

18 luglio 2023 15:18

CLIAPI 7: cosa c'entra Tesla con l'intelligenza artificiale?

di [Alessandro Pedone](#)



Domani Tesla presenterà al mercato i dati finanziari del secondo trimestre. L'azienda ha annunciato la realizzazione del primo veicolo CyberTruck destinato alla produzione in serie. Questo veicolo non è un prototipo ma è uscito dalla catena di produzione ed è in grado di scalare verso le decine di migliaia e poi i milioni di pezzi. Il CyberTruck è destinato a rimanere nella storia al pari del Ford Model T poiché è costruito con l'esoscheletro in acciaio, un modo radicalmente diverso rispetto a tutti gli altri veicoli.

Le innovazioni che Tesla sta portando avanti nel settore automobilistico stanno diventando sempre più note al pubblico. Lo scetticismo che c'era quando ho iniziato a scrivere di Tesla, nel 2020, da parte dei media e degli analisti finanziari sulla capacità di Tesla di rivoluzionare il settore auto, sta quasi completamente scomparendo. Ma Tesla continua ad essere percepita come un'azienda che produce auto. Questo è un errore madornale.

Il 14 febbraio scorso ho scritto il terzo articolo di questa serie dal titolo "Cosa c'entra Palantir con l'intelligenza artificiale?". All'epoca Palantir era considerata dalla maggioranza degli analisti come una grossa azienda di consulenza che lavora prevalentemente nel settore governativo sviluppando software di analisi dei dati estremamente personalizzati.

Questo era quello che si poteva vedere da un'analisi superficiale dell'azienda e dei suoi bilanci.

Quello che è successo nei mesi successivi è che il mercato ha compreso che Palantir è un player di primo piano nel settore dell'intelligenza artificiale ed è una società di software, non di consulenza.

Questo è quello che è accaduto al titolo dal giorno in cui ho pubblicato quell'articolo.



Tuttavia, questo articolo è dedicato a Tesla, non a Palantir.

Ho voluto dare un titolo simile perché anche Tesla è considerata dagli analisti solo una società che fa automobili, ignorando il fatto che Tesla è **la principale azienda nel settore dell'intelligenza artificiale applicata al mondo reale**.

Mentre Palantir ha tutte le possibilità di diventare - entro 5/10 anni - la società leader nella piattaforma per la realizzazione dei nuovi tipi di software che chiamiamo impropriamente intelligenza artificiale, Tesla ha la possibilità di diventare l'azienda che farà più profitti di qualsiasi altra azienda al mondo con servizi che porteranno l'intelligenza artificiale fuori dagli schermi dei dispositivi elettronici per interagire con il mondo reale, ovvero muovere fisicamente oggetti.

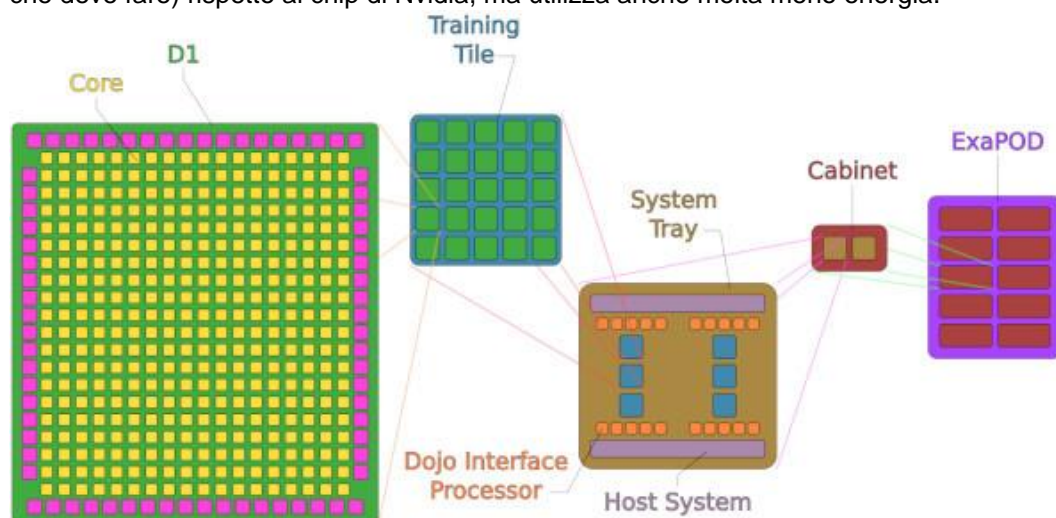
La tecnologia che Tesla sta sviluppando è letteralmente sconvolgente, ma la maggioranza degli analisti finanziari semplicemente la sta ignorando, come ignoravano ciò che ha sviluppato Palantir.

