



**Politecnico
di Torino**

Politecnico di Torino

Laurea Magistrale in
Ingegneria del Cinema e dei Mezzi di Comunicazione
Dipartimento di Automatica e Informatica

Sessione di Laurea Marzo/Aprile 2026

IA e Arte

Narrare al pubblico l'impatto emotivo delle performance musicali
mediante video generati con strumenti di intelligenza artificiale
generativa

Relatori:

Prof.ssa Tatiana MAZALI

Prof. Andrea BOTTINO

Candidata:

Chiara COLOSSEO

ABSTRACT

L'intelligenza artificiale generativa sta ridefinendo i processi di ideazione e produzione audiovisiva, aprendo nuove opportunità ma sollevando al contempo interrogativi sulla centralità dell'autore e sulla legittimità di considerare "creativi" i prodotti generati da algoritmi.

La presente tesi esplora il punto di incontro tra arte e intelligenza artificiale (IA), con l'obiettivo di realizzare un cortometraggio attraverso una pipeline basata prevalentemente su strumenti di IA generativa. In questa prospettiva, l'IA non viene considerata come sostituto dell'autore umano, bensì come strumento di supporto e co-creazione all'interno di un flusso di lavoro controllato e strutturato.

Il progetto si inserisce nell'iniziativa culturale *Narrazioni Parallele*, dedicata all'esplorazione del rapporto tra intelligenza naturale e artificiale in contesti musicali ed interattivi. A partire dai dati raccolti dal pubblico (emozioni, ricordi, pensieri) e a seguito di un'analisi comparativa svolta a individuare gli strumenti di IA generativa più efficaci per ciascuna fase della pipeline produttiva, è stato realizzato un cortometraggio animato della durata di circa cinque minuti, mediante l'impiego di modelli generativi per testi, immagini, video, musiche e suoni.

La ricerca evidenzia l'importanza del bilanciamento tra automazione ed intervento umano, fondamentale per garantire la coerenza narrativa e visiva del prodotto finale.

Il lavoro si presenta, in conclusione, come un esempio concreto di sperimentazione narrativa con le intelligenze artificiali generative, mettendone in luce sia le potenzialità che i limiti.

INDICE

Capitolo 1 – Introduzione	1
1.1 Panoramica della Tesi.....	1
1.2 Intelligenza artificiale oggi: panoramica e distinzione tra IA generativa e non generativa.	2
1.3 Intelligenza artificiale: impatto nelle arti audiovisive.....	4
1.4 Obiettivi del progetto	5
1.5 Associazione LUCAS e Narrazioni Parallele.....	6
1.5.1 Evento “Caro Mozart... Una lettura musicale delle lettere alla cugina di un genio inquieto”	7
1.5.2 Evento “Song writing 2 – Intelligenza naturale generativa”	8
1.5.3 Evento “In scena con Verdi”	9
1.6 Definizione del problema ed approcci adottati.....	10
1.6.1 Possibili criticità.....	10
1.6.2 Domande di ricerca	10
1.6.3 Approccio adottato	12
Capitolo 2 – Stato dell’arte.....	14
2.1 Intelligenza artificiale nella produzione audiovisiva	14
2.1.1 Evoluzione dell’intelligenza artificiale nel cinema.....	15
2.1.2 Esempi significativi di opere realizzate con l’intelligenza artificiale generativa.....	17
2.1.3 Esempi significativi di festival ed iniziative incentrati sull’intelligenza artificiale.....	20
2.1.4 Limiti tecnici ed operativi dell’intelligenza artificiale.....	20
2.2 Intelligenza artificiale e creatività: dibattiti ed implicazioni etiche.....	22
2.2.1 Creatività e ruolo dell’autore	23
2.2.2 Democratizzazione della produzione.....	24
2.2.3 Proprietà intellettuale e diritto d’autore	26
2.2.4 Bias algoritmici e distorsioni generate.....	27
2.2.5 Il fenomeno dell’Uncanny Valley	28
2.3 Strumenti di intelligenza artificiale per la produzione audiovisiva.....	31
2.3.1 Approcci operativi: online o offline.....	31
2.3.2 I benchmark: criteri e metriche di valutazione	32
2.3.3 Strumenti di intelligenza artificiale per la generazione testuale.....	34

