Universitat Oberta de Catalunya

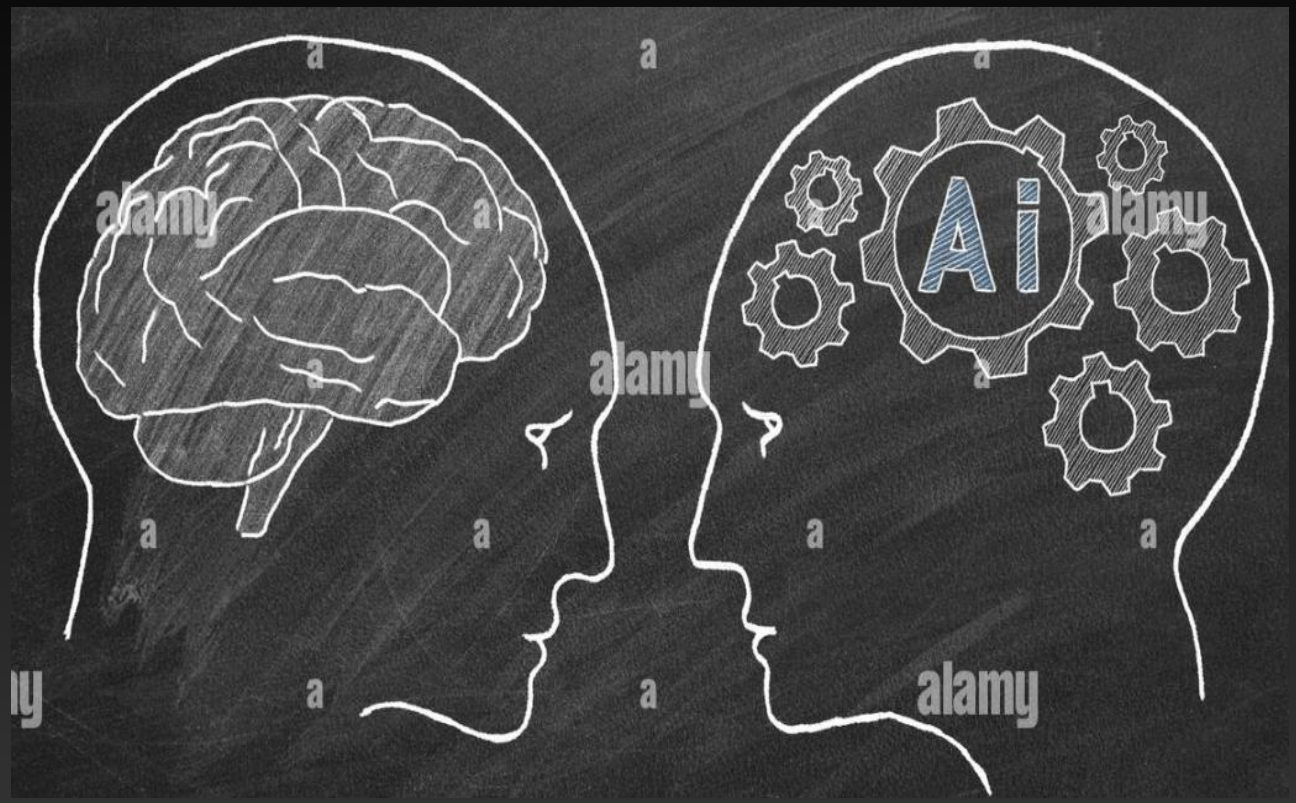
# La IA, nou gestor de la Ciberseguretat


TFM – Seguretat en sistemes operatius
Jordi Fenés Castell 06 / 2022

Màster interuniversitari de Seguretat de les TIC – MISTIC
Professors : Erik de Luis Gargallo i Jordi Serra Ruiz

# Guió de continguts

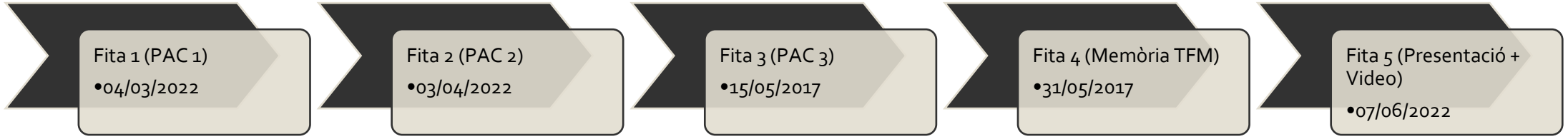
- Motivació, objectius i fites del treball
- Orígens de la IA
- Situació actual
- Bases de la IA
- Aprenentatge de la IA
- Classificació de les IA
- Punts forts de les IA
- Situació i reptes de la Ciberseguretat
- IA i Ciberseguretat
- Exemples de IA per Ciberseguretat
- Futur de la IA a la Ciberseguretat
- Conclusions
- Cloenda



# Motivació, objectius i fites del treball

Aquest treball intenta aportar unes bases mínimes de coneixement sobre les IA, explica la situació actual, bases, classificació i aprenentatge de les IA.  Mostra en que és bona, els punts forts de la IA, i que això es pot aprofitar per la gestió de la Ciberserguretat. Així, també es veu l'evolució d'eines de Ciberseguretat cap a automatismes, heurística, i alguna IA restringida. I també es revisen els reptes, les mancances, i les amenaces a les que s'exposa la Ciberseguretat actualment, i es veu com la IA pot ajudar a millorar certes situacions, augmentar la resiliència i corregir molts d'aquests problemes. S'afegeix alguns exemples de IA específiques per la Ciberseguretat, es visiona algunes prediccions i també futures possibles millores de les IA per a la gestió de la Ciberseguretat.

Aquest Treball, s'ha dividit en cinc fases, des de la data d'inici al 21 de Febrer de 2022, fins a la data final d'entrega el 7 de juny de 2022, quedant a part la defensa del TFM, de entre un dels cinc dies possibles del 13/06/2022  a 17/06/2022.

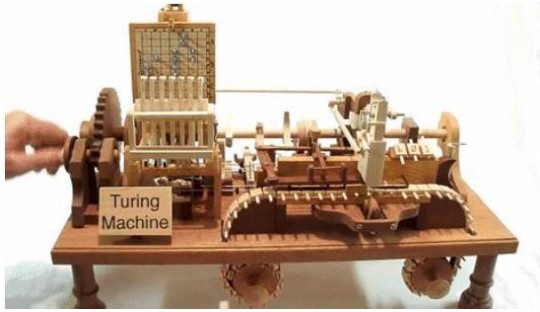


# Orígens de la IA (I)

En general es sol posar el naixement i fonaments de base de la IA a partir dels treballs de Alan Turing al 1950. Els més puristes diuen que és al 1956 a una conferència feta a la Universitat de Dartmouth. De totes formes, hi ha precedents en la història humana sobre emular el raonament humà de forma artificial, veiem quins son els antecedents:

- Mites, llegendes i textos (manuscrits, llibres, obres de teatre, ...)
  - De l'antiga Grècia tenim el mite de Galatea, escultura vivent de Pigmalion, els Trípodes (servents en festes i banquets) i les ajudantes de metall de Hefesto, fets tos a la seva forja.
  - A la edat mitja: el Golem de fang que ajuda el poble jueu, com crear vida artificial en la obra del àrab Jàbir Ibn Hayyan (pare de la química), i Paracelso que explica com fer homuncles.
  - Al segle XIX, la novela "Frankestein", al segle XX la obra teatral "R.U.R" introdueix el terme "robots", i les publicacions de Isaac Asimov precursor de les 3 lleis de la robòtica.

