NUMERO 1 IN ITALIA SU ELETTRONICA EMBEDDED E SEMICONDUTTORI

Firmware 2.0

#53
Maggio 2025

Artificial Intelligence Edge AI

IN QUESTO NUMERO:

CONOSCIAMO WHISPER DI OPENAI

CORSO DI ELETTRONICA PER RAGAZZI - PUNTATA 27

HARDWARE E SOFTWARE EMBEDDED SPINGONO L'INTELLIGENZA ARTIFICIALE VERSO L'EDGE

E MOLTI ALTRI ARTICOLI E PROGETTI!

it.emcelettronica.com

Elettronica Open Source

PROFESSIONAL & MAKER ZONE

Firmware 2.0 EDITORIAL PLAN 2025

COSA LEGGERAI NEL 2025?

TOPICS	MAKERS ZONE	DATA DI PUBBLICAZIONE
PCB Design	Power Management	1 Febbraio
Embedded	Microcontrollers	1 Marzo
Automotive	Sensors	1 Aprile
Artificial Intelligence	Edge AI	1 Maggio
Raspberry Pi	Wearable Projects	1 Giugno
Wireless/RF	Retrogaming	1 Luglio
Arduino	Open Source Projects	1 Settembre
IoT	Smart Monitoring	1 Ottobre
Industry 4.0	Automation Projects	1 Novembre
Test&Measurements	Connectors	1 Dicembre