2.3.4 Strumenti di intelligenza artificiale per la generazione di immagini.....	39
2.3.5 Strumenti di intelligenza artificiale per la generazione di video	44
2.3.6 Strumenti di intelligenza artificiale per la generazione audio.....	49
2.3.7 Strumenti di intelligenza artificiale per la post-produzione.....	51
2.3.8 Sintesi delle analisi sui benchmark.....	51
Capitolo 3 – Progetto di tesi.....	53
3.1 Struttura del workflow e impostazione della pipeline produttiva.....	53
3.2 Raccolta e analisi dei dati.....	56
3.2.1 Raccolta dei dati.....	56
3.2.2 Analisi preliminare e preparazione dei dati.....	59
3.2.3 Elaborazione dei dati con strumenti di IA generativa.....	62
3.3 Fase di Pre-produzione: scrittura.....	67
3.3.1 Astrazione dei dati.....	67
3.3.2 Scelte strutturali e narrative.....	70
3.3.3 Ideazione dei concept narrativi.....	73
3.3.4 Selezione e affinamento del concept finale.....	80
3.3.5 Sviluppo della sceneggiatura.....	82
3.3.6 Ideazione del titolo.....	84
3.4 Fase di Pre-produzione: stile visivo.....	85
3.4.1 Definizione dello stile visivo.....	85
3.4.2 Valutazione comparativa degli strumenti di generazione di immagini.....	88
3.4.3 Definizione dei personaggi.....	95
3.4.4 Generazione dello storyboard.....	99
3.5 Fase di Pre-produzione: osservazioni e conclusioni.....	103
3.6 Fase di Produzione: video.....	104
3.6.1 Valutazione comparativa degli strumenti di generazione video.....	105
3.6.2 Generazione delle clip video.....	109
3.7 Fase di Produzione: audio.....	114
3.7.1 Definizione dello stile musicale.....	114
3.7.2 Valutazione comparativa degli strumenti di generazione audio.....	115
3.7.3 Generazione della voce narrante.....	118
3.7.4 Generazione degli effetti sonori.....	121

3.7.5 Generazione della colonna sonora.....	123
3.8 Fase di Produzione: osservazioni e conclusioni.....	124
3.9 Fase di Post-produzione.....	127
3.9. DaVinci Resolve e strumenti IA integrati.....	127
3.9.2 Montaggio, color correction e ottimizzazione finale.....	130
3.9.3 Titoli di coda.....	134
3.10 Fase di Post-produzione: osservazioni e conclusioni.....	136
Capitolo 4 – Conclusioni e valutazioni finali.....	138
4.1 Pipeline IA: potenzialità e limiti.....	138
4.2 Dalla progettazione alla pratica: il compromesso creativo.....	141
4.3 Il ruolo dell'autore nei processi di co-creazione con l'IA.....	142
4.4 Implicazioni etiche e diritto d'autore.....	144
4.4.1 Bias e rappresentazione.....	144
4.4.2 Diritto d'autore delle opere generate.....	146
4.4.3 Democratizzazione e figura dell'autore.....	147
4.4.4 L'IA può comprendere le emozioni umane e creare?.....	148
4.5 Sviluppi futuri e prospettive.....	148
4.5.1 Sviluppi tecnologici attesi.....	149
4.5.2 Evoluzione delle pipeline produttive.....	150
4.5.3 Ridefinizione del ruolo dell'autore e democratizzazione.....	150
4.5.4 Questioni normative e diritto d'autore.....	151
4.5.5 Impatto sul linguaggio audiovisivo.....	151
4.6 Considerazioni conclusive.....	152
APPENDICI.....	I
APPENDICE A – Domande poste al pubblico.....	I
Concerto 1 – “Caro Mozart... Una lettura musicale delle lettere alla cugina di un genio inquieto”.....	I
Concerto 2 – “Song writing 2 – Intelligenza naturale generativa”.....	II
Concerto 3 – “In scena con Verdi”.....	III
APPENDICE B – Risposte del pubblico e dati raccolti.....	IV
APPENDICE C – Proposte concept narrativi.....	V
APPENDICE D – Sceneggiatura.....	VI

APPENDICE E – Character sheet.....	XV
APPENDICE F – Storyboard	XVI
APPENDICE G – Cortometraggio “The Paper Wolf”	XVII
Bibliografia e Sitografia.....	I
Ringraziamenti.....	II

CAPITOLO 1 – INTRODUZIONE

1.1 Panoramica della Tesi

Questo progetto di tesi si propone di esplorare il punto di incontro tra arte e intelligenza artificiale (IA), con l'obiettivo di realizzare un cortometraggio attraverso una pipeline basata prevalentemente su strumenti di IA generativa, affiancati da strumenti di IA per la post-produzione.

