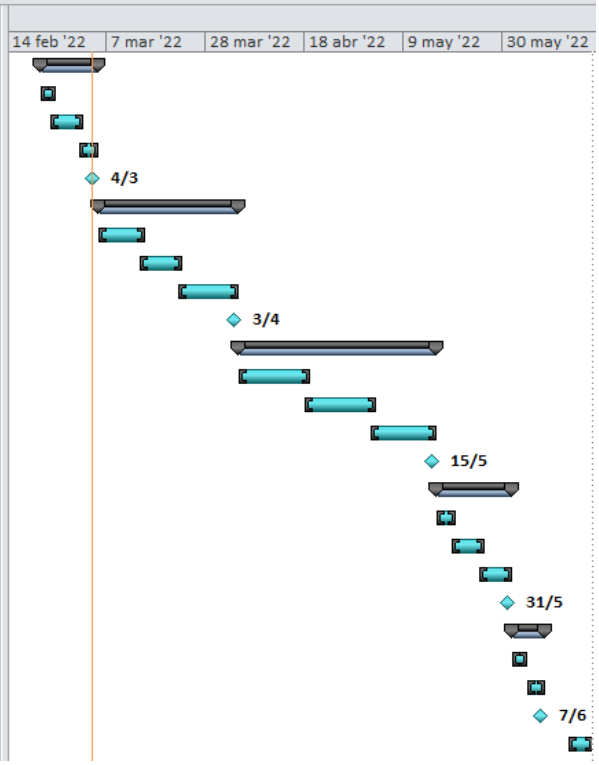
1.4. Planificació del treball

1.4.1. Establiment del calendari

Data d'inici del TFM: 21 de Febrer del 2022
Data de fi del TFM: 17 de Juny del 2022

1.4.2. Descomposició de tasques i Diagrama de Gantt

Nombre de tarea	Durac	Comienzo	Fin
1) Pla de treball	**10 días**	**lun 21/2/22**	**vie 4/3/22**
1.1) Llegir documentació TFM	3 días	lun 21/2/22	mié 23/2/22
1.2) Posar contexte i enfocar treball	5 días	mié 23/2/22	mar 1/3/22
1.3) Planificació del treball	4 días	mar 1/3/22	vie 4/3/22
1.4) Entrega PAC 1	0 días	vie 4/3/22	vie 4/3/22
2) Fonaments de IA i Ciberseguretat	**22 días**	**sáb 5/3/22**	**dom 3/4/22**
2.1) Origens, situació i bases IA	7 días	sáb 5/3/22	lun 14/3/22
2.2) Aprenentatge i classificació de les IA	7 días	lun 14/3/22	mar 22/3/22
2.3) Punts forts IA i refinar capitols	10 días	mar 22/3/22	dom 3/4/22
2.4) Entrega PAC 2	0 días	dom 3/4/22	dom 3/4/22
3) Relacionar IA i Ciberseguretat	**31 días**	**lun 4/4/22**	**dom 15/5/22**
3.1) Reptes i estat de IA i Ciberseguretat	11 días	lun 4/4/22	lun 18/4/22
3.2) Exemples i futur de IA i Ciberseguretat	11 días	lun 18/4/22	lun 2/5/22
3.3) Conclusions de IA i Ciberseguretat	11 días	lun 2/5/22	dom 15/5/22
3.4) Entrega PAC 3	0 días	dom 15/5/22	dom 15/5/22
4) Memòria final TFM	**12 días**	**lun 16/5/22**	**mar 31/5/22**
4.1) Revisió de la memòria	4 días	lun 16/5/22	jue 19/5/22
4.2) Bibliografia	5 días	jue 19/5/22	mié 25/5/22
4.3) Annexos	5 días	mié 25/5/22	mar 31/5/22
4.4) Entregar memòria final	0 días	mar 31/5/22	mar 31/5/22
5) Video + Presentació	**5 días**	**mié 1/6/22**	**mar 7/6/22**
5.1) Crear Presentacio	3 días	mié 1/6/22	vie 3/6/22
5.2) Crear Video	3 días	sáb 4/6/22	mar 7/6/22
5.3) Entregar Video + Presentació	0 días	mar 7/6/22	mar 7/6/22
6) Defensa TFM	**5 días**	**lun 13/6/22**	**vie 17/6/22**



1.4.3. Dates de control

Títol	Data
PAC 1	04/03/2022
PAC 2	03/04/2022
PAC 3	15/05/2022
Memòria Final TFM	31/05/2022
Vídeo + Presentació	07/06/2022
Defensa TFM	17/05/2022

1.5. Breu sumari del producte a obtindre

Com a productes finals de tot el explicat sobre la IA, la seva evolució i aplicació a problemes reals, tindrem per una banda, una visió del que és, del que pot fer, de les capacitats i punts forts d'una IA. I per l'altra, coneixements sobre la IA aplicada a la Ciberseguretat, del que pot millorar, que hi pot fer, millorar i solucionar. També, veure la importància d'un entrenament correcte de la IA, basat en uns principis de IA fiable. Així, tindrem una base de criteri sobre el important d'una IA com a gestor de Ciberseguretat. A més, també visionarem el futur de les IA a la Ciberseguretat, i com es pot assolir reptes de treballar amb més volums de dades, en menys temps i més eficiència.

1.6. Breu descripció dels altres capítols de la memòria

1. Introducció
   Plantejament del treball, objectius, planificació inicial i sumari dels productes a obtenir.
2. Orígens
   Referència històrica de l'aparició i evolució de la IA.
3. Situació actual
   Visió de l'ús actual de la IA en molts sectors, aportant beneficis, millores i afegint valor.
4. Bases de la IA

Exposició dels algorismes amb que sol treballar la IA, les seves variants i que resolen, en funció de model matemàtic.

5. Aprenentatge de la IA
   Exposició dels diferents mètode d'aprenentatge de les IA, detallar el automàtic (ML) i el profund (DL), i com tenir una IA fiable.
6. Classificació de les IA
   Mostrar les diferents classificacions en que s'agrupen les IA, en funció dels algorismes emprats, funcions i tipus d'aprenentatge.
7. Punts forts de les IA.
   Valoració dels principals trets de les IA, en que son bones i com poden millorar amb l'aprenentatge.
8. Situació i reptes de la Ciberseguretat
   Exposició de la situació actual de la Ciberseguretat, mancances, saturació d'alertes, capacitació, prioritats i volum de treball.
9. IA i Ciberseguretat
   Mostrar en que ens pot ajudar la IA a la Ciberseguretat, en la seva gestió, control, millora, reducció de risc i prevenció d'atacs.
10. Exemples de IA per Ciberseguretat
    Revisar alguns productes de Ciberseguretat que empren IA per donar solució i mostrar característiques o trets principals.
11. Futur de la IA a la Ciberseguretat
    Revisar tendències a curt i llarg termini de com pot evolucionar la IA a la Ciberseguretat, tenir visió dels beneficis i millores que pot oferir, i previsió d'una aplicació creixent a la Ciberseguretat.
12. Conclusions de IA a la Ciberseguretat
    Resum del que es important de les IA per la Ciberseguretat, que cal valorar, pros i contres de plantejaments, que ens pot interessar i quina millora de productivitat i resiliència tenim.
13. Glossari
    Explicació dels mots i passatges obscurs o difícils d'entendre.
14. Bibliografia
    Fonts d'informació utilitzades per a realitzar el TFM.
15. Annexos
    Altra informació de referència.

2. Orígens

De forma generalitzada, a la història es sol posar el naixement i fonaments de base de la IA, a partir dels treballs en 1950 de Alan Turing, matemàtic britànic, que obren una nova disciplina de les ciències de la informació. Publica l'obra "Computing Machinery and Intelligence", també un article acadèmic sobre si les màquines poden pensar, i també el conegut "Test de Turing". Però ja hi havia primer uns treballs sobre intel·ligència artificial, publicats en la década de 1940, que no van tenir gran repercussió. Encara que els més puristes, posen la fita de la fundació de la Intel·ligència Artificial (IA), com a terme i disciplina, a una conferència feta a la Universitat de Dartmouth (Dartmouth College) l'any 1956, organitzada per John McCarthy, Marvin Minsky, Claude Shannon i Nathaniel Rochester. En aquesta conferència els organitzadors van invitar a uns deu investigadors a formalitzar el concepte de intel·ligència artificial com un nou camp d'estudi científic. Ja sigui en una o altra data, es va donar o produir entre les dècades de 1940-1950, i va quedar definitivament establert el terme IA, en la segona dècada (1950), uns anys després del final de la segona guerra mundial (abril de 1945).

El cert es que aquesta idea, per emular el raonament humà de forma artificial, al llarg de la història, ha posat i aportat una sèrie de bases, i coneixements que son exposats amb una certa relació, de forma històrica en mites i llegendes, creences en idees o suposicions. I sobre tot, també en la literatura, en jocs, en teories, en invents, i descobriments o treballs que han donat suport al final, al naixement de la IA. Per tant, cal fer un repàs de tot això, ens cal veure quins son els precedents o antecedents, que hi ha al llarg de la història i que poden ser o estar relacionats amb dotar de raonament humà de forma artificial alguna cosa. I també ens cal veure quins son els coneixements i eines aparegudes a llarg de la història i que conformen les bases reals per al naixement de la IA.

Per altra banda, des del naixement del terme IA fins a avui en dia, la seva evolució o previsions de fites, a patit diferents entrebancs o períodes de congelació en el seu desenvolupament. Principalment sobre les prediccions que s'havien fet, i que no s'han acomplert per diferents motius. El problema sembla que ha estat el fet de proposar expectatives massa ambicioses, que la majoria de vegades no s'han pogut acomplir. De totes formes, a pesar de les ensopegades, hi ha hagut grans avenços, s'han aconseguit fites inesperades i actualment s'ha expandit el seu ús en molts camps i fins a sectors industrials. Així doncs, també ens caldrà veure quina ha estat la seva evolució, fins a l'actualitat, des de que es va definir el terme de IA.

2.1. Antecedents

Al llarg de la història hi ha una sèrie de mites, llegendes i textos, en que es fa referència al que podríem anomenar certa intel·ligència artificial, és a dir , en que es dona o s'atribueix cert raonament a objectes o coses. En aquest cas, podem trobar els següents:

- De l'antiga Grècia, tenim el mite de **Galatea**, escultura feta per **Pigmalión** a la que els deus condeixen el do de la vida.
- També a l'antiga Grècia, el mite dels **Trípodes** fabricats per **Hefesto**, que servien en les festes i banquets dels deus. I que també va crear unes ajudantes de metall que l'ajudaven en la seva feina diària.
- A la baixa Edat Mitja es troben varis textos místics jueus, que fan referència a la figura del **Golem**, una criatura de fang a la que el alè diví li dona vida per defensar el poble jueu.
- També a la baixa Edat Mitja, es pot trobar l'obra del àrab **Jàbir Ibn Hayyan** (en llatí: Geber, 721-813), filòsof i alquimista (és considerat el pare de la química), que en la seva obra explica con crear vida artificialment.
- A l'alta Edat Mitja (1493), tenim a **Paracelso** (àlies de Theophrastus Phillippus Aureolus Bombastus von Hohenheim) el qui ens descriu un procediment per crear petits sers humans (homuncles) a partir d'una sèrie de materials exòtics.
- Al segle XIX trobem la novel·la "Frankestein" (1818) de **Mary Shelley** (1797-1851), iniciadora del gènere de ciència ficció, i que és considerada com un antecedent de la intel·ligència artificial.
- A segle XX, tenim l'obra teatral "R.U.R." (Rossum's Universal Robots), al 1920, de **Karel Capek** (1890-1938), que introdueix per primer cop el terme "robot", en la que acaben eliminant als creadors i seran la nova humanitat.
- 1941, en **Isaac Asimov** (Isaac Ludovitx Asimov, 1920-1992) publica el conte "Cercle viciós", precursor que genera les 3 lleis de la robòtica.

D'altra banda, amb fonaments més seriosos o de forma més científica, tenim al llarg de la història, una sèrie de fets, plantejaments, treballs, invents, teories i descobriments, que fan unes aportacions tècniques i de coneixements per al naixement de la IA. De fet, alguns d'ells més aviat aporten alguns principis d'automatització, és a dir, ens aporten l'ús de sistemes de control automàtic racional. De entre tots ells, podem destacar els següents:

- 3000 aC[3] (aproximadament) "Torres de Hanoi", joc matemàtic antic, mostra interès en recerca resolutiva, guanyar amb mínims moviments possibles.

[3] aC: abans de Crist

- 300 aC (aproximadament) **Aristòtil** descriu de forma estructurada, un conjunt de regles (sil·logisme), que descriuen una part del funcionament de la ment, per obtenir conclusions racionals a partir de premisses donades.
- 250 aC (aproximadament) **Ctesibi** d'Alexandria, va construir la primera màquina autocontrolada, un regulador de flux d'aigua (racional però sense raonament).
- Segles VIII – IX, **Al Juarismi** amb el seu llibre "Al jabr", dona nom a la disciplina coneguda com àlgebra, aplicat en resoldre equacions de segon grau, i el seu mètode didàctic d'explicar les operacions o passos a fer, es va anomenar algorisme (o també algoritme).
- 1315, en **Ramon Llull** (1232-1316) al seu llibre "Ars Magna" va tenir la idea de que el raonament podia ser fet de forma artificial. Proposa la "Ars generalis ultima" una màquina mecànica que pot provar la veracitat o falsedat de certes afirmacions lògiques.
- 1703, en **Leibniz** (Gottfried Wilhelm von Leibniz , 1646-1716) publica "Explication de l'Aritmétique Binaire", fonamentat en la idea de Ramon Llull i en base a treballs d'altres matemàtics Indis i Xinesos. I també anuncia l'existència d'un llenguatge i màquina universals, que automatitzaria tot el mecanisme matemàtic de raonament, i per tant, el raonament humà.
- 1847, en **George Boole** (1815-1864) va establir la lògica proposicional (Booleana o àlgebra de Boole) amb el pamflet "Mathematical Analysis of Logic" desenvolupat el 1854 dins de la seva obra principal i més important.
- 1879, en **Gottlob Frege** (1848-1925) estén la lògica booleana i obté el que s'anomena "Lògica de Primer Ordre", que disposa d'un major poder d'expressió i aplicació d'aquesta lògica.
- 1842, la **Ada Lovelace** (Augusta Ada Byron King, comtessa de Lovelace, 1815-1852) crea el que es considera el primer algorisme que es va intentar aplicar a una màquina.
- 1903, en **Lee de Forest** (1873-1961) inventa el tríode, una vàlvula termoiònica de tres elèctrodes, base de l'electrònica en vàlvules de buit.
- 1937, en **Alan Turing** (1912-1954) publica un article sobre els nombres calculables, que estableix bases teòriques per a les ciències de computació. I introdueix el concepte de "Màquina de Turing" que formalitza el concepte d'algorisme, i que serà la precursora de les computadores digitals.
- 1940, en **Alan Turing i el seu equip** construeixen el que serà el primer ordinador electromecànic.
- 1941, en **Konrad Zuse** (Konrad Ernst Otto Zuse , 1910-1995) crea la primera computadora programable Z3, i el que es el primer llenguatge d'alt nivell (Plankalkül). Encara que del model Z1, es diu que va estar a punt el 1938. Les següents màquines que es creen, ja més potents, son l'ABC

(Atanasoff Berry Computer) i ENIAC (Electronic Numerical Integrator And Computer).

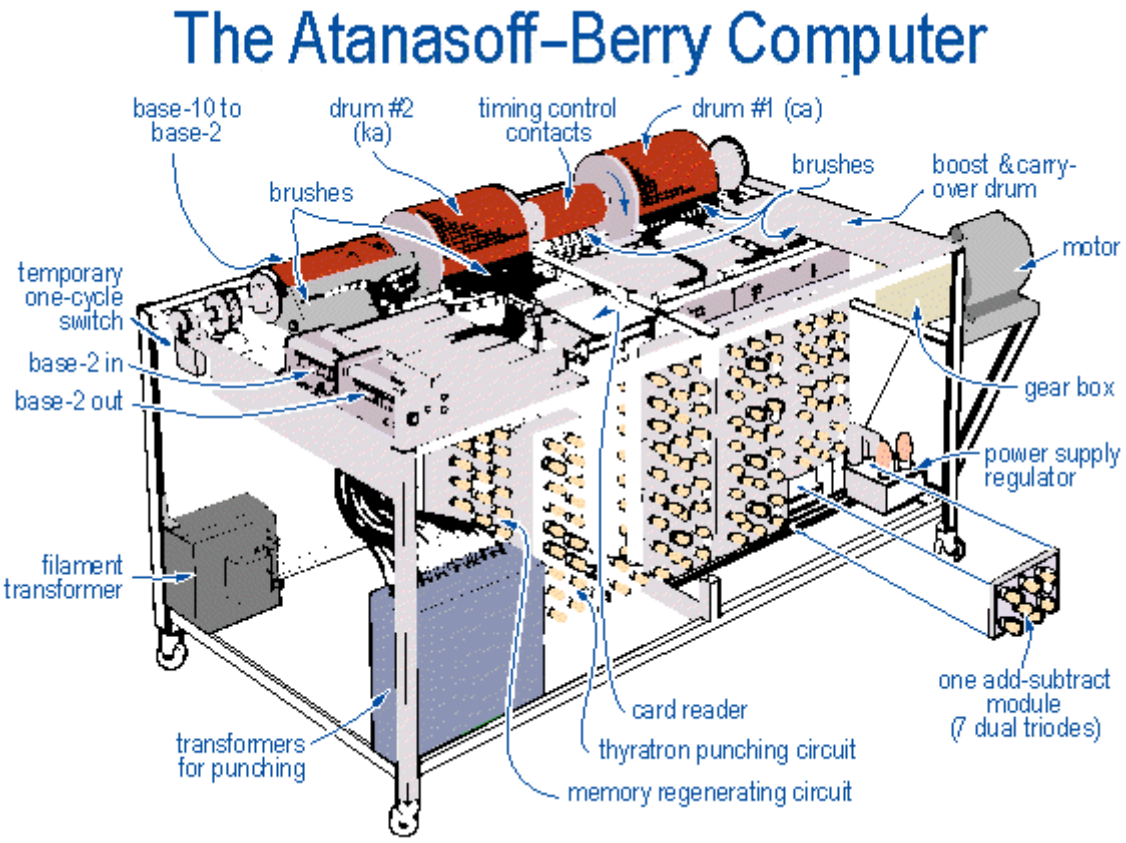


Figura 1: Ordinador ABC, font: Wikipedia

- 1943, en **Warren McCulloch i WalterPitts** presenten el seu model de neurones artificials, que es considera el primer treball de camp d'IA, a pesar de que el terme encara no existeix.
- 1950, en **Alan Turing** publica el seu article "Computing Machinery and Intelligence", que consolida el camp de la intel·ligència artificial amb la seva famosa "Prova o test de Turing", per tal de determinar si una màquina era intel·ligent o no.
- 1951, en **Willian Shockley** (William Bradford Shockley, 1910-1989) inventa el transistor d'unió, cosa que permet la creació d'ordinadors molt més ràpids i petits.
- 1956, a la **Universitat de Dartmouth**, es consolida el terme "intel·ligència artifical" (IA) a una conferència convocada per McCarthy, Minsky i altres, a la que assisteixen Allen Newell i Herbert Simon. Les seves previsions a 10 anys vista mai es van complir, pel que es va donar un abandonament de les investigacions durant 15 anys, van sorgir molts pocs avenços.

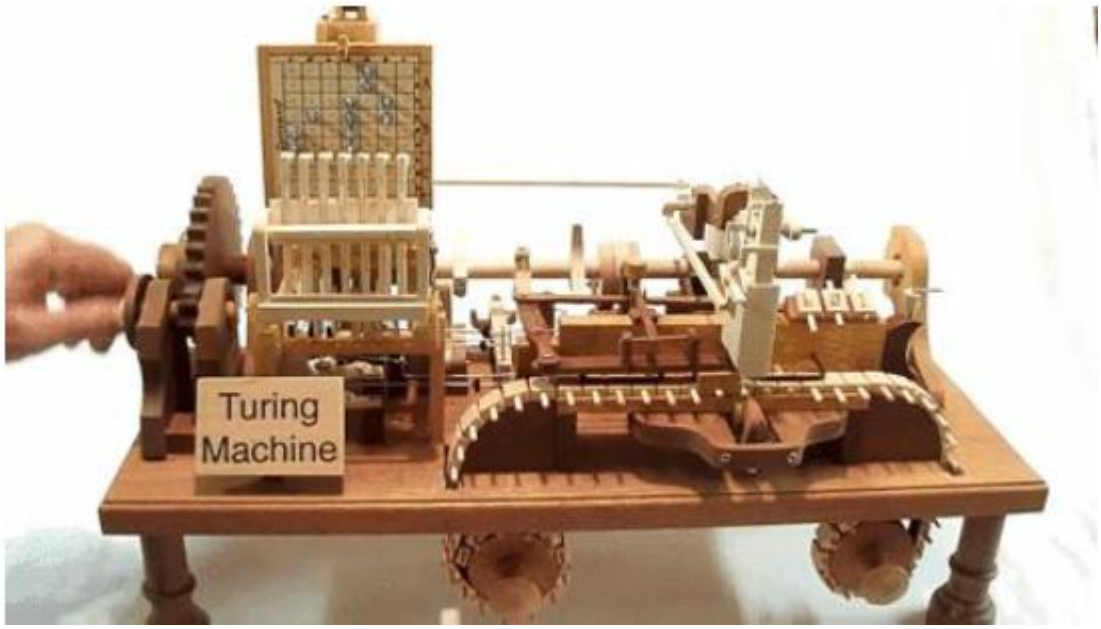


Figura 2: Màquina de Turing, font: http://www.cs.us.es/~fsancho/?e=221

2.2. Evolució

Agafem com a punt de sortida la consolidació del terme "intel·ligència artificial" (IA), i a partir d'aquest moment, es donen tota una sèrie d'avenços fins al moment actual, dels que anirem veient els considerats més significatius. Amb això, ens adonarem de que hi ha certs punts de congelació en els avenços al llarg de temps, degut a diversos motius. De totes formes, a partir de 1956, sempre a anat sortint algun que altre avenç, que ha tingut aplicació en el mon de les IA, fins arribar a l'actualitat, en que hi ha un creixement exponencial. Així doncs, a partir del moment en que es fixa el terme IA (intel·ligència artificial), principalment tenim els avenços següents:

- 1956, en **Allen Newell** (1927-1992) i **Herbert Simon** (1916-2001) coautors de "Logic Theorist", considerat el primer programa informàtic de IA. El any següent, treuen "General Problem Solver" (GPS) que resolt problemes.
- 1958, en **John McCarthy** (1927-2011) crea al MIT (Institut tecnològic de Massachusetts) el llenguatge LISP, el primer en processament simbòlic.
- 1959, en **Frank Rosenblatt** (1928-1971) crea el "Perceptró"[4], la primera unitat neuronal artificial que pot ser entrenada en base unes regles.
- 1960, en **Robert K. Lindsay** (?-?) desenvolupa "Sad Sam", programa de lectura d'oracions en anglès i la inferència de conclusions a partir de la seva interpretació.

[4] Perceptró: neurona artificial o unitat bàsica de inferència en forma de discriminador lineal, a partir de la que es desenvolupa un algorisme capaç de generar un criteri de selecció.

- 1961, en **George Devol** (George Charles Devol, 1912-2011), crea el robot "Unimate", el primer a nivell industrial que es posa a una línia de muntatge.
- 1964, en **Bertrand Raphael** (1936) construeix el sistema SIR (Semantic Information Retrieval) capaç d'inferir coneixement en base a la informació subministrada.
- 1964, en **Daniel G. Bobrow** (1935-2017) crea "STUDENT" amb LISP, programa que llegeix i resol el tipus de problemes de paraules que es troben als llibres d'àlgebra de secundària.
- 1966, en **Joseph Weizenbaum** (1923-2008) científic informàtic al MIT, crea "ELIZA" el primer chatbot[5] que conversava funcionalment en anglès amb una persona.
- 1965-1970, 2a. meitat dècada dels 60, apareixen els "**sistemes experts**"[6] que prediuen la probabilitat d'una solució sota un conjunt de condicions. En tenim alguns com "DENDRAL" (fet per Buchanan, Feigenbaum i Lederberg) que assistia a químics en estructures químiques complexes, ó "MACSYMA" (fet pel MIT) que assistia a enginyers i científics en solucions d'equacions matemàtiques complexes.
- 1968, en **Seymour Papert** (1928-2016), **Danny G. Bobrow** (1935-2017) i **Wally Feurzeig** (1927-2013) desenvolupen el llenguatge de programació LOGO. I en la mateixa data, **Marvin Minsky** (1927-2016) publica "Semantic Information Processing".
- 1969, en **Marvin Minsky** (1927-2016) i **Seymour Papert** (1928-2016) publiquen "Artificial Intelligence", que explora les fortaleses i limitacions dels "perceptrons" i determinen l'incapacitat d'implementar la funció lògica XOR.
- 1973, en **Alain Colmenauer** (1941-2017) i equip d'investigació de la Universitat de AIX-Marseille creen "PROLOG" (PROgrammation en LOGique) molt emprat a IA. En la mateixa data, **Shank** i **Abelson** creen els "scripts"[7], base de moltes tècniques actuals de IA i la informàtica en general.
- 1974, en **Edward H. Shortliffe** (1947) escriu la seva tesis amb "MYCIN" un "sistema expert" força conegut, que assistia a metges en diagnòstics i tractament d'infeccions a la sang.
- 1979, en **Hans Berliner** (1929-2017) professor de la Universitat Carnegie Mellon, crea al programa BKG 9.8, que guanya al "bakgammon"[8] al campió mundial "Luigi Villa".
- 1970-1980, creixement de l'ús de "sistemes experts" com: MYCIN, ABRL, R1/XCON, PXDES, PIP, PUFF, CASNET, INTERNIST/CADUCEUS, Cadet,

[5] Chatbot: assistent que es comunica amb els usuaris per missatges de text.
[6] Sistemes experts: emulen el raonament tal i com ho faria un expert en un àrea de coneixement.
[7] Scripts: guió o conjunt d'instruccions informàtiques
[8] Bakgammon: joc de taula per 2 persones que uneix atzar amb forts coneixements estratègics.

etc., del que alguns encara s'usen avui en dia com a "shells"[9], com: EMYCIN, EXPERT ó OPSS.