# Orígens de la IA (II)




- Bases mes serioses o de tipus científic (a part de jocs com les "Torres de Hanoi")
  - A l'antiga Grècia tenim el "sil·logisme" de Aristòtil i una màquina autocontrolada de Ctesibi.
  - Als segles VIII-IX, els "algorismes" de Al Juarismi per explicar els passos del llibre "Al jabr"
  - Bases de treball per la lògica:
    - 1315 Ramon Llull "Ars Magna" veracitat o falsedat d'afirmacions lògiques
    - 1703 Leibniz basat en altres treballs, publica "Explication de l'Aritmétique Binarie"
    - 1847 George Boole estableix la lògica proposicional
    - 1879 Gottlob Frege expandeix la lògica booleana amb la "Lògica de Primer Ordre"
  - 1842 Ada Lovelace crea el primer algorisme que s'intenta aplicar a una màquina
  - 1937-1940 Alan Turing publica articles, estableix la "Maquina de Turing" i junt al seu equip construeix el primer ordinador electromecànic.
  - 1900-1951 surten invents electrònics, com: el triode de Lee de Forest, primera computadora programable Z3 de Konrad Zuse, i es crea el transistor d'unió per William Shockley.
  - 1956 a la Universitat de Dartmouth es consolida el terme Intel·ligència Artificial (IA).

# Orígens de la IA (III)

- Evolució fins avui des de la consolidació del terme Intel·ligència Artificial (IA)
  - De 1956-1960, "Logic Theorist" primer programa informàtic de IA, el MIT crea el llenguatge LISP, apareix el "Perceptró" com primera unitat neuronal artificial, i apareix "Sad Sam" que interpreta oracions en anglès.
  - De 1961-1970, robot "unimate" a cadena de muntatge, apareix SIR que aprèn de informació, es crea "STUDENT" amb LISP que resol problemes de secundària, apareix "ELIZA" el primer Chatbot en anglès, i també apareixen els primers "Sistemes Experts".
  - De 1973-1980, es crea el llenguatge PROLOG, el programa BKG 9.8 guanya al Bakgammon al campió mundial, i apareixen mes "Sistemes Experts" i creix el seu ús.
  - De 1981-1988, auge de "Sistemes Experts" amb la "quinta generació de computadors", es publica treballs de xarxes neuronals i arquitectura de d'aprenentatge supervisada, es descriu un "agent intel·ligent" per IA, i apareixen llenguatges orientats a objectes (POO).
  - De 1997-2010, "Deep Blue" de IBM guanya als escacs a Garri Kaspàrov, apareixen diferents robots autònoms, i es crea una base de dades d'imatges per entrenar xarxes neuronals.
  - De 2011-2020, "Watson" de IBM guanya el concurs televisiu "Jeopardy!" (NLP), Google i Microsoft llencen asistents virtuals, les xarxes neuronals convolucionals creixen, apareixen llibreries de codi obert (TensorFlow i PyTorch) per IA, i la IA s'introdueix a molts sectors.

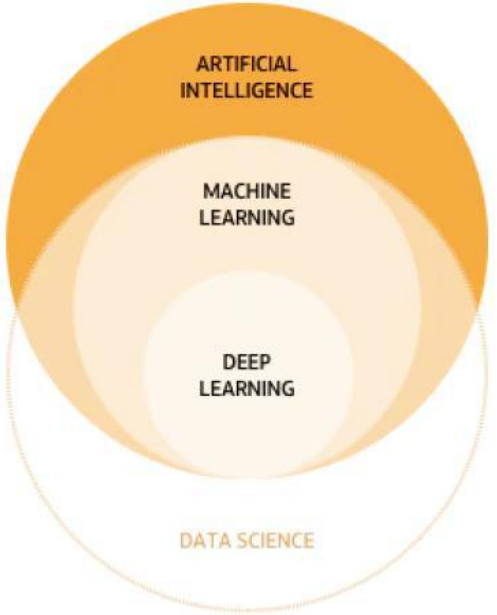
# Situació Actual (I)

L'actual expansió i ús de les IA en molts sectors i àrees professionals , es basa històricament, en tres esdeveniments en el decurs de l'evolució de les IA:

- 1987, descripció d'un agent intel·ligent, crea moltes i variades branques d'investigació
- 2011, aparició del NLP (Natural Language Processing)
- 2015, lliberies de codi obert (TensorFlow i Pytorch)

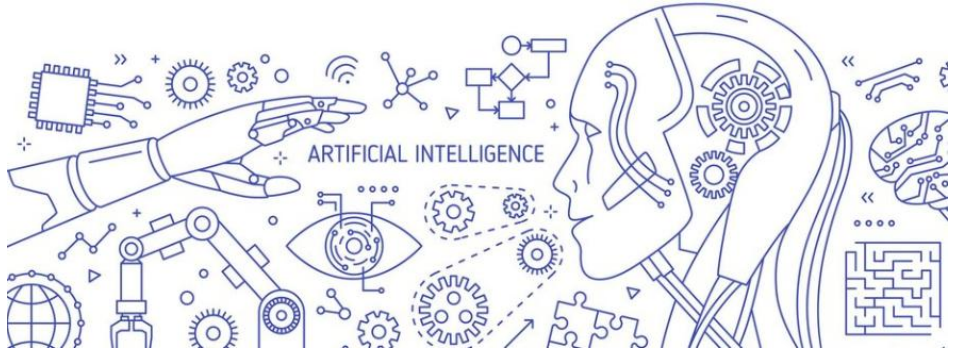
Cal afegir-hi principalment, la millora, el afinament i ús de l'aprenentatge ML (Machine Learning) i DL (Deep Learning), per sobre dels algorismes clàssics emprats inicialment en IA. A més, altres factors importants del seu entorn que impulsen les IA, son:

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



# Situació Actual (II)

Segons "Gartner", la analítica i la intel·ligència empresarial, son les tecnologies de diferenciació més importants d'una organització. I una publicació del "Harvard Business Review", diu que les empreses fan servir la IA principalment, per:



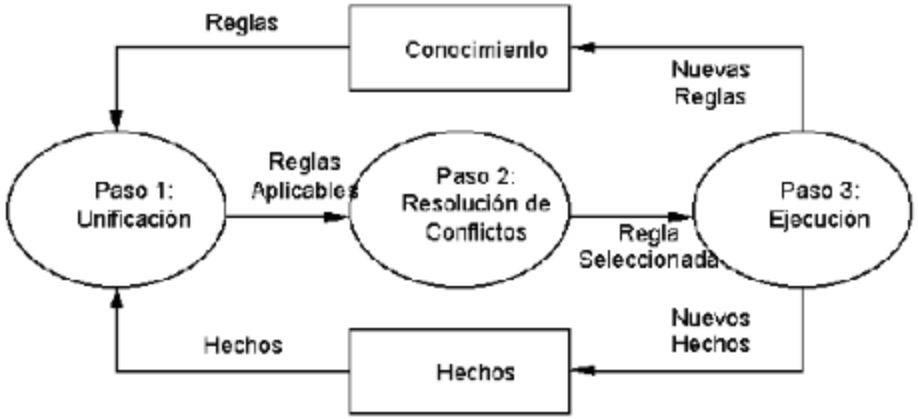

- Detectar i dissuadir intrusions de seguretat (44%).
- Resoldre problemes tecnològics dels usuaris (41%).
- Reduir el treball de la gestió de producció (34%).
- Medir l'acompliment intern en l'ús de proveïdors aprovats (34%).

En quant a l'aplicació de la IA a Catalunya, podem destacar dos programes o projectes:

- Catalonia.AI: estratègia d'intel·ligència artificial de Catalunya, que inclou un programa d'actuacions multisectorials, amb els suport de la Generalitat.
- AI4EU: projecte de la "Artificial Intelligence for Europe" a Catalunya, per a un ecosistema de IA d'àmbit Europeu, que proporciona eines, recursos i coneixements.

*També hi ha diferents centres que investiguen la IA: Barcelona Supercomputing Centre, Grup d'IA d'alt Rendiment, Institut d'investigació en IA, Centre de Visió per Computador, ... I fins a 20

# Bases de la IA (I)




En la investigació de la IA, es persegueix dos objectius centrals:

- Estudi de processos cognoscitius en general.
- Obtenir sistemes automàtics que facin feines reservades a humans.