Dal Cloud all'Edge: l'evoluzione intelligente dell'elettronica

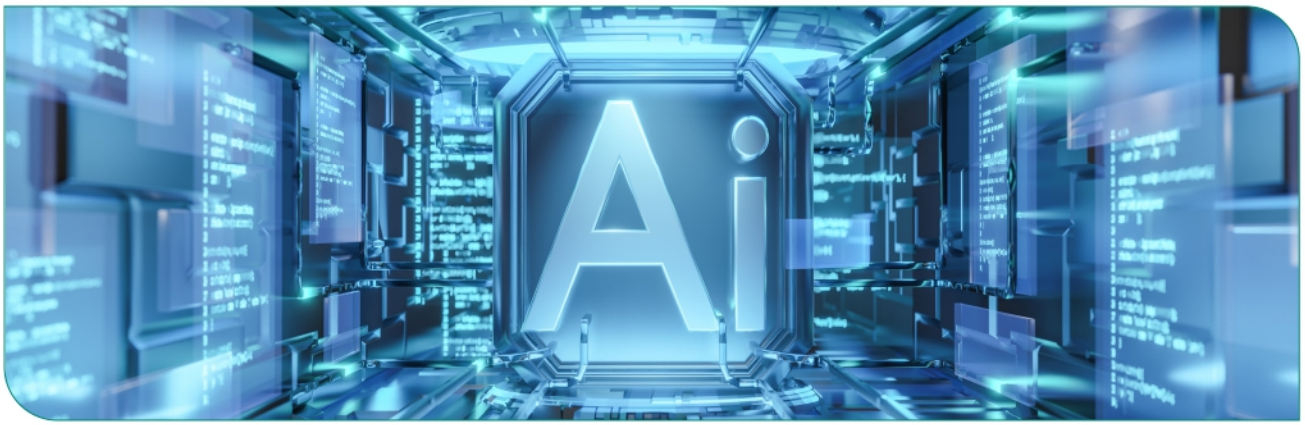
Cari lettori,

nel vasto panorama tecnologico contemporaneo, ***l'Intelligenza Artificiale non è più soltanto materia da laboratori accademici o grandi data center cloud****. Oggi, l'AI si avvicina sempre di più al punto in cui vengono generati i dati: il bordo della rete, o più semplicemente, l'Edge. È proprio da questa trasformazione che nasce il tema del nuovo numero di Firmware 2.0:* ***"Artificial Intelligence - Edge AI"****. L'Edge AI rappresenta un cambio di paradigma: se fino a poco tempo fa era necessario inviare dati a server remoti per ottenere analisi complesse o previsioni intelligenti, ora l'elaborazione può avvenire direttamente sul dispositivo. Che si tratti di una telecamera di sicurezza, di un sensore industriale o di uno smartphone, la* ***decentralizzazione*** *apre le porte ad una nuova era dell'elettronica, più intelligente, più efficiente e più autonoma. I vantaggi sono numerosi. Il primo è la riduzione della latenza, dal momento che elaborare dati in loco implica risposte più rapide ed essenziali per applicazioni in tempo reale come la guida autonoma o l'automazione industriale. C'è poi un miglioramento della privacy e della sicurezza, poiché i dati sensibili non devono lasciare il dispositivo. Inoltre, si riducono i costi legati alla trasmissione e allo storage nel cloud. L'evoluzione non sarebbe possibile senza i recenti progressi sia nell'hardware che nel software. Sul fronte hardware, assistiamo alla nascita di* ***chip specializzati****, progettati specificamente per eseguire modelli AI in ambienti a bassa potenza, basti pensare ai* ***microcontrollori con acceleratori neurali integrati****, come quelli basati su architettura ARM o RISC-V. Allo stesso tempo,* ***framework*** *come TensorFlow Lite, PyTorch Mobile o TinyML* ***stanno rendendo più semplice l'ottimizzazione e la distribuzione di modelli compatti****. Nel mondo industriale, l'Edge AI sta già trasformando la manutenzione predittiva, la visione artificiale, il controllo qualità e l'analisi energetica. Le macchine raccolgono dati e, allo stesso tempo, imparano da essi, si adattano, ottimizzano i processi e rilevano anomalie in tempo reale. Si tratta di un passo decisivo verso la fabbrica intelligente e la produzione sostenibile. Parallelamente, la diffusione dell'AI all'Edge richiede nuovi obiettivi per gli sviluppatori di firmware e sistemi embedded.* Come bilanciare le prestazioni con il consumo energetico? Come garantire l'affidabilità e la sicurezza di sistemi sempre più complessi? E come mantenere aggiornati i modelli AI una volta distribuiti sul campo? *Sono quesiti aperti che rendono questa fase di transizione tanto entusiasmante quanto impegnativa. In questo numero esploreremo da vicino le tecnologie, i casi d'uso e le strategie di implementazione che stanno rendendo possibile l'Edge AI. Vi presenteremo strumenti pratici e vi guideremo alla scoperta delle soluzioni che stanno modellando il futuro dell'elettronica intelligente.*

Buona lettura!

Giordana Francesca Brescia

Artificial Intelligence Edge AI



Founder&Editor
Emanuele Bonanni

CFO
Lidia Balica

Editorial Assistant
Maria Pisani

Maker in Chief
Giordana Francesca Brescia

Advertising & Marketing
Cristian Balica
cristian@contangosl.com

Graphic Designer
Marilde Mirra

Circulation
Users - 148.584
Social Network - 131.835

Espandi le tue competenze

Consigli utili, strumenti e articoli per i professionisti degli acquisti

resources.mouser.com/libreria-di-risorse-per-gli-acquisti

MOUSER ELECTRONICS

CONOSCIAMO WHISPER DI OPENAI

di **Andrea Garrapa**

Whisper di OpenAI è un sistema open source per il riconoscimento vocale automatico (ASR) progettato per trascrivere la lingua parlata in testo scritto, sfruttando tecniche di Deep Learning. Rilasciata nel settembre 2022, questa rete neurale è presto diventata uno strumento leggendario nell'elaborazione del linguaggio naturale, offrendo precisione e versatilità senza pari e dando origine a numerose applicazioni open source e commerciali. In questo articolo, faremo una panoramica completa sulle possibilità offerte da Whisper ASR.

