Imparare con gli strumenti di intelligenza artificiale multimodale

Introduzione ai concetti di base dell'intelligenza artificiale

Autore: Kacsur Annamária, ITStudy

Revisore: Hartyányi Mária, ITSTudy

M1-T1-MLC6

Obiettivo di apprendimento

Le più recenti tecnologie di intelligenza artificiale hanno superato i limiti della comunicazione basata soltanto sul testo. I modelli linguistici di grandi dimensioni, come ChatGPT, Gemini e Claude, inizialmente sviluppati per generare testi, oggi possono anche creare immagini, video e audio di qualità sempre più elevata. Sono inoltre in grado di elaborare informazioni digitali provenienti da fonti esterne, ad esempio dai sensori. I sistemi di IA che possono lavorare con diversi tipi di dati vengono chiamati **intelligenza artificiale multimodale**.

Questo materiale didattico presenta le opportunità offerte dall'IA multimodale nell'insegnamento e nell'apprendimento, spiegando in che modo si differenzia dall'utilizzo dei tradizionali materiali didattici multimediali, diffusi ormai da diversi decenni.

Come impariamo?

Gli scienziati studiano da molto tempo come funziona l'apprendimento, ma ancora oggi conosciamo solo in parte i processi attraverso cui i bambini piccoli imparano e trasformano le esperienze quotidiane in conoscenze. Quando un bambino gioca con una palla, ad esempio, utilizza contemporaneamente la vista, il tatto, il movimento e le parole associate all'oggetto. La comprensione non nasce semplicemente dall'insieme di queste esperienze, ma dalla capacità di riconoscere i collegamenti tra di esse.



Come imparano i bambini piccoli? (Fonte: ChatGPT)

L'apprendimento avviene in modo simile anche nei ragazzi più grandi e negli adulti. Non percepiamo il mondo attraverso "canali" separati: vediamo, ascoltiamo, agiamo, parliamo e leggiamo, e interpretiamo queste esperienze collegandole tra loro e con le nostre conoscenze pregresse. Ciò è coerente con la teoria costruttivista dell'apprendimento, secondo cui chi apprende non si limita a ricevere le informazioni che gli vengono trasmesse, ma le interpreta alla luce delle proprie esperienze e conoscenze pregresse, stabilisce collegamenti e dà loro un senso in modo autonomo.

Negli anni Novanta i ricercatori hanno approfondito sempre di più il modo in cui le diverse forme di espressione e di rappresentazione favoriscono la comprensione. Da qui è nata la teoria della **multimodalità**[1] ed è proprio in questo periodo che nell'istruzione sono comparsi per la prima volta i concetti di **apprendimento multimodale** e di **pedagogia multimodale**, che sottolineano come la conoscenza possa essere costruita anche riconoscendo e sviluppando collegamenti tra rappresentazioni diverse.

Dalla multimedialità alla multimodalità

Il linguaggio è stato a lungo il principale mezzo dell'insegnamento, in particolare il testo scritto e la parola: il docente spiega, mentre gli studenti leggono, prendono appunti e poi dimostrano le proprie conoscenze oralmente o per iscritto.



Didattica tradizionale: il docente spiega; gli studenti leggono, prendono appunti e poi riferiscono oralmente o per iscritto (Fonte: ChatGPT)

Con lo sviluppo delle tecnologie digitali, la **multimedialità** – l'uso combinato di testo, immagini, audio, video e animazioni – è diventata sempre più accessibile. Il termine "**modo**" o "**modalità**" indica la forma in cui un'informazione viene presentata o il tipo di dati che entra in un sistema. Ne sono esempi il testo, le immagini, l'audio, il video e i dati provenienti da sensori esterni, come la temperatura media giornaliera registrata da una stazione meteorologica.

[1] Nella parola multimodalità, "multi" è un prefisso di origine latina che significa molti o molteplici; "modalità" deriva dal latino modus (modo) e in questo contesto indica un modo di rappresentazione, di comunicazione o di espressione.

I sistemi digitali possono lavorare con:

- testo;
- immagini e grafica;
- parlato, musica e altri suoni;
- video;
- dati che descrivono movimenti e gesti;
- dati dei sensori, come misurazioni di posizione, temperatura o velocità.