Com mecanismes tècnics bàsics, per obtenir representació simbòlica del coneixement, tenim:

- La inferència simbòlica que inclou la deducció.
- L'heurística que escull el millor resultat possible per un objectiu.

Al tractar la informació, pels esquemes moderns de representació del coneixement, tenim:

- Esquemes lògics: empren sistemes basats en lògica de primer ordre.
- Xarxes semàntiques: adequats per classificacions i son gràfics.

I també, pels esquemes de representació de procediment o sistemes de producció, tenim:

- El Frame: descriu el atributs d'un determinat objecte o situació complexa.
- Els Scripts: descriuen seqüències d'esdeveniments en un context particular.
- Sistemes Experts: reprodueix el coneixement d'un expert humà en el seu domini competent.

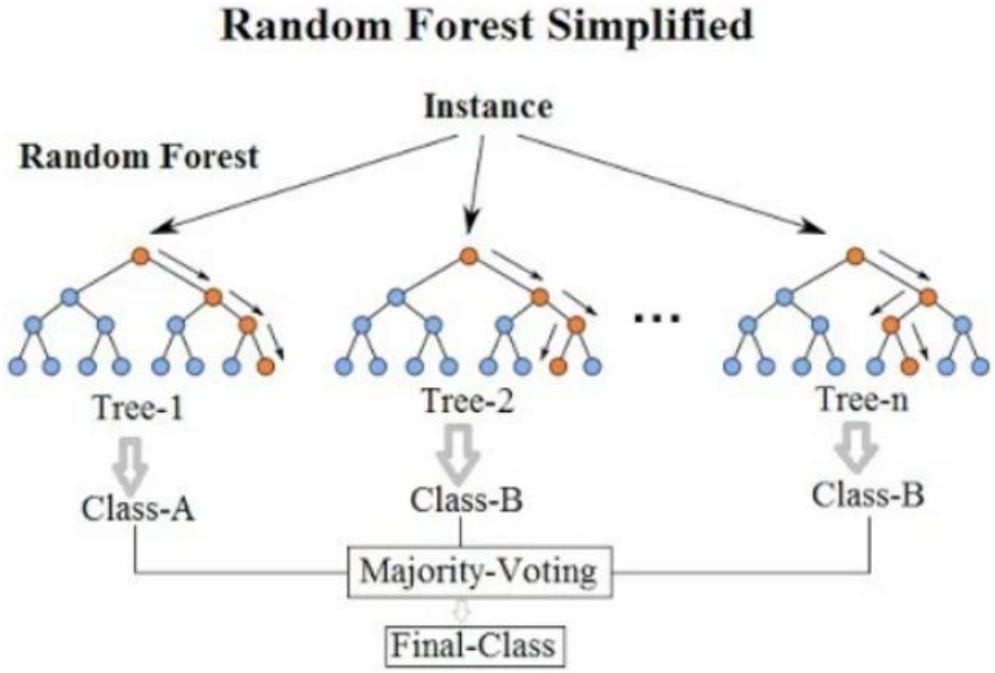
# Bases de la IA (II)

Les escoles de pensament sobre les IA, son les següents:

- Intel·ligència artificial convencional: És la IA simbòlic-deductiva basada en l'anàlisi formal i estadístic del comportament humà davant diferents problemes.
- Intel·ligència artificial computacional: És la IA subsimbòlic-inductiva, que comporta un aprenentatge interactiu i ajustament d'aquest, en base a dades empíriques.

Alguns dels diferents algorismes clàssics emprats inicialment en les IA, son:

- Regressió Lineal.
- Regressió Logística.
- Naive Bayes.
- K-nearest Neighbors
- Decision Tree
- Random Forest



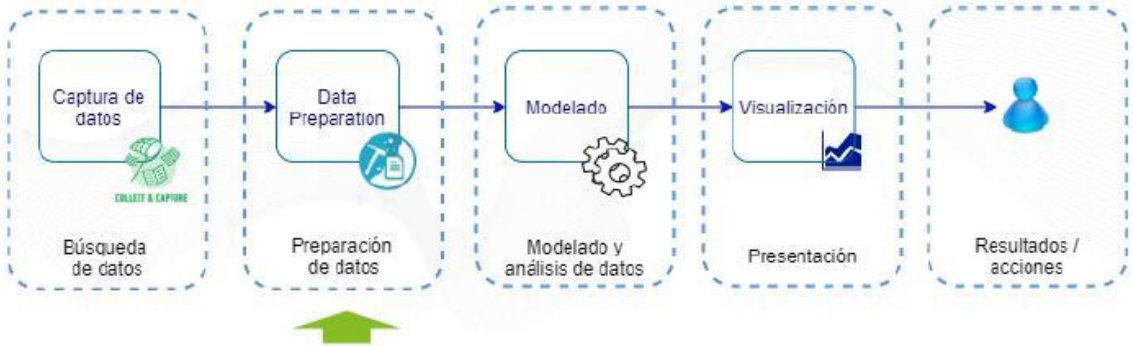
# Aprenentatge de la IA

Per a l'entrenament de la IA, tenim dos models d'aprenentatge:

- Aprenentatge automàtic (ML): intenta que un sistema aprengui i relacioni informació, igual que ho faria un humà, però amb un resultat millorat en el temps via algorismes adaptatius.
- Aprenentatge profund (DL): subconjunt de ML, inspirat en les xarxes neuronals, inclou l'experiència i el mateix sistema determina conclusions ell sol, amb un raonament no lineal.

Cal evitar el sobreajustament, principalment amb:

- Retenció de dades.
- Validació creuada.




Altres tècniques emprades junt a l'anterior, son: Oversampling, Stratified sampling, Early stopping, Train-validation-test sets, Cross-validation, Regularization, i Pruning.

L'objectiu de tot l'anterior es que les fonts d'informació emprades per la IA, siguin correctes, que no estiguin les dades esbiaixades, i així, amb el ML i el DL aconseguir una "IA fiable", segons es descriu a la Web de **ALTAI** (The Assessment List on Trustworthy Artificial Intelligence) per la UE.

# Classificació de les IA (I)

Clasificación de IA

Russell & Norving

| | inteligencia humana | racionalidad |
|---|---|---|
| pensamiento | Sistemas que piensan como humanos: *Enfoque cognocitivo* | Sistemas que piensan racionalmente: *Enfoque lógico* |
| comportamiento | Sistemas que actúan como humanos: *Enfoque "prueba de turing"* | Sistemas que actúan racionalmente: *Enfoque de agentes racionales* |

Segons experts en IA i ML, és dona les següents categories :

- **Sistemes que pensen com humans**.
- **Sistemes que actuen com humans**.
- **Sistemes que pensen racionalment**.
- **Sistemes que actuen racionalment** (idealment).

La classificació més habitual en el mon de les IA, sol ser:

- IA Dèbil (o estreta), que es tota la desenvolupada fins ara.
- IA Forta (o general), iguala o supera la intel·ligència humana, de moment es una utopia.

Una altra forma de classificar la IA, potser agrupant-les segons la tecnologia emprada, pel que es pot tenir grups com: Speech recognition, NLP, Visual recognition, Text recognition, Big Data, Sistemes Experts, Robòtica, ML, DL, Cognitive intelligence & cognitive services, etc.

# Classificació de les IA (II)

Per altra banda, en quant a l'aprenentatge ML o DL, és dona les següents categories :

- **Aprenentatge supervisat:** empra models predictius pels problemes de
  - Classificació
  - Regressió
- **Aprenentatge no supervisat:** els algorismes ajusten el seu model, segons entrada
  - Clustering.
  - Associació de regles.
  - Detecció d'anomalies.
  - Reducció de dimensions


- **Aprenentatge per reforç**: és base en la iteració constant via "prova i error", obtenint una recompensa. S'empra un agent que adapta comportament en funció de la recompensa i no li cal conèixer els processos de decisió. Els algorismes emprats, son: Criteri d'optimalitat, Força bruta, Atansament al valor de la funció, i Cerca política directa.