Il lavoro si inserisce nel programma di *Narrazioni Parallele* – iniziativa promossa dall'Associazione LUCAS – che unisce musica, creatività e tecnologie emergenti.

Il contesto in cui si colloca questa ricerca è quello di un periodo di transizione e trasformazione tecnologica. Negli ultimi anni, infatti, l'intelligenza artificiale ha compiuto progressi straordinari, passando da strumento di analisi riservato a pochi esperti a risorsa accessibile nella vita quotidiana.

Nel settore audiovisivo, il suo impiego non si limita più al solo supporto tecnico o all'ottimizzazione dei processi, ma si estende anche alla generazione di nuovi contenuti, ridefinendo così i confini tra creatività umana e automazione.

Questa trasformazione solleva molteplici questioni rilevanti: chi è l'autore di un'opera generata da algoritmi? L'intelligenza artificiale può realmente “comprendere” le emozioni umane e *creare* qualcosa di artistico (e con quali limiti)? E, infine, la democratizzazione promessa dagli strumenti generativi è reale o introduce nuove forme di esclusione?

Per rispondere a queste domande, il progetto adotta un approccio sperimentale volto ad automatizzare il più possibile l'intera pipeline produttiva, dalla scelta del soggetto e scrittura della sceneggiatura sino al montaggio finale, affidando all'intelligenza artificiale non solo l'esecuzione tecnica ma anche scelte tradizionalmente considerate “creative” ed “umane”.

La struttura della tesi riflette il percorso stesso della ricerca, partendo da una riflessione teorica sullo stato dell'arte per arrivare a una sperimentazione pratica volta a valutarne limiti e potenzialità, ed è così suddivisa:

- Il primo capitolo introduce il contesto del progetto, gli obiettivi e le domande di ricerca, oltre a presentare l'associazione LUCAS e il progetto *Narrazioni Parallele*.

- Il secondo capitolo offre una panoramica dello stato dell'arte, analizzando l'evoluzione dell'IA nel cinema e nelle arti visive ed affrontando le questioni etiche e percettive che ne derivano. L'analisi porterà ad un *benchmark* comparativo degli strumenti utilizzabili nelle diverse fasi della produzione.
- Il terzo capitolo descrive la realizzazione del progetto audiovisivo, documentando le fasi di sperimentazione, le scelte tecniche e le strategie adottate per automatizzare la pipeline di lavoro.
- Il quarto capitolo presenta le osservazioni conclusive e le criticità emerse, oltre a proporre una riflessione sulle prospettive future dell'utilizzo dell'intelligenza artificiale nella creazione artistica.

1.2 Intelligenza artificiale oggi: panoramica e distinzione tra IA generativa e non generativa

L'intelligenza artificiale (IA), oggi ampiamente diffusa nella vita quotidiana, appare spesso come una tecnologia recente; le sue origini risalgono però alla metà del Novecento. Il termine “Artificial Intelligence (AI)” fu infatti coniato nel 1955 da John McCarthy, informatico statunitense, che l'anno successivo, in occasione della Dartmouth Summer Research Project on Artificial Intelligence, venne nominato “founding father of artificial intelligence”¹.

McCarthy definì l'intelligenza artificiale come:

*«La scienza e l'ingegneria che si occupa di realizzare macchine intelligenti, in particolare programmi informatici intelligenti. [...] Consistente nel far sì che una macchina si comporti in modi che sarebbero definiti intelligenti se fosse un essere umano a comportarsi così»*².