In questo articolo desidero approfondire le due principali tecnologie legate all'intelligenza artificiale che Tesla sta sviluppando, nei limiti di un articolo divulgativo.

Dojo: il supercomputer per l'addestramento delle reti neurali

Tesla Dojo è un supercomputer progettato e costruito in casa da Tesla, pensato per l'addestramento di reti neurali, in particolare quelle che riguardano le immagini.

L'obiettivo di Dojo è di elaborare in **modo efficiente** (cioè risparmiando moltissima energia rispetto alla migliore tecnologia oggi disponibile, che è quella di Nvidia) milioni di terabyte di dati video catturati da situazioni di guida reali dalle circa 4 milioni di auto Tesla in giro per il mondo (che cresceranno enormemente nei prossimi anni). Questo obiettivo ha portato a un'architettura notevolmente diversa rispetto ai tradizionali design dei supercomputer. Dojo si basa sul chip D1, progettato da un team di Tesla guidato dall'ex progettista di CPU AMD Ganesh Venkataramanan. Il chip D1 ha 50 miliardi di transistor e una potenza di calcolo di 376 teraflop usando numeri a virgola mobile configurabili a 8 o 16 bit (CFloat8 e CFloat16). Ogni chip D1 è collegato a una rete a maglie bidimensionale con altri 24 chip D1, formando una "Training Tile". Sei Training Tile sono raggruppate in un System Tray, che è integrato con un'interfaccia host. Dojo impila i System Tray verticalmente in un armadio per minimizzare la distanza e il tempo di comunicazione tra loro. Il sistema Dojo ExaPod include 120 Training Tile, per un totale di 1.062.000 core utilizzabili, raggiungendo 20 exaflop. L'aspetto rivoluzionario di questo modello è che consente di ridurre in modo significativo il principale problema di queste architetture, ovvero i colli di bottiglia che si hanno quando una informazione deve passare da un singolo chip ad uno che sta distante. Con una serie di soluzioni ingegneristiche particolarmente ingegnose Dojo non solo è molto più veloce (per gli specifici tipi di calcolo che deve fare) rispetto ai chip di Nvidia, ma utilizza anche molta meno energia.