- 1981, en **Kazuhiro Fuchi** (1936-2006) anuncia el projecte japonès de la "quinta generació de computadors", que dona peu a un auge de "sistemes experts", però no va assolir objectius i genera una congelació als 90.
- 1986, en **James L. McClelland** (1948) i **David E. Rumelhart** (1942-2011), publiquen "Parallel Distributed Processing" que tracta sobre xarxes neuronals. En la mateixa data, **Michael I. Jordan** (1956) introdueix una arquitectura d'aprenentatge supervisada en seqüència de dades, RNNs (Recurrent Neural Networks).
- 1987, en **Martin Fischl** i **Oscar Firschein**, descriuen els atributs d'un agent intel·ligent, en un major àmbit, cosa que fa expandir les IA en moltes àrees, que han creat branques d'investigació enormes i diferenciades.
- 1988, apareixen el llenguatges orientats a objectes, POO (Programació orientada a objectes), en anglès OOP.
- 1997, la computadora autònoma "Deep Blue" de IBM, guanya al campió mundial d'escacs Garri Kaspàrov.

Figura 3: Partida d'escacs Garri Kaspàrov, font: Pinterest

- 2002, apareix "Roomba" el primer robot d'èxit comercial per a la llar, que és una aspiradora autònoma que porta aquest nom.
- 2005, un cotxe autònom desenvolupat per la Universitat de Stanford, guanya una competició de vehicles robot, amb 212 km de desert, sense suport humà.

[9] Shells: consola de comandes

- 2009, en **Fei-Fei Li** (1976), llença “ImageNet”, una base de dades gratuïta de 14 milions d’imatges, que molts investigadors d’IA fan servir per entrenar les xarxes neuronals, per catalogar fotos i identificar objectes.
- 2011, el superordinador “Watson” de IBM guanya el concurs de Televisió “Jeopardy!”, competint amb els dos màxims campions. En la mateixa data, Apple, Google i Microsoft llencen aplicacions mòbils amb assistents virtuals.

Figura 4: Assistent virtual Cortana, font: Pinterest

- 2012, s’utilitza una xarxa neuronal convolucional, i guanya un concurs de reconeixement d’imatges sobre “ImageNet”, amb rendiment espectacular. En la mateixa data, Google crea un ordinador capaç d’aprendre via Youtube a identificar cossos humans, i altres essers vius com gats.
- 2014, un ordinador desenvolupat a Rússia, de nom “Eugene” es va fer passar per un nen de 13 anys i supera el “Test de Turing”. En la mateixa data, **Ian J. Goodfellow** (1985), introdueix les xarxes generatives adversàries (GAN) que empra dos xarxes neuronals, enfrontant-les per tal de generar noves instàncies sintètiques de dades[10]. També en aquesta data “Amazon” llença al mercat, “Alexa”, assistent virtual intel·ligent per veu.
- 2015, apareixen dos llibreries de codi obert, ”TensorFlow” i “PyTorch” que passen a ser per defecte la base de projectes d’aprenentatge automàtic. És a dir, impulsen molt l’aprenentatge profund (DL – Deep Learning).
- 2016, Microsoft llença “Tay”, un “chatbot” capaç d’aprendre de la interacció amb persones, amb sols 1 dia d’interacció, es va tornar racista, xenòfob i homòfob. En la mateixa data, la IA de Google “Alpha Go”, guanya al campió mundial “Ke Jie” en el complex joc del Go.

[10] Instàncies sintètiques de dades: es fa servir abastament en la generació d’imatges, vídeo i veu.

- 2017, el motor d'escacs "Stockfish", considerat el millor del mon, amb 3.400 punts ELO[11], és guanyat per "AlphaZero" que sols coneixia les regles de joc i només tenia 4 hores d'entrenament, jugant amb ella mateixa. En la mateixa data, el software d'algorisme "Libratus" de la Universitat Carnegie Mellon, guanya al Poker contra rivals humans.
- 2018, Google desenvolupa "BERT", la primera representació de llenguatge bidireccional i sense supervisió, que es pot emprar en moltes tasques del llenguatge natural, com respondre a preguntes. En aquesta data, es llença el primer Televisor sobre IA, per part de "LG Electronics", sota la plataforma de nom "ThinQ". També en aquesta data, **s'inicia l'assentament de la IA en sectors importants del teixit productiu**, com la automoció en quant a la conducció autònoma per part de Tesla, Audi i altres.
- 2019, Google presenta "Doodle", una IA que representa un homenatge a "Johann Sebastian Bach", afegint una melodia simple de dos compassos, la IA crea la resta. En aquesta data, **s'intensifica la utilització de solucions basades en IA**, per part d'altres sectors productius com Turisme, Mobilitat o Assegurances, que fan servir algorismes de reconeixement d'imatges i de models predictius, per avaluar la probabilitat de tenir un accident.
- 2020-2021, la situació de pandèmia (COVID-19) afavoreix els **avenços de la IA en el camp de la Salut**, emprant sensors tèrmics automatitzats o aplicant eines de "Big Data"[12] per la detecció ràpida de pacients zero i el control de focus de contagi.

Figura 5: Humà i robot IA, font: Pinterest

[11] ELO: sistema de puntuació de mètode matemàtic, basat en càlcul estadístic sobre la habilitat relativa a jugadors d'esports com els escacs.

[12] Big Data: terme que descriu un gran volum de dades, estructurades i no estructurades, que permet el seu anàlisi massiu via mètodes no tradicionals.

En base a la cronologia anterior, podem observar que més o menys sempre a anat apareixent alguna idea, treball, invent, investigació, publicació o tendència al llarg del temps, emprada o aplicable a IA. De totes formes, tradicionalment, es consideren varies etapes en el temps, com a períodes d'hivern o congelació en la evolució de la IA, que son:

- 1974-1980 primer hivern evolució IA, degut a molts inicis en fals i carrerons sense sortida, deixen a la investigació en IA sense fons i amb poc interès.
- 1987-1993 segon hivern evolució IA, el fracàs d'assolir objectius de la "quinta generació de computadors", repercuteix en una forta caiguda d'interès i finançament en investigació en IA.

3. Situació actual

Revisant la cronologia anterior, es pot dir que per entendre l'actual expansió i penetració de la IA, en molts camps i sectors de producció, hi ha tres dates històriques amb aportacions clau:

- La primera el 1987, amb Martin Fischl i Oscar Firschein, que van descriure els atributs d'un agent intel·ligent, no sols en la comunicació, sinó en un major àmbit. Aquest fet, va provocar l'expansió de la IA en moltes àrees, cosa que va crear branques d'investigació enormes i diferenciades.
- La segona el 2011, s'aplica el "Processament de Llenguatge Natural" (PLN) que crea una nova disciplina dins de la IA, que uneix tècniques clàssiques amb rames molt dispars (Neurociència, Lingüística, Anatomia, Psicologia, etc.). I com a representant destacat tenim el sistema "Watson" de IBM que interactua i guanya un concurs televisiu ("Jeopardy!").
- La tercera el 2015, quan apareix dos llibreries de codi obert, "TensorFlow" i "PyTorch" que passen a ser per defecte la base de projectes d'aprenentatge automàtic. Cosa que fa que s'estengui l'ús de IA molt ràpidament en entorns educatius, en petites empreses tipus "startup"[13], en diferents investigacions públiques o privades, i altres camps i àrees.

Podem contrastar que a partir de la primera fita, comencen a sorgir unes IA que superen als humans en certes disciplines o jocs, i que van evolucionant en base a desenvolupaments privats de grans empreses i/o aplicació de "sistemes experts". En la segona fita, ja es dona una interacció per veu amb la IA, via el que es coneix com PLN (Processament de Llenguatge Natural, NLP en anglès), i que permet fins i tot, interpretar el doble sentit d'oracions dins de l'àmbit. A

[13] Startup: o empresa emergent, de nova creació que comercialitza productes i/o serveis que fa us intensiu de les TIC.

partir de la tercera fita, ja apareixen aplicacions de IA més professionals i de ventall més ampli. També es produeix una major adopció de la IA per part de grans empreses tecnològiques i altres sectors productius i educatius. A més, en aquesta data (2015) Elon Musk, Sam Altman i altres, creen una organització sense ànim de lucre, "Open AI", encarregada d'iniciar i promoure investigacions que permetin un avenç de la IA en pro d'un impacte positiu per la humanitat.

De fet, el gran canvi es dona gràcies al Aprenentatge Automàtic (ML – Machine Learning) i després al Aprenentatge Profund (DL – Deep Learning), ja que fins a les hores la forma més emprada per solucionar problemes de IA era amb la construcció d'algorismes, més o menys generals. Amb l'aprenentatge, s'empra una aproximació diferent per resoldre el problema, passa per emprar dades reals del comportament desitjat, per a construir una màquina que sigui capaç de reconèixer-ho i simular-ho. Encara que també es genera igualment uns algorismes genèrics de cerca en l'espai de solucions i algorismes d'optimització generals. Això, li permet ajustar correctament la màquina al comportament observat en les dades, independentment del domini del problema. Cal tenir en compte que a partir d'aquí, apareix la Intel·ligència Cognitiva, que es la associació amb èxit del NLP amb el ML. D'aquesta forma arriba el que es coneix com la "primavera" en l'evolució de les IA.

A partir de tot l'anterior, hi ha una gran explosió en l'aplicació de la IA, creix de forma constant, en molts camps, àrees i sectors productius. No sols s'aplica en grans tecnològiques o en l'educació i investigacions, molts desenvolupaments passen a ser un servei amb clients reals, en domini públic o privat. Podem trobar l'aplicació de IA en un munt d'àrees com:

- Assistents personals virtuals: llenguatge natural, reconeixement de veu, chatbots, agents experts, ...
- Llar: Domòtica o dispositius en la llar, ...
- Finances: filtres de SPAM, patrons de mercat, detecció de frau, ...
- Educació: plataformes "Machine Learning", ofertes personalitzades, ...
- Tecnologia: Big Data, anuncis personalitzats, videojocs, ...
- Comercial: automatització de processos robòtics, de marketing, ...
- Sanitat: diagnòstics mèdics, chatbots o altres, ...
- Climàtic: consum energia, previsions, ...
- Agrícoles: millora de rendiment, optimització tasques, previsions, ...
- Automoció, logística i transport: vehicles autònoms, optimització trànsit, ...
- TIC: defensa Cibernètica, reconeixement d'emocions i imatges, ...

Si mirem a fons, en quant a aplicació de IA, en podríem trobar moltes més, però amb aquesta llista, ja ens fem una idea. A més, aquesta adopció de la IA i

desenvolupament en tot tipus de sector, es troba impulsada principalment per tres factors, que serien:

- La capacitat de computació i d'alt rendiment ja està disponible.
- Es troba disponible grans volums de dades per l'aprenentatge.
- La IA aplicada proporciona un avantatge competitiu.

Figura 6: Cos humà analitzat, font: Pinterest

En el cas de les empreses, tenim que la tecnologia de IA millora el rendiment i la productivitat, per mitjà de l'automatització de processos o tasques que abans requerien esforç humà. La IA també pot donar sentit a dades a una escala que cap humà mai podria fer, cosa que pot generar importants avantatges per a l'empresa. Segons darrera enquesta de "Gartner"[14], la analítica i la intel·ligència empresarial, son les tecnologies de diferenciació més importants per a una organització, és a dir, son les tecnologies més estratègiques d'una empresa. En canvi, en la realitat, tenim que segons una publicació de "Harvard Business Review", les empreses fan servir la IA principalment per :

- Detectar i dissuadir intrusions de seguretat (44%).
- Resoldre problemes tecnològics dels usuaris (41%).
- Reduir el treball de la gestió de producció (34%)
- Medir l'acompliment intern en l'ús de proveïdors aprovats (34%).

[14] Gartner: empresa consultora i de investigació de les tecnologies de la informació, que fa una sèrie de publicacions molt importants i seguides fortament pel mon de les TIC.

Per altra banda, existeixen nombroses històries d'èxit que mostren el valor de la IA, i les empreses que incorporen l'aprenentatge automàtic i les interaccions cognitives, milloren l'experiència i la productivitat de l'usuari. En els casos que no, es degut a que la base no està prou consolidada i l'empresa no ho ha implementat d'una forma equilibrada, a part de que fa falta experiència en el tema. Per tant, saber quant, com i on incorporar la IA, o recórrer a un tercer, pot ajudar a minimitzar les dificultats, i així, assolir una implantació amb èxit. En sintonia amb tot l'anterior, **Andrew Y. Ng** (1976), director del laboratori de IA a la Universitat de Stanford, director de Landing AI i deeplearning.ai, declara: "La IA es tant important com la electricitat ho va ser en la seva època. Em resulta difícil imaginar un sector que la IA no vagi a transformar".

De totes formes, a pesar d'aquesta "primavera", cal tenir en compte un informe del 2018, escrit per 26 experts especialitzats en IA, que advertia dels perills que pot tenir aquesta tecnologia. Per exemple: casos de Cibercriminalitat, com uns terroristes o delinqüents poden modificar les IA disponibles al mercat o fer vídeos falsos, molt versemblants, per destruir reputacions.

En quant a l'aplicació de la IA a Catalunya, la Generalitat de Catalunya el Juliol de 2019 va realitzar un estudi, en el qual s'afirmava: "179 empreses treballen amb IA a Catalunya". La majoria d'aquestes empreses, desenvolupen software i dispositius, però també n'hi ha que creen algorismes, realitzen consultories o proveeixen serveis relacionats amb IA. Actualment podem destacar dos programes o projectes dins de l'àmbit de Catalunya, que son:

- **Catalonia.AI: Estratègia d'intel·ligència artificial de Catalunya** que consisteix en un programa d'actuacions multisectorials de suport a l'evolució de la IA a Catalunya, que està coordinat per Departament de Polítiques Administració i amb el suport del Govern de la Generalitat de Catalunya.
- **AI4EU: Projecte de plataforma europea a Catalunya** que és un projecte de la "Artifical Intelligence for Europe", que pretén crear un ecosistema en IA en l'àmbit europeu, que pugui desenvolupar i oferir als usuaris algorismes, eines, recursos i coneixement.

Per altra banda, a Catalunya es disposa de diferents centres tecnològics, que realitzen aportacions i investigacions en l'entorn IA. El nombre total d'ells, es pot aproximar a una vintena, i entre tots aquestos, podem destacar com els més coneguts / rellevants:

- Barcelona Supercomputing Centre (BSC)
- Grup d'Intel·ligència Artificial d'Alt Rendiment
- Institut d'investigació en intel·ligència Artificial (IIIA)
- Centre de Visió per Computador (CVC)

- Centre d'Investigació i Innovació Tecnològica de la UPC
- Institut de Robòtica i Informàtica Industrial (IRI)

4. Bases de la IA

En el camp d'investigació de IA, actualment hi ha dos objectius centrals que es persegueixen. El primer objectiu és l'estudi dels processos cognoscitius en general, que s'orienta cap a la consideració de la IA com a estudi de conducta humana intel·ligent. I el segon objectiu de IA, és el intent d'obtenir sistemes automàtics capaços de realitzar feines reservades als éssers humans. Apareix com una disciplina que persegueix construir màquines i programes que intenten fer tasques d'una forma igual o superior a la del ésser humà.

D'altra banda, com a mecanisme bàsic, tenim que la tècnica pròpia de la IA és obtenir una adequada representació simbòlica del coneixement. D'aquesta forma, tenim dos mecanismes principals per fer això, que son:

- **La inferència simbòlica**: els seus mecanismes de treball típics, inclouen la deducció. Per exemple, tenim el "modus ponens" en que un primer fet pot afirmar el segon (regles tipus: "si a llavors b"). I el "modus tollens" que és la anomenada abducció, a partir d'un segon fet podem postular un primer (regles tipus: "si d llavors c").
- **L'heurística**: és un conjunt de criteris, mètodes o principis que s'utilitzen per trobar entre diversos resultats possibles, quin o quins són els més efectius per aconseguir un objectiu. Està relacionat amb els mecanismes experimentals i empírics, amb la síntesi d'aquests s'acaba elaborant regles de l'experiència que es fan servir per seleccionar un camí d'acció davant tots els possibles casos.

La representació del coneixement emprat a la IA, no són sols unes estructures de dades que donen un sistema de representació eficient dels coneixements, sinó també quin s'ha de representar en cada ús. Els components essencials són les estructures de dades i procediments d'interpretació i maneig. Els coneixements a representar són els objectes, els processos i l'entorn en que existeixen objectes i processos. I a més, cal afegir la representació de diversos objectius, motivacions, elements de casualitat i temporalitat. Per gestionar aquest ús del coneixement, es consideren tres etapes: **l'adquisició**, intenta acumular coneixement nou i relacionar-lo amb el que ja tenia, **la recuperació**, que vol aconseguir el coneixement necessari per resoldre una qüestió en concret, i **el raonament**, en que es prova d'inferir alguna cosa a partir del coneixement ja seleccionat.

Les propietats que s'han de tenir en la representació del coneixement són: **l'adequació de representació**, que és la possibilitat de representar totes les classes de coneixement necessàries, **l'adequació inferencial**, és la possibilitat de manipular estructures de dades que serveixen per representar el coneixement, podent generar noves estructures que es corresponen a nous coneixements inferits dels anteriors, **l'eficiència inferencial**, és la capacitat d'incorporar nou coneixement addicional, que serveix per millorar l'ús del mecanisme d'inferència i optimitzar el funcionament de la IA, i **l'eficiència en l'adquisició del coneixement**, que és la facilitat d'adquirir informació nova incloent-hi el fet que el mateix sistema ha de ser capaç de controlar l'adquisició del nou coneixement.

Sobre l'anterior, un dels primers formalismes va ser "l'espai d'estats", que descriu tots els moments i passos del problema i de la solució. Actualment es fa servir esquemes més moderns sobre la representació del coneixement, i així tenim els esquemes de tipus declaratiu:

- **Esquemes lògics**: els fets i les formules son expressats el algun sistema basat en la lògica, emprant la lògica de primer ordre del càlcul de predicats.
- **Xarxes semàntiques**: son sistemes adequats per establir classificacions i són gràfics. Al graf que conforma una xarxa semàntica, els nodes hi posen objectes o conceptes, i els arcs narren les relacions entre els nodes.

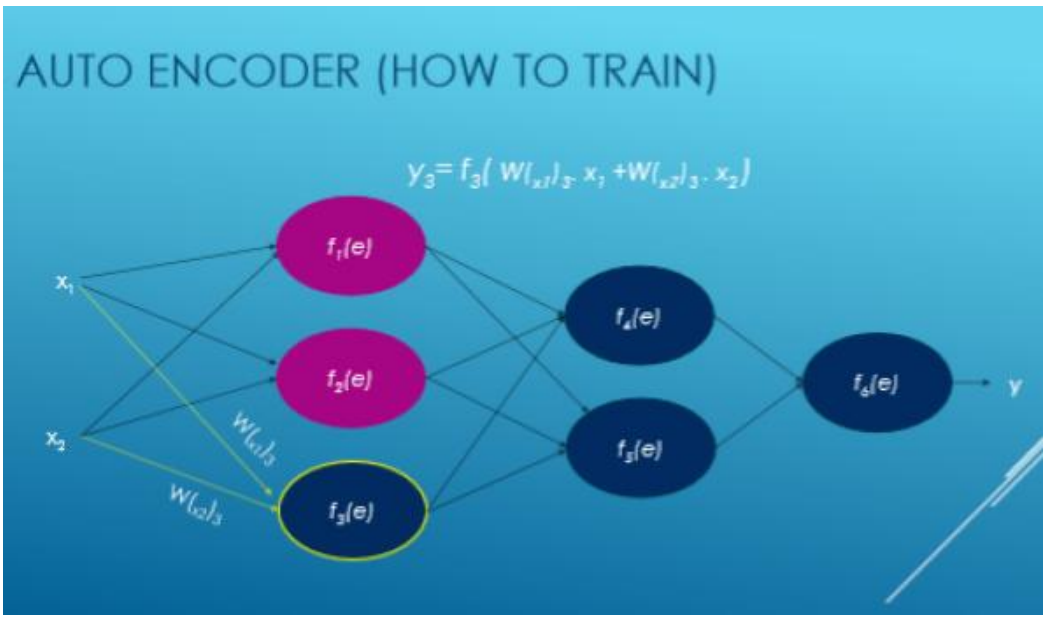


Figura 7: Graf representació neuronal, font: Pinterest

I també tenim els esquemes de representació de procediment o sistemes de producció, anomenats també "sistemes experts". Aquestos es basen en la idea d'emmagatzemar el coneixement en forma de procediments, enlloc de fer-ho

en proposicions. Això, és un conjunt de regles de producció que constitueix la base de coneixements. I com esquema de representació, tenim:

- **El frame**[15], que pot semblar una xarxa semàntica complexa d'estructures, i que s'utilitza per descriure el conjunt d'atributs que té determinat objecte o una situació complexa. També pot establir una jerarquia que és apta per fer classificacions.
- **Els Scripts**, que són estructures de representació mixtes especialitzades, que descriuen seqüències d'esdeveniments en un context particular.
- **Sistemes experts**, o sistemes basats en el coneixement, tenen la finalitat de reproduir correctament el coneixement d'un expert humà en el seu domini de competència. Se'ls sol classificar de forma exagerada, com: "la intel·ligència artificial amb èxit".

Per altra banda, fora del tractament i representació del coneixement, ens cal tenir en compte que hi ha dos escoles de pensament sobre les IA, que serien:

- **Intel·ligència artificial convencional**: és coneix com IA simbòlic-deductiva, i es basa en l'anàlisi formal i estadístic del comportament humà en front a diferents problemes. En aquest grup hi tenim :
    - Raonament basat en casos: ajuda a la presa de decisions mentre es resolent uns problemes concrets i requereix un bon funcionament.
    - Sistemes experts: infereixen una solució per mitjà del coneixement previ del context en que s'aplica i s'ocupa d'unes regles o relacions.
    - Xarxes bayesianes: proposa solucions mitjançant inferència de probabilitats.
    - IA basada en comportament: té certa autonomia i pot regular-se i controlar-se sola per tal de millorar.
    - Smart process management: facilita la presa de decisions complexes i proposa una solució a un determinat problema, igual que ho faria un especialista en el cas.
- **Intel·ligència artificial computacional**: És coneix com IA subsimbòlic-inductiva, i això implica desenvolupament o aprenentatge interactiu, com modificacions interactives dels paràmetres en sistemes de connexions, l'aprenentatge es fa en base a dades empíriques. Te dos objectius: el primer entendre els principis que possibiliten el comportament intel·ligent i el segon, especificar els mètodes per dissenyar sistemes intel·ligents.

A part de tot l'anterior sobre els objectius, tractament i representació del coneixement, per que això pugui ser funcional, s'acaba treballant via l'aplicació

[15] Frame: estructura de dades que conté una descripció general d'un objecte, que es deriva de conceptes bàsics i de l'experiència.

d'**algorismes** (o algoritmes), entre els diferents valors i nombres. En aquest punt, responem a una sèrie de dubtes o preguntes sobre els algorismes:

- Que és un algorisme (o algoritme)?: en poques paraules, son instruccions pas a pas que ajuden a una computadora a completar un càlcul. Sense ells no es podria iniciar un procés de programació.
- Per què es fan servir els algorismes?: donen instruccions descriptives que fan falta per dur a termes tasques específiques, i ens permet automatitzar processos que els humans solen fer a mà.
- On s'utilitzen els algorismes?: a tot arreu en l'ús d'eines tecnològiques, fins i tot l'accés a Internet es regeix per milers d'algorismes diferents.
- Com funciona un algorisme de IA?: bàsicament és un subconjunt estès d'aprenentatge automàtic que li diu a la computadora com aprendre a operar per si mateixa. El dispositiu sempre va adquirint coneixements per millorar els processos i executar tasques de forma més eficient. Quanta més interacció tingui, millor serà la seva capacitat de "notar" les preferències individuals.

En el nostre cas sobre les IA, podem veure varis dels algorisme clàssics que es fan servir, i alguns dels més bàsics i anomenats, són:

- **Regressió Lineal**: cerca una relació lineal entre les diferents variables, que tenen valors numèrics, pel que es tracta d'un problema de regressió o predicció. Sol ser un algorisme molt emprat en estadística i l'aprenentatge automàtic (ML). Té alguns models tipus, com: regressió simple o múltiple i rectes de regressió.
- **Regressió Logística**: ens permet veure els valors categòrics com un 0 o un 1. Es fa servir quan es vol conèixer si uns esdeveniment succeirà o no i per trobar els valors amb les variables independents es fa servir la estimació de "Maximun likelihood" o la màxima versemblança.

