

# Meta risponde a Google e Intel e presenta il suo nuovo chip AI

## Il nuovo chip per l'AI di Meta

Presentato poche ore fa il nuovo modello di **chip per l'intelligenza artificiale (AI)** di **Meta**, che secondo i suoi ingegneri sarà tre volte più efficiente del primo.

Si tratta della nuova versione del ***Meta training and inference accelerator*** (Mtia) *V1*, dedicata specificatamente allo sviluppo e il potenziamento dell'AI generativa dell'azienda.

Un lavoro fatto in casa, sia a livello software, sia hardware, che consente a Meta di controllare l'intero flusso *stack*, che va dal frontend (l'interfaccia utente e le interazioni con l'applicazione) al backend (l'esecuzione dell'applicazione e la comunicazione con eventuali altri app, compresa l'elaborazione dati).

*"Poiché controlliamo l'intero stack, possiamo ottenere una maggiore efficienza rispetto alle GPU disponibili in commercio"*, hanno scritto **Eran Tal, Nicolaas Viljoen, Joel Coburn e Roman Levenstein** nell'articolo di lancio della nuova versione dell'Mtia sul blog di Meta.

## Un lavoro a lungo termine

Il nuovo modello di chip però al momento non è ancora utilizzato per lo sviluppo diretto dell'AI proprietaria, più che altro è impiegato nel potenziamento degli algoritmi di

classificazione e raccomandazione.

Il prossimo passaggio, secondo quanto affermato dalla società nel testo di presentazione, sarà dedicato proprio al rafforzamento della sua AI, come fa ad esempio Google, con il suo chip AI TPU v5p che addestra i modelli di grandi dimensioni e la famiglia di chatbot Gemini 1,5 Pro.

*“Meta sarà una parte importante della nostra tabella di marcia a lungo termine per costruire e scalare l’infrastruttura più potente ed efficiente possibile per i carichi di lavoro IA unici di Meta”,* hanno affermato gli ingegneri di Meta.

## La corsa all’AI di Meta, Google, Intel, Microsoft e Amazon

L’annuncio del gigante tecnologico di San Francisco segue di un giorno quello dei rivali Google e Intel.

**Google** ha rivelato i piani per un nuovo processore basato su tecnologia Arm, che punta su consumi energetici più bassi. Si chiama **Axion** e offre prestazioni migliori del 30% rispetto agli altri chip con architettura Arm. Sarà disponibile per i servizi cloud che le aziende possono noleggiare e utilizzare, dagli annunci su YouTube all’analisi dei big data.

**Intel** ha lanciato una nuova versione del suo chip acceleratore di intelligenza artificiale. Si chiama **Gaudi 3** e promette prestazioni di calcolo doppie. L’azienda californiana punta a diventare un’alternativa a Nvidia che nel 2023 ha controllato l’83% del mercato dei chip per data center e che ha segnato una ultima trimestrale record.

Tutti questi annunci arrivano dopo che **Microsoft** nel novembre del 2023 ha rivelato i propri microprocessori personalizzati progettati per la sua infrastruttura cloud e per addestrare modelli linguistici di grandi dimensioni: l’Azure Maia AI Accelerator e l’Azure Cobalt CPU. Anche **Amazon**

nello stesso mese ha presentato il nuovo processore su tecnologia Arm **AWS Graviton4** e l'acceleratore **AWS Trainium2**.

L'obiettivo delle Big Tech è comunque sempre lo stesso, cioè ridurre la propria dipendenza da partner come Intel e **Nvidia**, competendo sui chip personalizzati che riescono a smaltire grandi carichi di lavoro sull'IA e il cloud.

Meta, Microsoft, Alphabet, Apple, Nvidia e Amazon hanno effettuato l'acquisto di 88 tra startup e imprese di AI dal 2010 ad oggi.

<https://www.key4biz.it/meta-risponde-a-google-e-intel-e-presenta-il-suo-nuovo-chip-ai/486395/>