Già qualche anno prima, nel 1950, Alan Turing aveva posto le basi teoriche per la nascita dell'Intelligenza Artificiale con la pubblicazione dell'articolo *Computing Machinery and Intelligence*³, dove viene posta la celebre domanda: «*Can machines think?*»

Nel corso dei decenni, l'intelligenza artificiale ha visto un'evoluzione significativa: dai primi modelli basati su regole logiche alle moderne reti neurali, capaci di apprendere e “comprendere”

¹ Dartmouth College, *Artificial Intelligence (AI) Coined at Dartmouth*, (1956/2024) [1]

² J. McCarthy, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (1955) [2]

³ A. M. Turing, *Computing Machinery and Intelligence*. (1950) [3]

grandi quantità di dati. Dai primi chatbot, come ELIZA (1966)⁴, si è giunti a modelli capaci di prendere decisioni e scrivere testi complessi, come ChatGPT e Claude, o generare contenuti multimediali, come Midjourney, Sora ed ElevenLabs.

Questa evoluzione ha portato alla nascita di diversi termini, spesso erroneamente usati come sinonimi. In questa fase preliminare al progetto è utile chiarirli rapidamente per evidenziarne le differenze.

L'*intelligenza artificiale (IA)* è l'insieme delle tecniche e tecnologie che permettono a una macchina di simulare capacità tipiche dell'intelligenza umana, quali l'analisi, il ragionamento, la percezione ed il problem solving.

Il *Machine Learning (ML)* rappresenta un suo sottoinsieme ed è basato su algoritmi che consentono ai sistemi di imparare dai dati e migliorare nel tempo, senza seguire rigide regole di programmazione. Il machine learning si basa sul riconoscimento di pattern e sulla capacità di fare previsioni partendo dai dati disponibili.

Da questo discende il *Deep Learning (DL)*, basato sull'utilizzo di reti neurali artificiali ispirate al cervello umano. Questi modelli utilizzano operazioni matematiche complesse (come la convoluzione) per elaborare le informazioni provenienti da dataset di grandi dimensioni e riconoscere pattern complessi. Il deep learning si dimostra particolarmente efficace anche nell'analisi di dati non lineari – come immagini, audio e testo – ed è alla base di molte applicazioni di IA, tra cui il riconoscimento vocale e visivo e l'elaborazione del linguaggio naturale. In questo campo di applicazione, i modelli di deep learning possono superare le performance degli approcci di machine learning tradizionali; richiedono però una notevole quantità di dati e risorse computazionali elevate.

Gli algoritmi di machine learning e deep learning sono alla base dell'*Intelligenza Artificiale Generativa e Non Generativa* attuale. La prima si riferisce a sistemi capaci di creare nuovi contenuti (immagini, video, testo, audio) partendo da prompt o dati di riferimento. La seconda si riferisce a sistemi finalizzati principalmente all'analisi, classificazione e previsione basate su dati esistenti (ad esempio riconoscimento di immagini, trascrizione audio, traduzione di testi).

⁴ J. Weizenbaum, *ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine*, (1966) [4]

1.3 Intelligenza artificiale: impatto nelle arti audiovisive

L'introduzione dell'intelligenza artificiale nel settore creativo rappresenta uno dei fenomeni più significativi e discussi dell'epoca recente.

Nel settore audiovisivo, l'IA è presente in forme diverse. Ad esempio, è ormai visibile quotidianamente con gli algoritmi di raccomandazione (*AI recommendation systems*) presenti su tutte le piattaforme di streaming, come Netflix, e orientati all'analisi dei gusti del pubblico per personalizzare la loro esperienza di visione. Parallelamente, si è affermata come strumento di supporto tecnico in grado di snellire il lavoro automatizzando le attività più onerose e di consentire a troupe dal budget o dalle risorse ridotte di realizzare contenuti di elevata qualità.

Negli ultimi anni, inoltre, stanno emergendo studi e sperimentazioni che rendono l'IA parte integrante del processo creativo, fino a configurarsi come elemento di creazione o co-creazione con gli autori umani. Un esempio emblematico è *Sunspring* (2016)^[5], il primo cortometraggio con sceneggiatura interamente generata da un'intelligenza artificiale, Benjamin, sviluppata da Ross Goodwin, *creative technologist* e ricercatore nel campo dell'IA, in collaborazione con il regista Oscar Sharp. Più recentemente, in Italia, EDI – Effetti Digitali Italiani e HAI – Human & Artificial Imagination hanno realizzato *The Last Image* (2025)^[6], un cortometraggio in cui l'IA è stata integrata in tutte le fasi della pipeline produttiva, dalla scrittura al montaggio, sotto la supervisione umana.