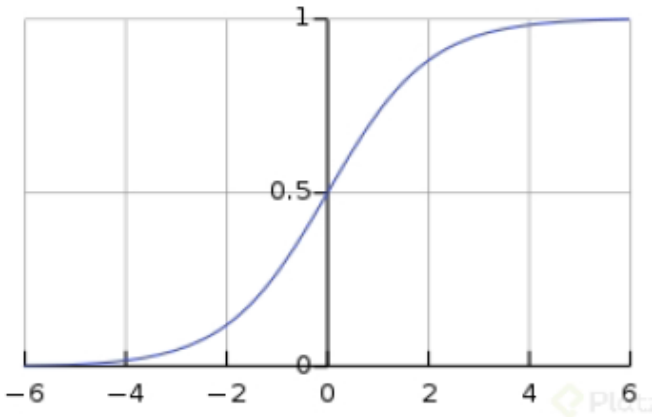


Figura 8: Gràfic Regressió Logística, font: Wikipedia

- **Naive Bayes**: conegut com classificador bayesià ingenu, aquest algorisme és específic per la classificació. Exemple: per classificar e-mails i decidir quins d'ells son spam o no.
- **K-nearest Neighbors**: serveix per predir un valor numèric i classificar un valor categòric. Pot ser emprat per resoldre tant problemes de regressió predictiva com de classificació. Treballa directament amb tot el conjunt de dades d'entrenament, si tenim un K igual a cinc, aquest serà el nombre de veïns propers i si es vol saber quin està més a prop, es pot mirar distàncies com Euclidiana, Hamming, o Manhattan.
- **Decision Tree**: o arbre de decisió alternatiu, que també ens serveix per predir i classificar via decisions. A mesura que es prenen decisions sobre situacions anirem avançant i sobre les rames del arbre de decisió, aquest es dividirà depenent de la quantitat de decisions disponibles o fetes.
- **Random Forest**: el mateix vist al punt anterior "Decision Tree" es aplicable a aquest algorisme, que ens permet la predicció de valors numèrics i categòrics. Es sol anomenar "Ensamble", ja que pot treballar amb un grup d'algorismes.

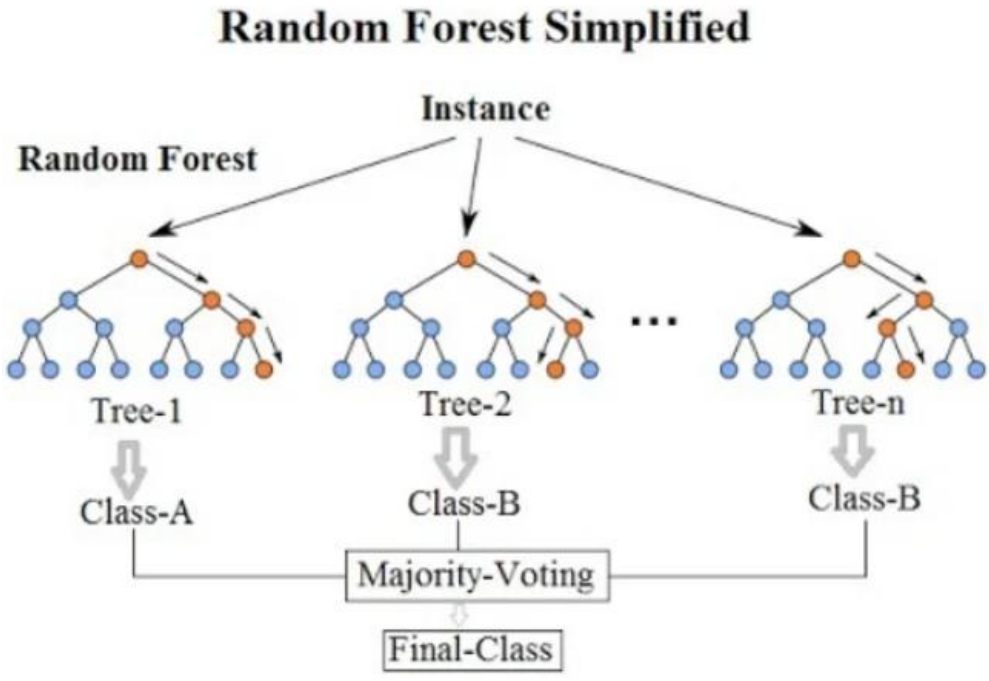


Figura 9: Gràfic Random Forest, font: Wikipedia

Cal tenir clar, que per que tot aquest conjunt de tècniques, funcionalitats i algorismes sobre el coneixement ens funcioni correctament, ens **cal fer un entrenament de la IA**. És a dir, cal fer un entrenament correcte, amb totes les variants possibles, i que ens permeti afinar el "raonament" de la IA, ja que si donem informació esbiaixada, obtindrem una IA tendenciosa, amb preferències.

5. Aprenentatge de la IA

En el punt anterior hem vist que per la gestió del coneixement, es consideren tres etapes: adquisició, recuperació i raonament. Això representa el que es diu "entrenament" de la IA, i que cal fer de forma correcta. Per tant, a part d'una fase inicial que pot ser intensiva i massiva sobre "entrenament", posteriorment hi haurà més aportacions d'informació, que han de ser igualment "entrenades". A les hores, en enfocaments de IA sobre un àrea o sector concret (com pot ser la Ciberseguretat), més que un "entrenament" per a tot el procés, és sol parlar i anomenar com **Aprenentatge de la IA** (**Revisar: Annex A). En aquest sentit, en punts anteriors, ja s'ha vist el nom dels diferents tipus d'aprenentatge considerats sobre IA, i que tot seguit expliquem:

- **Aprenentatge automàtic (ML – Machine Learning)**

Es considera una branca o disciplina dins de la IA, que vol aconseguir que un sistema aprengui i relacioni la informació de la mateixa forma que ho faria una persona, i que el resultat millori amb el temps. Una segona definició seria: és un subconjunt de la IA, que es centra en desenvolupar sistemes que aprenen o milloren el rendiment, en funció de les dades que consumeixen, tal com ho faria un humà. I una tercera definició seria: és un subgrup de la IA que permet a un sistema aprendre de les dades enlloc d'aprendre mitjançant una programació explícita. Per poder fer tot l'anterior, s'empra algoritmes capaços de detectar patrons en les dades prèvies, cosa que li permet crear prediccions futures o entreveure tendències. Aquests algorismes, es poden anar adaptant a mesura de que aprenen més sobre les dades que processen. Sovint, com que es pot fer tractament de dades amb models preparats per entrenar abans de fer la implementació final, i aquestes dades solen ser de tipus continu, es sol parlar també d'aprenentatge iteratiu, en aquest tipus de casos.

Altres definicions interessants sobre el ML, són les que fan diferents científics reconeguts en el món de la IA i ML, com:

- **Arthur Lee Samuel** (1901-1990), pioner en ML : "És el camp d'estudi que dona a les computadores la capacitat d'aprendre sense ser programades explícitament".
- **Tom M. Mitchell** (1951), reconegut per aportacions a ML : "Es diu que un programa de computadora aprèn de l'experiència E respecte a alguna classe de tasca T i mesura d'execució P, si la seva execució en les tasques T, mesurat per P, millora amb l'experiència E".

Per altra banda, aquesta vessant de la IA, actualment, és una de les que està més desenvolupada amb finalitat comercial o empresarial, ja que es fa servir per processar grans quantitats de dades ràpidament i son dipositades de forma entenedora pels humans. Com exemple de tot això, tindrem: les dades extretes de plantes de producció, que generen un flux constant de les màquines (com producció, funcionalitat, temperatura, etc) a un nucli central. Aquesta enorme quantitat de dades, representa el procés de producció que cal analitzar, per tal d'aconseguir una millora continua i una presa de decisions correcta. El gran volum de dades, fa que una persona tingui que emprar molt temps (dies), en el seu anàlisi i traçabilitat. Aquí, és on entra el ML, permet que s'analitzin les dades a mesura de que es van incorporant del procés productiu, i identifica patrons o anomalies d'una forma més ràpida i precisa. Així, es poden generar i enviar, avisos o alertes per una bona presa de decisions.

De forma resumida, per obtenir un bon sistema de Machine Learning (ML), i que sigui funcional i doni bons resultats, cal :

- Recursos de preparació de dades.
- Algoritmes, bàsics i avançats
- Automatització i processos iteratius
- Escalabilitat
- Modelat en conjunt

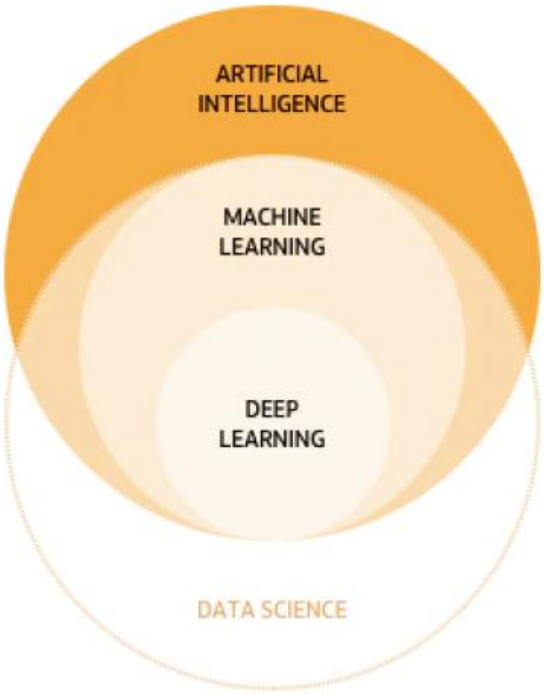


Figura 10: Conjunts dins de la IA, font: Web Oracle

- **Aprenentatge profund (DL – Deep Learning)**

Considerat una subdisciplina dins de l'aprenentatge automàtic (ML), és un sistema d'aprenentatge que s'inspira en el funcionament de les xarxes neuronals per processar la informació, i té una base matemàtica molt complexa al darrera. Encara que també fa servir l'experiència, no parteix d'indicacions estrictes que determinen el que es correcte i que no, així, el sistema pot determinar conclusions per si mateix. Una segona definició seria: és un conjunt d'algorismes de xarxes neuronals, pensats pel aprenentatge automàtic de les màquines i que participen en un raonament no lineal. I una tercera definició: és un conjunt de tècniques i procediments algorísmics, basats en ML, per fer que una màquina aprengui de la mateixa forma que ho fa un ésser humà.

Les xarxes neuronals complexes de DL estan dissenyades per emular com funciona el cervell humà, així que les computadores poden ser entrenades per bregar-se amb abstraccions i problemes mal definits. Aquesta tècnica, permet aprendre d'una forma "profunda" sense tenir un codi específic per fer-ho. Fins fa poc, només es solia emprar en el reconeixement d'imatges, veu i aplicacions de visió per computadora. Podem dir, que es troba en una primera fase per a desenvolupar tot el seu potencial, i que cada cop es farà servir més, convertint les dades en conjunts més detallats i escalables.

Figura 11: Acudit sobre IA, font: Dilbert.com

En general, les passes o procés a fer, per construir un model d'aprenentatge (ML o DL) serien les següents:

1. Recollir les dades: es pot treure dades des de moltes fonts, però és un punt dels que porta més complicacions i consumeix més temps.
2. Preprocessar les dades: cal assegurar-se que les dades tenen el format correcte per emprar-les a l'algorisme d'aprenentatge.
3. Explorar les dades: realitzar un preanàlisi per corregir casos de valors que falten o mirar de trobar a cop d'ull algun patró que faciliti el model.

4. Entrenar l'algorisme: aquí ja apliquem l'entrenament, i així es nodreix l'algorisme amb dades de les etapes anteriors.
5. Avaluar l'algorisme: es posa a prova la informació o coneixements que ha obtingut l'algorisme al pas anterior, cal avaluar precisió i prediccions.
6. Utilitzar el model: aquí ja posem el model a fer front al problema real, i també podem mesurar el seu rendiment.

Ja es faci servir un o altre model d'aprenentatge (ML o DL) sobre unes dades, segons el que es vulgui obtenir com resultat o aproximació, cal tenir molt en compte el sobreajustament o sobreentrenament. Això, es pot donar quant en l'entrenament sols es troben casualitats en les dades, que semblen patrons d'interès, però que no generalitzen. Anem a veure amb més detall això:

- **Sobreajustament** (o sobreentrenament):

és la tendència que tenen la majoria d'algorismes d'aprenentatge, a ajustar-se a unes característiques molt específiques de les dades d'entrenament, que no tenen relació causal amb la funció objectiu que s'està buscant per generalitzar. És pot donar per sobre, **Overfitting**: massa entrenament amb dades estranyes, o bé, per sota, **Underfitting**: poc entrenament o sols dades d'un mateix tipus. L'exemple més extrem d'un model sobreasjustat, seria un model que sols memoritzi les respostes correctes. Al emprar-lo amb dades que mai a vist abans, tindrà un rendiment compromès (o accidental), ja que mai va poder generalitzar un patró per predir. Les principals estratègies per bregar el sobreajustament, són: la retenció de dades i la validació creuada.

Retenció de dades: quan tenim un conjunt de dades, el dividim en un o més conjunts d'entrenament i altres conjunts d'avaluació. D'aquesta forma "retenim" una part de les dades d'entrenament, per fer una avaluació de la efectivitat del model. Així, es persegueix, evitar que les mateixes dades emprades per entrenar, siguin les mateixes emprades per avaluar.

Validació creuada: aquí, en lloc d'obtenir sols una simple estimació de la efectivitat de la generalització, es fa un anàlisi estadístic per obtenir altres mesures del rendiment estimat, com la mediana i la variància. I així es pot entendre com s'espera que el rendiment variï via els diferents conjunts de dades. Aquesta variació, es fonamental per l'avaluació de la confiança en la estimació del rendiment, ja que les estimacions es calculen sobre tot el conjunt de dades per mitjà de múltiples divisions, i intercanvis sistemàtics entre dades d'entrenament i dades d'avaluació.

En aquest sentit, hi ha tota una sèrie de tècniques, que es poden emprar junt amb l'anterior, per tal d'evitar el sobreajustament, i són:

- **Oversampling**: quan es fa la mostra, es força a que tingui dades de classe per assegurar-se de que el model s'entrenarà amb aquesta classe.
- **Stratified sampling**: es fa la mostra de forma que contingui la mateixa proporció sobre les dades per determinades columnes.
- **Early stopping**: per evitar que el model entreni amb els detalls de les dades d'entrenament, s'atura abans.
- **Train-validation-test sets**: s'utilitza un "validation set" per corregir errors en les prediccions produïdes en l'entrenament. El "test set" s'empra per testejar el model final.
- **Cross-validation**: empra el "train-validation-test", alternant les dades amb les que s'entrena.
- **Regularization**: penalitza els paràmetres per reduir "la importància".
- **Pruning**: es "poda" l'arbre d'entrenament. Sols és aplicable a CART[16]

Aquí, hem pogut veure la importància de fer un entrenament, i fer-lo de forma correcta, amb les dades que pertoquen, que no estiguin esbiaixades, ja que sinó, no es pot garantir el correcte funcionament de la IA, i per tant, no es pot obtenir el que es coneix com una "**IA fiable**". Podem prendre com a marc de "IA fiable" el que es troba definit a la web de **ALTAI** (The Assessment List on Trustworthy Artificial Intelligence) que va ser creat per la Comissió Europea i l'àmbit d'aplicació es tota Europa. On com quadre resum de directrius, tenim :

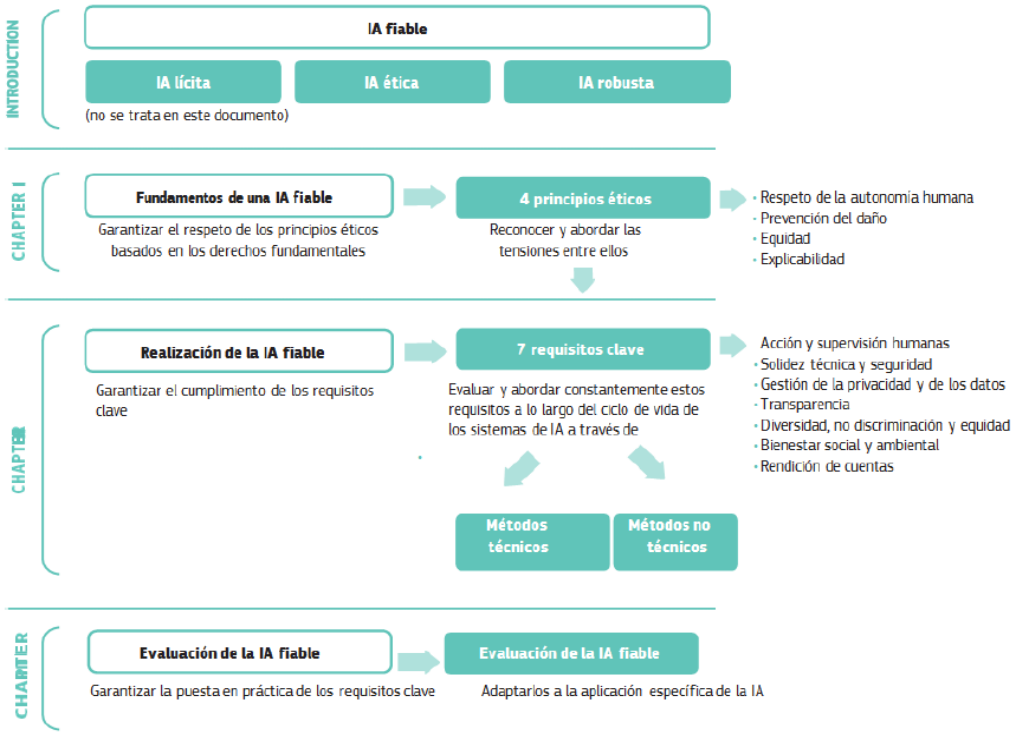


Figura 12: Resum directrius ALTAI, font: Web ALTAI

[16] CART: tipus d'algorisme basat en d'arbres de classificació i regressió.

6. Classificació de les IA

És complicat fer categories de la IA, el més fàcil potser seria fer categories segons els algorismes emprats en un sistema, però, alguns experts sobre el tema, com pot ser **Stuart J. Russell** (1962) i **Peter Norvig** (1956), fan la categorització següent sobre les IA:

- **Sistemes que pensen com humans**: Aquests sistemes tracten d'emular el pensament humà d'una forma bastant literal, per mitjà de xarxes neuronals artificials.
- **Sistemes que actuen com humans**: Aquests sistemes es centren sols en actuar com humans, estan més lligats a la robòtica clàssica i son molt menys flexibles.
- **Sistemes que pensen racionalment**: Aquests sistemes tracten d'aplicar lògica humana a l'hora de percebre, raonar i actuar. I no es centren en emular el comportament neuronal del cervell, sinó que son entrenats per actuar de forma humana en un entorn determinat, tipus agents experts.
- **Sistemes que actuen racionalment (idealment)**: Tracten d'emular de forma racional el comportament humà, obtenint conclusions pròpies a unes condicions d'entorn donades. El punt diferencial es el d'intentar que apliquin racionalitat a les seves decisions.

Per altra banda, de forma més habitual i comuna en el mon de les IA, es sol fer una divisió, en dos grans grups, que serien:

- **IA Dèbil** (o estreta): És coneix per les sigles **IAD**, i seria tota la Intel·ligència Artificial desenvolupada fins ara. És la IA dedicada a resoldre un problema específic o un conjunt de problemes, de forma optimitzada, però sense la possibilitat d'estendre's a problemes generals, sense la programació que pertoqui per fer-ho. Fins i tot els assistents virtuals més punters, entren en aquesta categoria.
- **IA Forta** (o general): És coneix per les sigles **IAF**, i seria tota la Intel·ligència Artificial que és capaç de igualar o superar la intel·ligència humana en capacitat de raonament i deducció. A dia d'avui això és una utopia, que sols existeix a la "ciència ficció". Encara que les màquines ens superen el moltes capacitats, no posseeixen sentiments reals ni capacitats cognitives pròpies, i tampoc consciència pròpia, i adaptació a qualsevol escenari.

Una altra forma, per fer grups o categories, pot ser agrupant-les **segons la tecnologia emprada** en la IA, i com que de fet, ja gairebé hem parlat d'elles en punts anteriors, sols posem una llista de quins grups serien:

- Reconeixement automàtic de la parla (Speech recognition)
- Processament del llenguatge natural PLN (NLP en anglès)
- Reconeixement visual (Visual Recognition)
- Reconeixement de text (Text Recognition).
- Big Data
- Sistemes experts
- Robòtica
- Aprenentatge automàtic (Machine Learning)
- Aprenentatge profund (Deep Learning)
- Cognitive Intelligence & Cognitive Services

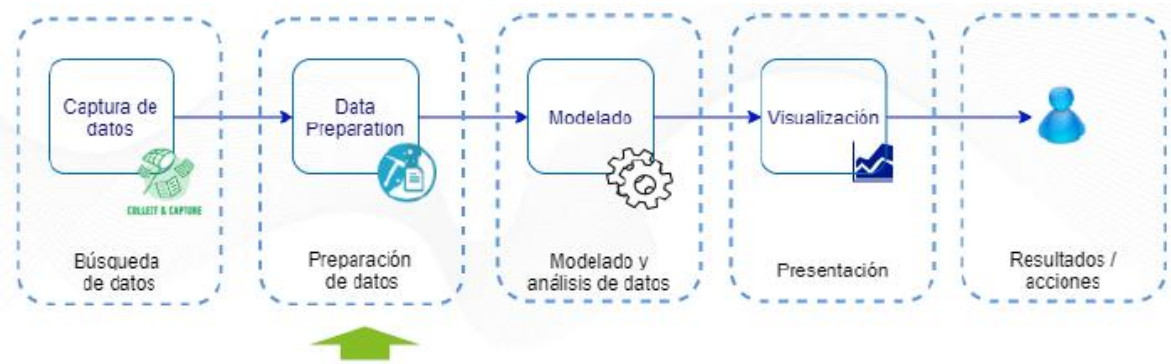


Figura 13: Procés d'anàlisi de dades, font: Pinterest

En quant a l'entorn de l'aprenentatge, ja sigui ML o DL, també es pot fer una classificació **segons els algorismes emprats**, d'aquesta forma, tenim tres grups principals, que són:

- **Aprenentatge supervisat** (Supervised machine learning): es basa en models predictius, que fan ús de dades d'entrenament. Donat un conjunt conegut de dades, es pretén que el sistema sigui capaç d'aconseguir una determinada sortida, de forma el model es ajustat o entrenat fins aconseguir els resultats adequats. Aquí, les mostres de dades estan etiquetades i l'algorisme genera una funció que relaciona les variables d'entrada amb les de sortida desitjada. És fa servir en dos tipus de problemes: Classificació i Regressió.
    - Classificació: aquest model prediu valors discrets, les dades es separen en un cert nombre de classes. Com algorismes d'aquest tipus, tenim: Logistic Regression, Decision Trees, Random Forest i Suport Vector Machines (SVMs). Per tal de validar aquest model, algunes mesures més emprades, són: Matriu de confusió, Mètrica d'exactitud, Métrica d'exhaustivitat, Mètrica de precissió, i la puntuació F1.

- Regressió: aquest model prediu valors continus. Com algorisme d'aquest tipus, tenim: Linear Regression, Regression Trees, Non-Linear Regression. Bayesian Linear Regression, i Polynimical Regression. Per tal de validar aquest model, algunes de les mesures més emprades, solen ser: Error mitjà absolut, Error quadràtic mitjà, i Arrel de l'error quadràtic mitjà.

- **Aprenentatge no supervisat** (Unsupervised machine learning): és similar a l'anterior, però els algorismes emprats, ajusten el seu model únicament en funció de les dades d'entrada. És a dir, l'algorisme fa un autoentrenament sense indicacions externes. Aquí, les dades no estan etiquetades, així que aprèn sol de les dades d'entrada, sense disposar de la informació de les sortides. És fa servir en quatre tipus de problemes: Clustering, Associació de regles, Detecció d'anomalies, i Reducció de dimensions.

  - Clustering: Aquest model s'utilitza per agrupar dades existents de les que es desconeix les característiques en comú o es vol trobar-les. És prova de crear punts centrals per diferenciar grups i descobrir les característiques comunes per proximitat.
  - Associació de regles: s'utilitza per trobar associacions i relacions entre grans conjunts d'elements de dades. Les regles mostren amb quina freqüència es produeix un conjunt d'elements en una transacció.
  - Detecció d'anomalies: és refereix a la identificació d'elements o esdeveniments que no s'ajusten a un patró esperat, o a altres elements en un conjunt de dades que generalment no son detectables per un humà.
  - Reducció de dimensions: son tècniques per poder reduir la dimensió de les dades. Això es pot emprar per evitar sobreajustos o per poder interpretar les dades, ja sigui mitjançant factors o gràficament.

- **Aprenentatge per reforç** (RL, Reinforcement Learning): consisteix en la iteració constant i basada en "prova i error" que una màquina es capaç de fer en un temps record, davant determinades condicions o entorn donat (exemple un joc) i amb un objectiu específic, anomenat "recompensa". Aquí es desenvolupa un sistema (agent) que es vol millorar en eficiència fent certa tasca via la interacció amb el seu entorn. Així, el "agent" rep una "recompensa" que li permet adaptar el seu comportament. La diferència amb els anteriors, es que aquest no li es necessari el coneixement dels processos de decisió (MDP[17] – Markov Decision Process), i per tant, es fa

[17] MDP: acrònim anglès per "Markov Decision Process", és un procés estocàstic de control en temps discret, que modela la presa de decisions en situacions parcialment aleatòries i parcialment fixades.

servir en els processos de decisió en que els mètodes exactes no son viables. Com algorismes d'aquest tipus, tenim: Criteri d'optimalitat, Força bruta, Atansament al valor de la funció, i Cerca política directa.

