Università degli Studi di Milano
Facoltà di Scienze e Tecnologie
Corso di Laurea in Informatica (L-31)

ANALISI QUANTITATIVA E PERCETTIVA DI VIDEO CREATI CON GENERATORI AI

Relatore: Prof. Raffaella Lanzarotti

Correlatore: Prof. Andrea Gaggioli

Tesi di:
Federico COSCIA
Matricola: 977772

Anno Accademico 2023-2024

a Celestina

Indice

Introduzione

Per chiunque abbia visitato qualunque sito di divulgazione o social media negli ultimi due anni, sarà stato impossibile non imbattersi in contenuti, fotografici o video, generati dall'Intelligenza Artificiale (IA). L'argomento è vasto, così come la quantità di contenuti diversi generabili tramite IA. Noi ci concentreremo su video basati su "avatar".

Un video basato su avatar è un video parlato rappresentante una persona, il cui movimento della bocca, della faccia, e talvolta anche del corpo, è stato generato tramite IA. Spesso gli avatar sono persone reali, le quali hanno messo a disposizione la loro figura affinché potesse essere animata. In altri casi, un avatar può anche essere creato a partire da immagini di persone non reali, generate a loro volta attraverso strumenti di IA generativa (es. Stable Diffusion, Midjourney, DALL-E). Si crea di fatto una copia digitale della persona raffigurata, con il potere di fargli dire qualsiasi cosa. Vi è un testo di riferimento, e a partire da una fotografia o un breve video del soggetto, viene generato un video rappresentante il soggetto che espone ad alta voce il testo indicato.[1]

Si identificano quindi due nuove categorie di video, i video reali, raffiguranti una persona reale registrati fisicamente, e i video "fake", raffiguranti a loro volta una persona reale, ma la cui voce e il cui movimento del corpo sono stati realizzati tramite generatori IA. Questa tecnologia ha trovato grande fortuna nel mondo della pubblicità e della istruzione, dove sempre maggiori sono i costi per la registrazione di video in presa diretta. L'utilizzo di tali sistemi permette di realizzare video senza doverli registrare fisicamente, riducendo di molto i costi di produzione. L'obiettivo di questo studio è valutare se tali video fake possono avere la stessa efficacia comunicativa di un video reale, oppure se la natura artificiale di tali video può risultare un ostacolo abbastanza grande nella comprensione e fruizione del contenuto video tale da poter essere misurato attraverso un esperimento scientifico, rendendo questa sorgente tecnologia non ancora pronta, o inadatta.

[1] Naturalmente la realizzazione di questi video richiede anche la presenza di una voce parlata. Per questo si utilizzano modelli generativi di tipo TextToSpeech, ma la voce utilizzata tipicamente non è la voce della persona rappresentata, bensì una voce di servizio. Nel nostro studio sono stati utilizzati modelli di voce di servizio.

Struttura dell'esperimento

L'esperimento è stato strutturato come segue:

1. Identificazione di brevi video educativi reali (max. 5 minuti)
2. Generazione di doppioni fake a partire dai video reali identificati, dove il contenuto informativo è lo stesso, ma il video, così come la voce del soggetto, sono stati generati tramite IA
3. Presentazione di un esperimento fittizio, con proposta in double blind[2] di due video a soggetti volontari, i quali sono ignari della vera natura dell'esperimento e in particolare della natura dei video[3]
4. Sessione di domande a risposta aperta sui contenuti discussi nei video, per valutare il grado di comprensione dei contenuti proposti
5. Acquisizione di dati multimodali durante l'esperimento e in particolare durante la visione dei video, tra cui espressioni del viso, battito cardiaco, ritmo respiratorio e tracciamento degli occhi.
6. Analisi ed elaborazione dei dati acquisiti, alla ricerca di cluster indicativi di un fattore di rilevanza/non rilevanza della natura dei video per la comprensione e la fruizione del video visionato

Inizieremo con una breve spiegazione di come questi tipi di video vengono generati, secondo la letteratura attuale, per poi entrare nel dettaglio della ricerca condotta. Infine, verranno tratte le dovute conclusioni in base a quanto trovato.

[2]Vengono proposti due video a un soggetto volontario, ma quest'ultimo non sa che uno dei due video è fake

[3]A ogni soggetto viene sempre mostrato un video reale e un video fake.

Capitolo 1

Generazione dei video fake

1.1 Funzionamento

1.2 Soluzioni attualmente disponibili

1.3 Video real

1.3.1 Scelta dei video real

1.3.2 Pre-Processing

1.4 Generazione dei video fake

Capitolo 2

Setting di acquisizione

2.1 Modalità di acquisizione

2.2 Estrazione delle feature

2.2.1 Video

2.2.2 Dati fisiologici

2.2.3 Eye-tracking

Capitolo 3

Protocollo di acquisizione

3.1 Stesura del protocollo

3.2 Sviluppo dell'interfaccia

3.3 Salvataggio dei dati raccolti

Capitolo 4

Analisi dei dati acquisiti

Conclusioni

Bibliografia

Ringraziamenti