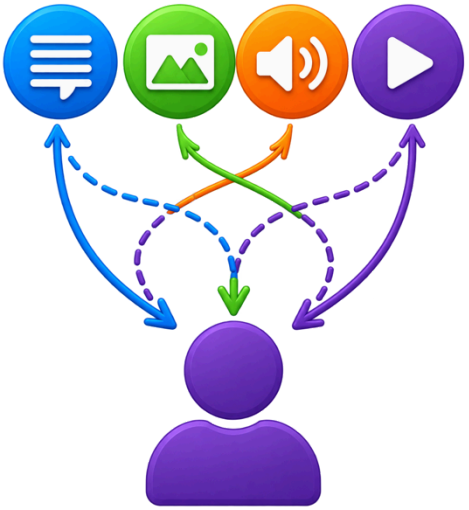
Ciò non significa che una modalità coincida con un formato di file. JPEG e PNG, ad esempio, sono formati di immagine diversi, ma appartengono entrambi alla stessa **modalità visiva**.



La multimedialità nella didattica: più media usati insieme – le informazioni sono presentate una accanto all'altra.

L'esperienza degli ultimi decenni ha dimostrato che i CD didattici interattivi e le presentazioni colorate arricchite da immagini, animazioni e suoni favoriscono l'apprendimento, ma non producono un miglioramento qualitativo della comprensione.

Ancora oggi la maggior parte delle lezioni si concentra soprattutto sul testo e sulla parola (libri di testo, spiegazioni frontali alla classe e quaderni).

	
Presentazione su più canali	**Costruire significati collegando rappresentazioni diverse**
MULTIMEDIALITÀ	MULTIMODALITÀ

Se gli studenti continuano soprattutto a ricevere informazioni già pronte e il collegamento tra i significati veicolati dai diversi media in una spiegazione dipende interamente dall'iniziativa del docente, gli elementi multimediali trasmettono lo stesso messaggio senza contribuire alla comprensione.

In un approccio didattico multimodale, gli studenti non si limitano a ricevere informazioni in più forme, ma confrontano, collegano, trasformano e creano essi stessi rappresentazioni diverse. L'accento si sposta così dalla trasmissione di informazioni tramite la multimedialità alla costruzione multimodale di significati da parte degli studenti.



Costruzione multimodale della conoscenza (Fonte: ChatGPT)

La formula di una funzione, il suo grafico e una storia che descrive un movimento reale non sono semplicemente tre “formati” diversi dello stesso contenuto didattico. Gli studenti devono riconoscere come i cambiamenti descritti nella storia compaiono nel grafico e come possono essere espressi attraverso relazioni matematiche. La comprensione si sviluppa man mano che gli studenti riconoscono i collegamenti tra le rappresentazioni.

Intelligenza artificiale multimodale

L’intelligenza artificiale multimodale accetta in input diversi tipi di dati (testo, immagini, audio e video) e, su richiesta, genera output di tipo diverso.



Intelligenza artificiale multimodale (Fonte: ChatGPT)

L’**intelligenza artificiale multimodale**[2] è in grado di elaborare e interpretare più tipi di dati – testo, immagini, audio, video e altro ancora – all’interno di un unico modello. Il termine “multimodale” si riferisce alla varietà dei dati di input e di output che il modello può gestire.

Non si limita a combinare formati multimediali diversi, ma “impara” anche le relazioni tra di essi

[2] **Codifica delle modalità**: ogni tipo di dato in input viene prima elaborato da un codificatore (encoder) specializzato. Il codificatore delle immagini converte gli elementi dell’immagine (pixel) in dati numerici, il codificatore linguistico trasforma il testo in token, mentre il codificatore audio estrae le caratteristiche importanti dal parlato o dal suono. Questi codificatori colgono le caratteristiche essenziali di ciascuna modalità, filtrando i dettagli irrilevanti.

Integrazione multimodale (fusione): il modello combina le informazioni provenienti da più modalità per formare una rappresentazione unificata dei dati di input. I modelli di base (foundation model) multimodali realizzano di solito questa integrazione attraverso strati transformer condivisi e meccanismi di attenzione, che permettono alle informazioni di una modalità di influenzare l’interpretazione di un’altra. Ad esempio, interpretando l’espressione del viso di una persona insieme al suo tono di voce e alle parole pronunciate, il modello può dedurne le emozioni o le intenzioni.

Ambiti di applicazione

Sanità

L'intelligenza artificiale può analizzare immagini mediche, come radiografie o risonanze magnetiche, insieme alle informazioni scritte contenute nella documentazione clinica del paziente. Esaminando queste fonti in modo combinato, può individuare elementi utili e contribuire a formulare possibili ipotesi diagnostiche, che devono comunque essere valutate dal personale sanitario.