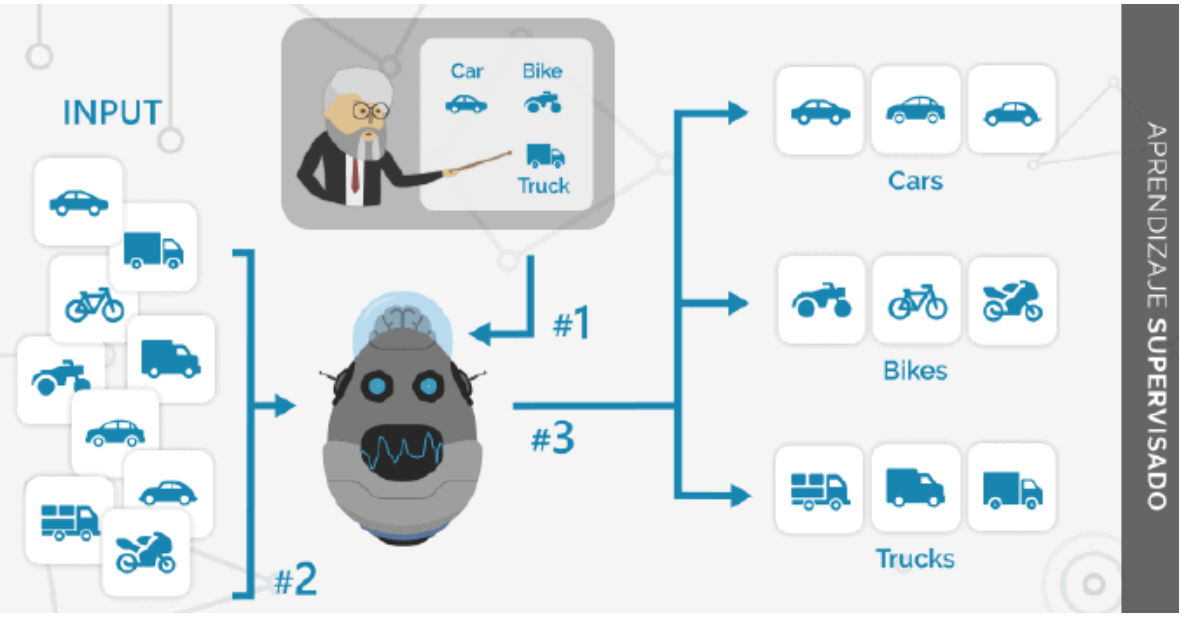


Figura 14: Aprenentatge Supervisat, font: Auraquantic.com

7. Punts forts de les IA

Fins ara, ja hem vist que el principi fonamental de la IA, és replicar i després, superar la forma en que els humans perceben i reaccionen davant el mon. Per tant, això s'està convertint el la base principal de la innovació, ja que la IA gràcies a varies formes de ML (Machine Learning) que reconeixen patrons en les dades, li permet fer prediccions i pot afegir valor a un negoci. D'aquesta forma, una IA ens pot ajudar en el següent:

- Proporcionar una comprensió més complerta sobre l'abundància de dades que hi ha disponibles.
- Confiar en les prediccions per automatitzar tasques massa complexes o mundanes.

Actualment, la tecnologia IA, millora el rendiment i la productivitat de l'empresa per mitjà de l'automatització de processos o tasques que abans requerien un gran esforç humà. A més, la IA també pot donar sentit a les dades a una escala que mai cap humà podria. D'aquesta forma, les empreses, han fet de la ciència de dades una prioritat i estan fent grans inversions sobre això. Segons una de les darreres enquestes de "Gartner", és classifica l'analítica i la intel·ligència

empresarial com les tecnologies que proporcionen més diferenciació per una organització.

Per altra banda, en el punt "3. Situació actual", ja hem vist que aquesta adopció de la IA i fort desenvolupament en tot tipus de sector, es troba impulsada principalment per tres factors, que serien:

- La capacitat de computació i d'alt rendiment ja està disponible.
- Es troba disponible grans volums de dades per l'aprenentatge.
- La IA aplicada proporciona un avantatge competitiu.

En aquest sentit, tenim a les empreses varis casos d'ús i èxit amb varies IA, que donen solucions a diferents àrees, com:

- Assistència virtual: pot concertar reunions, proposar articles d'interès, recomanar contactes o fer el seguiment precís de les tasques d'un empleat.
- Millora de l'atenció al client: amb "chatbots", es podem resoldre dubtes o incidències amb clients, que suposa un estalvi en suport i atenció al client.
- Augment de la productivitat: no sols pot detectar "colls d'ampolla", sinó que també pot deduir sota quines circumstàncies es donen.
- Anàlisi intel·ligent: si es configura així, pot analitzar dades (estructurades o no), obtenint conclusions més ràpid que una persona amb "Data Mining".
- Predicció de dades: pot millorar la predicció de dades, per exemple, pot anticipar quins clients pagaran o no les pròximes factures.
- Marketing i vendes intel·ligents: pot esbrinar quin producte es ven més, en quina època de l'any, en quin país, ciutats, edats, perfils, sense gaire esforç.
- Detecció d'anomalies: pot detectar casos fora de la tendència lògica de l'empresa, independentment del departament en que passi.
- Anàlisi de sentiments: dins del llenguatge natural PLN (NLP en anglès) és pot avaluar l'estat anímic de la persona, en funció del text.

De totes formes, a nivell més representatiu, per part de persones reconegudes en el mon IA, tenim el que comenta **Andy Chan**, Product Manager de "Infinia ML", i **Kai-Fu Lee**, fundador de "Sinovation Ventures". Segons ells, com a punts forts, tenim que els principals beneficis o avantatges d'una IA aplicada a l'empresa, serien:

1. **Automatització de processos**: la IA permet que els agents desenvolupin tasques repetitives, rutinàries i d'optimització de processos d'una forma automàtica i sense la intervenció humana.
2. **Potenciació de les tasques creatives**: la IA allibera a les persones de tasques rutinàries i repetitives, això permet que puguin dedicar el temps a fer funcions creatives.

3. **Aporta precisió**: la IA aporta major precisió que el ser humà, i pot prendre decisions que abans sense la IA es feia de forma manual o monitorada.
4. **Redueix l'error humà**: la IA redueix les falles per limitacions del ser humà, pot emprar sensors, per detectar falles de fabricació, que no són detectables pels sentits humans.
5. **Redueix el temps emprat en anàlisi de dades**: la IA permet que l'anàlisi i explotació de dades, per exemple de producció, es faci en temps real.
6. **Manteniment predictiu**: la IA permet fer un manteniment d'equipament segons el temps i condicions d'ús o funcionament d'aquest, millorant rendiment i cicle de vida.
7. **Millora en presa de decisions a nivell producció i de negoci**: la IA al disposar d'informació estructurada, permet a cada responsable poder prendre decisions de forma més ràpida i eficient.
8. **Control i optimització de processos productius i línies de producció**: la IA aconsegueix processos més eficients, lliures d'errors, pel que té un major control sobre les línies de producció.
9. **Augment de productivitat i qualitat a la producció**: la IA no sols fa augmentar la productivitat a nivell màquina, sinó que també ho fa a nivell dels treballadors, i de la qualitat de la feina que fan.

Per altra banda, hi ha una sèrie de persones, com: Stephen Hawking, Bill Gates o altres investigadors, que veuen un risc en les IA, sobre tot, si no es limiten sols a reproduir tasques humanes, i això els preocupa. De totes formes, segons alguns, encara hi ha varies limitacions que impedeixen de moment, una forta expansió de les IA. Aquestes limitacions, poden ser corregides, i serien:

- **Disponibilitat de dades**: moltes vegades les dades es presenten de forma aïllada o son inconsistents i de baixa qualitat. Per superar això, cal definir una estratègia clara des del inici, per poder extreure les dades per la IA d'una forma organitzada i consistent.
- **Falta de professionals qualificats**: ara per ara, hi ha manca de perfils amb habilitats i experiència en aquest tipus d'implementacions amb IA. Per engegar aquest procés, cal comptar amb professionals que hagin treballat en projectes d'igual abast.
- **El cost i temps d'implementació dels projectes IA**: això es un factor molt important a l'hora de decidir d'executar aquest tipus de projectes. Si a l'empresa no hi ha coneixement sobre sistemes IA, cal valorar una possible externalització, tant de la implantació com del manteniment, per obtenir uns resultats amb èxit del projecte.

Una altra cosa a tenir en compte sobre les IA, és el fet de que gairebé totes les grans tecnològiques han adoptat i implementat aquesta tecnologia, oferint molts

serveis que poden ser consumits per Internet, en el "Cloud"[18]. Per tant, a l'hora de desenvolupar un projecte IA, és un factor més a tenir en compte. Està clar, que escollir l'opció "in situ" o al "cloud", dependrà dels diferents interessos que tingui l'organització. Però en molts cassos, si hi ha alguna de les limitacions que hem comentat, pot ser un solució força viable que cal tenir en compte. Anem a veure alguns dels més representatius en el "cloud" i que potser cal valorar segons el projecte:

- **Microsoft Cognitive Services**: és un conjunt de serveis d'intel·ligència cognitiva, posats al públic per Microsoft al seu "cloud" (Azure). Està dividit en categories com: veu, visió, llenguatge, decisió i cerca.
- **Google Cloud AI**: es un conjunt de serveis que Google ofereix als seus plans "cloud". Abasta o compren des del consum de serveis cognitius, fins al desenvolupament d'aplicacions de IA (pròpiament dites).
- **Watson AI**: plataforma de IBM, que permet integrar el que es considera el IA més potent actualment, a la nostra pròpia aplicació.

Cal tenir en compte, que gràcies a aquest serveis que ofereixen aquestes grans tecnològiques, alguns de forma gratuïta, altres com a plans limitats o d'altres formes, s'ha disparat el nombre de TFM[19], Tesis o altres aportacions de tipus acadèmic. Pel que tenim que hi ha tot un ventall de solucions, si bé moltes de tipus laboratori o experimental, basades en l'aplicació de la IA en "cloud" sobre diferents problemes.

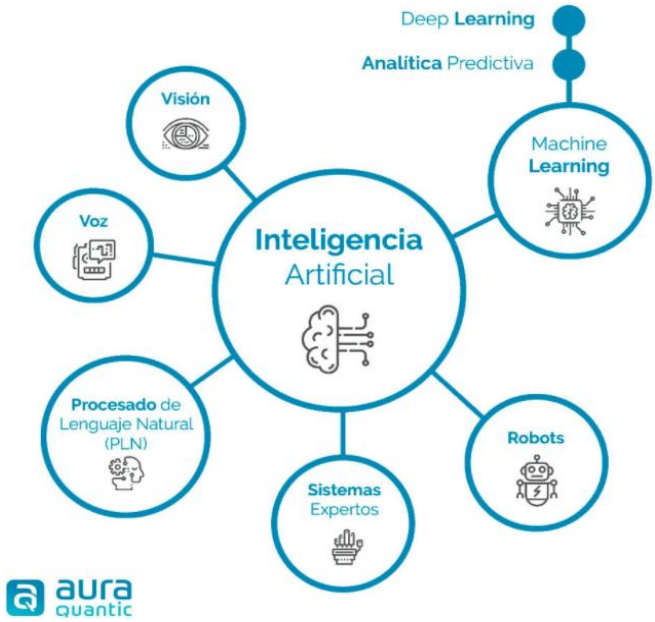


Figura 15: Quadre aplicacions IA, font: Auraquantic.com

[18] Cloud: anomenat "cloud computing", servei de fitxers o recursos ofert via Internet.
[19] TFM: simplificació o acrònim de Treball Final de Màster

8. Situació i reptes de la Ciberseguretat

Avui en dia, gràcies a la informació publicada a molts i diferents mitjans de comunicació, es pot constatar un increment exponencial de diferents i tot tipus d'atacs per Internet i la Xarxa, ja sigui a persones particulars, organitzacions o empreses i fins i tot, a governs o Institucions d'aquestos. I per altra banda, tenim que, en certs mitjans especialitzats, s'entreveu també un increment (potser no de forma exponencial, però si lineal amb forta pendent) del nombre de Vulnerabilitats que es detecten en diferents aplicacions, maquinari IoT[20], Drivers[21] i Firmware[22]. D'aquesta forma, es pot considerar, que tenir els Sistemes d'Informació actualitzats es una de les primeres fites, que tant usuaris, com organitzacions o empreses i governs han de tenir en el seu punt de mira. Això, ha de ser una tasca o feina constant que s'ha de fer per mitigar les possibles Vulnerabilitats que vagin sorgint. Podem prendre com exemple les constants actualitzacions que proporciona Microsoft i altres cases de Software.

En aquest sentit, per donar una visió sobre les principals amenaces a les que podem estar exposats, hi ha diferents estudis, que en fan un recull i manifesten que els Usuaris en general, es troben exposats a les amenaces següents:

- Ignorància
- Malware[23] i Bots[24]
- Comptes hackejats per Phishing[25]
- Spam[26] (correu no desitjat).
- Llar insegura amb xarxa Wi-Fi
- Dades perdudes (robatori o desfer-se de dispositius sense esborrar-los)
- Atacs per Wi-Fi

Per altra banda, a nivell organització o empresa, l'anterior també afecta, ja que el personal com individus, també hi estan exposats i més amb el teletreball. A més, en aquest cas, tenim varis estudis que comenten que les tendències i amenaces sobre la Ciberseguretat per al 2022, poden ser les següents:

[20] IoT: acrònim en anglès de "Internet of Things", Internet de les coses, dispositius remots amb certa computació, com sensors, mesures de nivells, temperatura o altra informació.
[21] Drivers: o controladors, és un component de software que permet al Sistema Operatiu i un dispositiu comunicar-se entre ells.
[22] Firmware: és un programà bàsic que es carrega al iniciar un dispositiu i que permet controlar els circuits electrònics d'aquest.
[23] Malware: terme anglès per software maliciós, programa o codi maliciós, perjudicial pels sistemes.
[24] Bot: és un programa que fa tasques repetitives, predefinides i automatitzades, son de molts tipus.
[25] Phishing: terme anglès de pescar, consisteix en enganyar les persones per a que comparteixin informació confidencial com contrasenyes, números i dades de la targeta de crèdit, etc.
[26] Spam: terme anglès per a definir el correu no desitjat, amb finalitats publicitàries o comercials.

- Ransomware[27] (increment de perill pel RaaS[28])
- Amenaces associades al treball remot (Malware, Phising, Smishing[29], ...)
- Business Email Compromise (BEC) i estafa per PDF
- Deepfakes[30] (per suplantar identitats)
- Amenaces internes dins de l'empresa (per descuits o a propòsit)
- Pèrdua de recursos legals de l'empresa per a fer accions il·legals
- Vulnerabilitats del sistema
- Robatori de dades
- Atac i segrest de dispositius IoT.

Donades aquestes situacions, i també l'evolució constant d'aplicacions, dels serveis i funcionalitats al Cloud, Wi-Fi d'empresa, les IoT, etc. podem veure que la idea o els conceptes clàssics sobre Ciberseguretat, en algun punt han deixat de tenir un sentit fixa o inamovible i han passat a ser dinàmics. S'han tingut que anar adaptant i evolucionant a cada pas sobre els canvis funcionals, aparició de noves amenaces i també vulnerabilitats, que ha sofert l'organització o empresa sobre els seus sistemes d'informació.

8.1. Visió tradicional

Fins ara, les empreses es pensen que estan segures, o que tenen un bon nivell de seguretat, amb una bona gestió de les polítiques de seguretat i l'ús i gestió de diferents elements com : Firewalls, WAF (Web Advanced Firewall), Analitzadors de Xarxa actius/passius com IDS/IPS (Intruder Detection System / Intruder Prevention System), Antivirus locals i centralitzat en servidor, Routers i Switchos amb filtratge i Vlans[31] , i també un SIEM[32].

Tots aquests elements s'han de gestionar, ja sigui per Alertes, Monitors, canvis de configuració o altres motius. De forma general, es sol fer que tots els esdeveniments i la informació generada en LOG, vagi a parar al SIEM, i que aquest gestioni tota la informació rebuda, correlant-la, per tal de preveure amenaces, possibles atacs, i ens alerti. Tot això, fa que es necessiti certa

[27] Ransomware: terme anglès per malware de rescat, és un tipus de malware que impedeix accedir al sistema o fitxers, en que es sol demanar el pagament d'un rescat, per fer-ho i sense garanties de res.
[28] RaaS: en anglès, Ransomware as a Service, model de negoci de part de guanys, de desenvolupadors de malware pel que ofereixen serveis i eines per a una campanya de ransomware contra una victima.
[29] Smishing: terme anglès, per una tècnica d'enviament d'un SMS simulant ser legítim, amb l'objectiu de robar informació confidencial.
[30] Deepfakes: terme anglès per a la edició i falsificació de vídeos amb persones aparentment reals amb l'objectiu d'enganyar al destinatari.
[31] Vlan: terme anglès per "Virtual Lan", xarxa de àrea local virtual que permet crear xarxes lògiques virtuals dins una mateixa xarxa física.
[32] SIEM: acrònim angles de "Security Information and Event System", el SEM detecta patrons fora del normal en temps real, i el SIM centralitza els registres de seguretat per interpretar-los i guardar-los.

capacitació, uns coneixements avançats i força experiència per tal de fer-ho correctament. I això, sol ser un dels problemes en la gestió d'aquests elements de Seguretat, la falta de preparació, falta de coneixements i també l'elevat nombre de alertes, falsos positius i altra informació que es genera i cal tractar. Per tant, sol haver-hi manca de personal preparat, i una saturació d'informació que pot ser rellevant o no, i que d'una forma o altra s'ha de tractar i classificar.

Un nou pas que s'ha fet per monitorar, detectar i auditar el trànsit de la xarxa, de forma que es millori la seguretat de les organitzacions o empreses, ha estat l'aparició de diferents "Frameworks"[33], que permeten gestionar molt millor les diferents perspectives de la Ciberseguretat, com es veu en la imatge següent:

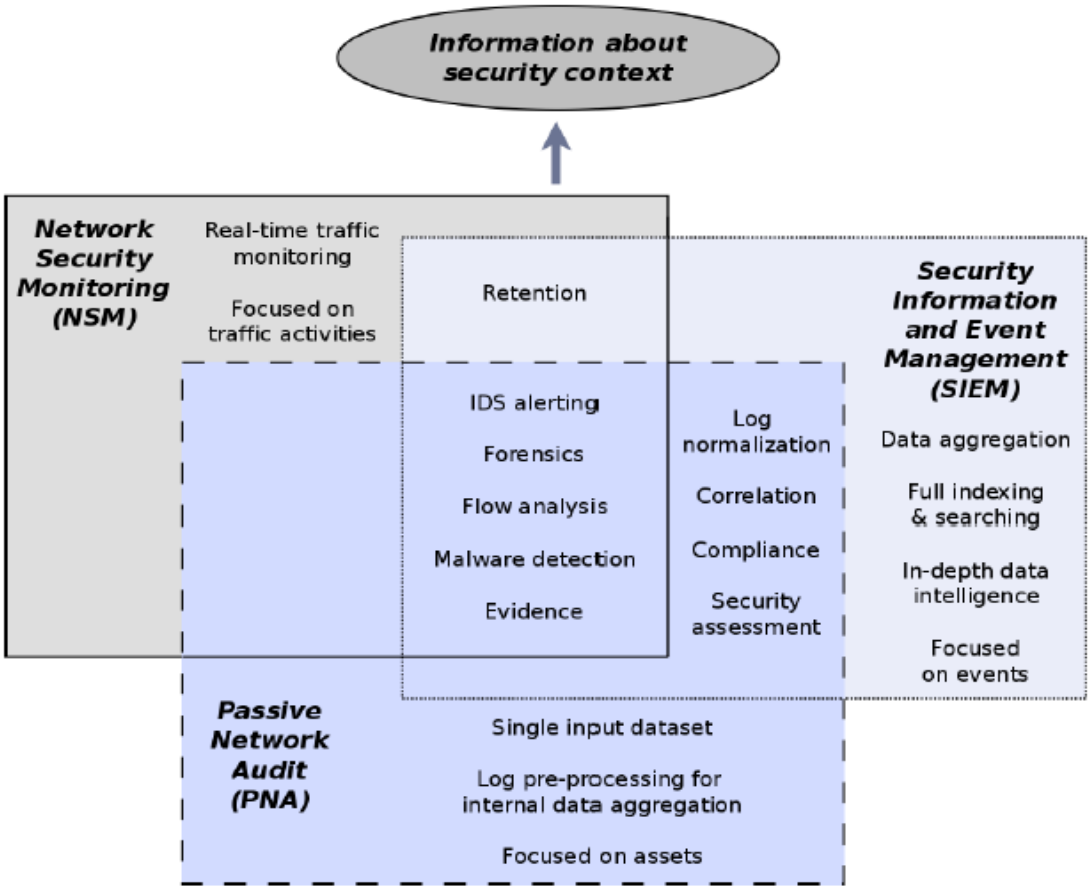


Figura 16: Models de monitor i anàlisi, font: Santillan,2014

- **NSM (Network Security Monitoring):** Model d'anàlisi del trànsit de xarxa que dona lineaments per desenvolupar un framework que inclogui tècniques i eines pel monitoratge, detecció i retenció d'evidències sobre incidents de seguretat. Aquest model es basa en l'anàlisi de dades generades per eines

[33] Framework: esquema o marc de treball que ofereix una estructura base per elaborar un projecte amb objectius específics, seria com una plantilla o punt de sortida per a un desenvolupament.

com els IDS, Firewall, analitzadors de fluxos i altres. Aquí es fa èmfasis en les tècniques a emprar per aconseguir la millor detecció, és a dir, no sols es descriu quines eines poden ser emprades, també el com, quan i on emprar-les, dins del context de la xarxa, així com consideracions d'implementació, zones i punts de monitoreig.

- **SIEM (Security Information and Event Management):** La mineria de dades s'aplica en els SIEM per la identificació de patrons, amb propòsits de detecció, auditoria i interpretació de la informació. Per tant, en aquest sentit, es pot emprar models d'anàlisi estadístics i és un bon candidat pel ML (Machine Learning). Les fonts de dades a analitzar poden ser eines com: IDS, IPS, Firewalls, routers, bitàcoles[34] de sistema, etc. El SIEM sol dur a terme un correlació de dades amb l'objectiu de filtrar informació per identificar i interpretar els esdeveniments relacionats amb incidents de seguretat. Aquí es combinen les característiques dels SIM (Security information Manager) i dels SEM (Security Event Manager), per dur a terme un anàlisi en temps real i emmagatzematge a llarg termini de registres d'esdeveniments.
- **PNA (Passive Network Audit):** De forma similar als SIEM, el PNA inclou l'anàlisi de bitàcoles i correlació de dades, però com a model d'anàlisi, a diferència dels SIEM, empra sols el trànsit de xarxa com la seva única font principal, per l'obtenció d'informació i la posterior generació d'informes. Això ho fa de forma passiva, no realitza cap acció o intervenció sobre la xarxa que analitza. L'acumulació de dades es fa de forma interna, ja que inclou el processament i descodificació prèvia de dades per la generació de bitàcoles o dades que podrien ser per un SIEM, però sols empra dades del trànsit de xarxa, pel que s'aproximen a bitàcoles reals. Aquesta extracció implica una interpretació i extrapolació de dades que representen una "signatura" sobre determinades activitats o sistema (per exemple trànsit de capçaleres HTTP). Així, amb el pre-processament es pot identificar i descodificar protocols, versions de software, dominis, alertes de IDS, fluxos, etc, de forma passiva a la xarxa, sense accedir a les bitàcoles de sistemes o altres dispositius.

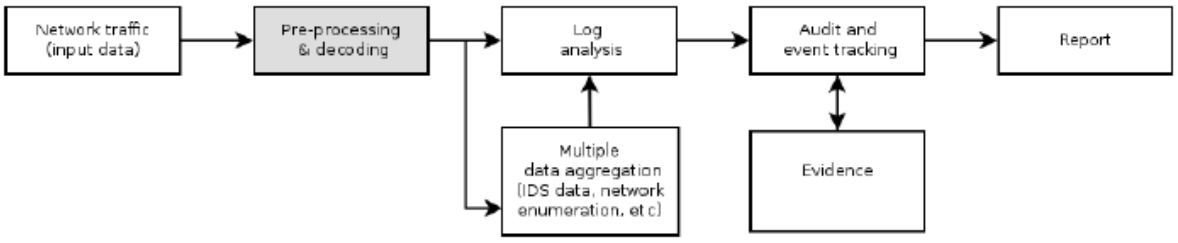


Figura 17: Diagrama auditoria passiva trànsit de xarxa, font: Santillan,2014

[34] Bitàcoles: designa un registre escrit d'accions i esdeveniments que surten al dur a terme certa feina.

En general, es sol definir un SOC (Security Operations Center), que pot ser intern si es disposa de recursos, però normalment sol ser extern, amb personal especialitzat i que pot tenir funcions limitades. Aquest equip, és el responsable de garantir la seguretat de la informació, emprant eines per la supervisió i administració de la seguretat dels sistemes d'informació, com el SIEM, per tal de recollir, correlar esdeveniments, i si cal i està habilitat, intervenir de forma remota. L'objectiu d'un SOC, es detectar, analitzar i corregir incidents de Ciberseguretat, emprant solucions tecnològiques i diversos enfocaments. En general, poden supervisar l'activitat de xarxes, servidors, terminals, bases de dades, aplicacions, llocs Web o altres sistemes cercant senyals febles o comportaments anòmals que poden indicar un incident de seguretat o compromís. I el SOC, ha de garantir que els possibles incidents de seguretat s'identifiquin, s'analitzin, es defensin, s'investiguin i s'informin adequadament.

Igual que el SOC, que pot ser extern, parcial o totalment, també apareix el MSSP (Managed Security Service Provider), que no es més que un Proveïdor de Serveis de Seguretat Gestionada. Aquest, es cuidaria de monitorar els esdeveniments de seguretat a la xarxa, enviar les alertes quan s'identifiquin anomalies, informes d'auditoria, acompliments i tot un treball forense si es dona un esdeveniment d'incident.

En resum, en aquesta visió, a pesar de certs avenços, de disposar d'alguns automatismes de protecció a certes alertes, millores de gestió i/o externalització sobre alguns serveis, es detecta una falta de preparació de les organitzacions o empreses, per fer front a totes les noves amenaces i/o reptes que apareixen. Així, en general es veu una desprotecció i falta de resiliència als reptes, degut als següents motius:

- Falta de formació o recursos per afrontar-los
- Falta de proves per fer un test de la capacitat real de prevenció i/o reacció
- Falta de mesures tècniques especifiques per mitigar-los
- Falta de sistemes de qualitat per detenir atacs sofisticats.