INTRODUZIONE

Whisper di OpenAI, azienda nota per lo sviluppo di **ChatGPT**, è un modello AI/ML, in particolare un modello ASR (*Automatic Speech Recognition*). Più precisamente, Whisper è un nome generico per diversi modelli di diverse dimensioni, che vanno da 39 milioni a 1,55 miliardi di parametri, con i modelli "più grandi" che offrono una migliore precisione a scapito di tempi di elaborazione più lunghi e costi computazionali più elevati.

Lo scopo principale di Whisper è **trascrivere il parlato in testo**. Può anche tradurre il parlato da una qualsiasi delle lingue supportate in testo inglese. Oltre a queste capacità fondamentali, Whisper può essere ottimizzato e messo a punto per compiti specifici, ad esempio, per eseguire funzioni aggiuntive come la trascrizione in live streaming. Il modello può anche essere perfezionato per riconoscere e trascrivere nuove lingue, dialetti e accenti, e può essere reso più sensibile a domini specifici per riconoscere il gergo ed i termini tecnici del settore. La flessibilità consente agli sviluppatori di adattare Whisper ai loro casi d'uso specifici.

I NUMERI DI WHISPER

Whisper è addestrato su un vasto set di dati supervisionati pari a circa 680.000 ore, rendendolo uno dei sistemi ASR più completi disponibili. Il set di dati, proveniente da Internet e da risorse accademiche, comprende un'ampia varietà di ambiti e condizioni acustiche, garantendo che Whisper possa trascrivere accuratamente il parlato in diversi scenari del mondo reale. Inoltre, 117.000 ore di questi dati di pre-formazione riguardano parlato multilingue, permettendo dei *checkpoint* (set di parametri nelle varie fasi dell'addestramento) che possono essere applicati a 99 lingue, molte delle quali sono considerate con *scarse risorse* (lingue con risorse digitali insufficienti).

La vastità dei dati di training contribuisce alla capacità di Whisper di *generalizzare* (essere accurato con dati mai visti) e di funzionare in modo efficace in varie applicazioni. Essendo un modello pre-addestrato direttamente sul compito supervisionato del riconoscimento vocale, **il suo livello medio di precisione è superiore alla maggior parte degli altri modelli open source**.

Detto questo, data la natura generalista del suo set di dati di addestramento iniziale, il modello è matematicamente più sbilanciato verso frasi che non hanno nulla a che fare con i dati audio professionali, il che significa che normalmente richiederebbe almeno qualche messa a punto per produrre risultati costantemente accurati in ambienti business.

Whisper si distingue come il miglior sistema ASR della categoria grazie alla sua eccezionale precisione di base e alle prestazioni nella gestione di lingue diverse. La sua adattabilità a condizioni acustiche difficili, ad esempio audio rumoroso e multilingue, lo distingue dagli altri sistemi di riconoscimento vocale. Secondo la Open ASR Leaderboard, il tasso medio di errore delle parole è dell'8,06%, ovvero è accurato al 92% per impostazione predefinita.

Ci sono cinque dimensioni di modello, quattro con versioni solo in inglese, che offrono compromessi di velocità e accuratezza. Di seguito, in **Tabella 1** sono riportati i nomi dei modelli disponibili ed i loro requisiti di memoria approssimativi e la velocità di inferenza relativa al modello di grandi dimensioni; la velocità effettiva può variare a seconda di molti fattori, tra cui l'hardware disponibile.

Whisper consente, inoltre, agli sviluppatori di bilanciare costi computazionali, velocità e precisione, rendendolo estremamente versatile e utile in una vasta

Dimensione	Parametri	VRAM richiesta	Velocità relativa
minuscola (en)	39 M	1 GB	32x
base (en)	74 M	1 GB	16x
piccola (en)	244 M	2 GB	6x
media (en)	769 M	5 GB	2x
larga	1550 M	10 GB	1x

Tabella 1: Modelli disponibili in Whisper

gamma di applicazioni. La velocità media della trascrizione di Whisper varia da 8 a 30 minuti, a seconda del tipo di audio, utilizzando una GPU. Richiede due volte più tempo se la trascrizione viene eseguita solo su CPU.