L'emergere di questa nuova frontiera non è tuttavia privo di critiche. Alcuni registi affermati, tra cui Guillermo del Toro, hanno espresso preoccupazioni riguardo all'utilizzo dell'intelligenza artificiale nel settore audiovisivo⁵. Si sono verificate anche mobilitazioni del settore volte a richiedere maggiori tutele normative e puntare l'attenzione sul rischio di affievolimento del talento umano e sulla possibilità di rimpiazzo di professionalità creative. Un esempio rilevante è rappresentato dallo sciopero del 2023 degli sceneggiatori degli Stati Uniti⁶.

Allo stesso tempo, stanno nascendo festival ed iniziative dedicate a questo nuovo linguaggio audiovisivo, come il *Reply AI Film Festival*^[9] e l'*AI International Film Festival*^[10], concorsi internazionali che premiano opere realizzate con strumenti di intelligenza artificiale generativa.

L'evoluzione degli strumenti di IA sta quindi portando ad una rivoluzione del settore audiovisivo, ma resta da comprendere se questa si tradurrà in una vera democratizzazione della

⁵ A. Watercutter, *Guillermo del Toro Hopes He's Dead Before AI Art Goes Mainstream*, (2025) [7]

⁶ S. Cosimi, *Cosa prevede l'accordo per l'intelligenza artificiale dopo lo sciopero degli autori e sceneggiatori*, (2023) [8]

produzione, rendendo la creazione di contenuti accessibile a chiunque, o se invece introdurrà nuove barriere tecnologiche e competenze specialistiche.

Il Capitolo 2 approfondirà questi temi, analizzando l'evoluzione dell'intelligenza artificiale nel settore multimediale, le implicazioni etiche e sociali che ne derivano ed i principali strumenti oggi disponibili per la generazione di contenuti audiovisivi.

1.4 Obiettivi del progetto

L'obiettivo generale di questo progetto è esplorare il punto d'incontro tra arte, musica e intelligenza artificiale attraverso la creazione di un cortometraggio generato con una pipeline basata su IA. Il lavoro intende analizzare fino a che punto le tecnologie generative possano intervenire nel processo creativo audiovisivo, dalla raccolta dei dati fino alla post-produzione, individuando opportunità, limiti e implicazioni etiche.

Questo macro-obiettivo è stato affiancato dai seguenti propositi:

- Sperimentare metodologie di raccolta dati durante gli eventi musicali, coinvolgendo attivamente sia il pubblico che i performer.
- Analizzare e categorizzare i dati raccolti per trasformarli in materiale utile per costituire la base per la generazione dei contenuti tramite IA.
- Progettare e testare una pipeline il più possibile automatizzata, che utilizzi gli strumenti di intelligenza artificiale in tutte le principali fasi (pre-produzione, produzione e post-produzione) per generare un video interamente con IA, partendo da una sceneggiatura costruita sulla base dati raccolti.
- Indagare le implicazioni etiche e sociali legate all'utilizzo di strumenti di IA generativa, con particolare attenzione ai temi dell'autorialità, dei *bias* (distorsioni degli algoritmi) e della rappresentazione inclusiva.
- Raggiungere un equilibrio tra automazione e controllo umano, valutando in che misura la creatività possa essere condivisa tra artista e intelligenza artificiale.
- Definire un'identità estetica autonoma del prodotto finale, capace di restituire in modo coerente e suggestivo sia i dati raccolti sia la riflessione sul rapporto tra uomo e macchina.

1.5 Associazione LUCAS e Narrazioni Parallele

Come già accennato, il progetto si è svolto all'interno di *Narrazioni Parallele*^[11], iniziativa che unisce performance musicali e sperimentazione tecnologica, promossa dall'associazione LUCAS.

Fondata nel 2002 da Roberto Beccaria, LUCAS (Libera Unione di Cultura Arte e Spettacolo)^[12] ha l'obiettivo di contribuire alla creazione di strategie per il marketing, lo sviluppo e la riqualificazione del territorio attraverso l'innovazione culturale, con particolare attenzione alla sperimentazione tra arte, scienza e nuove tecnologie. Mediante collaborazioni con istituzioni culturali, accademiche e realtà del territorio, LUCAS promuove un approccio multidisciplinare che considera la creatività come strumento di partecipazione sociale e benessere collettivo.