8.2. Nous reptes i evolució d'eines i serveis

Com hem comentat al inici de l'apartat, darrerament, les amenaces i els atacs, tant a particulars com organitzacions i empreses, estan creixent, també el nombre de Vulnerabilitats detectades. Per altra banda, es dona una evolució de la xarxa de l'empresa, incorpora nous tipus de "endpoints"[35], serveis externs al Cloud, accessos remots per teletreball o de proveïdors i Wi-Fi d'empresa, així,

[35] Endpoint: és qualsevol dispositiu remot que sigui físicament la part final d'una xarxa

s'expandeix la superfície exposada a les amenaces. Per tant, tenim una evolució de la situació i uns nous reptes que apareixen per la Ciberseguretat, que principalment podem dir que son:

- Les mancances comentades en la part final del punt anterior
- Increment de les amenaces i atacs
- Increment de les Vulnerabilitats detectades
- Manca de personal especialitzat o amb coneixements suficients
- Incorporació d'elements IoT a la xarxa (sensors diversos amb computació)
- Incorporació de càmeres de seguretat a la xarxa
- Possible Incorporació de mòbils / tablets d'empresa a la xarxa
- Incorporació de telefonia per IP
- Incorporació de la Wi-Fi d'empresa
- Increment de serveis externs al Cloud (normalment via VPN[36] permanent)
- Increment de VPN's temporals per teletreball i proveïdors externs
- Impossibilitat de gestionar el increment de totes les Alertes, falsos positius i altra informació per part de l'equip de seguretat de l'empresa.
- Impossibilitat de mantenir tot el Software, Drivers i Firmware al dia, per manca de recursos (ja sigui de personal, diners, coneixements o de temps).

Un cas preocupant, son els dispositius remots externs, com IoT (inclou aparells de domòtica, neveres, controladors industrials, ...), càmeres de seguretat externes, i els AP[37] del Wi-Fi. Donat que en la majoria de casos, el seu Firmware, és el Software de funcionament, i no és fàcil tenir-lo actualitzat, la majoria de vegades pot requerir un desplaçament al lloc i executar els procediments definits. A part de que els seus acompliments sobre requeriments de seguretat, solen ser molt bàsics i molt poc verificats, prefereixen oferir o donar funcionalitats. De vegades, si hi ha un controlador centralitzat per aquests "endpoints", hi sol haver algun procés o utilitat que permet mantenir-los actualitzats, i també pot oferir altra informació i serveis. El problema apareix per tenir-los controlats i analitzats si son externs i no treballen via una VPN (cosa difícil). El fet es que segons fonts especialitzades en Ciberseguretat, a la "Dark Web"[38] ja es pot trobar Firmware per a diferents dispositius, preparat per actuar com "bots" d'una xarxa "botnet", i a més, solen oferir els mateixos serveis que l'original. Donat que aquests dispositius no solen enviar esdeveniments i/o informació sobre intents d'accés, es pot efectuar un atac de força bruta (o més sofisticats), per tal de tenir-hi accés. Això, pot permetre el segrest del dispositiu

[36] VPN: acrònim anglès de "Virtual Private Network", xarxa privada virtual via un túnel de xifrat.
[37] AP: acrònim angles de "Access Point", punt d'accés que permet la connexió Wi-Fi.
[38] Dark Web: terme anglès per a la Web Fosca, està per sota la "Deep Web" o Web profunda o Web oculta.

via un Firmware modificat per actuar com a "bot maliciós" sense que aparentment es noti cap canvi per part de l'usuari, a part d'un possible canvi de la contrasenya d'accés i/o bloqueig d'actualitzacions. Per ara, es difícil i força complicat donar solució a aquesta problemàtica, dins uns costos i el temps invertit, que siguin raonables.

Per tal de millorar la Ciberseguretat d'una part de tot l'anterior, apareix el SOAR (Security Orchestration, Automation and Response), que és una orquestració, automatització i resposta de seguretat. Fa referència a tres funcions clau del software emprat pels equips de seguretat: la gestió de casos i fluxos de treball, l'automatització de tasques i un mitja centralitzat per accedir a la informació sobre les amenaces, realitzar consultes i compartir aquesta informació. Aquest terme va ser creat per "Gartner", però "IDC" el descriu: anàlisis, intel·ligència, resposta i organització de la seguretat (AIRO), en canvi "Forrester" empra el terme automatització i organització de la seguretat (SAO). Normalment, el SOAR es sol implementar en col·laboració amb el Centre d'Operacions de Seguretat (SOC) de les empreses. Les plataformes que empren aquest concepte poden supervisar les fonts d'informació sobre amenaces i generar respostes automàtiques per mitigar els problemes de seguretat. En resum, és podria considerar al SOAR, com una evolució del SIEM que s'actualitza sobre amenaces, i al que a part de saber que passa en la xarxa, pot actuar en base a aquesta informació.

Figura 18: Plataforma SOAR, font: A2Secure

Per cada una de les tres funcions comentades, es pot dir que comprenen o abasten, els següents objectius o intencions:

- **Security Orchestration:** implica la integració d'eines de seguretat que solen ser dispars, i la automatització dels seus processos, per tal de reduir la complexitat i augmentar l'eficàcia de les operacions de seguretat.
- **Security Automation:** te com objectiu reduir la participació humana en les tasques de seguretat per mitjà de l'ús de tecnologia per detectar, prioritzar i posar remei automàticament a les amenaces.
- **La resposta de seguretat:** fa referència a la planificació, la gestió, el seguiment i la notificació de les accions de resposta a incidents un cop que es detecta una amenaça.

A més, els beneficis clau que pot aportar el SOAR, no sols son una oportunitat per consolidar solucions i funcions de seguretat, és un canvi en la forma d'evitar atacs de manera proactiva, obtenint informació sobre les accions de l'amenaça, i una resposta precisa i ràpida a les amenaces quan en donin. Alguns dels beneficis clau que es pot veure, inclouen:

- **Temps mig de resposta reduït (MTTR):** els equips de SOC i de seguretat poden donar resposta més ràpid amb accions de resposta automatitzades. El factor humà, sol ser un retard pels casos d'amenaces conegudes i verificades amb un grup específic i definit d'accions per remeiar l'atac. Així, SOAR redueix el temps de resposta, per mitjà de la feina conjunta de les seves funcionalitats.
- **Reducció del impacte de les amenaces:** gràcies a una detecció i resposta més ràpides, les amenaces es tracten més aviat, cosa que evita la seva capacitat de prosperar i les atura abans no facin més mal.
- **Millor intel·ligència sobre amenaces:** les solucions SOAR, poden tenir una certa intel·ligència sobre amenaces, poden estar actualitzades a les amenaces més recents, cosa que permet als equips de seguretat dissenyar respostes adients a amenaces específiques.
- **En lloc de sols informació (Security Insight):** Donat que SOAR consolida tot un ventall de detalls de seguretat, amb fonts variades, les capacitats d'anàlisi i informes integrats, faciliten que els equips de seguretat entenguin la naturalesa i abast de les amenaces, amb informació que es pot processar en accions i respostes automàtiques.
- **Operacions optimitzades:** ja sigui per mitjà d'un equip de seguretat intern o extern, l'ús d'una solució SOAR ajuda a automatitzar la consolidació de dades de seguretat, la gestió d'amenaces de menor prioritat per mitjà de l'automatització, i elimina conjectures de la resposta a incidents de plans.
- **Millor rendiment i productivitat:** per mitjà de l'automatització SOAR, els equips de SOC i seguretat funcionen de forma més eficient, cosa que

permet prioritzar millor les tasques i respondre a més problemes en el mateix període de temps.

- **Costos reduïts:** el factor d'automatització, per si mateix, fa evident que l'ús d'una solució SOAR, serà més econòmica que realitzar la detecció i donar la resposta, de forma manual. Això, fa que la identificació i tractament de les amenaces sigui simple i ràpida pel SOAR.

Vist que el SOAR es més proactiu, ja no va tenir sentit un servei MSSP, per tant, apareix el MDR (Managed Detection and Response), que es un servei de Detecció i Resposta Gestionat. Aquest servei MDR, sol incloure un SOAR, ja que per tenir una detecció i respostes ràpides i precises, només es pot donar per mitjà de respostes de flux de treballs automatitzats definits, que gestionen les amenaces detectades.

Per altra banda, per millorar la protecció, donar resposta a amenaces més sofisticades i mitigar les febleses dels Usuaris, sorgeix l'eina EDR (Endpoint Detection Response), que és una Detecció i Resposta de Punt final. Combina l'antivirus tradicional amb eines de monitorització i possible intel·ligència artificial (heurística i possible IA), per donar una resposta ràpida i eficient als riscos i les amenaces complexes. Dona protecció i monitora la xarxa interna de l'empresa i equips endpoint (emprats per empleats, va des d'ordinadors fins a tablets i smartphones). Així, tenim que la informació registrada es porta a una Base de Dades central, on es realitzen anàlisis, detecció, investigació, informes i alertes, d'aquesta forma no sols es detecten esdeveniments al moment, sinó també per mitjà d'anàlisis de les dades ja registrades.

En general, el EDR s'integra amb altres eines de seguretat, per fer tasques com: millorar la visibilitat dels comportaments i processos en els endpoints, administrar els actius físics i de informació, millorar la resposta, i ajudar a recollir dades d'anàlisi de dispositius per TI[39]. Algunes de les aplicacions i eines que incorpora, a part de l'antivirus tradicional, solen ser:

- Eines d'anàlisi que poden emprar heurística i/o ML (Machine Learning) per millorar la detecció de les amenaces.
- Una Sandbox[40], que permet provar el comportament d'arxius i/o programes en un sistema virtual i aïllat de proves.
- Escaneig de IOCs[41] i regles YARA[42], que permeten analitzar i detectar les amenaces complexes donades en temps real.

[39] TI: acrònim de Tecnologies de la Informació.
[40] Sandbox: terme en anglès que literalment vol dir caixa de sorra, però fa referència a un entorn segur i aïllat on es pot testar una aplicació o programa.
[41] IOC: acrònim anglès per "Indicator of Compromise", que vol dir Indicador de Compromís

- Ús de llistes blanques i negres de correus electrònics, pagines Web i IP.
- Interacció i operativitat amb altres eines de seguretat com: SIEM, IDS/IPS o eines "antimalware"[43].

Figura 19: Endpoints segurs amb EDR, font: Pinterest

D'aquesta forma, els EDR proporcionen una sèrie d'avantatges per a les organitzacions o empreses que els fan servir, i en general son les següents:

- Millor capacitat d'anticipació sobre els atacs dirigits. Degut als models preventiu (pre-infecció) i detectiu (post-infecció) s'analitzen patrons de comportament que permeten anticipar amenaces.
- Reducció del temps d'exposició a les amenaces. Donat el seu enfocament reactiu que permet actuar en qüestió de segons, o minuts.
- Dona visió global de les amenaces contra els endpoints.
- Recull informació exhaustiva i detallada del endpoints (O.S.[44], hardware, processos actius, ...).
- Permet creat patrons de detecció automatitzats i recollir i guardar informació de forma automàtica.
- Monitora la integritat dels sistemes i arxius de configuració que son claus i genera alertes per accés o modificacions sospitoses.
- Dona accés a la informació en un sol punt, permet una ràpida investigació d'un incident.

[42] YARA: acrònim anglès per "Yet Another Ridiculous Acronym", la traducció no importa, però és una eina que empra regles que permeten detectar "malware", basat en firmes com els antivirus tradicionals.
[43] Antimalware: terme anglès per a un programa dissenyat per prevenir, detectar i remeiar software perjudicial o maliciós, dins de dispositius informàtics.
[44] O.S.: acrònim anglès per "Operating System", és el Sistema Operatiu.

Abans d'adoptar una solució EDR, donat que no totes son iguals, cal valorar una sèrie de característiques que son clau, com les següents:

- **Bona capacitat de filtrat:** ha de ser capaç de diferenciar els falsos positius.
- **Bloqueig avançat d'amenaces:** ha d'evitar les amenaces no sols quan es detecten, sinó també mentre dura l'atac.
- **Bona capacitat de resposta a incidents:** el temps de resposta a de ser ràpid i no entorpir el funcionament normal de la xarxa.
- **Protecció vers múltiples amenaces:** el EDR ha de poder gestionar varies amenaces avançades al mateix temps, sense perdre rendiment.
- **Funcionament en gran nombre de S.O.:** molts EDR es troben limitats a funcionar en un parell de sistemes operatius, la part central de l'eina no es problema, però si que ho es pel cas dels endpoints no estàndard i els IoT.

9. IA i Ciberseguretat

Aquí, desenvoluparem l'apartat per veure que ens pot aportar una IA a la Ciberseguretat, tenint en compte tot el que s'ha dit i comentat en els punts anteriors. De forma resumida, segons l'anterior i informació de publicacions especialitzades, a nivell particular i des de l'empresa podem veure importats increments en: amenaces, atacs, vulnerabilitats, la superfície exposada (xarxes de comunicació, VPN's, nombre i varietat de endpoints, serveis gestionats), i les actualitzacions (ja sigui de S.O., Software, Drivers o Firmware). Per altra banda, la Ciberseguretat continua tenint falta o manca de personal, de formació i coneixements, d'experts i de temps disponible per a gestionar totes les alertes, falsos positius, informació i monitors, que genera o tenen totes les eines i serveis destinats a la Seguretat. Encara que el SOAR i el EDR millorin una mica el panorama, amb els automatismes, resposta ràpida, integració d'eines, cobertura i resposta proactiva, continua faltant temps per revisar-ho tot, poder fer proves de configuracions, de contingència sobre atacs i altres temes com la documentació de processos, de incidències, acompliments i altres.

Recordem alguns dels grans avantatges que te la IA, que es poden traduir com objectius, i que es poden aplicar en el l'entorn de la Ciberseguretat:

- **Gestió massiva d'informació:** La IA prioritza quines situacions i possibles atacs tenen prioritat de gestió i també quins son falses amenaces, reduint la càrrega de treball dels sistemes.
- **Resposta en temps real:** La IA permet tenir accions immediates com resposta als atacs, per a minimitzar riscos, en base a infinitat de dades i del seu context.

- **Automatització:** la resposta a moltes amenaces es automàtica, minimitza el seu cost en quant a detecció i resposta.
- **Predicció:** La IA ajuda amb un millor anàlisi forense d'atacs previs, el que es tradueix en una millor defensa.

D'aquesta forma els objectius o avantatges anteriors, que ens donen robustesa, resiliència i resposta, poden ser aplicats a diferents àrees de la Ciberseguretat, cosa que ja es dona actualment, i algunes de les àrees poden ser les següents:

- **Threat hunting[45]:** permet identificar amenaces i neutralitzar ciberatacs. Les tècniques tradicionals en base a identitat o IOC (Indicators of Compromise) poden ser millorades, tancant forats de seguretat al gestionar i interpretar els indicadors de comportament.
- **Gestió de Vulnerabilitats:** el nombre de Vulnerabilitats creix cada dia, i no podem permetre que siguin explotades. Així, UEBA[46] (User and Entity Behavior Analiytics) permet identificar comportaments estranys que indica l'activitat d'atacs, fins i tot abans de disposar dels pegats que corregeixen la vulnerabilitat.
- **Data centers:** La IA facilita l'optimització i monitorització dels centres de processaments de dades, a més d'ajudar a detectar les amenaces i els comportaments anòmals. Millora l'aprofitament d'aquests recursos i la seva evolució, estalviant costos i reduint riscos com pot ser caiguda de serveis o l'execució de software maliciós.
- **Seguretat en les xarxes:** Encara que es tingui polítiques d'actuació front a comportaments d'usuaris, i identificats els processos de cada aplicació, la IA pot aprendre els patrons de comportament del trànsit de la xarxa, i fer recomanacions d'agrupació de càrregues de treball i l'aplicació de polítiques de seguretat.
- **Identificació segura d'usuaris:** Tant per la protecció dels usuaris que accedeixen als nostres serveis com al conjunt d'elements emprats, la IA pot identificar l'ús d'identitats falses o atacs de força bruta o altres, de forma que crea una barrera addicional al accés fraudulent als nostres serveis.
- **Privacitat de la informació i "compliance":** La IA ajuda a classificar de forma automàtica la informació segons el seu nivell crític de cara a diferents regulacions com GDPR (General Data Protection Regulation - RGPD). Això implica un estalvi en quant als esforços actuals que es fan manualment, i s'evita les riscos que això suposa.

[45] Threat hunting: terme anglès, que significa caça d'amenaces, és un procés de cerca iterativa i proactiva a traves de les xarxes per detectar i aïllar amenaces avançades que son capaces d'evadir les solucions de seguretat existents.

[46] UEBA: acrònim anglès de "User and Entity Behavior Analiytics" posat per Gartner, es tradueix per anàlisi del comportament de les persones connectades a la xarxa d'una organització o empresa.

- **Bloqueig de bots segons comportament:** l'activitat de serveis "bot" (no perillosos), encara que sigui lícita, consumeix ample de banda dels nostres servidors, i perjudica la "experiència d'usuari" de clients. La IA pot classificar la seva activitat, i arribar a limitar la seva acció segons necessitats.

Figura 20: Connexió IA i humà, font: Amazon AWS

En aquest sentit, també veiem que la IA pot millorar molts altres punt febles de la Ciberseguretat, o accelerar la millora d'altres. Hi ha molts processos i aspectes que s'han arribat a acceptar com normals (que no ho son realment), i que la IA pot ajudar a millorar. D'aquesta forma, podem mostrar els següents:

- **L'error humà en la configuració:** les falles humanes son una part important de les febleses de la Ciberseguretat. La configuració adequada d'un sistema, pot ser difícil d'administrar, fins i tot per grans equips de TI. A part, podem tenir un elevat nombre de canvis de maquinari i com afegit els serveis al Cloud i dispositius externs. Aquí, per mitjà de la automatització intel·ligent i adaptativa, els equips poden rebre el assessorament oportú sobre els problemes acabats de descobrir. Es pot rebre assessorament per les opcions a aplicar o permetre ajustar la configuració de forma automàtica si es convenient.
- **L'eficiència humana amb activitats repetides:** aquest també és un dels aspectes problemàtics a la Ciberseguretat. No hi ha cap procés manual que es pugui repetir a la perfecció cada vegada, i més en entorns dinàmics. Molt sovint, després d'una configuració inicial, es torna a revisar el mateix equip per ajustar configuracions obsoletes o incorrectes. Aquí, suposem el cas d'amenaces, la IA pot accelerar els canvis amb un retard mínim.
- **La fatiga per excessos d'alarmes d'amenaces:** aquest també és un dels aspectes de feblesa a la Ciberseguretat. Cada cop hi ha més superfície

exposada a possibles atacs, i les capes de seguretat es tornen cada cop més elaborades i extenses, i aquestes poden generar una pluja d'alertes sense resoldre. Això, fa que l'equip humà sigui qui les ha d'analitzar, i pren les decisions i les mesures necessàries. Aquí, la IA pot ajudar a administrar un major nombre d'aquestes amenaces, de forma efectiva i practica, pot agrupar-les amb etiquetat automàtic, i en pot resoldre algunes.

- **El temps de resposta a les amenaces:** això es un valor clau pels equips de Ciberseguretat. Des de l'explotació fins a la implementació, els atacs maliciosos avancen molt ràpidament, i poden estar fins i tot automatitzats. La resposta humana pot anar darrera de l'atac inicial, però es sol centrar en els atacs amb èxit per prevenir els intents d'atac, pel que els atacs no descoberts son un perill. Aquí, la IA pot extreure les dades del atac, per agrupar-la i deixar-la a punt pel seu anàlisi, també pot donar informes simples pel fer més senzill el processament i pressa de decisions, i pot donar recomanacions sobre mesures a prendre per evitar danys i preveure futurs atacs.
- **La identificació i predicció de noves amenaces:** influeix en el termini de resposta a atacs de Ciberseguretat. Normalment, les amenaces existents es detecten amb cert retràs, i en el cas d'atacs desconeguts (comportament i eines desconegudes) la reacció de l'equip pot ser encara més lenta. I fins i tot, amenaces silencioses de tipus robatori de dades, a vegades poden passar desapercebudes. Per altra banda,cada dia apareixen nous atacs amb exploits[47] de "dia zero" que sol ser un problema greu. Aquí, la IA pot cercar elements comuns entre la nova amenaça i anteriors identificades en la detecció d'un atac, poden trobar patrons o comportaments similars, marcs de treball i codis emprats via ML, i així, pot predir noves amenaces i reduir el temps de reacció a elles.
- **La capacitat de dotació de personal:** sol ser uns dels problemes actuals a la Ciberseguretat. La quantitat o nombre de professionals qualificats d'una organització o empresa sol ser limitada. A part, de que s'ha de formar i certificar si cal, però principalment per tenir-lo al dia sobre Ciberseguretat. Aquí, la IA pot fer reduir el nombre d'especialistes que son necessaris, però caldrà que estigui al dia en temes de IA, ML i DL.
- **L'adaptabilitat:** aquest no es tan obvi com els anteriors, però pot tenir un fort impacte en les capacitats del servei de seguretat. Pot ser que a l'equip li falti capacitat per adaptar el seu conjunt d'habilitats als requisits específics que es necessiten, i per tant, es pot reduir l'eficàcia de l'equip. Aquí, la IA amb el conjunt de dades adequat, i els algorismes altament capacitats, pot esdevenir una solució a aquest problema.

[47] Exploits: terme anglès, per definir un programa informàtic, una part de software o Script que s'aprofita d'un error o vulnerabilitat, per provocar un comportament no previst en un dispositiu

En general, es considera que la IA a la Ciberseguretat pot abastar un gran conjunt de disciplines, junt a ML (Machine Learning) i DL (Deep Learning), però més aviat sembla que pot tenir o tindrà el seu propi paper. En principi, la IA es centra en el "èxit" mentre que la "precissió" té menys pes, ja que el seu objectiu es donar una resposta natural a tasques complexes, en base a unes decisions reals, independents, imparcials i transparents. La seva programació està dissenyada per trobar la solució ideal d'una situació, en lloc de sols la difícil conclusió lògica del conjunt de dades. Per tant, cal comprendre i entendre el mode de funcionament d'una IA moderna i les seves disciplines subjacents.

Encara que els sistemes autònoms de IA, no estan massa estesos a la Ciberseguretat, la seva feina no necessita d'accessos i interferència externa. No obstant, els sistemes de IA que ajuden o incrementen els nostres serveis de protecció, son pràctics i ja estan disponibles. El paper ideal de la IA a la Ciberseguretat, és la interpretació de patrons descoberts pels algorismes de ML, encara que la IA moderna no es capaç d'interpretar els resultats com ho faria un humà. Així, encara es desenvolupa de forma activa, algorismes similars al pensament humà, però encara hi ha molt per fer, les IA encara han d'aprendre a replantejar-se situacions en base a conceptes abstractes. És a dir, poder arribar a assolir aquell nivell de creativitat i pensament crític que tenen els humans.

De forma resumida, si estem decidits a implantar una IA per Ciberseguretat, cal implementar una bona estratègia per fer-ho, centrar-se en ella, i passar per les següents etapes:

- **Crear una plataforma de dades:** cal identificar les fonts de dades i/o crear plataformes de dades i lligar-les al funcionament de la IA.
- **Seleccionar casos d'ús d'alt impacte:** cal seleccionar un conjunt de casos d'ús rellevants per tal d'accelerar i maximitzar els beneficis.
- **Col·laborar externament:** cal col·laborar amb els Partners[48] estratègics per millorar la intel·ligència d'amenaces.
- **Valorar implementar SOAR / EDR:** donat que molts dels actuals es poden integrar amb una IA això pot millorar la gestió de la seguretat.
- **Formar a Ciberanalistes:** cal capacitar i formar els analistes de seguretat per que dominin la IA.
- **Governança eficaç:** cal establir un model d'administració de la IA per la Ciberseguretat, per tal d'oferir millores a llarg termini de forma transparent i ètica.

[48] Partners: terme anglès per socis de negoci, també s'empra per als serveis externs contractats.

10. Exemples de IA per Ciberseguretat

Donat de que ja existeix al mercat solucions de IA per Ciberseguretat, que apliquen molts dels algorismes, mètodes i funcionalitats que hem anat explicant en el decurs d'aquest treball, anirem a veure algunes d'elles. Principalment veurem com es defineixen elles mateixes, les característiques que remarquen com punts forts, informació de com treballen (si s'exposa), detalls curiosos o d'interès, i altra informació que es consideri rellevant de mostrar. Treballarem com exemple d'aquestes solucions, les següents: Darktrace "Cyber AI", IBM "Dr. Watson", i NVIDIA "Morpheus".

- **Cyber AI de Darktrace:**

Dins de la seva Web principal : https://www.darktrace.com/es/ podem trobar el següent: "El Darktrace Immune System es la plataforma de ciberdefensa autònoma líder mundial. La seva guardonada "Cyber AI" protegeix la seva força laboral i les seves dades contra atacs sofisticats, al detectar, investigar i respondre a les Ciberamenaces en temps real, independentment del lloc on es donin". Si anem a visitar l'enllaç del text anterior, tenim que defineixen "Cyber AI" com una tecnologia d'autoaprenentatge, que li permet detectar patrons rars i inèdits a la informació, en mig del soroll generat per l'activitat diària dels sistemes digitals d'una organització. Per mitjà de desviacions subtils del "patró de vida" de l'organització, pot distingir entre el amic del enemic, i mostrar atacs o ciberamenaces reals, que d'altra forma haurien passat desapercebudes. Els seu principals productes, amb els beneficis clau que mostren, son:

- Enterprise Immune System:
  - Detecció basada en l'autoaprenentatge (s'adapta continuament)
  - Anàlisis automatitzat (genera informes d'amenaces en segons)
  - 100% de visibilitat (protecció al núvol, SaaS, e-mail, OT[49] i on-prem)
  - Instal·lació ràpida (no requereix configuració manual)
- Industrial Immune System:
  - Autoaprenentatge (detecta noves amenaces a mesura que surten)
  - Cobertura i visibilitat total (véu: OT, IT, IoT, núvol, SaaS)
  - Identifica totes les formes d'amenaça (malware, falles, ...)
  - Sense punts de referència fixes (independent de protocols i tecnologies)
- Antigena Network:
  - Deté tots els atacs (fins a campanyes dirigides a desconeguts)

[49] OT: acrònim angles per "Operational Technology", és <tecnologia operativa> emprat en xarxes industrials.