# Punts forts de les IA (I)

En un negoci, la IA ens pot ajudar en varies coses, com:

- Tenir una millor comprensió sobre l'abundància de dades que hi ha disponibles
- Confiar a les prediccions la automatització de tasques complexes o mundanes

En aquest sentit, ja dit abans, la IA està impulsada per tres factors:

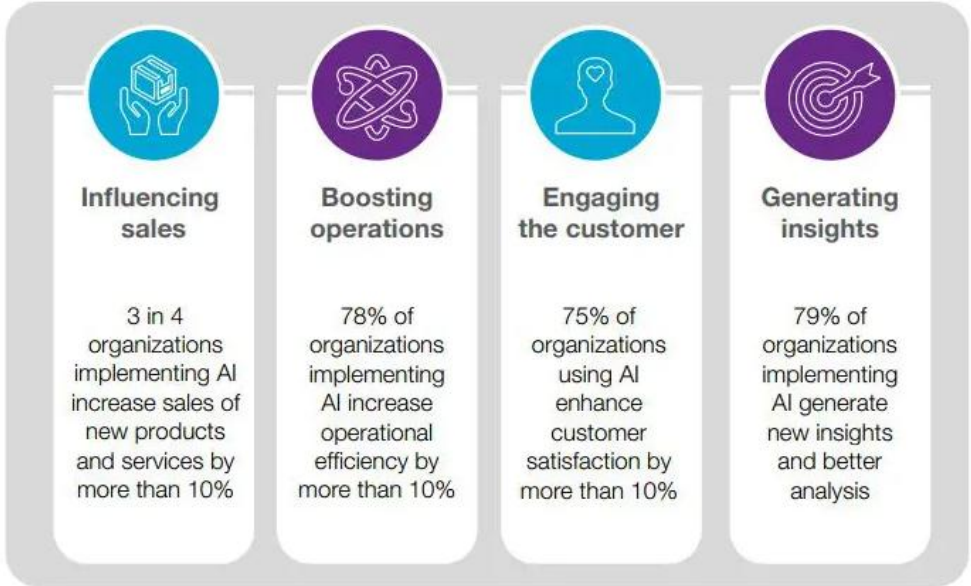
- La capacitat de computació i d'alt rendiment ja està disponible
- Hi ha disponibles grans volums de dades per l'aprenentatge
- La IA aplicada proporciona un avantatge competitiu

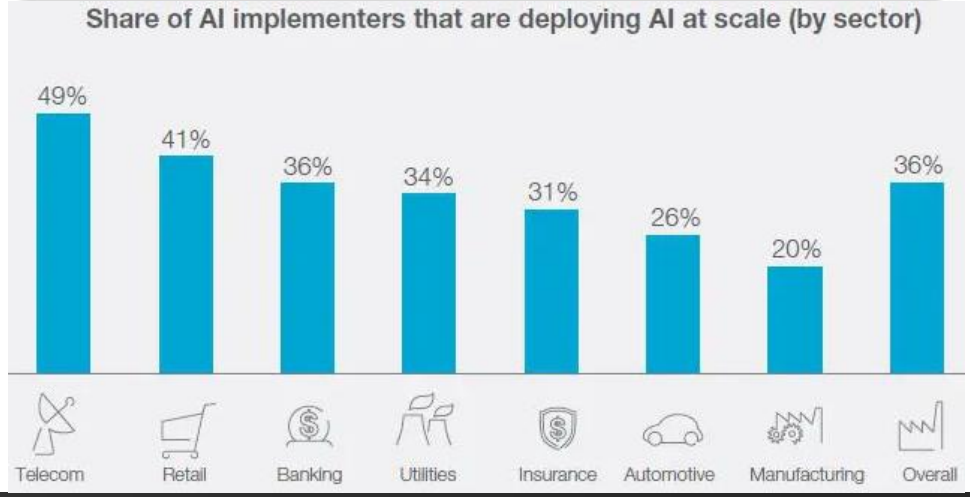
** Un detall a tenir en compte, es que totes les grans empreses tecnològiques d'avui en dia, i també les emergents, han adoptat la IA per a moltes tasques, i n'ofereixen serveis en línia a Internet, o en els seus Clouds o com productes Software per empreses.

# Punts forts de les IA (II)

Segons varies persones reconegudes en el mon de les IA, com **Andy Chan** i **Kai-Fu Lee**, les IA ens poden aportar el beneficis o avantatges següents:

- Automatització de processos
- Potenciació de les tasques creatives
- Aporten precisió
- Redueixen l'error humà
- Redueixen temps emprat en anàlisi de dades
- Manteniment predictiu
- Millora la presa de decisions en producció i negoci
- Control i optimització de processos i línies de producció
- Augment de productivitat i qualitat a la producció



# Situació i reptes de la Ciberseguretat (I)

De forma esquemàtica i resumida, a la Ciberseguretat ens trobem actualment amb el següent:

- Gran creixement dels Ciberatacs
- Increment d'amenaces i vulnerabilitats
- Increment del tipus i varietat de "endpoints"
- Increment de serveis externs al Cloud
- Expansió de la superfície exposada
- Falta de personal amb coneixements/capacitats

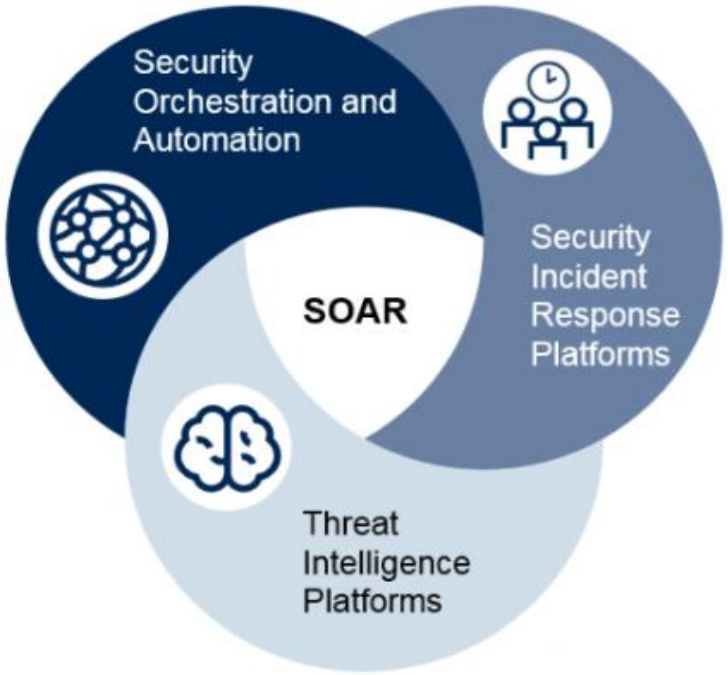
- Impossibilitat de gestionar totes les Alertes, avisos i notificacions
- Impossibilitat de gestionar totes les actualitzacions (S.O., Drivers, Firmware, ...)

Els plantejaments tradicionals de Ciberseguretat amb "frameworks" de gestió, polítiques de seguretat i eines (Firewall, WAF, IDS/IPS, Antivirus, Routers, Switchos, SIEM, ...) s'han quedat curts per la protecció necessària. Aquests plantejaments es basen principalment en informar, però la reacció o actuació en front una amenaça es de forma manual, per tant, estan obsolets o mancats d'actualització i millora.