Narrazioni Parallele è un progetto ideato da Roberto Beccaria e Claudia Lupo, realizzato in partenariato con l'Accademia Corale Stefano Tempia e l'associazione Cultura e Salute. Vi è inoltre una collaborazione con la Filarmonica TRT, del Teatro Regio di Torino, le Biblioteche Civiche Torinesi ed il Politecnico di Torino.

Il progetto nasce da una domanda semplice ma ambiziosa:

«Può l'intelligenza artificiale incontrare la musica classica e contemporanea, creando qualcosa di nuovo insieme al pubblico che partecipa ai concerti?»

Attraverso un percorso triennale (2023–2026), *Narrazioni Parallele* esplora il possibile dialogo tra Intelligenza Naturale e Intelligenza Artificiale Generativa, con l'idea di unire innovazione tecnologica e partecipazione culturale.

Il progetto si articola in due anime principali:

- *Narrazione Contemporaneità* coinvolge giovani compositori e compositrici a livello nazionale, attraverso le *call for composers*. I partecipanti seguono percorsi di *mentorship*, guidati da figure di rilievo, che culminano nella presentazione dal vivo di opere inedite.
- *Narrazione Partecipazione* mette al centro il pubblico coinvolto nei concerti, invitandolo a co-costruire visioni nate dall'ascolto di musica e parole, condividendo emozioni, pensieri ed immagini.

Come conferma della sua rilevanza nel panorama culturale tecnologico, l'iniziativa *Narrazioni Parallele* ha vinto il bando triennale *Torino che Cultura!* della Città di Torino.

Un ulteriore elemento distintivo è l'attenzione all'accessibilità: grazie alla collaborazione con APIC (Associazione Portatori di Impianto Cocleare) e MED-EL Italia, i concerti e gli eventi sono progettati per essere fruibili anche da persone con ipoacusia. È stato infatti predisposto un sistema di ascolto assistito basato su cuffie FM e connettività wireless LOLA, a supporto dei partecipanti con disabilità uditive, garantendo così l'inclusività dell'evento.

Il progetto di tesi si colloca nella linea di *Narrazione Partecipazione*, sviluppando un esperimento audiovisivo che intende trasformare le percezioni del pubblico in narrazione visiva.

Durante tre eventi, sono stati raccolti dei dati qualitativi attraverso un sondaggio realizzato con la piattaforma Mentimeter, rivolto sia al pubblico che ai performer. Queste informazioni (emozioni, parole, colori, ricordi ed impressioni) costituiranno la base narrativa ed emotiva del video, venendo utilizzate come input per le intelligenze artificiali generative. La co-creazione tra artisti, pubblico e IA farà quindi nascere un prodotto audiovisivo traducendo in immagini e suoni la dimensione emotiva dell'ascolto musicale.

1.5.1 Evento “Caro Mozart... Una lettura musicale delle lettere alla cugina di un genio inquieto”

Il primo evento della rassegna, svoltosi il 29 settembre alla Biblioteca civica Italo Calvino di Torino, ha proposto un viaggio tra musica, epistole ed umorismo alla scoperta dei lati più umani ed inaspettati di Wolfgang Amadeus Mozart.

Il pubblico, composto principalmente dai seniores frequentatori della biblioteca, è stato guidato in questo viaggio intergenerazionale alla scoperta di Mozart da Eleonora Savini – violinista, performer ed insegnante – accompagnata dall'ensemble della Filarmonica Teatro Regio Torino, composto da un quartetto d'archi e due corni.

Il percorso narrativo ed emotivo si è articolato in due momenti distinti. Nella prima parte, attraverso la lettura di estratti delle lettere bizzarre del giovane Mozart alla cugina e l'esecuzione di *Ein Musikalischer Spass KV 522* – una composizione satirica in cui il compositore inserisce dissonanze e corni stonati per parodiare i musicanti di villaggio privi di esperienza – è emerso un ritratto irriverente e giocoso del giovane artista. Nella seconda parte, le lettere al padre hanno dipinto un Mozart adulto e riflessivo, che affronta con rispetto e profondità la malattia e la morte del genitore; l'esecuzione del mottetto *Ave Verum Corpus K 618*, composto negli ultimi mesi di vita, ha invitato alla riflessione sulla caducità della stessa.

Durante l'evento, il pubblico è stato invitato a partecipare attivamente alla costruzione di un identikit di Mozart, intrecciando conoscenze pregresse, impressioni emerse durante l'ascolto e fantasia (ad esempio, per immaginare quale personaggio odierno potrebbe incarnare al meglio la figura del compositore).

