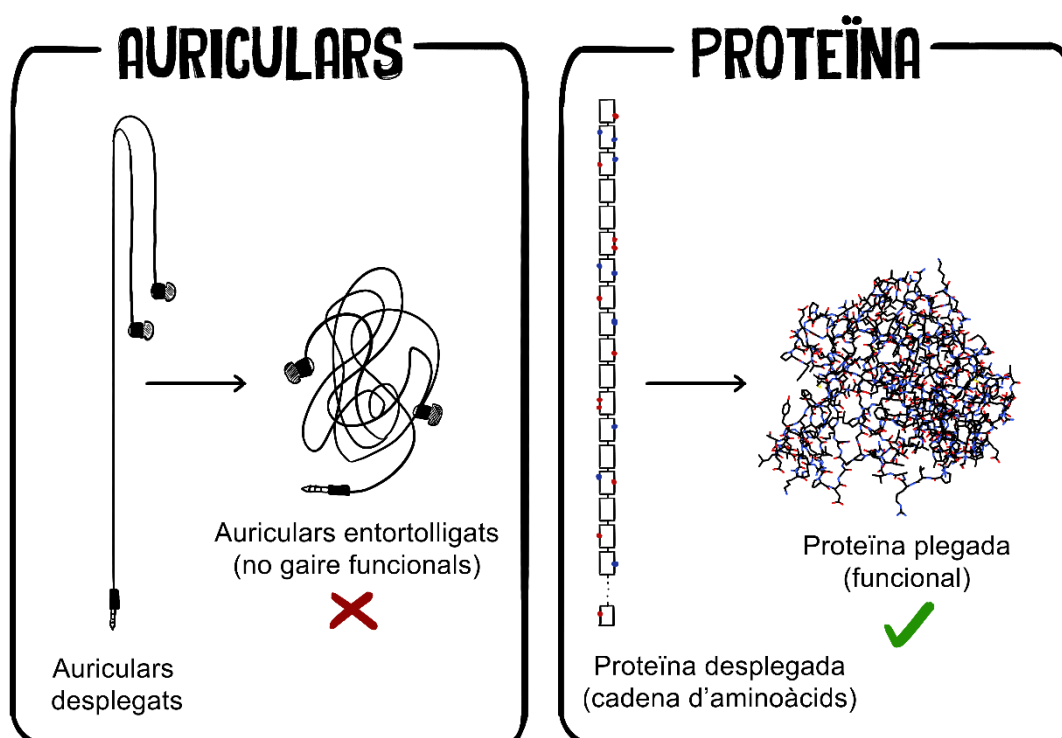


## AlphaFold i la revolució de les proteïnes.

Les proteïnes són molècules grans i complexes que realitzen una àmplia gamma de funcions en els organismes vius, des de catalitzar reaccions químiques fins a transportar molècules dins de les cèl·lules. Cada proteïna està formada per una seqüència única d'aminoàcids, que s'uneixen com perles en un collaret. La seqüència d'aquests aminoàcids és el que determina l'estructura tridimensional de la proteïna final, que a la seva vegada determina quina serà la seva funció.

Perquè ens entenguem, una proteïna és com aquells auriculars amb cable que deixem oblidats en alguna butxaca i queden plegats i entortolligats sobre si mateixos. Amb la diferència que la proteïna pot exercir la seva funció un cop està plegada, i els auriculars no tant. Tal com els cables dels auriculars es van embolicant i girant uns al voltant dels altres, els aminoàcids que formen una proteïna també es poden torçar i girar de moltes maneres. Això dificulta entendre la seva estructura!



Precisament, determinar l'estructura d'una proteïna és un pas crític per entendre la seva funció, però també és una tasca increïblement difícil. De fet, es poden necessitar anys d'assaig i error per determinar l'estructura d'una sola proteïna mitjançant mètodes experimentals com la cristal·lografia de raigs X o l'espectroscòpia de ressonància magnètica nuclear (RMN).