# Situació i reptes de la Ciberseguretat (II)

Per respondre a l'anterior, apareix el SOAR(Security Orchestration, Automation and Response) i el EDR (Endpoint Detection Response), son RPA. Així, obtenim els beneficis següents amb el SOAR:

- Temps mig de resposta reduït (MTTR)
- Reducció del impacte d'amenaces
- Millor intel·ligència sobre amenaces
- Security Insight (no sols informació)
- Operacions optimitzades
- Millor rendiment i productivitat
- Costos reduïts



SOAR = SOA + SIR + TIP

I els beneficis que ens aporta el EDR (com antivirus centralitzat), son:

- Millor capacitat d'anticipació sobre els atacs dirigits
- Reducció del temps d'exposició a les amenaces
- Visió global de les amenaces contra els endpoints i accés a la informació en un sol punt
- Recull informació exhaustiva i detallada dels endpoints (O.S., Hardware, processos, ...)
- Pot crear patrons de detecció automatitzats i recollir informació automàticament
- Monitora la integritat dels sistemes i arxius de configuració

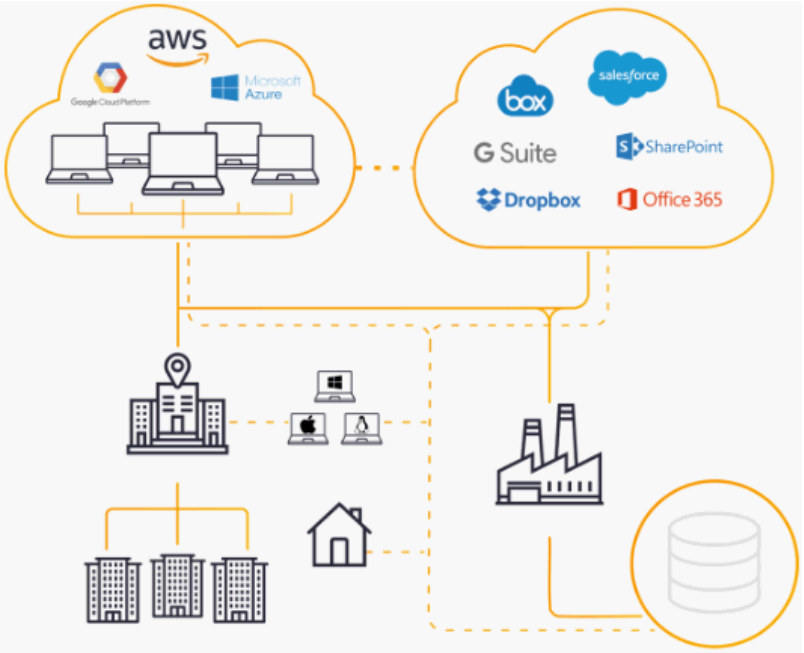
# IA i Ciberseguretat (I)


Recordem alguns avantatges que te la IA:

- Gestió massiva d'informació
- Resposta en temps real
- Automatització
- Predicció.

Per tant, la IA dona una resposta, robustesa i resiliència, que poden aplicar-se a diferents àrees de la Ciberseguretat com:

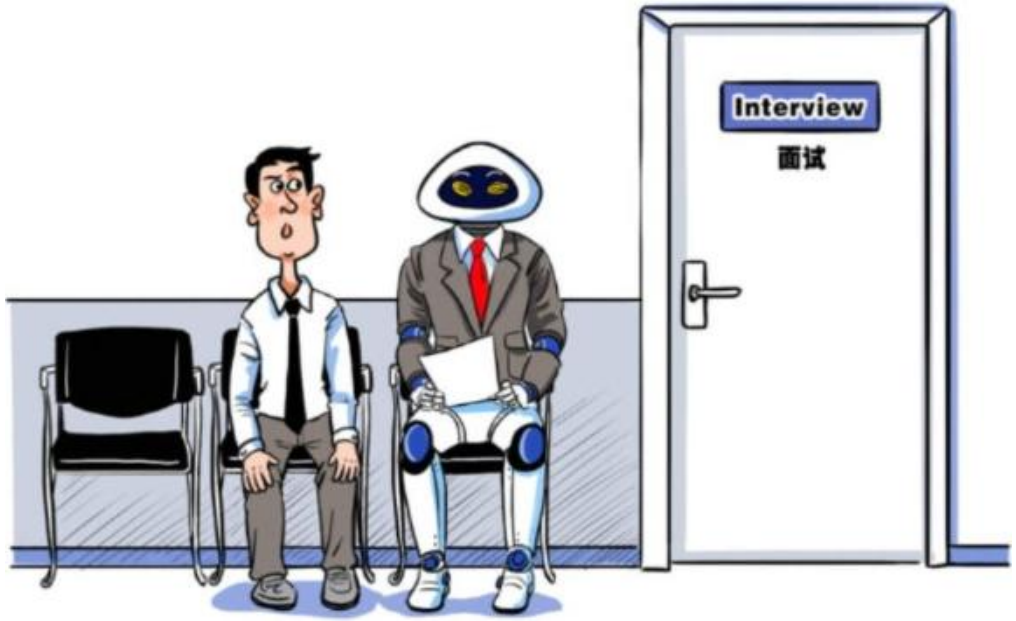
- Threat hunting (identificació i neutralització d'amenaces)
- Gestió de Vulnerabilitats
- Data Centers
- Seguretat en les xarxes
- Identificació segura d'usuaris
- Privacitat de la informació i "compliance"
- Bloqueig de bots segons comportament



# IA i Ciberseguretat (II)

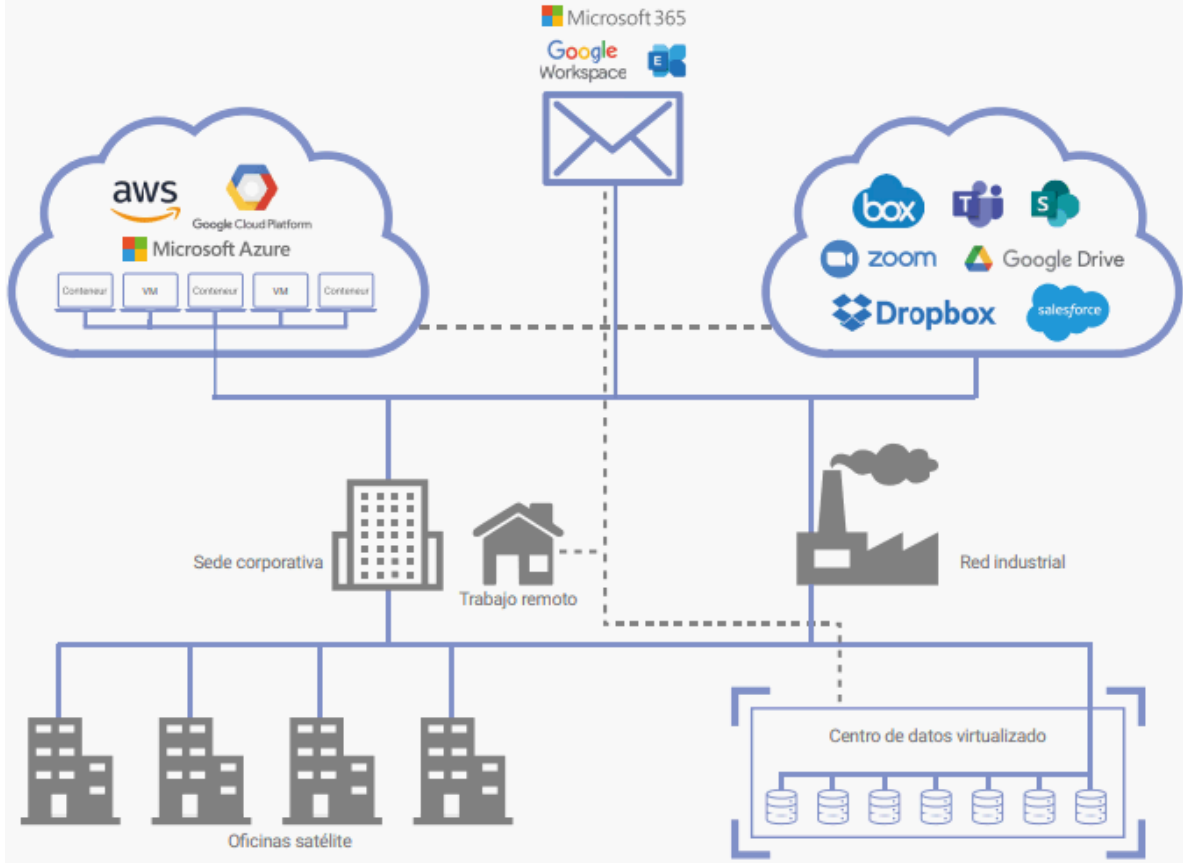
Per altra banda, la IA també pot millorar molts punts febles de la Ciberseguretat, hi ha processos i aspectes que es consideren com normals, però que en realitat no ho son, i la IA els pot corregir i els pot millorar. I tindríem els següents:

- L'error humà en la configuració
- L'eficiència humana en activitats repetides
- La fatiga per excessos d'alarmes d'amenaces
- El temps de resposta a les amenaces
- La identificació i predicció de noves amenaces
- La capacitat de dotació de personal
- L'adaptabilitat



# Exemples de IA per Ciberseguretat (I)

Com primer exemple, tenim el anomenat "Cyber AI" de l'empresa Darktrace:

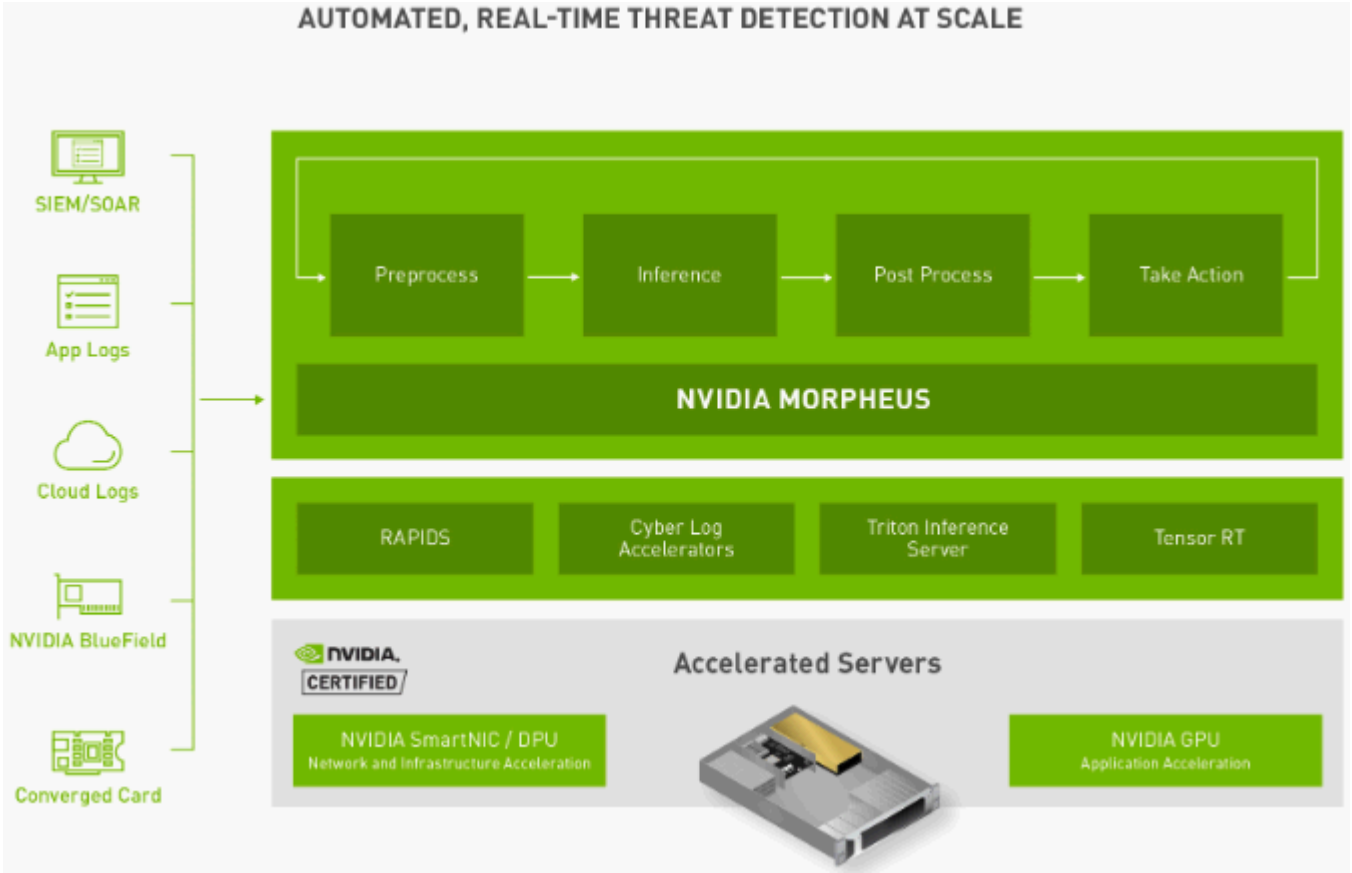
# Exemples de IA per Ciberseguretat (II)

Com segon exemple, tenim el "Dr. Watson" (o sols Watson) de l'empresa IBM:

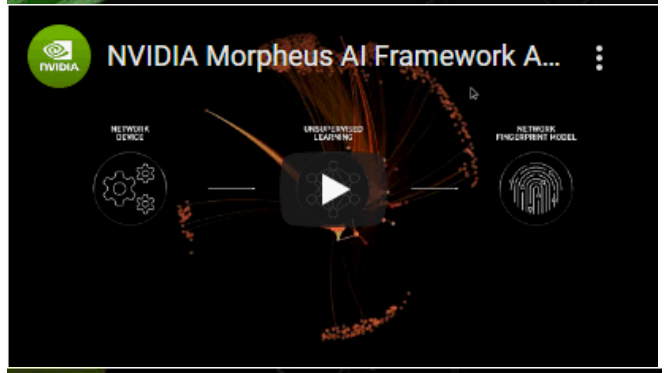
| CASOS DE USO POR CATEGORÍA | Descripción | Ventas | Marketing | Operaciones | Finanzas | Atención al Cliente | Recursos Humanos | IT |
|---|---|---|---|---|---|---|---|---|
| Visualización de datos interactiva | Encuentre rápidamente nuevos conocimientos sobre datos con visualizaciones interactivas, lenguaje natural, puntos de partida inteligentes y plantillas. | • Calidad del pipeline<br>• Incentivos de ventas<br>• Ajuste del territorio por rep.<br>• Estrategia de ventas<br>• Rentabilidad de las ventas<br>• Eficacia de las ventas | • ROI de campañas<br>• Tendencias de los clientes<br>• Rentabilidad de los clientes<br>• Coste de adquisición<br>• Valor del ciclo de vida del cliente | • Análisis de la cadena de suministro<br>• Asignación de activos<br>• Uso<br>• Abastecimiento de materiales<br>• Compliance<br>• Informes de reclamaciones<br>• Benchmarking de la competencia | • Análisis del ROI<br>• Riesgo<br>• Análisis programado<br>• Estrategia de precios<br>• Planificación: análisis actual<br>• Análisis de rentabilidad | • Rendimiento de los agentes<br>• Complejidad de los casos<br>• Escalado de los casos<br>• Repetición de llamadas<br>• Satisfacción de los clientes | • Planificación salarial<br>• Eficacia del plan de incentivos<br>• Eficacia de la formación<br>• Análisis de los beneficios a los empleados<br>• Eficacia de la gestión laboral | • Salud del sistema<br>• Análisis de tickets<br>• Análisis y cumplimiento de SLA<br>• Rechazar ciberataques |
| Búsqueda automática de patrones de comportamiento | Construir automáticamente modelos estadísticos de sus datos.<br>Actuar con confianza sobre los hallazgos gracias a visualizaciones y textos descriptivos que generan resultados. | • Retención de representantes<br>• Clasificación de oportunidades<br>• Patrones de contratos Ganados/Perdidos | • Cross-sell/Up-sell<br>• Mejora de los ratios de respuesta<br>• Pérdida de clientes<br>• Optimización del Tráfico Web | • Planificación de inventario & stock<br>• Optimización de la producción<br>• Priorización de las rutas de distribución<br>• Predicción de interrupciones | • Colecciones<br>• Ingresos faltantes<br>• Fraude<br>• Previsión de ingresos | • Resoluciones con éxito<br>• Duración de las llamadas<br>• Pérdida de clientes | • Reclutamiento<br>• Desgaste de los empleados<br>• Características de los mejores empleados | • Predicción de interrupciones<br>• Planificiación de la capacidad<br>• Previsión de la demanda |
| Análisis de Social Media | Obtenga información sobre las marcas, organizaciones o temas de su elección en las redes sociales. | • Conocimiento de la competencia<br>• Intención de compra | • Salud, atributos y monitorización de marca<br>• Eficacia de las campañas | • Conocimiento de los proveedores | • Conocimiento del vendedor<br>• Sentimiento del inversor | • Sentimiento de clientes<br>• Servicio proactivo de atención al cliente<br>• Programas de fidelidad | • Reclutamiento<br>• Supervisión de la actividad de los empleados | • Evaluación de la amenaza |