IL FUNZIONAMENTO DI WHISPER

Whisper è un modello di Deep Learning end-to-end basato su un'architettura **Transformer** codificatore-decodificatore. I modelli Transformer si distinguono per la loro capacità di tenere traccia di come più parole e frasi si relazionano tra loro, consentendo di tenere conto delle dipendenze a lungo termine. In altre parole, i Transformer possono "ricordare" ciò che è stato detto in precedenza per contestualizzare le parole, il che aiuta ad aumentare la precisione della trascrizione.

Nel caso specifico di Whisper, trascrive il parlato in un

a livello di frase, la trascrizione vocale multilingue e la traduzione vocale in inglese. L'architettura del trasformatore pre-addestrato di Whisper consente al modello di dedurre il contesto più ampio delle frasi trascritte e di "riempire" le lacune nella trascrizione in base a questa comprensione. In questo senso, **si può dire che Whisper ASR sfrutti le tecniche di Intelligenza Artificiale generativa per convertire il linguaggio parlato in testo scritto**.

OpenAI ha reso disponibile il modello *large-v2* che offre prestazioni più veloci rispetto al modello open source e ha un prezzo di 0,006 $/minuto di trascrizione. Esistono anche API basate su Whisper che si basano su un'architettura ibrida e migliorata di Whisper per offrire un insieme più esteso di capacità e caratteristiche rispetto all'API OpenAI ufficiale.

CORSO DI ELETTRONICA PER RAGAZZI – PUNTATA 27

di Fulvio De Santis

In questa puntata continueremo la trattazione sui diodi in cui spiegheremo come progettare un circuito elettronico con diodi, utilizzati come regolatori di tensione in corrente continua. Il progetto del regolatore sarà preceduto da una spiegazione teorica passo-passo su come definire le specifiche ed i calcoli di progetto, poi seguirà un esempio pratico del progetto con simulazione per la verifica dei risultati teorici.

IL REGOLATORE DI TENSIONE

Il regolatore di tensione in corrente continua è un dispositivo che deve garantire una tensione di uscita continua costante a fronte di variazioni della tensione che alimenta il dispositivo, della corrente assorbita dal carico applicato alla sua uscita, e della corrente assorbita dal dispositivo di regolazione stesso. Il diodo è un elemento circuitale che quando è polarizzato direttamente può essere utilizzato per realizzare un semplice regolatore di tensione sfruttando la sua caduta di tensione fissa, che in un comune diodo è di circa 0,7 V. Nella seguente esercitazione teorica, analizzeremo il progetto di un circuito regolatore di tensione realizzato con quattro diodi in serie. La **Figura 1** mostra lo schema elettrico del regolatore di tensione in cui l'elemento che determina la regolazione è costituito dalla serie di quattro diodi 1N4148.

Con questo circuito vogliamo realizzare un **regolatore di tensione** in corrente continua che fornisca una tensione continua di 2,8 V, ad esempio, per alimentare un circuito integrato che funziona con una tensione di alimentazione di 2,8 V. Desideriamo ottenere una tensione costante di 2,8 V supponendo che: (a) la sorgente esterna di alimentazione del regolatore di 12 V non sia proprio stabile ma possa variare di ±10% rispetto al suo valore nominale; (b) all'uscita del regolatore sia applicato un resistore di carico; (c) la caduta di tensione su ogni diodo sia esattamente pari a 0,7 V; (d) in corrispondenza della tensione diretta di 0,7 V in ogni diodo 1N4148 deve scorrere una corrente di 5 mA (0,005 A) alla temperatura ambiente di 25 °C, come si può rilevare dalla tabella estratta dal datasheet del diodo riportata in **Figura 2**.