- Resposta quirúrgica (manté el funcionament normal de la resta)
- Actua en segons (neutralitza propagació d'atacs en temps real)
- Supervisió total (envia notificacions al mòbil per actuacions)

- Antigena Email
  - Autoaprenentatge (actualitza contínuament segons noves evidències)
  - Compren al ser humà (es basa en patrons de comportament únics)
  - Resposta proporcionada (segons natura i gravetat de l'amenaça)
  - Instal·lació en 5 minuts (instal·lació virtual o amb hardware)

De totes formes, encara que no ho posen com a producte, degut als esquemes que mostren, sembla que hi ha com una eina anomenada "Cyber AI Analyst" que potser s'està integrant dins del producte "Enterprise Immune System" o que potser és o era una opció addicional a part, o que pot anar per separat.

Figura 21: Diagrama treball Darktrace, font: Darktrace

També, a la seva pàgina inicial, dins de l'enllaç "Cyber AI", podem descarregar un Informe Tècnic, en PDF, que dona més informació sobre l'eina. Aquí, presenten alguns eslògans com: "Líder mundial en Cyber AI autónoma" ó "Més de 6800 organitzacions en tot el mon confien en la plataforma de Cyber AI de Darktrace". I també declaracions com la del Dr. Nick Jennings, president de IA de la Imperial College de Londres (UK), que diu: "una important innovació en enginyeria ... essencial per tractar el volum, singularitat i velocitat dels incidents cibernètics moderns", o la del Dr. Andrew Herbert, membre de la Real Acadèmia d'Enginyeria, que diu: "Darktrace ha identificat una nova forma de seguretat cibernètica que promou tota la industria cap endavant més enllà dels actuals models de defensa. Per mitjà de la implementació de mètodes

d'aprenentatge de màquina avançats en una nova aplicació de software, ha establert una empresa líder sense competència significativa".

En l'informe tècnic comentat abans, del que hem tret la imatge anterior, es basa sobre el "Darktrace Immune System" amb tres apartats principals: Enfocament basat en l'autoaprenentatge, Resposta autònoma, i Cyber AI Analyst. Aquí, donen informació general enfocada a la seva IA (sobre autoaprenentatge, resposta autònoma, i anàlisis i interpretació de dades), però podem destacar una funció interessant, el "Threat Visualizer". Això, dona una visibilitat en temps real de tota la infraestructura digital, mostrant informació útil sobre el núvol, e-mail i xarxa corporativa, dins uns sol panell gràfic intuïtiu i fàcil d'emprar. I que simplifica la visualització i investigació de ciberamenaces, a més, pot retrocedir en el temps, fins al inici d'un incident, i veure cada esdeveniment en temps real.

La resta de punts de l'informe tècnic, fan referència a tenir protecció per a tota l'empresa amb els seu/s producte/s, comentant els detalls de "Cyber AI", sobre: el núvol i SaaS, e-mail, IoT, OT (xarxes industrials), i la xarxa. I en alguns d'aquests apartats mostra casos d'èxit de la seva eina. Un ràpid resum de l'anterior és el següent:

- Cyber IA pel núvol i SaaS[50]: Enfocat als serveis externs, núvol (cloud) i/o SaaS, per falta de visibilitat, control situació i possible relaxació de permisos d'accés, que mostra falta de solidesa i unificació en cobertura de seguretat. La seva eina cobreix aquest buit amb l'autoaprenentatge, sobre el que es normal en cada capa (comportament) i protecció de xarxa, que li permet detectar desviacions subtils, i indicadors d'amenaces. Aquí mostra cinc casos d'èxit: "M365 compromès i infiltració de SharePoint", "Configuració errònia al núvol", "Descarrega sospitosa d'arxiu Box", "Atac eludeix regla de 'viatge impossible' al Microsoft 365", i el de "Microsoft 365 i Microsoft Teams compromesos".
- Cyber IA per correu electrònic: Enfocat al correu electrònic, primer comenta els problemes que pot tenir, després diu que analitza els missatge de forma aïllada i els correla amb llistes negres, firmes o regles establertes, però que això es molt bàsic. Per tant, afegeixen la cerca de "patró de vida" per cada usuari, per trobar indicadors dèbils d'amenaces, cosa que permet aturar molts i variats atacs avançats de correu electrònic. Aquí mostra dos casos d'èxit: "Atac de Spoofing coordinat" i el de "Atac per correu electrònic de presa de control de la cadena de subministre".
- Cyber AI pel Internet de les Coses (IoT): Enfocat en dispositius IoT, tota la gama (inclou cafeteres, neveres i altres), comenta que no solen estar creats

[50] SaaS: acrònim angles per "Software as a Service" que vol dir solucions software com un servei via Internet o xarxa.

pensant en la seguretat. Poden estar coberts per la seguretat "endpoint" tradicional, per amenaces conegudes, però no es suficient. Comenten la impossibilitat d'executar un antivirus o una solució "endpoint" a aquests dispositius, i una falta de visibilitat de xarxa a les empreses, per dir la quantitat exacta que hi ha de "endpoints", i menys de IoT amb IP. Per tant, cal una visió més ampla, no sols considerar vulnerabilitats o el dispositiu gestionat, cal revisar els comportaments complexes que s'hi donen. Així, la seva solució permet monitorar tots els dispositius a la xarxa, aprendre el "patró de vida", detectar atacs dirigits a IoT, i donar resposta en temps real. Aquí mostra dos casos d'èxit: "Espionatge corporatiu per mitjà d'un hack de CCTV", i el de "Dades confidencials filtrades amb un locker intel·ligent".

- Cyber IA per xarxes industrials (OT): Enfocat en dispositius OT dels ICS[51] (Sistemes de Control Industrial), que solen estar aïllats de Internet, però que darrerament es solen integrar a la xarxa de TI. Això, genera desafiaments de seguretat, ja que molts dels OT van ser dissenyats fa dècades, pel que poden tenir moltes vulnerabilitats, i poden ser emprats com ponts per atacs més greus, a part del dany al maquinari sota el OT. Comenten que cal anar a una IA per donar una resposta forta i efectiva, que pot ser amb informació i anàlisis unificats sobre OT i TI, per detectar les amenaces inicials. Això es pot apuntalar amb "Cyber AI Analyst", ràpids resums d'alt nivell d'incidents. Aquí mostra dos casos d'èxit: "Virus Shamoon detectat", i el de "Eines de escaneig dirigides a ICS".
- Cyber IA per la xarxa: Enfocat en la protecció de la xarxa, entorns dinàmics i independent de la ubicació de treball. Primer comenta que a diferència dels sistemes tradicionals, Darktrace te una comprensió del comportament que es normal de la xarxa, i junt amb les anteriors eines, això li permet detectar tot un ventall d'amenaces. Afegeix que l'anterior amb Darktrace Antigena, li permet fer intervencions quirúrgiques sobre les amenaces emergents, molt ràpidament. Això no sols és una protecció dinàmica, intel·ligent i quirúrgica sinó que té un gran abast, i neutralitza automàticament el ransomware, criptomineria, i amenaces internes dirigides. Comenta que un cop feta una detecció o amenaça, es prenen mesures per a tota la xarxa. Aquí mostra dos casos d'èxit: "El ransomware Sodinokibi infecta una empresa de serveis financers", i el de "Mineria de Bitcoin encoberta".

Fins al final podem trobar diferents eslògans, avinents a l'apartat, i al final de tot hi ha les dades de contacte de les seves oficines internacionals. De totes formes, no ho diuen directament, però s'entreveu que els seus productes treballen o poden treballar en combinació, entre tots ells.

[51] ICS: acrònim anglès per "Industrial Control System", que vol dir Sistema de Control Industrial.

- **Dr. Watson de IBM:**

En aquest cas, IBM ha estat un pioner en temes de IA amb el seu Dr. Watson, a part de tenir molts productes relacionat amb la seguretat, i que se solen poder combinar entre ells, com el SIEM QRadar, QRadar SOAR, QRadar Advisor i Watson. Podem obtenir més informació d'aquests productes a:

- https://www.ibm.com/es-es/products/cognitive-security-analytics/details
- https://www.ibm.com/es-es/qradar/security-qradar-soar

Per tant, es veu que les eines estan escalades fins arribar a la IA "Watson" i s'entreveu que la base de tot plegat, és el SIEM, ja que IBM les defineix i mostra com a possibles complements d'aquest, encara que alguns empren IA i ML. Així, els complements comentats, serien els següents:

- IBM Qradar User Behavior Analytics: dona visibilitat sobre les amenaces internes, comportaments anòmals, usuaris en risc, i genera coneixement significatiu aplicant ML de comportament a les dades de QRadar.
- IBM QRadar Advisor with Watson: fa ús de IA i el ML, de forma que l'equip es pot centrar en les qüestions crítiques de seguretat. L'assessor es cuida de les amenaces repetitives del SOC, impulsa investigacions coherents i exhaustives i redueix temps de durada per aconseguir un procés d'escalat més decisiu.
- IBM QRadar Incident Forensics: repassa les accions dels Ciberdelinquents per tenir informació detallada sobre el forat provocat, reconstrueix les dades afectades per l'incident, per veure el delicte pas a pas. Dona una major visibilitat als equips de TI, encara que no disposin de coneixements o de la formació necessària.
- IBM QRadar Data Store: recopila, analitza i emmagatzema de forma rentable grans volums de dades de seguretat i operacions de TI. Empra la IA per generar coneixement detallat en les investigacions i crea ràpidament aplicacions personalitzades per resoldre qualsevol problema de seguretat.
- IBM QRadar Data Synchronization: millora la resiliència de TI, i recuperació després de desastre. Permet copiar les dades (successos i fluxos) i arxius de configuració de forma fàcil i rentable, per als diferents desplegaments i la recuperació per QRadar. També pot gestionar quin desplegament està actiu per desastre, error humà ó al provar capacitats de resiliència de dades.

A la Web anterior de "QRadar Advisor with Watson", es pot descarregar un document PDF: "Beyond the Hype: AI in your SOC", que respon a set qüestions molt bàsiques sobre els riscos, seguretat, IA, ML, l'equip de seguretat i la seva implementació. Sembla que està pensat principalment per a directius o alts càrrecs desconeixedors del tema. Així i tot, presenta un quadre resum com:

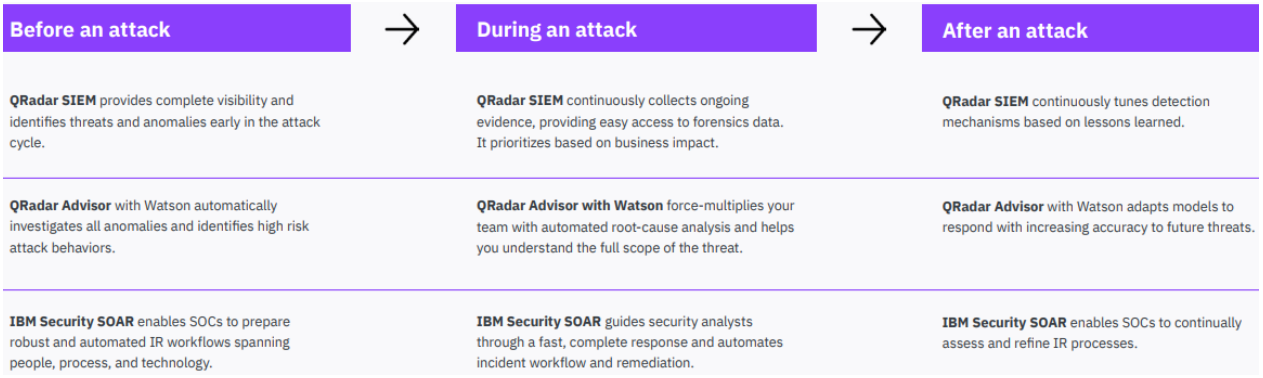


Figura 22: Quadre eines IBM QRadar, font: IBM

De totes formes, al document, en el punt 6, es respon a la pregunta: "Quins avenços de seguretat puc esperar amb IA?, i ho fa amb vuit punts, que son els següents:

1. Encadena diferents incidents potencials, automàticament: comenta que la IA destaca en l'automatització i la integració de l'anàlisi de causes arrel, capta connexions per conèixer amenaces i riscos, no es cansa, i a més mostra les interrelacions. Evita tancar una alerta perquè que era una única instància d'atac, troba els punts en comú entre incidents via el raonament cognitiu i proporciona retroalimentació accionable amb el context. La IA recopila informació sobre amenaces externes per ajudar a afegir més context al nostre anàlisi i detectar allò que els altres poden perdre.
2. Soluciona el problema de personal: La IA determina l'anàlisi de la causa arrel i pot orquestrar els passos següents basant-se en el coneixement que ha construït sobre les amenaces i la nostra organització. No se'n va de vacances, no et deixa mai per a una altra feina, i no ens hem de preocupar de no reconèixer un IOC significatiu.
3. Condueix investigacions coherents i més fondes, cada vegada: La IA pot llegir dades estructurades i no estructurades, superant els humans. Aprèn i ofereix la informació que necessària per reduir el temps mitjà en detectar i el temps mitjà en respondre (MTTD i MTTR), amb un procés d'escalada més ràpid i decisiu. La IA ofereix analítiques avançades per detectar amenaces conegudes i desconegudes. Impulsa investigacions coherents i més fondes cada cop, i permet als analistes prendre decisions basades en dades en lloc de confiar en el seu instint.
4. Realitza investigacions més exhaustives i coherents en una fracció del temps: L'aprofitament de la IA per fer una extracció automàtica de dades d'investigació/intel·ligència d'amenaces permet als analistes de seguretat fer investigacions més exhaustives, coherents i en una fracció del temps. I els permet centrar-se en investigar amenaces estratègiques i caça d'amenaces.

La IA correlaciona la intel·ligència d'amenaces amb les investigacions donant als analistes una visió més completa de l'amenaça.

5. Es centra primer en les alertes més importants: La priorització d'alertes ajuda els analistes a classificar les alertes de forma eficaç, agafa primer les alertes més crítiques, descobreix falsos negatius i falsos positius i redueix en gran mesura les possibilitats de que es perdin incidents crítics.
6. Aprofita MITRE ATT&CK[52] per a investigacions d'amenaces més efectives: El mapa de les accions de l'atacant al marc MITRE ATT&CKTM representa visualment una línia de temps d'esdeveniments que mostra la progressió d'una amenaça, donant lloc a investigacions d'amenaces més ràpides i precises, que al seu torn redueixen el temps d'espera.
7. Dona una visió completa de la investigació: L'anàlisi d'investigació creuada reflecteix una visió més completa de la investigació més enllà de la infracció actual. Identifica i connecta alertes vinculades al mateix atac que semblen no tenir relació aparent, redueix el nombre d'alertes i duplicació de treballs.
8. Ofereix un flux de treball de resposta a incidents robust i automatitzat que abasta persones, processos i tecnologia: La IA guia els analistes de seguretat amb una resposta ràpida i completa que es basa en dades i proves. Automatitza el flux de treball i la correcció, i permet als SOC avaluar i perfeccionar els processos de resposta a incidents de forma contínua.

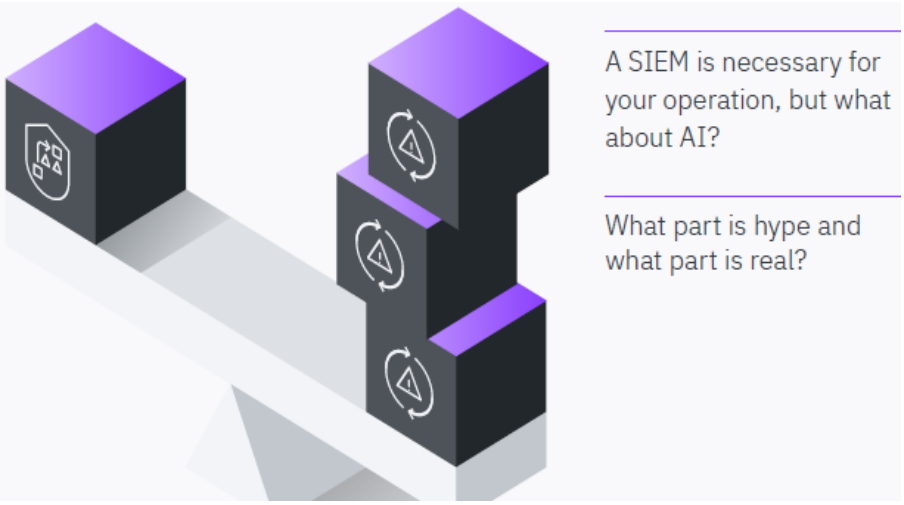


Figura 23: Imatge al inici del document PDF, font: IBM

En general, podem veure que l'anterior, es centra en combinar diferents productes que son ajudats o millorats amb la IA IBM Watson, o eines amb coneixements derivats de la IA Watson. De totes formes, a la seva Web també podem trobar els productes: "Watson Assistant" i "IBM Watson Studio", que en

[52] MITRE ATT&CK: acrònim anglès de "Adversarial Tactics, Techniques, and Common Knowledge" és una guia per classificar i descriure els ciberatacs i les intrusions.

principi treballen desvinculats del SIEM. Tenim més informació d'aquests productes a:

- https://www.ibm.com/es-es/products/watson-assistant/enterprise-security
- https://www.ibm.com/cloud/watson-studio

El "Watson Assistant", es descriu com una IA conversacional preparada per a l'empresa, i que protegeix als nostres clients amb seguretat i escalabilitat. Te la certificació ISO 27001, 27017 i 27018, la certificació SOC 2, acompleix HIPAA (Healt Insurance Portability & Accountability Act dels EE.UU.), i acompleix el GDPR de la Unió Europea. En principi, està destinat a organitzacions grans i complexes amb dades confidencials, que necessiten funcions de seguretat i escalabilitat, per estar protegides de l'ús indegut de dades de clients i donar suport al agent virtual en hores punta. Es pot instal·lar en l'entorn de l'empresa via el "IBM Cloud Pak® for Data", i ofereix una disponibilitat del 99,9% segons el pla SLA[53] (Service Level Agreement) marcat.

En canvi el "IBM Watson Studio" **permet crear, executar, gestionar i escalar una IA** de confiança en qualsevol núvol (Cloud) i permet automatitzar el seu cicle de vida per a ModelOps[54], i també optimitzar les decisions de qualsevol lloc de "IBM Cloud Pak® for Data". Fa servir marcs de codi obert com PyTorch, TensorFlow i scikit-learn amb IBM i les seves eines d'ecosistema per a la ciència de dades visual i basades en codi. Per fer-ne ús, recomana una sèrie de passos o accions com les següents:

- Implementar MLOps i IA de confiança
- Optimitzar les decisions
- Desenvolupar models visualment
- Accedeix a Watson NLP
- Accelera el desenvolupament de la IA amb "AutoAI"
- Aprenentatge federat

Comenta que te opcions flexibles, i mostra uns models on poden estar les nostres dades com: "IBM Cloud Pak for Data", "IBM Cloud Pak for Data as a Service", i "IBM Cloud Pak for Data System". I també mostra els beneficis que podem obtenir, i que son els següents:

- **Optimitza la IA i l'economia del núvol:** IA multinúvol d'empresa amb models de consum flexible, i es pot crear i implementar a qualsevol lloc.

[53] SLA: acrònim anglès per "Service Level Agreement", que és l'Acord de Nivell de Servei contractat.
[54] ModelOps: terme anglès, segons Gartner: model que es centra principalment en la governança i la gestió del cicle de vida d'una àmplia gama de model de IA i decisions operacionals (ML, DL, agents, ...).

- **Predir resultats i prescriure accions:** optimitza els horaris, els plans i l'assignació de recursos per mitjà de prediccions. Simplicfica el modelatge d'optimització amb una interfície de llenguatge natural (NLP).
- **Sincronitza aplicacions i IA:** permet unir i entrenar desenvolupadors i científics de dades. Rep models per mitjà d'una API REST a qualsevol núvol i estalvia temps i costos en gestió d'eines diferents.
- **Unifica les eines i augmenta la productivitat per ModelOps:** IA per a qualsevol núvol, governa i assegura projectes de ciència de dades a escala.
- **Ofereix una IA justa i explicable:** els esforços de supervisió es redueixen entre un 35% i un 50%, augmenta la precisió del model entre un 15% i un 30%, per tant, augmenta els beneficis nets d'una plataforma de dades i IA.
- **Gestiona els riscos i el compliment normatiu:** protegeix vers l'exposició i les sancions regulatòries, i simplifica la gestió del risc del model de IA per mitjà de la validació automatitzada.

Per altra banda, mostren i comenten del "IBM Watson Studio" característiques com les següents:

- "AutoAI" per una experimentació més ràpida
- Refineria de dades avançada
- Suport de bloc de notes de codi obert
- Eines visuals integrades
- Model de formació i desenvolupament
- Amplis marcs de codi obert
- Optimització integrada de decisions
- Gestió i seguiment de models
- Model de gestió de riscos

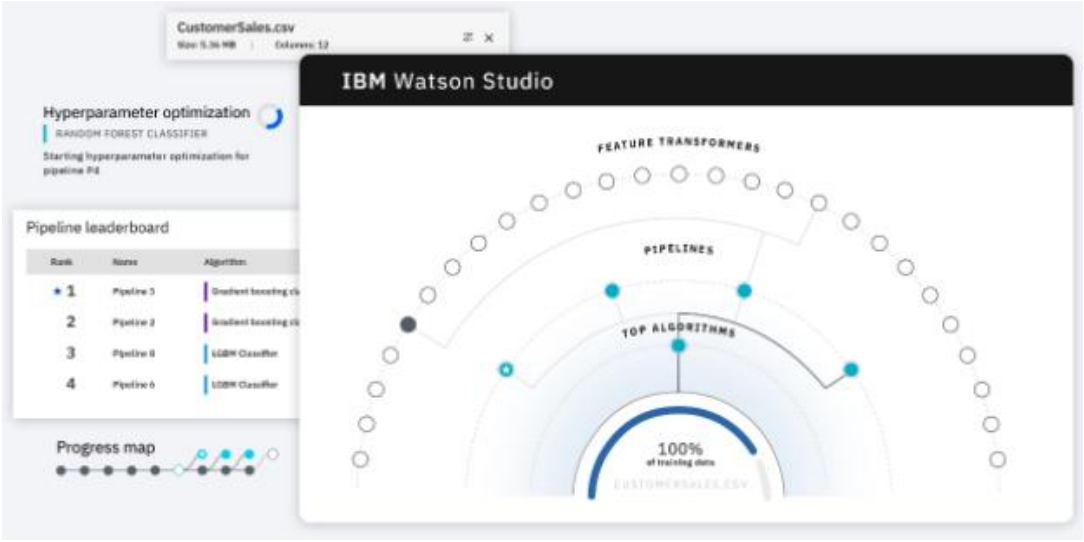


Figura 24: Imatge IBM Watson Studio, font: IBM

- **Morpheus de NVIDIA:**

Tenim la seva Web a: https://developer.nvidia.com/morpheus-cybersecurity i es presentat com: "Open AI Framework per a proveïdors de Ciberseguretat". Diu que NVIDIA Morpheus és un marc d'aplicacions obert que permet als desenvolupadors de Ciberseguretat crear canalitzacions de IA optimitzades per filtrar, processar i classificar grans volums de dades en temps real. Els kits per a desenvolupadors de AWS (amazon), de Red Hat, o que s'executen en servidors certificats per NVIDIA, admeten models de IA prèviament entrenats, permetent als clients inspeccionar contínuament la telemetria de la xarxa i del servidor a escala. Morpheus, ofereix un nou nivell de seguretat de la informació als centres de dades, permet la protecció dinàmica, la telemetria en temps real i les defenses adaptatives, i això li permet detectar i corregir les amenaces a la Ciberseguretat.

Els seus principals trets sobre els beneficis que pot aportar, son els següents:

- **Detecció d'amenaces en temps reals basada en IA a escala:** les empreses poden observar totes les seves dades a tota la xarxa, i aplicar la inferència de IA i la supervisió en temps real a tots els paquets i fluxos de dades necessaris, a una escala que abans era impensable d'aconseguir.
- **Analitzar el comportament de cada usuari:** les millores que te, permet als desenvolupadors implementar fluxos de treball que empren cada usuari, servei, compte i màquina de forma única al centre de dades de l'empresa. És fa servir l'aprenentatge no supervisat per marcar quan canvien els patrons d'activitat de l'usuari i la màquina.
- **Detectar amenaces a la Ciberseguretat a l'instant:** ofereix als equips de seguretat una visibilitat complerta de les amenaces de seguretat, i combina un processament de IA inigualable amb la supervisió en temps real de cada servidor i paquet a tot el centre de dades.

De la mateixa forma, com eina que és "Morpheus", també es mostra quines son les seves principals característiques, i que son les següents:

- **Desenvolupament i desplegament ràpid:** integra eines per facilitar la creació de solucions de Ciberseguretat. És basa en les biblioteques "RAPIDS™" (aprofita execució processos a la GPU[55] del sistema), els marcs d'aprenentatge profund (DL) i el servidor d'inferències "NVIDIA Triton™". Així, simplifica l'anàlisi dels registres i la telemetria per ajudar a detectar i mitigar les amenaces de seguretat.

[55] GPU: acrònim anglès per "Graphics Processing Unit", que seria la unitat de procés gràfic del sistema, és sol explotar molt en la mineria de Bitcoin, es fa per aprofitar recursos del maquinari.

- **Capacitats de Ciberseguretat de IA:** es poden implementar els nostres propis models per mitjà de marcs d'aprenentatge profund (DL) comuns. O també emprar un dels models prèviament entrenats i provats de NVIDIA, per iniciar-se en la creació d'aplicacions, per identificar informació confidencial filtrada, detectar programari maliciós o frau, fer mapes de xarxa, marcar canvis de comportament dels usuaris o identificar errors via els registres.
- **Telemetria en temps real:** pot rebre telemetria de xarxa, en temps real de tots els servidors accelerats amb NVIDIA® BlueField® DPU[56] i aplicació basada en NVIDIA DOCA™, inclosos "Telemetry Flow Inspector" i App Shield al centre de dades, sense afectar el rendiment. La integració del marc en una oferta de Ciberseguretat de tercers, aporta la millor informàtica de IA del món a les xarxes de comunicació.
- **Preparat per DPU:** la unitat de processament de dades (DPU) de NVIDIA BlueField, descarrega, accelera i aïlla les funcions crítiques de la infraestructura de centre de dades. També amplia el registre de seguretat estàtica a un model sofisticat de telemetria dinàmica, en temps real, que evoluciona amb noves polítiques i intel·ligència d'amenaces.