1.5.2 Evento “Song writing 2 – Intelligenza naturale generativa”

Il secondo evento si è svolto il 30 ottobre presso la Biblioteca civica Italo Calvino di Torino, con la conduzione di Arianna Gallo, cantante e formatrice.

L'incontro ha proposto un'esperienza partecipativa di co-creazione musicale e testuale, ispirata alla poesia *Il paese con la S davanti* di Gianni Rodari e al tema “un giorno senza regole”.

Attraverso un percorso basato su musica, scrittura e improvvisazione, il pubblico – composto principalmente dai seniores frequentatori della biblioteca – è stato guidato per esplorare le potenzialità espressive della voce e del corpo, con l'obiettivo di realizzare una breve canzone.

Durante l'attività, i partecipanti hanno risposto a domande ispirate allo stile surrealista e giocoso di Gianni Rodari.

Un esempio di una domanda, con una delle risposte ritenute più belle, è:

D: “Se potessi inventare una nuova regola per vivere meglio, quale sarebbe?”

R: “Regalerei un fiore invece di litigare”.

Anche in questo caso è stato utilizzato Mentimeter, il quale ha permesso la raccolta e la visualizzazione in tempo reale delle risposte. Queste hanno fornito la base testuale per la scrittura collettiva della canzone, che è stata poi accompagnata da semplici strumenti ritmici – tamburelli, mani, piedi – ed interpretata coralmente dal pubblico.

La canzone finale, composta attraverso questo processo di intelligenza naturale generativa (in contrapposizione ironica all'intelligenza artificiale generativa) è diventata espressione di una creazione autenticamente umana, basata sull'ascolto reciproco e condivisione di idee.

La metodologia adottata ha richiesto ai partecipanti curiosità e disponibilità a mettersi in gioco e di tornare bambini per un paio di ore, scrivendo il testo del loro “giorno senza regole”.

*In un giorno tutto senza le regole,
abbracciamo un albero quando ci va.*

*Da supereroi, i cuori leggeremmo,
una pizza al volo con John Cage.*

*Se dentro al mondo si potesse,
indietro nel tempo vorremmo viaggiar.*

*Tra le strade, odore di libertà.
Se fossimo musica, "boom boom clap"!*

Testo della canzone

1.5.3 Evento "In scena con Verdi"

Il terzo ed ultimo incontro si è svolto l'11 novembre presso la Biblioteca civica Italo Calvino di Torino e ha proposto un viaggio cronologico attraverso alcune delle opere più significative di Giuseppe Verdi, eseguite dal Quartetto d'archi della Filarmonica Teatro Regio Torino e dalle voci del Consort Tempia.

Al pubblico sono state proposte alcune domande tramite la piattaforma Mentimeter, per raccogliere le loro impressioni e le suggestioni scaturite durante l'ascolto. Inoltre, è stata data la possibilità di mettere in gioco la fantasia riscrivendo le prime righe della celebre *Va', pensiero*.

Una delle riscritture è stata infine scelta, per coerenza in metrica e in stile verdiano, dal coro per essere cantata.

*Va, pensiero, sull'ali dorate
Va, ti posa sui clivi, sui colli
Ove olezzano tepide e molli
L'aure dolci del suolo natal*

Testo originale

*Parla, afferma le tue idee
grida il tuo dissenso,
sii sempre coerente con le tue idee,
ma soprattutto, combatti per un ideale!*

Testo riscritto

1.6 Definizione del problema ed approcci adottati

La realizzazione di un video tramite una pipeline completamente basata su strumenti di intelligenza artificiale generativa rappresenta una sfida complessa.

1.6.1 Possibili criticità

Le difficoltà principali derivano da due fattori.

In primo luogo, ci troviamo in un contesto in cui l'automazione dei processi creativi audiovisivi è ancora parziale: le tecnologie disponibili sono molteplici, eterogenee e in continua evoluzione, ma non ancora integrate in un flusso di lavoro unificato e completamente autonomo. L'attuale ecosistema degli strumenti di intelligenza artificiale generativa è infatti caratterizzato da un'ampia varietà di soluzioni, ciascuna con specifiche capacità, limitazioni tecniche e costi. Come verrà approfondito nel capitolo successivo, poiché esistono numerosi strumenti altamente specializzati e adatti per la realizzazione della pipeline produttiva, sarà richiesta un'attenta valutazione per individuare le combinazioni più efficaci e coerenti per gli scopi del progetto.