Aquí és on entra en joc el protagonista d'avui, AlphaFold. AlphaFold és un algorisme d'aprenentatge profund desenvolupat per DeepMind, una empresa que va ser comprada per Google el 2014. Aquest algorisme anomenat AlphaFold va ser creat per predir l'estructura d'una proteïna en funció de la seva seqüència d'aminoàcids. És a dir, en funció de les peces que la conformen, però sense necessitat de tenir la proteïna per estudiar-la. El programa és capaç de fer aquestes prediccions sobre plegament

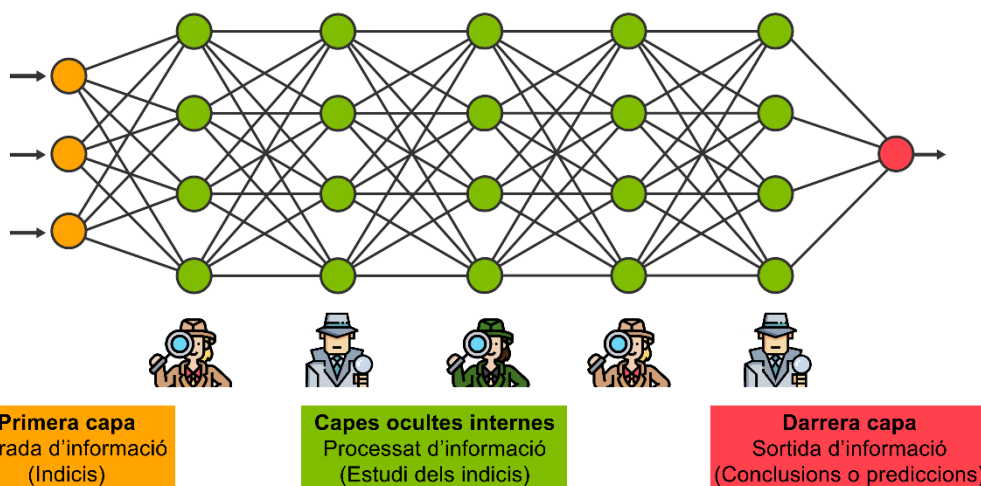
mitjançant una base de dades massiva que conté les estructures ja resoltes de proteïnes conegudes, així com una varietat d'altres dades d'interès.

El programa funciona només després d'haver estat entrenat amb aquest conjunt de dades d'estructures de proteïnes conegudes que acabem de mencionar. Més precisament, utilitza una xarxa neuronal profunda per identificar els patrons i les relacions entre les seqüències d'aminoàcids i les estructures de proteïnes corresponents. Un cop s'ha entrenat la xarxa neuronal, es pot utilitzar per predir l'estructura d'una nova proteïna basant-se únicament en la seva seqüència d'aminoàcids.

## PERÒ... QUÈ ÉS UNA XARXA NEURONAL PROFUNDA?

És com un grup de detectius que treballen junts per resoldre un crim. Com si cada detectiu té un paper específic i està especialitzat en determinats tipus d'indicis, les diferents capes d'una xarxa neuronal també tenen una tasca específica al analitzar i interpretar les dades.

De la mateixa manera que els detectius fan servir els indicis i proves que recopilen per reconstruir una imatge del que va passar en un crim, una xarxa neuronal fa servir les dades amb les que ha estat entrenada per fer prediccions o prendre decisions sobre noves dades que se li presenten. Com més dades ha après la xarxa, millor es torna a l'hora de fer prediccions precises!

