

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 Lanza

**Correlatore:** Prof. Andrea Gaggioli

Tesi di:  
Federico COSCIA  
Matricola: 977772

Anno Accademico 2023-2024

*a Celestina*



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Generazione dei video fake</b>	<b>3</b>
1.1 Funzionamento . . . . .	3
1.2 Soluzioni attualmente disponibili . . . . .	3
1.3 Video real . . . . .	3
1.3.1 Scelta dei video real . . . . .	3
1.3.2 Pre-Processing . . . . .	3
1.4 Generazione dei video fake . . . . .	3
<b>2 Setting di acquisizione</b>	<b>4</b>
2.1 Modalità di acquisizione . . . . .	4
2.2 Estrazione delle feature . . . . .	4
2.2.1 Video . . . . .	4
2.2.2 Dati fisiologici . . . . .	4
2.2.3 Eye-tracking . . . . .	4
<b>3 Protocollo di acquisizione</b>	<b>5</b>
3.1 Stesura del protocollo . . . . .	5
3.2 Sviluppo dell'interfaccia . . . . .	5
3.3 Salvataggio dei dati raccolti . . . . .	5
<b>4 Analisi dei dati acquisiti</b>	<b>6</b>
<b>Conclusioni</b>	<b>7</b>
<b>Bibliografia</b>	<b>8</b>

# Introduzione

Per chiunque abbia visitato siti di divulgazione e 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, 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, a partire da una immagine o un breve video dello stesso. 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 il video generato rappresenta il soggetto che espone ad alta voce il testo indicato.<sup>1</sup> 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 quindi di realizzare video senza doverli registrare fisicamente, riducendo di molto i costi di produzione.

Si identificano quindi due nuove categorie di video, i video "real", video raffiguranti una persona reale registrato fisicamente, e i video "fake", video raffiguranti a loro volta una persona reale, ma la cui voce e il cui movimento sono stati appunto realizzati tramite generatori IA.

Visto l'ampio potenziale applicativo di tale tecnologia nella comunicazione in generale, 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.

---

<sup>1</sup>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 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 di due video a soggetti volontari, i quali sono ignari della vera natura dell'esperimento e in particolare della natura dei video (reali e fake) <sup>2</sup>
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
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

Vedremo innanzitutto 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.

---

<sup>2</sup>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