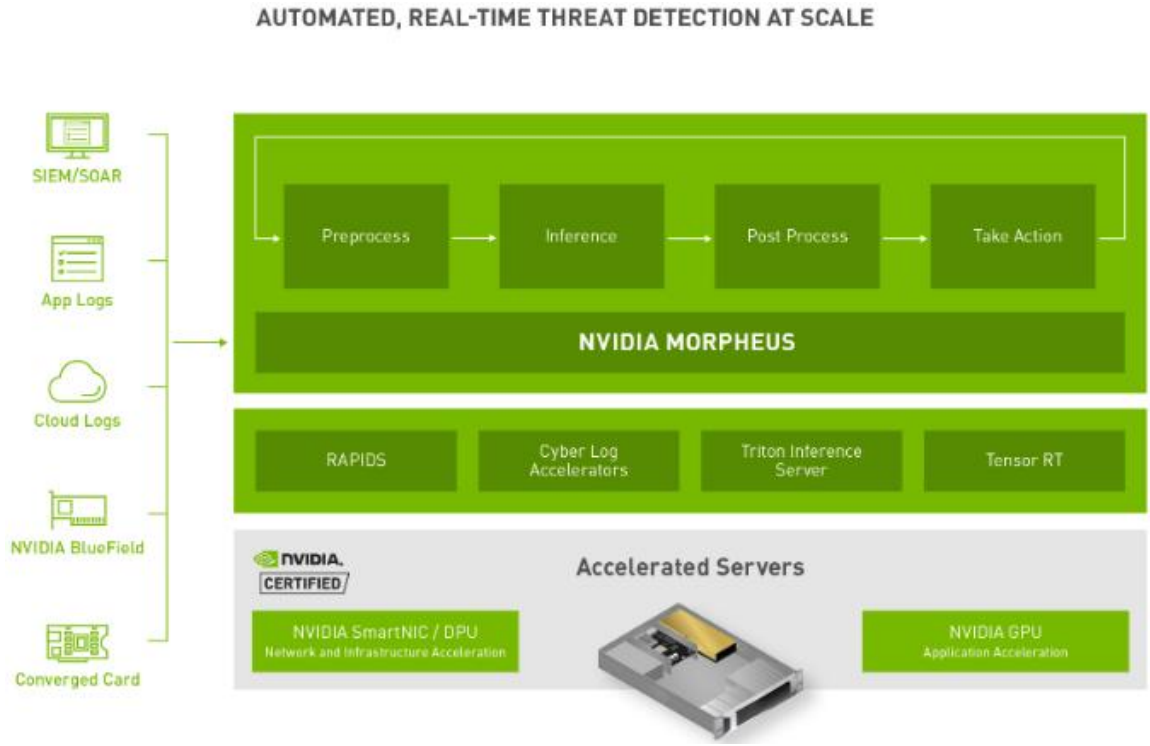


Figura 25: Arquitectura de NVIDIA Morpheus, font: NVIDIA

[56] DPU: acrònim anglès de "Data Processing Unit", son les unitats de processament de dades, una nova classe de processador programable, especialitzat en moure dades en els centres de dades.

Per altra banda, a la informació final, disposen d'un apartat, per mostrar les capacitats clau de la Ciberseguretat de IA, segons el seu model. D'aquesta forma, en dos columnes mostren primer els seu models preformats i en l'altra columna opcions per personalitzar les nostres aplicacions, en detall això és:

Models preformats:

- Empremta digital: pot generar una empremta digital única de cada usuari, servei, compte i màquina del centre de dades de l'empresa. Això pot ser emprat per l'aprenentatge no supervisat, per marcar quan canvien els patrons d'activitat de l'usuari i de la màquina.
- Classifica le dades sensibles filtrades: cerca i classifica credencials, claus i contrasenyes, números de targetes de crèdit, números de comptes bancaris i molts més filtrats.
- Anomalies de comportament del perfil: detecta les anomalies perfilant comportaments per detectar codis maliciosos o configuracions incorrectes. Pot emprar diversos tipus de registres per orientar casos d'ús específics.
- Detectar intents de pesca (Phishing): pot emprar un model de IA amb NLP (processament del llenguatge natural) per analitzar correus electrònics en cru (raw), i classificar-los automàticament com entrada, brossa o Phishing.
- Identificar errors als registres del servidor: escaneja els registres del servidor amb el model de manteniment predictiu basat en NLP, per identificar errors i falles potencials que no es marcarien amb les regles de filtratge de registres existents.

Personalitzar les nostres aplicacions:

- Integrar els desplegaments i el codi existents: permet integrar fàcilment els nostres propis models ja entrenats amb "Morpheus" per fer inferències, i es pot canviar per models més nous, sense interrompre la pipeline. També es pot integrar sense esforç amb eines SIEM, SOAR, visualització o gràfics existents.
- Formats comuns compatibles: empra marcs d'aprenentatge profund (DL) comuns coms ONNS, PyTorch, TensorFlow, i NVIDIA TensorRT™ amb el suport del Triton Inference Engine. La inferència d'arbre també s'admet via la "RAPIDS Forest Inference Library (FIL)".
- Supervisar el rendiment del model: permet tenir mètriques en temps real sobre el rendiment del model amb MLOps integrats.
- Afinament dels guions (Scripts): pot augmentar la precisió i reduir els falsos positius emprant Scripts bàsics, que es poden personalitzar pel nostre entorn. Permet oferir als nostres científics de dades i experts una avantatge en IA personalitzada per a les nostres aplicacions úniques.

11. Futur de la IA a la Ciberseguretat

Anteriorment, ja hem comentat que la IA a la Ciberseguretat pot abastar un gran conjunt de disciplines, junt a ML (Machine Learning) i DL (Deep Learning), i que com disciplines, continuen la seva evolució. D'entrada, ja veiem que es barreja models i conceptes de IA convencional i IA computacional, apareix la IA cognitiva com resultat de la associació de ML i el NLP (Natural Language Processing) i es fan servir models de Xarxes Neuronals amb el DL. Per altra banda, per donar resposta a amenaces conegudes i que estan documentades, no es comenta directament, però de segur que s'empra Sistemes Experts integrats a la IA (no cal que s'aprengui, ja sap que cal fer) i que s'actualitzen.

Tot l'anterior es continua desenvolupant i millorant en quant a funcionalitats, capacitats i abast sobre sistemes, el núvol, xarxes i endpoints. De totes formes, **amb els diferents aprenentatges, queda palesa una forta dependència de la IA sobre les dades**, no sols per tenir una "IA fiable", sinó per acomplir tota una sèrie de requeriments, com pot ser el GDPR. D'aquesta forma, per dissenyar, desenvolupar, validar i desplegar sistemes de IA, **s'ha de valorar la qualitat de la dada, del model i del resultat**. Per tant, hi ha tot un seguit d'aspectes a tenir en compte a l'hora de desenvolupar un sistema de IA, aplicables actualment i en un futur, degut a regulacions i privacitat de dades, que serien els següents:

- Privacitat: assegurar-se que el model de IA pot emprar qualsevol model de dades i que es manté la privacitat de la dada.
- Equitat: s'ha de vetllar per que el model de IA no sigui afectada per biaixos implícits en la mostra de dades, i no discrimini o afavoreixi determinades sortides sense raó aparent.
- Traçabilitat: si es dona una falla del sistema, s'ha de poder analitzar quin ha estat el motiu, depurar responsabilitats i posar els mitjans necessaris per que no es repeteixi.
- Robustesa: en cas d'un Ciberatac, s'ha de conèixer fins a quin punt podem fiar-nos o no de la sortida del sistema.
- Fiabilitat: s'ha de conèixer si la sortida del model es fiable i que passa si la seva entrada canvia mínimament.
- Causalitat: cal conèixer si podem influir en la sortida del model, actuant sobre les dades d'entrada.
- Explicabilitat i transparència: s'ha d'intentar que l'usuari pugui entendre com funciona el model i que ha vist el sistema en les dades per obtenir certes conclusions.
- Governança de la dada: s'ha de garantir l'ús lícit, eficient i eficaç de la dada

També abans, ja s'ha comentat que hi ha manca de coneixements i d'experts en Ciberseguretat, però també **falten molts experts en IA amb ML**, personal que pugui treballar en la programació d'aquest àmbit. La seguretat de la xarxa de ML es pot beneficiar molt de personal que la pugui mantenir i ajustar-la segons necessitats. Podem veure que el conjunt global de persones qualificades i capacitades és molt petit i que no pot cobrir la demanda mundial de personal que pugui donar solució a aquest problema. Per tant, es pot preveure, que en el futur anirà creixent encara més la demanda sobre personal expert en IA i ML, a mesura que aquesta tecnologia es va introduint i aposentant en les organitzacions i empreses.

Figure 1: Total AI Jobs Posted in Top 12 Countries by GDP, July 2015 Through March 2019

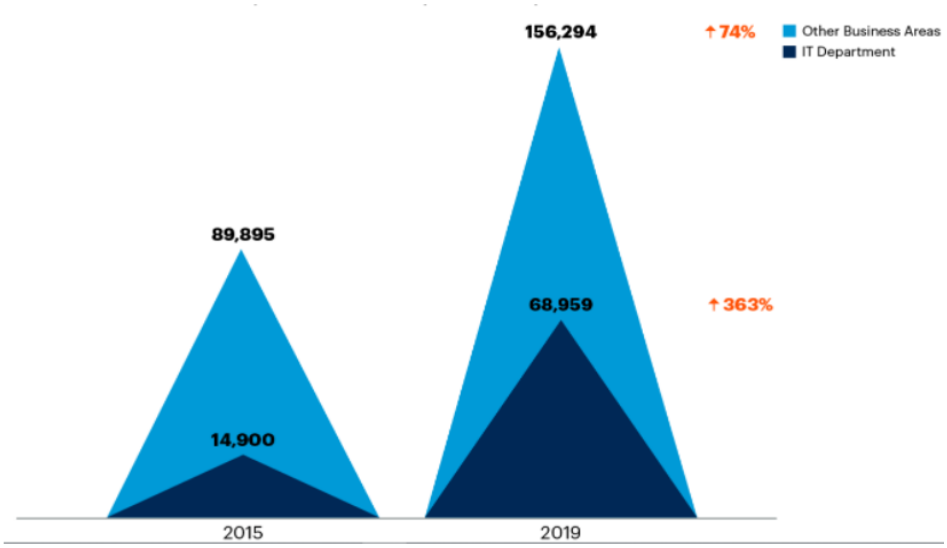


Figura 26: Gràfic demanda talent IA, font: Gartner

A més, abans s'ha comentat que la IA i ML, pot suplir algunes mancances del personal de seguretat, cosa que pot portar a pensar que es pot prescindir de certes funcions. Realment, més aviat es diversificarà, hi haurà altres rols nous i s'agruparan d'altres, **els equips humans seguiran sent essencials** per la continuïtat i resiliència de la Ciberseguretat, treballant amb IA i ML. Així, el pensament i la creativitat seran fonamentals en la presa de decisions, ja que la IA no està preparada per això. A més, cal triar les dades, preparar tot el procés de ML, revisar resultats, validar sortides, etc. això es tot un procés que l'equip de Ciberseguretat ha de gestionar, junt al seu manteniment i ajustos necessaris en el temps. És a dir, apareixen noves tasques a gestionar, que es poden prioritzar o no, altres es poden deixar a la gestió de la IA, i altres poden passar sols a una supervisió i control del que informen. Per tant, es pot intuir un reajustament organitzatiu i realineament de com funciona la Ciberseguretat a una organització o empresa, que funcioni amb IA i ML.

Per altra banda, segons surt a un informe realitzat pel institut d'investigació de "Capgemini", suggereix enfortir les defenses de la Ciberserguretat amb IA. Diu que **és un assumpte que s'ha de resoldre urgentment en les empres modernes**, degut a que alguns Ciberdelinqüents ja estan emprant tecnologia IA per fer atacs cibernètics. Del seu informe, podem destacar certs comentaris, com:

- "3 de cada 4 executius diu que la IA permet que la seva organització respongui a violacions més ràpidament".
- "El 69% de les empreses creu que la IA es necessària per respondre als Ciberatacs".
- "3 de cada 5 empreses diu que l'ús de la IA millora la precisió i eficiència dels analistes Cibernètics".

En aquest sentit, pensen i creuen que la IA i ML, faran un paper fonamental i important en diversos sectors, properament, en un futur pròxim. Amb l'ajuda d'aquestes eines a la Ciberseguretat, les organitzacions estaran molt més protegides, però hauran d'estar atentes als atacs Cibernètics avançats i ajuntar aquesta tecnologia amb la força de treball humà. A més, a mesura que la xarxa creix, tenim més superfície exposada, i les dades son més complexes. La IA, ofereix millors solucions a les necessitats de Ciberseguretat de les empreses, i és converteix en una aliada inevitable per la Ciberseguretat moderna.

Finalment, **anem a veure deu tendències** sobre Ciberseguretat de les que es pot millorar la gestió o mitigar incidents o explotar informació, amb l'ús de la IA amb ML, degut a les seves aplicacions i funcionalitats que hem explicat en el decurs d'aquest treball. I que serien les següents:

- **Resposta al Ransomware**: el perill del ransomware encara existeix, a més hi ha el RaaS (Ransomware as a Service) i el nombre de formes de fer-se vulnerable a un atac d'aquest tipus creix. Cal seguir unes normes bàsiques i educar els Usuaris per evitar aquest potencial risc d'infecció. Aquí, la IA ens ajuda amb el correu electrònic, el reconeixement de processos i "patrons" a la xarxa, encara hi ha molt camí per recórrer degut a noves vulnerabilitats.
- **Combinar Desenvolupament d'aplicacions amb Ciberseguretat**: si la nostra empresa desenvolupa aplicacions, s'ha de fer que la Cibersegurtat formi part del cicle de vida del desenvolupament. El problema es dona per que sovint s'empra un desenvolupament àgil per estalviar temps, o per pressió als programadors per acabar-ho. Així, codi amb errors o sense verificació i validació acaba sent el de producció. Cal fer les proves pertinents, encara que sigui via una empresa externa, ja que pot donar una visió millor que la del programador que pot estar pressionat. Aquí la IA ens pot ajudar a minimitzar els perills informant d'ús fora del normal.

- **Emprar DL per detectar "DGA[57]-Generated Domains"**: es sol emprar DGA, per generar i tenir noms de domini com origen d'un atac, i es poden donar desenes de milers de noms de domini en un sol atac. Posar el noms en una llista negra, es una pèrdua de temps, això es pot emprar per entrenar una xarxa neuronal, que ven entrenada, sol tenir èxit en detectar-ho. Els canvis d'adaptació vistos sobre DGA, continuen sen aturats per DL.
- **Detectar amenaces que no son de programari maliciós**: hi ha molts vectors d'atac que no son de programari (Locky, CryptoWall, CTBLocker, ...) que empren vulnerabilitats de navegadors, Microsoft Office, PowerShell i WMI (Windows Management Instrumentation). Per detectar-ho, cal revisar el comportament de l'ordinador, pel que es pot emprar xarxes neuronals i algorismes d'aprenentatge automàtic per controlar comportament normal. S'ha d'entrenar la IA per que diferenciï les dos situacions, i així tenir millors mètodes de detecció.
- **Honeypots[58] adaptatius i Honeytokens**[59]: això és una trampa pels "hackers", per tal de recopilar informació sobre un atac i l'atacant. Donat que això ja és un truc vell, cal emprar models adaptatius que canvien el que fan en funció de l'atac. Així, es llença defenses per veure la reacció de l'atacant i aconseguir més informació sobre habilitats i eines emprades en cada nou repte. Una IA pot aprendre aquest comportament i el pot reconèixer en un futur, a part de tenir identificada informació rellevant de l'atacant.
- **Aconseguir millor comprensió de com funcionen les xarxes neuronals**: en la Ciberseguretat anirà augmentant l'ús de xarxes neuronals per detectar atacs. Pel que cal pensar en una explicació assequible a tothom i que s'entengui com funciona. Els resultats de senyals processats per moltes capes de neurones, és que només aquells que han continuat complint el llindar, han continuat fins la següent capa, fins que finalment, es pren una acció. En una IA, el DL fa servir o te com a base, les xarxes neuronals.
- **Utilització de xarxes neuronals a capsules (CapsNet)**: Per poder processar grans volums d'informació, s'empra les xarxes neuronals convolucionals, que filtren les dades que flueixen entre les capes d'una xarxa neuronal, per reduir la quantitat de dades processades en cada capa posterior. Les CapsNet principalment s'han utilitzat en reconeixement d'imatges, però la idea de les relacions i la informació de l'entitat farà que

[57] DGA: acrònim anglès de "Domain Generation Algorithms", que és Algorismes de Generació de Dominis, s'empra per generar noms de domini pseudoaleatoris com orígens d'atacs.

[58] Honeypot: terme anglès per "Pot de mel", sol ser un sistema preparat especial, amb vulnerabilitats que atreuen als atacants en una trampa, per tal de recollir informació sobre l'atacant.

[59] Honeytoken: terme anglès per "Testimoni de mel ", es tracta d'una informació aparentment valuosa, que es troba amagada, però a la que es pot arribar amb una mica de "hacking" i que et porta a un lloc monitorat.

les xarxes neuronals siguin més intel·ligents, podent identificar patrons complexos en atacs de Ciberseguretat.

- **Aprenentatge de reforç profund**: Aquí, per resoldre problemes complexos, s'afegeixen més capes a les xarxes neuronals, passant a ser xarxes d'aprenentatge profund. Se sol emprar un procés anomenat "Q-Learning" que implica els termes estat i acció, que s'utilitzen per trobar una solució òptima. Això, ofereix a la IA la capacitat d'aprendre sense posar-hi atenció, quan s'enfronta a nous reptes. A la Ciberseguretat , el futur ofereix una gran varietat d'escenaris en que la IA haurà d'aprendre per si sola la millor forma de derrotar un atac. La informàtica cognitiva donarà una visió de les millors respostes que pot tenir una analista.
- **Protecció de IoT (i de OT industrial)**: com ja hem comentat, els IoT fan expandir el nombre de "endpoints", que cal tenir en compte al protegir tota la xarxa. Cal endurir el sistema operatiu dels IoT, actualitzar-lo, fer proves de seguretat, i endurir la xarxa a la que es connecten. Aquí, la IA ens pot ajudar amb el reconeixement de comportament i patrons de la xarxa.
- **Predicció del futur**: de moment la informàtica cognitiva ajuda en la investigació de les amenaces Cibernètiques i identifica la causa arrel d'una atac. Però, la Xina ha fet un pas més, ha començat a emprar IA per intentar predir delictes abans que es produeixin. El viceministre de Ciència, Li Meng, va anunciar que amb l'ús de IA i reconeixement facial, volen recopilar informació sobre persones i la seva activitat. Amb Big Data, la IA per lluitar contra el crim, està creant un sistema de classificació, per etiquetar grups de persones altament sospitoses. Ni tant sols intentar enganyar el sistema amb una mascara serà viable, això farà la IA més intel·ligent, "reidentificant" un individu. Això representa un ús potencial per a màquines cada cop més intel·ligents, però en els països més democràtics, això de moment es impensable.

De totes formes, publicacions més recents al Web, mostren **cinc tendències sobre la IA i la Ciberseguretat**, que potser son més genèriques o d'àmbits al voltat de la IA i Ciberseguretat. Per exemple: la firma d'investigació "Technavio" espera que el mercat de la Ciberseguretat basat en IA creixi en 19 mil milions de dòlars del 2021 al 2025. Ara, anem a veure en detall aquestes tendències :

- **Reduirà la càrrega i la necessitat d'experts a Ciberseguretat:** donat que amb la pandèmia molt personal va treballar en remot i al tornar a la nova normalitat, molts s'han quedat teletreballant, s'augmenta la superfície exposada. Això es tradueix en més feina de gestió i informació o alertes pels equips de Ciberseguretat, d'altra banda ja coneixem la manca d'experts en Ciberseguretat, i sabem que cap humà pot estar al dia amb totes les amenaces viables. Per tant, la IA juga un paper important en aquests casos,

pot reconèixer patrons d'atacs, activitats sospitoses i els punts de xarxa més vulnerables. Pot fer front a tasques repetitives i propenses a errors, i fer informes automatitzats per revisar analistes humans. Així, la IA gestiona i alleugereix el increment de possibles amenaces i la feina de Ciberseguretat.

- **Automatitzarà les mesures de seguretat en gestió d'identitats i accés:** ara, la IAM[60] (Identity and Access Management) és més important que mai, per marcs de seguretat amb confiança zero, es requereix que tot usuari de la xarxa estigui autenticat, autoritzat i validat contínuament. La IA por reduir la feina a fer, introduint automatització intel·ligent als sistemes de seguretat, pot analitzar i supervisar activitats de l'usuari. I també la IA pot millorar la seguretat en l'experiència d'autenticació del client, des de creació compte, inici de sessió i interacció amb els comptes de servei.
- **Millora el Blockchain[61] (si s'empra):** l'ús de Blockchain s'ha incrementat els darrers anys, no sols pels "Bitcoin" sinó en aplicacions de registres mèdics, o registres de votació electrònica. La IA pot diferenciar l'ús d'uns casos i altres, i millorar la seguretat dels casos lícits, també pot analitzar les cadenes de dades a granel.
- **Millorarà esforços en compliment normatiu:** la IA pot aplicar normes i requisits reguladors a les dades per mitjà de xarxes complexes, que és un sistema millor que els processos de cerca manuals. Es pot emprar la IA per fer un seguiment de les agències reguladores del món, i així ajudar a controlar i mantenir el compliment continu, a mesura que les regles canviïn.
- **Millora de la seguretat de la xarxa al núvol (Cloud):** moltes empreses traslladen al núvol part de gestió i també les dades, la Ciberseguretat s'ha tornat més complexa. Les noves IA estan dissenyades per treballar al núvol i apareixen noves solucions híbrides, que inclou la capacitat de supervisar i analitzar dades en diferents entorns.

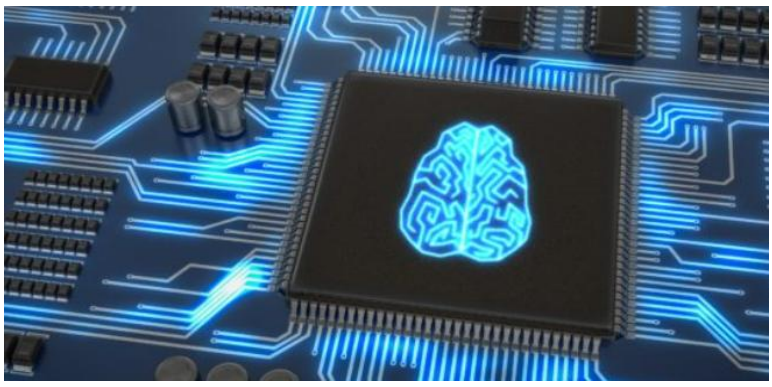


Figura 27: Intel·ligència artificial en un xip, font: Pinterest

[60] IAM: acrònim anglès per "Identity and Access Management" , que es la gestió d'identitat i accés.
[61] Blockchain: terme anglès per Cadena de Blocks, no sols s'empra en mineria de dades sinó que moltes noves aplicacions n'estan fent ús.

12. Conclusions de IA a la Ciberseguretat

En el decurs d'aquest treball, hem parlat dels orígens, fonaments i bases de les IA, així com de la seva evolució, expansió i millora, fins a l'actualitat. Hem fet notar l'aparició de tres punts clau en la història que han permès el creixement i expansió de les IA: la definició d'agents intel·ligents, l'aparició del NLP (Natural Language Processing) i la disponibilitat de llibreries de codi obert (PyTorch i TensorFlow). Entrant al detall, hem vist alguns dels algorismes emprats, les classificacions de les IA, els beneficis que comporta una IA amb ML (Machine Learning) i DL (Deep Learning), i la importància de fer i tenir un bon aprenentatge. Cal evitar que aquest aprenentatge sigui erroni, amb dades esbiaixades o que pugui ser manipulat per Ciberdelinqüents, ja que malmetria la IA. Així, també hem comentat el marc de treball definit per **ALTAI** (The Assessment List on Trustworthy Artificial Intelligence), per tal d'obtenir una "IA fiable". Aquí ens hauria agradat aprofundir en aquest marc a nivell Europeu, però ens desviaríem del tema, i això ja queda pels casos de reglamentació al implementar una "IA fiable" segons normativa de la EU.

També hem vist que el NLP (Natural Language Processing) junt a les xarxes neuronals, ha donat pas als sistemes cognitius que es fan servir actualment, i que treballen amb "patrons" i cert raciocini, així la IA s'aproxima a la forma de pensar humana. D'aquesta forma, també s'ha vist que es fan servir diferents algorismes en una sola IA, alguns son per unes tasques fixes o comunes i altres son adaptatius, en funció de l'aprenentatge donat. De totes formes, això no s'atura a aquí, hem mostrat que cada branca de la ciència sobre les IA, continua avançant pel seu compte, i alguns avenços ofereixen noves i millors solucions, que properament, es poden anar aplicant a les IA actuals.