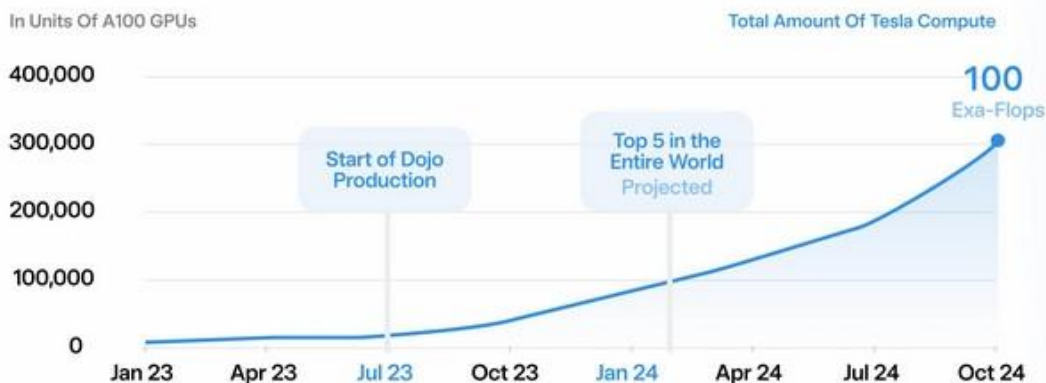
Tesla ha annunciato

Dojo per la prima volta ad Agosto del 2021 durante un evento chiamato "Ai Day". Ancora un volta, la quasi totalità degli analisti ha completamente ignorato la notizia pensando che fosse qualcosa di irrilevante e che comunque non avesse ancora nessun impatto sui ricavi dell'azienda.

Questo mese inizia la produzione effettiva del supercomputer. Attualmente Tesla, come tutti quelli che fanno intelligenza artificiale, utilizza i chip di Nvidia ed ha uno dei più grandi supercomputer al mondo. Ciò è necessario per elaborare l'immensa quantità di immagini raccolte quotidianamente dalle 8 telecamere presenti in ciascuna delle milioni di auto Tesla in giro per le strade.

Con l'inizio della produzione del nuovo sistema, Tesla ha annunciato l'intenzione di raggiungere entro la fine dell'anno la capacità di calcolo per entrare nei primi 5 supercomputer al mondo ed entro la fine del 2024 di triplicare la potenza fino a raggiungere i 100 exa-flops (1)

Trained On Extremely Large Compute



Si tratta di una potenza di calcolo strabiliante e completamente costruita in casa. Dojo costerà molto meno di quanto sarebbe costato avere questa stessa potenza di calcolo con i chip di Nvidia, e costerà meno anche in termini di energia consumata durante il funzionamento. Ma il vantaggio principale è che consentirà di fare cose che non sarebbe stato possibile realizzare senza tutta questa potenza di calcolo.

Tesla, infatti, ha realizzato Dojo semplicemente perché sul mercato non esisteva niente di simile. Per lo sviluppo degli ambiziosi progetti che vedremo nel prossimo paragrafo aveva bisogno di un hardware che non esisteva ancora.

E' molto probabile che un giorno Dojo possa essere un servizio che verrà affittato in modo simile a quello che fa Amazon con AWS e ciò costituirà un'altra fonte di ricavi per l'azienda.

Un modello di mondo (“Large World Model”)

Con tutta questa enorme potenza di calcoli e di dati, Tesla sta realizzando quello che potremmo chiamare un Large World Model, per utilizzare una analogia con i Large Language Model come il famoso ChatGPT.

[Nel primo articolo di questa serie](#), ho spiegato come funzionano questi modelli che sono essenzialmente predittivi. Si fornisce in input al modello un testo e questo restituisce un altro testo che rappresenta la più probabile continuazione del testo precedente.

Tesla sta facendo la stessa cosa, ma invece di fornire al modello in input un testo, fornisce dei filmati e il modello predice la più probabile continuazione del filmato stesso.

Non si tratta di una trasformazione delle immagini inserite, si tratta proprio di rigenerare l'immagine da zero. In modo simile a quello che fanno i famosi servizi come Dall-E o Midjourney, ma con la differenza che il modello di Tesla non genera immagini sulla base di un prompt, ma sulla base dei filmati precedenti.

Il modello di Tesla, in un certo senso, “comprende” ogni singolo dettaglio del filmato in input e rigenera il più probabile mondo immediatamente successivo che si può prevedere sulla base di cosa sta accadendo nelle immagini di input (gli oggetti - compreso il robot stesso - che si muovono) e di eventuali istruzioni.

Un esempio di istruzioni nel caso di un robot automobile: come sarà il mondo se decido di cambiare carreggiata? Per fare questo, il modello di mondo che sta sviluppando Tesla deve avere una rappresentazione non del linguaggio, ma proprio del mondo intero: oggetti (case, automobili, strade...), persone, animali, vegetazione, paesaggi.