IA multimodale in ambito sanitario (Fonte: ChatGPT)

Assistenza clienti e accessibilità

L'IA multimodale consente agli utenti con disabilità visive o motorie di porre domande a voce e di ricevere le risposte nella modalità più adatta a loro.

Creazione di contenuti

Modelli come ChatGPT e Gemini utilizzano input combinati di testo e immagini per modificare un'immagine o creare nuovi contenuti creativi.

Robotica e sistemi autonomi

I veicoli a guida autonoma e i robot possono utilizzare l'IA multimodale per combinare le immagini delle telecamere con le informazioni raccolte dai sensori. In questo modo riescono a interpretare meglio l'ambiente circostante e a prendere decisioni per muoversi e svolgere le proprie attività.

Analisi di documenti

Quando riceve un documento che contiene grafici, diagrammi e immagini incorporate, l'intelligenza artificiale multimodale elabora tutti gli elementi visivi insieme al testo e risponde alle domande interpretandoli congiuntamente.

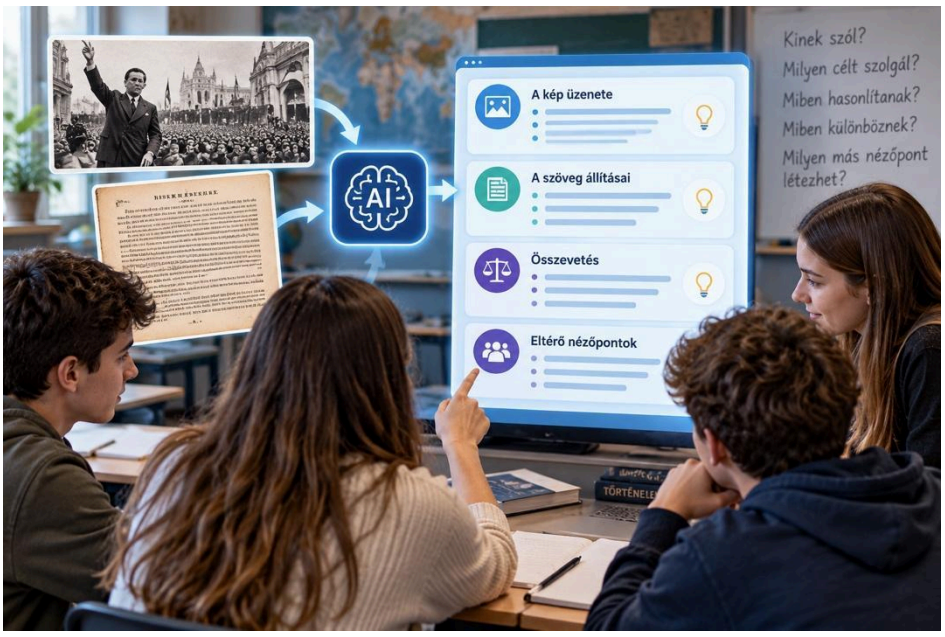
Esempi dall'istruzione e formazione professionale

Durante un progetto ambientale, gli studenti possono caricare in uno strumento di IA un video che mostra le condizioni di un fiume e un documento con i risultati delle misurazioni effettuate. L'IA può aiutarli a collegare ciò che osservano nel video con i dati raccolti e a individuare possibili conclusioni attraverso l'analisi delle due fonti.



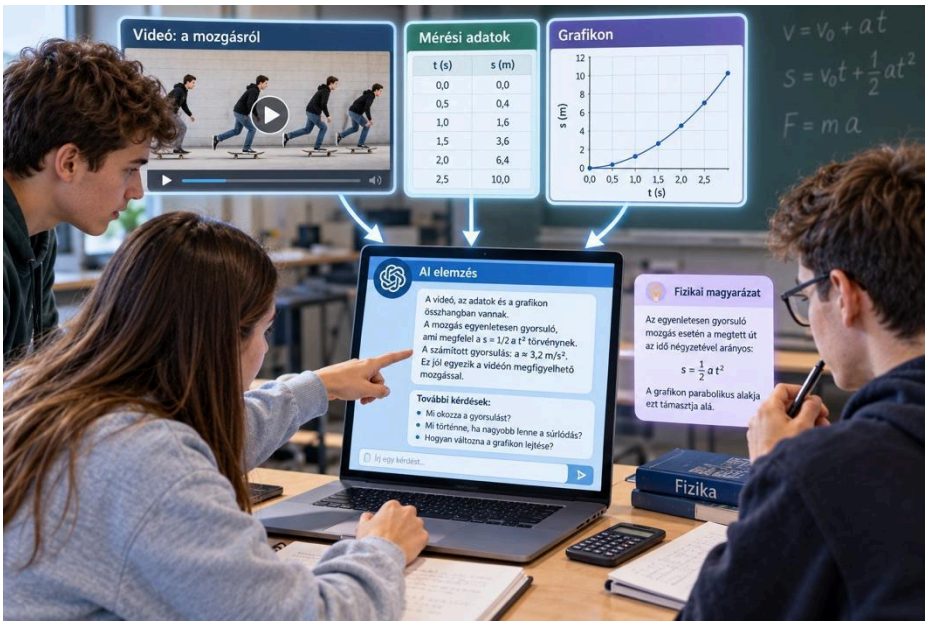
IA multimodale in un progetto ambientale (Fonte: ChatGPT)