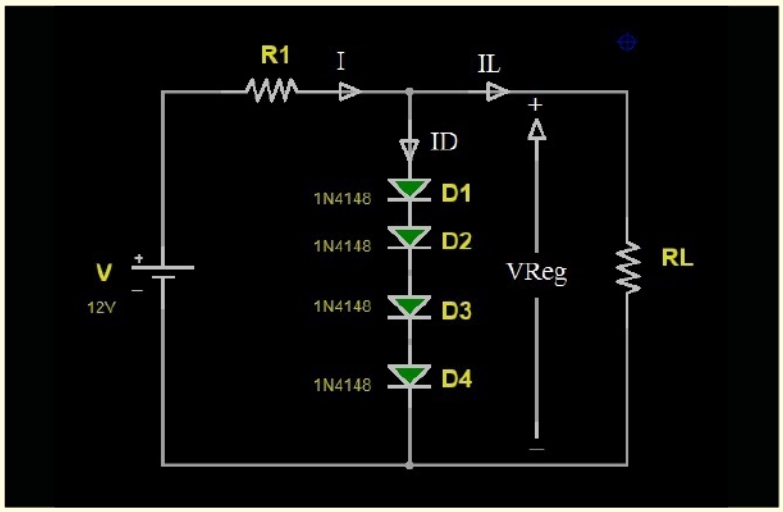


Figura 1: Schema elettrico del regolatore di tensione

Nella prima fase del progetto calcoleremo la variazione della tensione di uscita del regolatore in funzione della variazione della tensione d'ingresso considerando disconnesso il carico.

CALCOLO DELLA VARIAZIONE DELLA TENSIONE DI USCITA DEL REGOLATORE IN FUNZIONE DELLA VARIAZIONE DELLA TENSIONE D'INGRESSO

Innanzitutto, calcoliamo il valore di resistenza del resistore R1 applicando la legge delle tensioni di Kirchhoff (LVK) alla maglia V-R1-VReg:

V-R1*I-VReg=0

Utilizzando questa equazione possiamo ricavare il valore di resistenza del resistore R1. Quindi, isoliamo il termine contenente R1 a sinistra del simbolo = e portiamo a destra gli altri due termini, poi cambiamo segno a tutti i termini (ricordiamo che ciò corrisponde a moltiplicare per -1 tutti i termini):

R1*I=V-VReg

R1=(V-VReg)/I

Sostituiamo i simboli con i valori numerici:

1N4148; 1N4448 High-spe...

ELECTRICAL CHARACTERISTICS

T_j = 25 °C unless otherwise specified.

SYMBOL	PARAMETER	CONDITIONS	MIN.	MAX.	UNIT
V_F	forward voltage	see Fig.3			
	1N4148	I_F = 10 mA	–	1	V
	1N4448	I_F = 5 mA	0.62	0.72	V
		I_F = 100 mA	–	1	V
I_R	reverse current	V_R = 20 V; see Fig.5		25	nA
		V_R = 20 V; T_j = 150 °C; see Fig.5	–	50	µA
I_R	reverse current; 1N4448	V_R = 20 V; T_j = 100 °C; see Fig.5	–	3	µA
C_d	diode capacitance	f = 1 MHz; V_R = 0 V; see Fig.6	–	4	pF