# Exemples de IA per Ciberseguretat (III)

Per acabar, tenim el "Open AI framework Morpheus" de l'empresa NVIDIA:



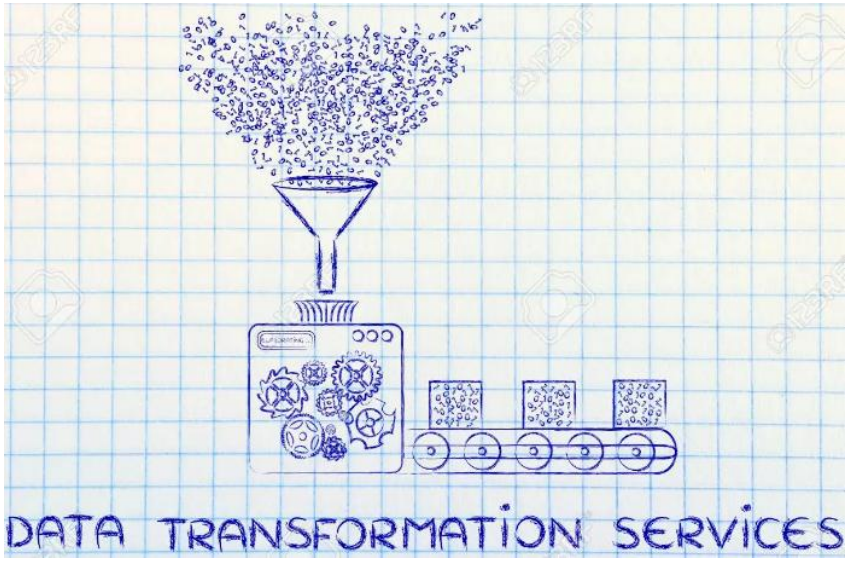





# Futur de la IA a la Ciberseguretat (I)

Amb l'aprenentatge, s'observa una gran dependència de la IA sobre les dades, i segons diferents regulacions i privacitat de dades, cal vigilar la qualitat de la dada, el model i el resultat, en aspectes com:

- Privacitat: emprar qualsevol model, mantenint privacitat
- Equitat: evitar afectació biaixos implícits de mostres
- Traçabilitat: si hi ha una falla, s'ha de poder analitzar
- Robustesa: fins on ens podem fiar de la IA en cas de Ciberatac
- Fiabilitat: si canvia l'entrada, la sortida continua fiable
- Causalitat: saber si es pot influir sortida, actuant sobre entrada
- Explicabilitat i transparència: enteniment del model pel usuari
- Governança de la dada: ús lícit, eficient i eficaç

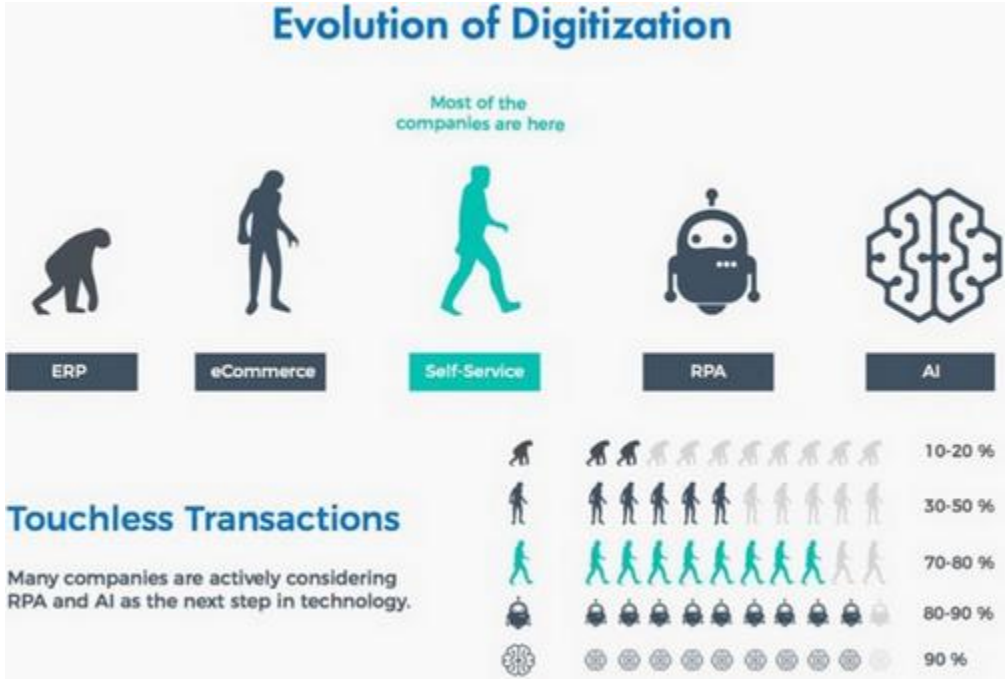



**Una cosa a tenir en compte, és que falten molts experts de IA en ML i s'haurà d'emprar experts externs, a part de que els equips humans en Ciberseguretat amb IA seguiran sent essencials.

# Futur de la IA a la Ciberseguretat (II)

Segons diferents experts en IA i Ciberseguretat, les deu tendències futures, seran les següents:

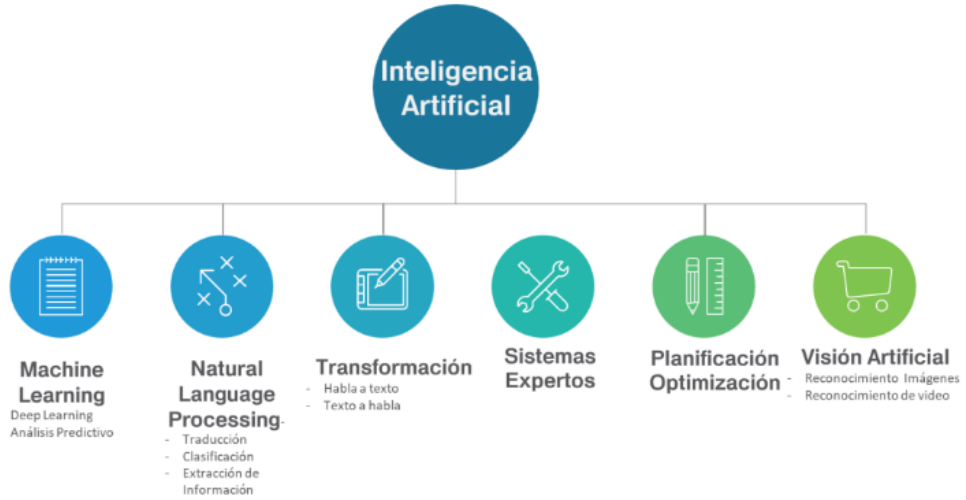
- Millor resposta al Ransomware
- Combinar desenvolupament d'aplicacions amb Ciberseguretat
- Emprar DL per detectar "DGA-Generated Domains"
- Detectar amenaces que no son de programari maliciós
- Honeypots adaptatius i Honeytokens
- Millor comprensió del funcionament de les xarxes neuronals
- Utilització de xarxes neuronals a capsules (CapsNet)
- Aprenentatge de reforç profund (Q-Learning)
- Protecció de IoT (i de OT industrial)
- Predicció del futur