Due settimane fa, Ashok Elluswamy, il capo del progetto AutoPilot di Tesla, ha presentato al CVPR Workshop on Autonomous Driving 2023 una relazione nella quale spiega questa tecnologia.

[Il video della presentazione è presente su YouTube:](#)

Non molti sanno che la nostra mente funziona, concettualmente, in modo simile a quello che sta facendo Tesla per

i suoi robot.

La maggioranza degli esseri umani è convinta che le immagini ricostruite dai nostri cervelli corrispondano esattamente a ciò che esiste nel mondo “reale”. Non è così! Il prof. emerito dell’Università della California, Donald Hoffman, ha pubblicato nel 2020 un eccellente libro, tradotto anche in italiano con il titolo “L’illusione della realtà. Come l’evoluzione ci inganna sul mondo che vediamo” dove spiega come la realtà che vediamo non è affatto come la vediamo, ma è una utilissima rappresentazione che non può essere reale.

In primo luogo, fisicamente, al centro dei nostri occhi, esiste un punto nel quale non esistono i ricettori della luce e quindi non ci sono informazioni fisiche su cosa ci sia in quello spazio. Se noi vedessimo esattamente ciò che colpisce i nostri occhi in quel preciso momento, dovremmo vedere un punto nero al centro di ogni immagine. Il nostro cervello ricostruisce l’immagine in tempo reale per sostituire quel punto cieco con ciò che è più probabile che vi sia sulla base delle immagini precedenti (i micromovimenti oculari, detti nistagmo fisiologico, servono anche a questo).

Più in generale il nostro cervello è una macchina predittiva che immagina ciò che accadrà nell’immediato futuro e sulla base di queste previsioni (parliamo di pochi secondi o frazioni) predispose l’intero organismo nella migliore condizione per fronteggiare la situazione prevista.

Il modello di mondo che sta sviluppando Tesla simula esattamente questa stessa cosa.

Nel prossimo paragrafo vedremo a cosa può servire una simile tecnologia.

Cosa farà Tesla con tutta questa tecnologia?

La prima applicazione pratica di questa tecnologia è chiaramente la guida autonoma. Salvo una manciata di analisti finanziari che ha compreso l’imminente rivoluzione che Tesla porterà nei prossimi 2-3 anni, la totalità degli analisti finanziari è convinta che la guida autonoma sia qualcosa che non impatterà significativamente nel business per ancora molti anni.

I pochi che si occupano di questo settore sono convinti che l’azienda leader sia **Waymo**, un’azienda di proprietà di Google, che produce taxi a guida autonoma effettivamente presenti sul mercato già oggi.

Il problema di questo approccio alla guida autonoma è che oltre ad essere incredibilmente costoso, non è scalabile perché si basa sulla disponibilità di mappe di estrema precisione che devono essere continuamente aggiornate e che sono presenti solo in aree molto ristrette. Se prendi uno dei costosissimi veicoli di Waymo e lo porti fuori da quell’area diventa utile come un ferro da stiro in mezzo alla strada.

Tesla sta realizzando la guida autonoma risolvendo i problemi fondamentali della visione e “comprensione” del mondo reale!

Elon Musk è da sempre stato (comprensibilmente) criticato perché da anni sostiene che entro l’anno il problema della guida autonoma sarà risolto dal punto di vista tecnico.

Come nella favola di Esopo, “Al lupo! al lupo!”, dopo un po’ di volte che Elon Musk ha detto questa cosa e non si è realizzata, nessuno lo prende più sul serio. Se analizziamo i progressi raggiunti da questa tecnologia, però, ci accorgiamo di due cose: la prima è che i progressi sono esponenziali, la seconda è che siamo arrivati ad un livello nel quale la guida autonoma è già più sicura di quella della media umana. La media umana non è certamente ancora un livello accettabile per considerare il problema risolto, è necessario che il software sia almeno 10 volte più sicuro di un essere umano, ma considerato il primo punto (cioè la crescita esponenziale) questo livello verrà raggiunto molto prima di quanto si possa pensare.