Figura 2: Corrente diretta in funzione della tensione diretta del diodo 1N4148

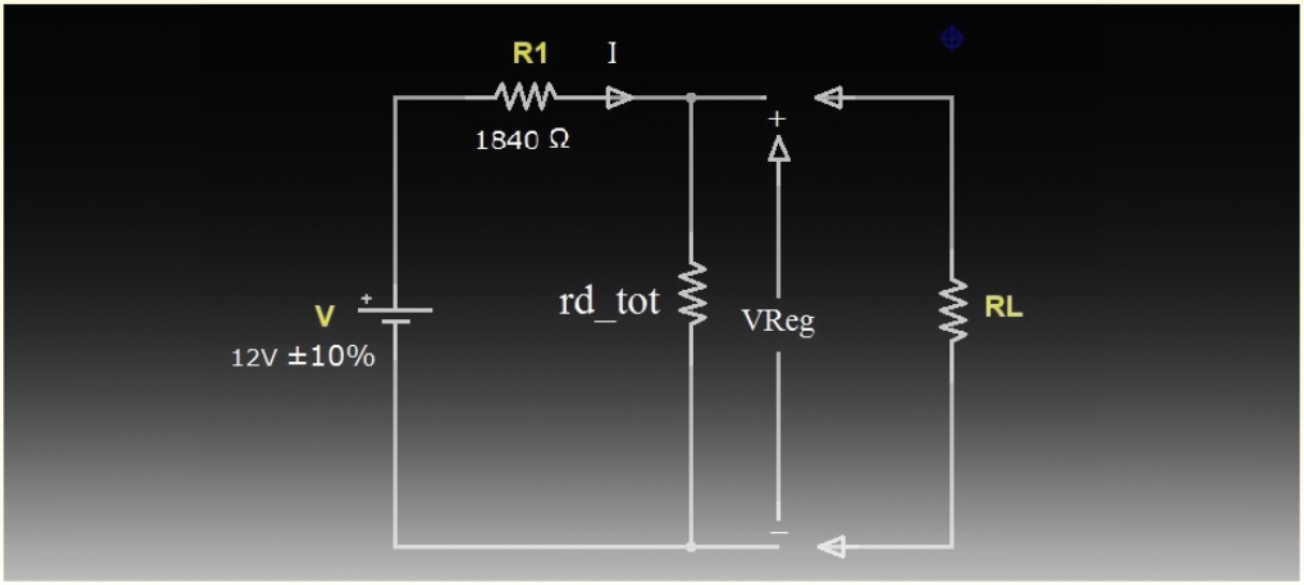


Figura 3: Circuito con il partitore R1-rd_tot

R1=(12-2,8)/0,005=1840 Ω

Nello schema di Figura 3 riportiamo il circuito in cui i

HARDWARE E SOFTWARE EMBEDDED SPINGONO L'INTELLIGENZA ARTIFICIALE VERSO L'EDGE

di **Fulvio De Santis**

Questo articolo analizza l'obiettivo ed i vantaggi dell'implementazione diretta di modelli e algoritmi di Intelligenza Artificiale (IA) ai confini della rete, e spiega come i progressi nell'hardware e nei software embedded stanno agevolando l'applicabilità dell'IA.

INTRODUZIONE

L'Intelligenza Artificiale si è trasformata da una tecnologia di nicchia in qualcosa con cui le persone interagiscono quotidianamente, espandendosi oltre i settori ingegneristici e tecnologici, una tendenza che ha portato le aziende di quasi tutti i settori a considerare come potrebbero sfruttare l'Intelligenza Artificiale per aumentare l'efficienza, ridurre i costi e aumentare le capacità dei loro prodotti. L'accessibilità e la facilità d'uso delle soluzioni IA basate sul cloud ampiamente disponibili, hanno reso più facile per quasi chiunque interagire con modelli e strumenti progettati per l'Intelligenza Artificiale. Tuttavia, non tutte le innovazioni legate all'Intelligenza Artificiale avvengono nel cloud. Con i progressi tecnologici nella progettazione dei dispositivi embedded, le capacità di elaborazione si stanno facendo strada nei prodotti di consumo come computer e cellulari, nonché in altri dispositivi elettronici alimentati a batteria, applicazioni come videocitofoni, elaborazione della visione nei sistemi automobilistici e motori per infrastrutture energetiche e sistemi industriali.

L'Intelligenza Artificiale edge, ovvero la capacità di eseguire modelli localmente, vicino alla fonte dei dati, sta migliorando la reattività, l'efficienza, l'affidabilità e la sicurezza dell'elettronica. I processori embedded che rendono possibile questa trasformazione dal cloud all'edge, integrano al loro interno componenti come core specializzati per l'elaborazione del segnale digitale (**DSP**), e sono supportati da strumenti basati su **GUI** (**Graphical User Interface**), facili da usare e che riducono al minimo il tempo e le competenze necessarie per portare l'Intelligenza Artificiale ai bordi.