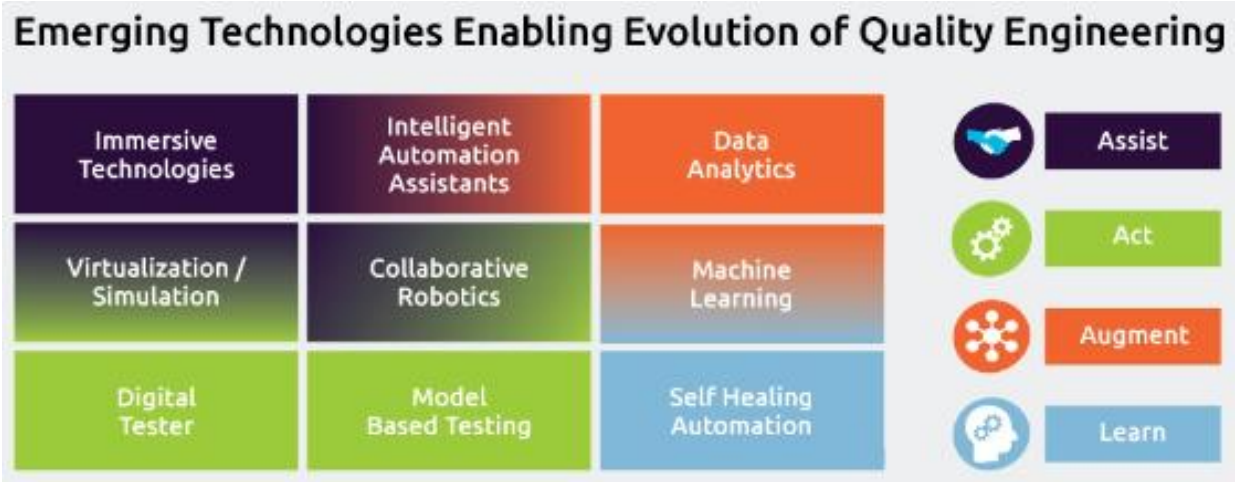


# Futur de la IA a la Ciberseguretat (III)

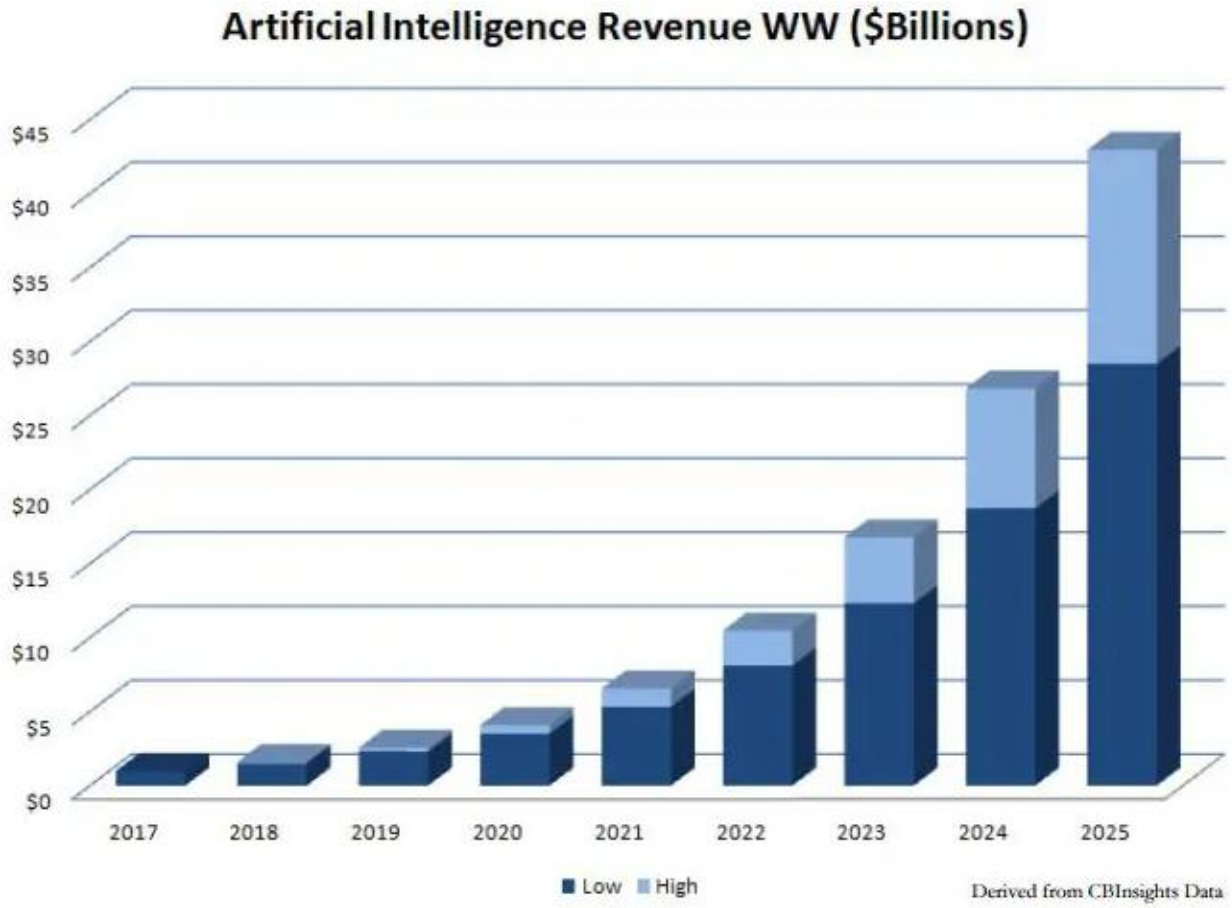
I segons altres experts en IA i Ciberseguretat, les cinc tendències futures, seran les següents:

- Reducció càrrega i necessitat d'experts a Ciberseguretat
- Automatitzar mesures seguretat en gestió d'identitats i accés
- Millora seguretat procés de Blockchain (si s'empra)
- Millorar esforços en acompliment normatiu
- Millorar la seguretat de la xarxa al núvol (Cloud)



# Futur de la IA a la Ciberseguretat (IV)





Nuevos puesto de trabajo en el campo de la inteligencia artificial

# Conclusions (i línies de treball futur)

Segons **Andrew Y. Ng**, 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".

Com a tot treball, hi ha coses o parts, que es podrien canviar, modificar, ampliar o revisar, pel que identifiquen alguns punts de treball futur, com :

- Aprofundir en les directrius de **ALTAI**, a nivell Europeu, per aconseguir una "IA fiable", tal com la defineixen ells.
- Revisar les millores més recents que es donen el mon de les IA, en els sistemes cognitius (com l'ús de GPU) i que es poden aplicar per gestionar la Ciberseguretat.
- I per últim, fer algun estudi sobre com pot quedar i millorar el mapa de rols de Ciberseguretat al afegir una IA com gestor.

# Cloenda

- Fi de la presentació, i moltes gràcies per la seva atenció
- Preguntes i dubtes del TFM: jorfencas@uoc.edu