

# Analisi quantitativa e percettiva di video creati con generatori AI

Federico Coscia - 977772

24 Marzo, 2025

## 1 Ente, contesto iniziale, obiettivi

Il tirocinio è stato svolto presso il Phuselab - DI Unimi, sotto la supervisione della docente Raffaella Lanzarotti, in collaborazione con il dipartimento di psicologia dell'Università Cattolica del Sacro Cuore, collaborando con i docenti di psicologia Andrea Gaggioli e Maurizio Mauri, e un team di loro tesisti, Matteo Scarinzi, Paolo Galimberti ed Emanuela Ardore. Il progetto affronta l'utilizzo di video generati tramite Intelligenza Artificiale (IA) nell'apprendimento, e nasce dall'interesse a dare una risposta scientifica all'applicabilità di tali sistemi di generazione video per la generazione di video-lezioni, come alternativa efficace alla tradizionale video-lezione con presentatore umano, comportando una riduzione non indifferente dei tempi e i costi di produzione di questo tipo di contenuti.

## 2 Lavoro svolto

Il lavoro è stato suddiviso nelle seguenti fasi: produzione di due video-lezioni (in forma reale e sintetica), stesura di un protocollo sperimentale, sviluppo di un interfaccia per l'acquisizione dei dati, selezione di un campione, con conseguente acquisizione dei dati, e analisi dei dati acquisiti. Per la produzione dei video, sono stati innanzitutto valutati tre servizi di IA generativa diversi, DupDub<sup>1</sup>, Synthesia.io<sup>2</sup> ed HeyGen<sup>3</sup>. Tra questi è stato scelto HeyGen, poiché

---

<sup>1</sup><https://www.dupdub.com>

<sup>2</sup><https://www.synthesia.io/>

<sup>3</sup><https://www.heygen.com>

valutato come il più avanzato tra i tre, in grado di generare un avatar a partire da un video reale di riferimento, ricreando l'aspetto, i gesti, i movimenti e il suono della voce del presentatore reale. Il tesista dell'Università Cattolica Matteo Scarinzi si è occupato della scrittura di due brevi testi educativi, e presso la sede in Buonarroti dell'Università Cattolica del Sacro Cuore, sono state registrate due video-lezioni, in collaborazione con i dottorandi in psicologia Michele Paleologo e Marta Pizzolante, i quali hanno esposto i testi prodotti. A partire dai video registrati, è stato utilizzato il servizio scelto per generare le controparti sintetiche.

È stato selezionato un campione di 20 partecipanti, selezionati tra studenti e lavoratori, i quali hanno preso visione delle due video-lezioni, una con presentatore reale, e l'altra con presentatore generato tramite IA, in ordine controbilanciato su tutto il campione, seguendo un disegno di tipo **within-subject**. Durante la visione dei video, sono stati acquisiti diversi segnali multimodali, quali il battito cardiaco, il livello di sudorazione della pelle, le espressioni facciali, il movimento dello sguardo sullo schermo, e a seguito di ogni video sono stati somministrati dei questionari di valutazione: valutazione dell'esperienza di apprendimento, valutazione del presentatore, valutazione del coinvolgimento con il video (Video Engagement Scale, VES), e autovalutazione emotiva (Positive and Negative Affect Schedule, PANAS), insieme a un breve quiz di tre domande di comprensione a risposta multipla sui contenuti proposti.<sup>4</sup>

Per facilitare la fase sperimentale, è stata sviluppata un'interfaccia ad-hoc in Python 3.9, la quale si occupa di selezionare le video-lezioni da mostrare, somministrare i questionari previsti dal protocollo sperimentale e coordinare la registrazione di tutti i segnali, acquisendo il volto dei partecipanti tramite webcam, predisponendo la calibrazione e gestendo la registrazione dei dati di eye-tracking, e salvando le risposte fornite ad ogni questionario. L'acquisizione dei segnali di battito cardiaco e sudorazione della pelle è stata effettuata con un braccialetto EmbracePlus di Empatica, gestito e controllato indipendentemente dallo sperimentatore tramite smartphone. Per la predisposizione dei dati all'analisi, sono stati prodotti una serie di script che si occupano del download dei segnali fisiologici dai server di Amazon AWS messi a disposizione da Empatica, la sincronizzazione dei dati acquisiti, l'estrazione di feature, e la suddivisione dei dati in due gruppi: i dati acquisiti durante la visione dei video reali, e i dati acquisiti durante la visione dei video generati tramite IA.

---

<sup>4</sup>I questionari sono stati selezionati/redatti dal docente Andrea Gaggioli.

### 3 Tecnologie coinvolte

Per svolgere il lavoro, è stato utilizzato un EyeLink 1000 di SR Research Std. per l'eye-tracking, messo a gentile disposizione dal docente Andrea Gaggioli, un braccialetto EmbracePlus di Empatica per l'acquisizione dei segnali fisiologici, di proprietà del PhuseLab, e gli applicativi prodotti per l'acquisizione e l'analisi dei dati sono stati sviluppati in **Python**. Le acquisizioni sono state svolte presso la sede di Buonarroti dell'Università Cattolica del Sacro Cuore.

### 4 Competenze e risultati raggiunti

I risultati mostrano che, a livello fisiologico, non emergono differenze significative tra i due tipi di video, suggerendo che l'utilizzo di queste tecnologie non alteri l'attivazione emotiva degli spettatori. Non sono state rilevate differenze nel comportamento dello sguardo, suggerendo che non vi siano forti elementi di distrazione nei video sintetici, come ad esempio artefatti o imprecisioni dovute alla natura sintetica degli stessi. Tuttavia, le lezioni con presentatore IA sono state riportate come significativamente meno chiare, coinvolgenti e stimolanti. Inoltre, l'analisi delle espressioni facciali evidenzia una maggiore presenza di micro-espressioni riconducibili all'espressione del disgusto, suggerendo un possibile effetto di percezione inconscia della natura sintetica del contenuto proposto. Nonostante questi risultati, non sono state misurate differenze significative nel grado di apprendimento, suggerendo che queste limitazioni non sono abbastanza severe da risultare un ostacolo concreto per l'apprendimento, almeno nel breve termine. Considerando anche la rapidità e il basso costo di produzione, l'uso di avatar IA appare vantaggioso in contesti di formazione aziendale o corsi brevi online, mentre potrebbe essere meno adatto per percorsi accademici di lunga durata, dove il coinvolgimento e la interazione con il docente sono fondamentali. Il progetto ha permesso l'acquisizione di competenze significative nell'acquisizione, l'organizzazione e l'analisi di un insieme di dati complesso. Le esperienze svolte in laboratorio hanno fornito competenze significative nell'utilizzo di sistemi di eye-tracking e di cattura di segnali fisiologici. La collaborazione con l'Università Cattolica del Sacro Cuore ha stimolato lo sviluppo di competenze significative nel lavoro in team in un progetto di grandi dimensioni, il quale ha richiesto una comunicazione efficiente tra i due enti, forti capacità decisionali, e l'apprendimento e l'applicazione di competenze multidisciplinari, spesso in un contesto time-sensitive. Infine, lo sviluppo dell'interfaccia utilizzata per l'analisi dei dati ha conferito competenze significative nel processo di sviluppo di un applicativo di grandi dimensioni.