In una lezione di storia, gli studenti possono chiedere all'IA di rispondere alle loro domande analizzando insieme una fotografia, un manifesto e una fonte storica scritta dell'epoca.



IA multimodale in una lezione di storia (Fonte: ChatGPT)

In fisica, gli studenti possono esaminare il video di un movimento insieme alle relative misurazioni e al grafico: il grafico corrisponde a ciò che vediamo nella registrazione? E come si può spiegare la relazione con le leggi della fisica?



IA multimodale in una lezione di fisica (Fonte: ChatGPT)

Gli studenti dell'istruzione e formazione professionale caricano una fotografia o un video di un componente guasto insieme alla relativa documentazione tecnica e ai dati di misurazione. Analizzando congiuntamente le fonti, l'IA può rispondere alle domande degli studenti sulle possibili cause del guasto.



IA multimodale in classe (Fonte: ChatGPT)

Indicazioni per i docenti

Le applicazioni di intelligenza artificiale di oggi consentono ai docenti di creare in modo rapido e semplice immagini, video, presentazioni e infografiche di grande impatto. Sviluppare materiali didattici diventa più facile, la loro rapida obsolescenza diventa un problema minore e adattarli alle esigenze individuali degli studenti può diventare realtà.

L'intelligenza artificiale multimodale può produrre un cambiamento più profondo se incoraggiamo gli studenti a confrontare contenuti creati da loro stessi o generati con l'IA, a farli analizzare, a chiedere un riscontro sulle proprie idee e a chiedere le motivazioni delle risposte ricevute.

L'IA multimodale apre nuove possibilità per l'apprendimento attivo. Con la guida del docente, può aiutare gli studenti a diventare partecipanti costruttivi del proprio apprendimento. Il ruolo del docente è progettare e creare un ambiente di apprendimento in cui l'IA multimodale stimoli gli studenti a pensare, interpretare, confrontare e costruire i propri significati. Oggi queste tecnologie possono ridurre il predominio dell'elaborazione delle informazioni basata sul testo e avvicinare l'istruzione alla vera natura dell'apprendimento.

Bibliografia

- Simamkele Matuntuta (2026), How AI is Transforming Vocational Education and Training, articolo su cloudassess.com (ultima consultazione: 3/7/2026)
- Ayush Chauhan (2025): Multimodal AI: Beyond Text and Images, articolo su https://www.techaheadcorp.com,
- Kamalmeet Singh (2026): How Multimodal AI Works?, articolo di blog su https://medium.com, ultima consultazione: 8/7/2026
- Amol Yadhav (2024): Multimodal AI: The Next Frontier in Personalized Education, articolo di blog su https://medium.com/, ultima consultazione: 8/7/2026
- Peraschi Selvan Subramanian (2025): Multimodal AI: The future of integrated intelligence, articolo pubblicato su World Journal of Advanced Engineering Technology and Sciences, ultima consultazione: 9/7/2026

Link

- IA multimodale https://www.llamaindex.ai/glossary/multimodal-ai
- Come funzionano i modelli di IA multimodale? Un video su YouTube: https://www.youtube.com/watch?v=WkoytIA3MoQ

Copyright © 2026 Consorzio VITA

Quest'opera è concessa in licenza secondo i termini della Creative Commons Attribuzione-Condividi Allo Stesso Modo 4.0 Internazionale (CC BY-SA 4.0).



Finanziato dall'Unione europea. Le opinioni espresse sono esclusivamente quelle degli autori e non riflettono necessariamente quelle dell'Unione europea o dell'Agenzia nazionale.