L’attuale software FSD (Full Self Driving) è in Beta ed è arrivato alla sua 11esima versione (ciascuna delle quali composte da tante altre sotto-versioni). L’attuale versione non è stata realizzata con Dojo e molto parzialmente include il “modello di mondo” di cui abbiamo scritto precedentemente (include le occupancy network, ma qui non vogliamo entrare troppo nei dettagli tecnici).

Considerato cosa è in grado di fare adesso il FSD 11, senza questa tecnologia, cosa sarà in grado di fare fra un anno quando le utilizzerà pienamente?

Un indizio può derivare dal fatto che Elon Musk ha annunciato che la versione 12 di FSD non sarà più in beta, ma sarà un prodotto considerato definitivo.

La guida autonoma è solo l’applicazione più vicina nel tempo di queste tecnologie, non la più importante! Sebbene la guida autonoma sia un mercato enorme (stimato 10 volte più grande rispetto al mercato della produzione di auto), c’è un mercato potenziale nel quale queste tecnologie possono essere applicate che possiamo considerare sostanzialmente senza un limite definibile: quello dei robot umanoidi.

Il modello di mondo al quale sta lavorando Tesla è stato certamente realizzato in primo luogo per la guida autonoma, ma quest’ultima è solo un caso ristretto di applicazione. Il caso di applicazione generale è quello dei robot umanoidi che Tesla sta sviluppando.

Fino a pochi mesi fa ero convinto che questa tecnologia fosse qualcosa che avremmo visto almeno fra cinque se non dieci anni.

Adesso, avendo dedicato decine di ore di studio a questa cosa, mi sto convincendo che inizieremo realmente a vedere qualcosa già nel 2024, con i primi robot presenti in quantità elevate nelle fabbriche di Tesla. Ritengo che la prima produzione per il pubblico sarà entro cinque anni, non più dai cinque ai dieci anni, come pensavo solo pochi mesi fa.

Conclusioni

Tesla è un'azienda unica. E' riuscita ad imporre una totale rivoluzione di un settore così importante come quello della produzione di auto e parallelamente sta preparando una rivoluzione in tre mercati ciascuno dei quali sono più grandi del precedente: 1) la guida autonoma, 2) la produzione e distribuzione di energia e 3) i robot umanoidi. Non esiste al mondo un'azienda che ha un potenziale assoluto (non percentuale) di crescita come Tesla. Nel breve termine, considerato che siamo all'inizio dell'emergere di una bolla sull'intelligenza finanziaria, il prezzo delle azioni Tesla possono fare di tutto: aumentare in modo parabolico o diminuire tantissimo se per qualche ragione la bolla dovesse scoppiare molto presto. Vista da una prospettiva di cinque anni, però, il prezzo attuale delle azioni Tesla sarà comunque considerato un affare, così come oggi consideriamo un prezzo favoloso quello che nel febbraio 2020 avevo chiamato "[Folle corsa di Tesla](#)". Per le condizioni del febbraio 2020, il prezzo di Tesla era sicuramente da considerarsi una bolla finanziaria (e nei mesi successivi si è sgonfiata!), ma se un investitore ha compreso la tecnologia che Tesla sta sviluppando e può aspettare che questa tecnologia si traduca in prodotti reali, il prezzo di oggi sarà comunque considerato un affare straordinario.

NOTA

1 - FLOPS è l'acronimo di Floating point Operations Per Second. In informatica, indica il numero di operazioni in virgola mobile eseguite in un secondo dalla CPU. Un exa-flop è un'unità di misura della potenza di calcolo di un computer. Un exa-flop corrisponde a un trilione di operazioni in virgola mobile al secondo

CHI PAGA ADUC

l'associazione non **percepisce ed è contraria ai finanziamenti pubblici** (anche il 5 per mille)

La sua forza economica sono iscrizioni e contributi donati da chi la ritiene utile

DONA ORA (<http://www.aduc.it/info/sostienici.php>)