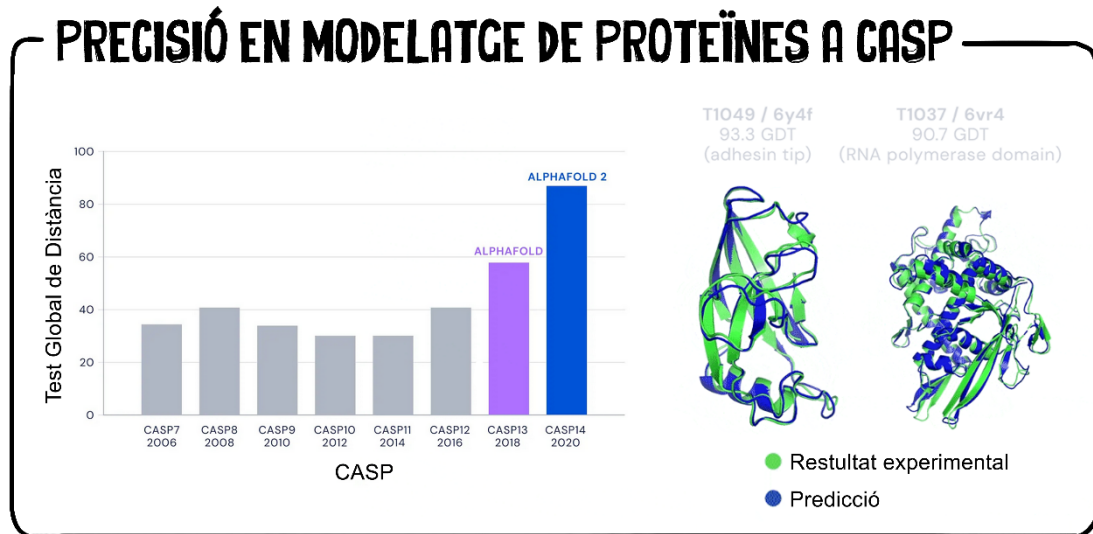


En general, una xarxa neuronal profunda és un tipus d'algorisme d'aprenentatge automàtic dissenyat per imitar la forma en què el cervell humà processa la informació. Utilitzant diverses capes de nodes interconnectats, i podent analitzar dades complexes i fer prediccions amb una precisió impressionant.

Una de les innovacions clau d'AlphaFold és la forma en què integra múltiples fonts de dades per fer les seves prediccions. A més de la seqüència d'aminoàcids, el programa té en compte una àmplia gamma d'altres dades, incloses les distàncies entre parells d'aminoàcids de la cadena de proteïnes, els angles entre els aminoàcids adjacents i les posicions relatives de diferents parts de la proteïna.

Mitjançant aquest enfocament, AlphaFold és capaç de fer prediccions molt precises de l'estructura d'una proteïna. De fet, amb un nivell de precisió que rivalitza o supera el de mètodes experimentals com la cristal·lografia de raigs X o l'espectroscòpia RMN, que hem mencionat anteriorment. De fet, a l'experiment de predicció a cegues "CASP" del 2020, AlphaFold va poder predir l'estructura de gairebé totes les proteïnes del conjunt

de proves amb un nivell de precisió que era competitiu amb aquests mètodes experimentals, una gesta increïble.



Les implicacions d'aquest avenç són molt importants. Amb la capacitat de predir l'estructura de les proteïnes amb més precisió i eficàcia que mai, els científics poden ja obtenir nous coneixements sobre els mecanismes subjacents de moltíssimes malalties i desenvolupar tractaments més efectius per a condicions com el càncer, l'Alzheimer i el VIH (virus causant de la SIDA).

A més, la tecnologia darrere d'AlphaFold podria tenir implicacions més àmplies per al camp de la intel·ligència artificial i l'aprenentatge automàtic. El programa és un bon exemple de com els algorismes d'aprenentatge profund es poden utilitzar per resoldre problemes científics complexos i podria obrir el camí per a nous avenços en una àmplia gamma de camps, des de la ciència dels materials fins a la modelització del clima.

En general, AlphaFold representa un avenç important en la nostra capacitat d'entendre i manipular els components de la vida. Tot i que les implicacions d'aquesta tecnologia encara s'estan explorant, té el potencial de capgirar completament com abordem actualment els problemes científics complexos en camps de la medicina o altres ciències de la vida. De fet, de ben segur que ja heu sentit a parlar sobre altres algorismes basats en xarxes neuronals com el famós chatGPT. De fet, la versió més recent (versió 4) de ChatGPT va aprovar l'examen de llicència mèdica dels EUA amb molt bona nota.

Agafem-nos fort que venen temps de canvis! Ens veiem a la propera entrada de Xarrup de Ciència!

## Referències

- Senior, A.W., Evans, R., Jumper, J. et al. Improved protein structure prediction using potentials from deep learning. Nature 577, 706–710 (2020). <https://doi.org/10.1038/s41586-019-1923-7>
- Estructura d'una Ribokinasa al PDB: <https://www.rcsb.org/structure/8CQX>
- Portal web de deepmind: <https://www.deepmind.com/open-source/alphafold>
- Resultats del CASP: <https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>
- Article de Wired: How AI is Unlocking the Secrets of the Protein Universe" <https://www.wired.com/story/how-ai-is-unlocking-the-secrets-of-the-protein-universe/>
- Informe tècnic de GPT-4 arXiv:2303.08774 [cs.CL]. <https://doi.org/10.48550/arXiv.2303.08774>