Per altra banda, també s'ha mostrat el panorama actual de la Ciberseguretat en moltes organitzacions i empreses, i en detall s'ha mostrat les mancances i reptes als que s'enfronta. S'observa principalment falta de recursos humans, de capacitats i de coneixements, de temps per atendre totes les alertes, alarmes o notificacions que es generen, i els errors humans. A més, actualment també es dona un increment de vulnerabilitats i amenaces als sistemes, i a la xarxa, que any rere any, va creixent cada cop més. Així, s'ha fet un repàs de les eines tradicionals que s'empren per la Ciberseguretat, de l'evolució d'aquestes eines i dels nous programaris i serveis que s'ofereixen per la Ciberseguretat. Hem pogut apreciar que moltes de les noves eines, a part d'oferir automatismes per mitigar amenaces, també porten certa intel·ligència incorporada o sinó, com a mínim empren la heurística. Algunes d'aquestes eines, fins hi tot, donat que ja no cal aprendre el que ja es sap o es coneix, afegeixen Sistemes Experts i

també la intel·ligència cognitiva, en canvi altres, es basen en evolucions del SIEM o aplicacions de la IA sobre ell, i també automatismes per mitigar els incidents, atacs i les amenaces trobades en base a deteccions.

Arribats a aquí, ja som conscients del increment d'amenaces i vulnerabilitats, del increment de la superfície que s'exposa en les xarxes actuals, i de les mancances i reptes que ha d'afrontar la Ciberseguretat. I per altra banda, ja sabem el que es una IA, del que pot fer per ajudar a la Ciberseguretat, i també, que es poden combinar diferents tipus de IA com Sistemes Experts i Sistemes Cognitius o altres tipus sota una sola denominació de IA. D'aquesta forma, s'ha vist que es poden tenir unes IA pre-entrenades, tenir una cobertura per IoT i OT (tots els "endpoints"), i protecció en el núvol i el treball remot. També, que tot això ens permet una reducció important dels errors humans i de les feines repetitives i manuals a la Ciberseguretat. I que l'anterior, pot propiciar l'aparició d'uns nous rols i l'eliminació d'altres, portant a terme una reestructuració del mapa de la Ciberseguretat. A més, s'ha vist que es pot emprar la GPU i les DPU per accelerar algorismes i tasques de la IA, si bé, això sols ho remarca un sol proveïdor. De fet, és una idea interessant, ja que cal tenir en compte que les GPU's en els servidors i centres de dades estan molt desaprofitades, i emprar-les pels càlculs i processos de la IA, representa un bon aprofitament dels recursos disponibles. Potser en un futur, totes les IA podran aprofitar com a mínim, la GPU per a les seves operacions, processos i tasques.

Fins aquí, tot el que hem vist sembla beneficis, però cal pensar que les IA també poden ser vulnerables a Ciberatacs, i que cal planificar molt bé el seu desplegament i el seu aprenentatge, per que aquest no sigui esbiaixat. A més, cal pensar també en la seva protecció des de totes les vessants, en el model d'aprenentatge, en la seva realimentació de dades per aconseguir una "IA fiable". Per tant, cal revisar la disponibilitat i ús de les dades, ja que encara que la IA pot estar pre-entrenada, a de continuar aprenent de la nostra xarxa i prendre decisions en base a aquesta i els seus "patrons". I també, cal valorar acuradament el cost i temps que s'ha d'invertir en la seva implementació. Ja s'ha comentat la mancança d'experts en entorns IA i ML, pel que també s'ha recomanat de tenir un assessorament expert extern sobre IA, i si es possible, alguns exemples d'implementacions amb èxit, que es puguin consultar, ja que això, ajuda molt en aquestes situacions.

Com a comentari sobre la documentació pel treball, ens ha constat una mica obtenir informació amb facilitat sobre alguna part de les IA, ja que la majoria de llibres, publicacions, i altres formats més recents, es troben en idioma anglès, i molts son força tècnics, pel que amb certes expressions hem tingut que cercar el seu significat real. Per altra banda, en quant a la Ciberseguretat si que hem

trobat més publicacions i més assequibles, si bé, moltes ja tenen uns quants anys i no eren aprofitables, altres si que han estat punts de referència per elaborar aquest treball. Al final, de fet, l'esforç en idioma anglès, ens ha permès ampliar els coneixements sobre les IA, conèixer els acrònims, termes i definicions emprats en aquest idioma, i la seva correspondència al català.

En resum, sobre el tema d'aquest treball i valorant tot el que s'ha exposat fins a aquest punt, arribem a la conclusió final de que l'aplicació de la IA a la Ciberseguretat, amb tots els seus pros i els seus contres, es la millor solució per mitigar les amenaces, mancances i reptes que es presenten actualment i els que sortiran en un futur pròxim, a la Ciberseguretat. Per tant, a nivell personal, creiem que properament anirà apareixent més varietat i tipus de IA enfocades a la Ciberseguretat. Aquests noves IA, aniran afegint totes les millores o beneficis que apareixen en les diferents branques d'estudi de les IA, cobriran més i millor l'expansió de superfície de xarxa exposada (IoT, OT, ...), i aprofitaran millor els recursos del maquinari (GPU i DPU). A més, pensem que s'integrarà millor amb altres eines de tercers per Ciberseguretat, seran millors en quan a resposta i automatismes. I podran oferir actualitzacions, si més no, per als Sistemes Experts que integrin (com ho fa un antivirus). A part, veiem una major facilitat d'ús i implementació, més accés a les IA pre-entrenades, i un increment del nombre de formats de dades estructurades i no estructurades emprats en l'aprenentatge pels algorismes adaptatius. Així, la IA segons el nostre parer, serà o esdevindrà "el nou gestor" a la Ciberseguretat.

Figura 28: Imatge IA i Humà, font: https://www.istockphoto.com/

13. Glossari

aC: acrònim de Abans de Crist.

AI (Artificial Intelligence): acrònim anglès de Intel·ligència Artificial (IA).

Algorismes: (o algoritmes) conjunt finit d'instruccions o passos que serveixen per executar una tasca o resoldre un problema.

Antimalware: terme anglès per a un programa dissenyat per prevenir, detectar i remeiar software perjudicial o maliciós, dins de dispositius informàtics.

AP: acrònim angles de "Access Point", punt d'accés que permet la connexió Wi-Fi.

Bakgammon: joc de taula per 2 persones que uneix atzar amb forts coneixements estratègics.

Big Data: terme que descriu un gran volum de dades, estructurades i no estructurades, que permet el seu anàlisi massiu via mètodes no tradicionals.

Bitàcoles: designa un registre escrit d'accions i esdeveniments que surten al dur a terme certa feina.

Blockchain: terme anglès per Cadena de Blocks, no sols s'empra en mineria de dades sinó que moltes noves aplicacions n'estan fent ús.

Bot: és un programa que fa tasques repetitives, predefinides i automatitzades, son de molts tipus.

CART: tipus d'algorisme basat en d'arbres de classificació i regressió.

Chatbot: assistent que es comunica amb els usuaris per missatges de text.

Ciberseguretat: protecció d'actius d'informació mitjançant el tractament de les amenaces existents per a la informació que és processada, emmagatzemada i transportada per sistemes d'informació que es troben connectats amb Internet.

Ciberdelinqüent: persona que comet un acte il·legal i sovint fraudulent per mitjà de les tecnologies de la informació i la comunicació, generalment Internet.

Cloud: anomenat "cloud computing", servei de fitxers o recursos ofert via Internet.

Dark Web: terme anglès per a la Web Fosca, està per sota la "Deep Web" o Web profunda o Web oculta.

Deepfakes: terme anglès per a la edició i falsificació de vídeos amb persones aparentment reals amb l'objectiu d'enganyar al destinatari.

DGA: acrònim anglès de “Domain Generation Algorithms”, que és Algorismes de Generació de Dominis, s’empra per generar noms de domini pseudoaleatoris com orígens d’atacs.

DL (Deep Learning): acrònim anglès de Aprenentatge Profund.

DPU: acrònim anglès de “Data Processing Unit”, son les unitats de processament de dades, una nova classe de processador programable, especialitzat en moure dades en els centres de dades.

Drivers: o controladors, és un component de software que permet al Sistema Operatiu i un dispositiu comunicar-se entre ells.

ELO: sistema de puntuació de mètode matemàtic, basat en càlcul estadístic sobre la habilitat relativa a jugadors d’esports com els escacs.

EDR (Endpoint Detection Response): acrònim anglès sobre Detecció i Resposta de Punt final

Endpoint: és qualsevol dispositiu remot que sigui físicament la part final d’una xarxa

Exploit: terme anglès, per definir un programa informàtic, una part de software o Script que s’aprofita d’un error o vulnerabilitat, per provocar un comportament no previst en un dispositiu

Firmware: és un programà bàsic que es carrega al iniciar un dispositiu i que permet controlar els circuits electrònics d’aquest.

Frame: estructura de dades que conté una descripció general d'un objecte, que es deriva de conceptes bàsics i de l'experiència.

Framework: esquema o marc de treball que ofereix una estructura base per elaborar un projecte amb objectius específics, seria com una plantilla o punt de sortida per a un desenvolupament.

Gartner: empresa consultora i de investigació de les tecnologies de la informació, que fa una sèrie de publicacions molt importants i seguides fortament pel mon de les TIC.

GDPR: acrònim anglès per “General Data Protection Regulation”, que és la Regulació General de Protecció de Dades (RGPD).

GPU: acrònim anglès per “Graphics Processing Unit”, que seria la unitat de procés gràfic del sistema, és sol explotar molt en la mineria de Bitcoin, es fa per aprofitar recursos del maquinari.

Honeypot: terme anglès per “Pot de mel”, sol ser un sistema preparat especial, amb vulnerabilitats que atreuen als atacants en una trampa, per tal de recollir informació sobre l’atacant.

Honeytoken: terme anglès per "Testimoni de mel ", es tracta d'una informació aparentment valuosa, que es troba amagada, però a la que es pot arribar amb una mica de "hacking" i que et porta a un lloc monitorat.

IAM: acrònim anglès per "Identity and Access Management" , que es la gestió d'identitat i accés.

ICS: acrònim anglès per "Industrial Control System", que vol dir Sistema de Control Industrial.

Instàncies sintètiques de dades: es fa servir abastament en la generació d'imatges, vídeo i veu.

IOC: acrònim anglès per "Indicator of Compromise", que vol dir Indicador de Compromís.

IoT: acrònim en anglès de "Internet of Things", Internet de les coses, dispositius remots amb certa computació, com sensors, mesures de nivells, temperatura o altra informació.

IDS/IPS: acrònim anglès per "Intruder Detection System / Intruder Prevention System", que en català seria: sistema de detecció d'intrusos i sistema de prevenció d'intrusos, respectivament.

Malware: terme anglès per software maliciós, programa o codi maliciós, perjudicial pels sistemes.

MDP: acrònim anglès per "Markov Decision Process", és un procés estocàstic de control en temps discret, que modela la presa de decisions en situacions parcialment aleatòries i parcialment fixades.

MDR (Managed Detection and Response): acrònim anglès pel Servei de Detecció i Resposta Gestionat.

MITRE ATT&CK: acrònim anglès de "Adversarial Tactics, Techniques, and Common Knowledge" és una guia per classificar i descriure els ciberatacs i les intrusions.

ML (Machine Learning): acrònim anglès de Aprenentatge Automàtic.

ModelOps: terme anglès, segons Gartner: model que es centra principalment en la governança i la gestió del cicle de vida d'una àmplia gama de model de IA i decisions operacionals (ML, DL, agents, ...).

MSSP (Managed Security Services Provider): acrònim anglès per Proveïdor de Serveis de Seguretat Gestionada.

MTTR (Mean Time To Repair): acrònim anglès pel Temps Mig de Reparació, que també s'empra com temps mig de resposta.

NLP (Natural Language Processing): acrònim anglès de Processament del Llenguatge Natural (PLN).

O.S.: acrònim anglès per "Operating System", és el Sistema Operatiu.

OT: acrònim angles per "Operational Technology", és <tecnologia operativa> emprat en xarxes industrials.

Partners: terme anglès per socis de negoci, també s'empra per als serveis externs contractats.

Perceptró: neurona artificial o unitat bàsica de inferència en forma de discriminador lineal, a partir de la que es desenvolupa un algorisme capaç de generar un criteri de selecció.

Phishing: terme anglès per pescar, consisteix en enganyar les persones per a que comparteixin informació confidencial com contrasenyes, números i dades de la targeta de crèdit, etc.

RaaS: en anglès, Ransomware as a Service, model de negoci de part de guanys, de desenvolupadors de malware pel que ofereixen serveis i eines per a una campanya de ransomware contra una victima.

Ransomware: terme anglès per malware de rescat, és un tipus de malware que impedeix accedir al sistema o fitxers, en que es sol demanar el pagament d'un rescat, per fer-ho i sense garanties de res.

Resiliència: capacitat de recuperar-se d'una falla i conservar la confiança del servei, ha de garantir la protecció de les operacions de forma que una amenaça o incompliment no afecti a la resta de negoci.

SaaS: acrònim angles per "Software as a Service" que vol dir solucions software com un servei via Internet o xarxa.

Sandbox: terme en anglès que literalment vol dir caixa de sorra, però fa referència a un entorn segur i aïllat on es pot testar una aplicació o programa.

Scripts: guió o conjunt d'instruccions informàtiques.

Shells: consola de comandes.

SIEM: acrònim angles de "Security Information and Event System", el SEM detecta patrons fora del normal en temps real, i el SIM centralitza els registres de seguretat per interpretar-los i guardar-los.

Sistemes experts: emulen el raonament tal i com ho faria un expert en un àrea de coneixement.

Smishing: terme anglès, per una tècnica d'enviament d'un SMS simulant ser legítim, amb l'objectiu de robar informació confidencial.

SLA: acrònim anglès per “Service Level Agreement”, que és l’Acord de Nivell de Servei contractat.

SOC: acrònim anglès de “Security Operations Center”, Centre d’operacions de Seguretat.

Spam: terme anglès per a definir el correu no desitjat, amb finalitats publicitàries o comercials.

Startup: o empresa emergent, de nova creació que comercialitza productes i/o serveis que fa us intensiu de les TIC.

TFM: simplificació o acrònim de Treball Final de Màster.

Threat hunting: terme anglès, que significa caça d’amenaces, és un procés de cerca iterativa i proactiva a traves de les xarxes per detectar i aïllar amenaces avançades que son capaces d’evadir les solucions de seguretat existents.

TI: acrònim de Tecnologies de la Informació.

UEBA: acrònim anglès de “User and Entity Behavior Analiytics” posat per Gartner, es tradueix per anàlisi del comportament de les persones connectades a la xarxa d’una organització o empresa.

Vlan: terme anglès per “Virtual Lan”, xarxa de àrea local virtual que permet crear xarxes lògiques virtuals dins una mateixa xarxa física.

VPN: acrònim anglès de “Virtual Private Network”, xarxa privada virtual via un túnel de xifrat.

WAF: en anglès “Web Advanced Firewall”, tipus de Firewall especialitzat en el protocol Web i la seva funcionalitat (SSL/TLS, peticions, formularis, Scripts, ...).

YARA: acrònim anglès per “Yet Another Ridiculous Acronym”, la traducció no importa, però és una eina que empra regles que permeten detectar “malware”, basat en firmes com els antivirus tradicionals.

14. Bibliografia

1. **Llibre:** Ted Coombs. “Artificial Intelligence & Cybersecurity For Dummies”, IBM Limited Edition. 2018 by John Wiley & Sons, Inc.
2. **Llibre:** Judith Hurwitz i Daniel Kirsh. “Machine Learning For Dummies”, IBM Limited Edition. 2018 by John Wiley & Sons, Inc.
3. **Llibre:** Lawrence C. Miller. “Cybersecurity Automation For Dummies”, Juniper Networks Edition. 2019 by John Wiley & Sons, Inc.
4. **Llibre:** Joseph Steinberg. “Cybersecurity For Dummies”, A Wiley Brand. 2020 by John Wiley & Sons, Inc.
5. **Llibre:** Lawrence C. Miller, CISSP. “Cybersecurity For Dummies”, Palo Alto Networks Edition. 2014 by John Wiley & Sons, Inc.
6. **Llibre:** Rudolph Russell. “Machine Learning: Guia Paso a Paso Para Implementar Algoritmos De Machine Learning Con Python”. 2018 Rudolph Russell.
7. **EBook:** Stack Overflow contributors. “Aprendizaje machine-learning”. 2020 https://riptutorial.com/es/home
8. **Publicació:** Julio Villena, Raquel M. Crespo Garcia, i José Jesús Garcia Rueda. “Historia de la Inteligencia Artificial” Inteligencia en Redes de Comunicaciones. 2012 Universidad Carlos III de Madrid.
9. **Publicació:** Aldo Valdez Alvarado. “Machine Learning para Todos”. Puno – 2019.
10. **Publicació:** AED (Asociación Española de Directivos). “Machine Learning, Inteligencia Artificial y Big Data: Lo que todo directivo debe saber”. 2019
11. **Publicació:** Management Solutions 2018. “Machine Learning, una pieza clave en la transformación de los modelos de negocio”. 2018 Area de I+D
12. **Publicació:** Management Solutions, i Universidad Politécnica de Madrid. “Algoritmos de Machine Learning”. 2021 Cátedra iDANAE.
13. **Publicació:** Micah Muser i Ashton Garriott. “Machine Learning and Cybersecurity, Hype and Reality”. Center for Security and Emerging Technology (CSET). June 2021
14. **Publicació:** DataRobot (datarobot.com). “Driving Efficiency in Cyberspace Through AI”. 2021 DataRobot, Inc.
15. **Publicació:** Lorenzo Pupillo, Stefano Fantin, Afonso Ferreira i Carolina Polito. “ Artificial Intelligence and Cybersecurity, Technology, Governance

and Policy Challenges". Final Report of CEPS Task Force. Centre for European Policy Studies (CEPS). Brussels, May 2021.

16. **Publicació:** Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial. "Directrices Éticas para una IA Fiable". Creado por la Comisión Europea en Junio de 2018. Publicado el X de abril de 2018.
17. **Publicació:** Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial. "Una Definición de la Inteligencia Artificial: Principales Capacidades y Disciplinas Científicas". Creado por la Comisión Europea en Junio de 2018. Publicado el X de abril de 2018.
18. **Publicació:** Katanosh Morovat i Brajenda Panda. "A Survey of Artificial Intelligence in Cybersecurity". 2020 International Conference on Computacional Science and Computacional Intelligence (CSCI). 2020 IEEE.
19. **Publicació:** Kim Andreasson. "Artificial Intelligence & Cybersecurity: Balancing Innovation. Execution and Risk". Team of EIU (Economist Intelligence Unit) Pillsbury Winthrop 202.
20. **Publicació:** SAS and all other SAS Institute Inc. "The AI Bussiness Case Guide". 2021 SAS Institute Inc.
21. **Publicació:** Kirti Raj Bhatele, Harsh Shrivastava, i Neha Kumari. "The Role of Artificial Intelligence in Cyber Security". 2019, IGI Global.
22. **Publicació:** Juan J. Padial, Universidad de Málaga. "Técnicas de programación 'Deep Learning': ¿Simulacro o realización artificial de la inteligencia?". Naturaleza y Libertad: Numero 12, 2019
23. **Publicació:** Fernando Berzal, Universidad de Jaén, CEATIC, V Jornadas Doctorales del programa TIC."Deep Learning".2019 DECSAI, Departamento de Ciencias de la Computación e I.A.. Universidad de Granada.
24. **Publicació:** Erik Leonel Otero, Andres Ezequiel Figueroa, Tom Fiszon, Guido Segesso, i Yago Graña De Brasi. "Deep Learning, la tecnologia del mañana". 2018 Universidad Tecnológica Nacional, Facultad Regional Buenos Aires.
25. **Tesis:** Christoffer Sjöblom. "Artificial Intelligence in Cybersecurity and Network Security". Computer Engineering, Faculty of Science and Engineering (FNT) Äbo Akademi University. Spring 2021.
26. **Article:** Andrés Abeliuk i Claudio Gutiérrez. "Historia y evolución de la inteligencia artificial". 2021 Inteligencia Artificial (Chile).
27. **Article:** Nisha Rawindaran, Ambikesh Jayal i Edmond Prakash. "Machine Learning Cybersecurity Adoption in Small and Medium Enterprises in Develop Countries". Computers 2021 by MDPI.

28. **Article:** Lilian Judith Sandoval. “Algoritmos de aprendizaje automático para anàlisis y predicción de datos”. ITCA-FEPADE 2018.
29. **Article:** Feng Tao, Muhammad Shoaib Akhtar i Zhang Jiayuan. “The future of Artificial Intelligence in Cibersecurity: A Comprehensive Survey”. EAI Endorsed Transactions on Creative Technologies, July 2021.
30. **Article:** Akash Hebbar i Dr S Anupama Kumar. “Artificial Intelligence in Cyber Security”. 2021 JETIR, May 2021, Volume 8, Issue 5.
31. **Web:** https://www.cesce.es/es/w/asesores-de-pymes/breve-historia-la-inteligencia-artificial-camino-hacia-la-empresa Març 2022
32. **Web:** https://www.nationalgeographic.com.es/ciencia/breve-historia-visual-inteligencia-artificial_14419 Març 2022
33. **Web**: https://www.elternativa.com/blog-elternativa/historia-inteligencia-artificial Març 2022
34. **Web**: http://www.cs.us.es/~fsancho/?e=221 Març 2022
35. **Web**: https://ca.wikipedia.org/wiki/Intel%C2%B7lig%C3%A8ncia_artificial Març 2022
36. **Web**: https://es.wikipedia.org/wiki/Inteligencia_artificial Març 2022
37. **Web**: https://www.zendesk.com.mx/blog/historia-inteligencia-artificial/ Març 2022
38. **Web**: https://www.crehana.com/es/blog/desarrollo-web/historia-de-la-inteligencia-artificial/ Març 2022
39. **Web**: https://www.cice.es/blog/articulos/historia-evolucion-la-inteligencia-artificial/ Març 2022
40. **Web**: https://www.oracle.com/mx/artificial-intelligence/what-is-ai/ Març 2022
41. **Web**: https://nexusintegra.io/es/ventajas-y-desventajas-de-la-inteligencia-artificial/ Març 2022
42. **Web**: https://www.sas.com/es_es/insights/analytics/what-is-artificial-intelligence.html Març 2022
43. **Web**: https://www.ibm.com/co-es/analytics/journey-to-ai Març 2022
44. **Web**: https://platzi.com/blog/algoritmos-para-el-machine-learning/?utm_source=google&utm_medium=paid&utm_campaign=14603491644&utm_adgro%E2%80%A6 Març 2022
45. **Web**: https://enorcerna.com/tecnologia/algoritmo-de-inteligencia-artificial-todo-lo-que-necesita-saber-al-respecto Març 2022

46. **Web:** https://www.futurespace.es/machine-learning-los-origenes-y-la-evolucion/ Març 2022
47. **Web:** https://www.oracle.com/es/data-science/machine-learning/what-is-machine-learning/ Març 2022
48. **Web:** https://www.ibm.com/co-es/analytics/machine-learning Març 2022
49. **Web:** https://www.sas.com/es_es/insights/analytics/machine-learning.html Març 2022
50. **Web**: https://iaarbook.github.io/ Març 2022
51. **Web:** https://datos.gob.es/ca/blog/como-aprenden-las-maquinas-machine-learning-y-sus-diferentes-tipos Març 2022
52. **Web:** https://medium.com/datos-y-ciencia/introduccion-al-machine-learning-una-gu%EDa-desde-cero-b696a2ead359 Març 2022
53. **Web:** https://latam.kaspersky.com/resource-center/definitions/ai-cybersecurity Abril 2022
54. **Web:** https://blog.enzymeadvisinggroup.com/inteligencia-artificial-nuevo-hito-ciberseguridad Abril 2022
55. **Web:** https://www.realinstitutoelcano.org/analisis/la-ciberseguridad-y-su-relacion-con-la-inteligencia-artificial/ Abril 2022
56. **Web:** https://blog.conzultek.com/ciberseguridad/inteligencia-artificial-en-la-ciberseguridad Abril 2022
57. **Web:** https://digital.la.synnex.com/por-que-la-inteligencia-artificial-es-esencial-para-la-ciberseguridad-moderna Abril 2022
58. **Web:** https://www.uv.mx/infosegura/general/noti_traficored/ Abril 2022
59. **Web:** https://www.ibm.com/security/artificial-intelligence Abril 2022
60. **Web:** https://www.sailpoint.com/identity-library/how-ai-and-machine-learning-are-improving-cybersecurity/?elqct=PaidMedia&elqchannel=GoogleSearch&elqcta=EAIaIQobChMI6f7PiYSg9wIVBBEGAB2dTAYjEAAYBCAAEg%E2%80%A6 Abril 2022
61. **Web:** https://blog.smartekh.com/que-es-soar-y-que-beneficios-tiene-para-tu-organizacion Abril 2022
62. **Web:** https://www.redhat.com/es/topics/security/what-is-soar Abril 2022
63. **Web:** https://www.viewnext.com/siem-vs-soar/ Abril 2022

64. **Web:** https://www.a2secure.com/blog/plataformas-soar-revitalizacion-de-las-plataformas-siem/ Abril 2022
65. **Web:** https://protecciondatos-lopd.com/empresas/que-es-edr/ Abril 2022
66. **Web:** https://www.arsys.es/blog/edr-seguridad Abril 2022
67. **Web:** https://www.checkpoint.com/es/cyber-hub/what-is-endpoint-detection-and-response/ Abril 2022
68. **Web:** https://www.incibe.es/protege-tu-empresa/blog/sistemas-edr-son-y-ayudan-proteger-seguridad-tu-empresa Abril 2022
69. **Web:** https://www.darktrace.com/es/plataforma-de-cyber-ai/ Maig 2022
70. **Web:** https://www.ibm.com/es-es/products/cognitive-security-analytics/details Maig 2022
71. **Web:** https://www.ibm.com/es-es/qradar/security-qradar-soar Maig 2022
72. **Web:** https://www.ibm.com/es-es/products/watson-assistant/enterprise-security Maig 2022
73. **Web:** https://www.ibm.com/cloud/watson-studio Maig 2022
74. **Web:** https://developer.nvidia.com/morpheus-cybersecurity Maig 2022
75. **Web**: https://www.datamation.com/security/artificial-intelligence-ai-in-cybersecurity-trends/ Maig 2022

15. Annexos

15.1. **Annex A:** diferències Big Data, IA, ML i DL