Gli strumenti IA edge, in particolare, consentono l'implementazione di modelli IA/Machine Learning direttamente sui dispositivi periferici, come microcontrollori, microprocessori e sensori intelligenti, che svolgono una funzione chiave nel portare l'elaborazione all'edge, consentendo ai dispositivi di elaborare dati localmente senza necessariamente fare affidamento sui servizi cloud, con una migliore efficienza energetica e minori costi dei dispositivi stessi. Gli strumenti IA, locali o basati sul cloud, sono essenziali per creare modelli efficienti e ottimizzati che possono essere eseguiti localmente su dispositivi ottimizzati. Ai produttori ed ai progettisti dell'Intelligenza Artificiale o di dispositivi embedded, questi strumenti consentono di creare soluzioni intelligenti che migliorano prodotti e servizi. La **Figura 1** mostra un dispositivo edge, un minicomputer Android IIoT.

L'Intelligenza Artificiale è la capacità di una macchina di esibire una sorta di intelligenza o ragionamento. Quando oggi la maggior parte delle persone pensa all'Intelligenza Artificiale, spesso immagina generatori di testo e immagini, o avversari virtuali nei videogiochi pronti a sfidare le capacità umane. Ma anche il più semplice degli algoritmi è tecnicamente un esempio di IA in senso letterale. L'ampiezza dell'IA ed i suoi molteplici casi d'uso hanno portato a diversi **sottodomini**, tra cui i più noti sono l'apprendimento automatico e il Deep Learning.

La maggior parte dell'Intelligenza Artificiale utilizzata per le applicazioni embedded è l'apprendimento automatico, il sottodominio in cui macchine e algoritmi "imparano" a risolvere un problema; ad esempio, un veicolo che riconosce un pedone rispetto ad un ostacolo analizzando i dati delle immagini per individuare schemi comuni. Un modello di apprendimento automatico può anche apprendere ricevendo grandi quantità di dati di addestramento o dati già etichettati. Il processo di addestramento consente ai modelli di apprendimento automatico di distinguere schemi nei dati, che possono utilizzare per fare inferenze future.

Figura 1: Dispositivo edge Android IIoT

Nel campo dell'apprendimento automatico, il **Deep Learning** è diventato una delle sue implementazioni più popolari, data la sua capacità di risolvere problemi altamente complessi in modo accurato, sebbene ciò richieda molte risorse informatiche. Il Deep Learning utilizza reti neurali multistrato, che sono modelli di dati ispirati ai neuroni nel cervello umano. Consente ai progettisti di

getica dei processori embedded.

L'Intelligenza Artificiale **edge** e l'Intelligenza Artificiale **cloud** differiscono per quanto riguarda il modo in cui ricevono ed elaborano i dati e interagiscono con le risorse basate su cloud. Nell'IA implementata su dispositivi periferici, l'inferenza si concretizza mediante l'elabora-

ABBONATI A

Firmware 2.0

PER AVERE **TUTTA L'ELETTRONICA A PORTATA DI CLICK** E RESTARE SEMPRE **AGGIORNATO SULL'ELETTRONICA EMBEDDED, I MICROCONTROLLORI** E **L'INNOVAZIONE TECNOLOGICA**

Elettronica Open Source

+ 145.000
REGISTERED USERS

7.414 AVERAGE DAILY PAGEVIEWS (FEB2020)

830.610 2020 ANNUAL VISITORS

THE BIGGEST **EMBEDDED COMMUNITY** IN ITALY

SOCIAL CONNECTIONS

+ 83.000

+ 23.000

CATEGORIES

PROFESSIONALS
53 %

ACADEMICS/STUDENTS
25 %

MAKERS/HOBBYISTS
22 %

Elettronica Open Source