In secondo luogo, il progetto parte dai dati raccolti durante i concerti, introducendo una difficoltà aggiuntiva: far interpretare correttamente i dati alle IA. Questi, infatti, sono molto eterogenei e diversificati tra loro, oltre che fortemente legati al contesto dell'evento in cui sono stati raccolti. Trovare un metodo per selezionarli e farli interpretare diventa quindi essenziale, poiché questi saranno la base del lavoro, che dovrà essere coerente con gli eventi e l'esperienza del pubblico. L'IA non dovrà solo "leggere" i dati, ma comprenderli e tradurli in elementi narrativi coerenti e con significato.

1.6.2 Domande di ricerca

Queste criticità hanno portato alla formulazione di una serie di domande di ricerca che hanno guidato l'intero sviluppo del progetto, dalla fase di analisi iniziale sino alla realizzazione attiva del cortometraggio.

La prima questione fondamentale riguarda l'interpretazione dei dati da parte dell'IA. I dati relativi ai concerti, come già evidenziato, sono eterogenei e profondamente contestuali; risulta quindi cruciale capire come selezionare quelli rilevanti senza omettere elementi che potrebbero rivelarsi significativi per la realizzazione del progetto. Si pone inoltre la questione se questa fase di selezione e contestualizzazione debba essere gestita manualmente o se possa essere a

sua volta automatizzata attraverso strumenti di IA, individuando criteri di verifica adatti per controllare che gli output siano fedeli alle esperienze vissute.

Una seconda questione fondamentale riguarda la selezione degli strumenti di IA da utilizzare. Come emergerà nel capitolo successivo, esistono molteplici soluzioni disponibili per ciascuna fase della produzione audiovisiva. Questa abbondanza di opzioni richiede di stabilire criteri chiari di selezione – qualità degli output, costi, tempo, facilità di utilizzo e di integrazione con gli altri elementi della pipeline – per la scelta degli strumenti più adatti.

Poiché il prodotto finale deve risultare unitario e coerente, emerge anche la questione di come garantire coesione stilistica e narrativa tra gli output di strumenti diversi utilizzati in fasi successive della pipeline. Ciascun modello di IA possiede infatti peculiarità estetiche e operative proprie, che potrebbero generare discontinuità nel passaggio da una fase all'altra. A questo si aggiunge un'ulteriore criticità legata alla mancanza di continuità e memoria tra i diversi strumenti: ogni fase del workflow di una produzione audiovisiva si fonda su scelte compiute in precedenza che influenzano quelle successive ma le IA, lavorando per compartimenti stagni, non sono in grado di mantenere una visione d'insieme del progetto. Ne deriva che, sebbene gli attuali modelli siano molto performanti nelle singole attività, la difficoltà nel mantenere coerenza narrativa e stilistica lungo l'intero processo è ancora elevata.

Una terza questione riguarda la progettazione di un flusso di lavoro che sia al contempo efficace, riproducibile e flessibile.

L'efficacia richiede che la pipeline produca un contenuto di qualità e comprensibile, rispettando vincoli di tempo e budget. Questo implica identificare le tecniche di formulazioni dei prompt più performanti e determinare quando e come utilizzare immagini di reference o altri materiali di input per guidare la generazione e garantire coerenza visiva.

La riproducibilità esige che il processo sia sufficientemente documentato e standardizzato per poter essere replicato in progetti futuri. Tuttavia, questa esigenza si scontra con la natura intrinsecamente variabile dei sistemi generativi: a parità di prompt, i modelli possono produrre risultati diversi; inoltre, i sistemi evolvono nel tempo attraverso nuovi cicli di *training* e possono modificare i propri comportamenti in base alla memoria conversazionale.

Infine, la flessibilità richiede che la pipeline possa adattarsi al contesto specifico di ciascun progetto, lasciando spazio per scelte creative e artistiche proprie di ogni contenuto audiovisivo.

Queste ultime due domande pongono l'interrogativo sulla possibilità di costruire una pipeline modulare, in cui singoli componenti possano essere sostituiti o aggiornati senza compromettere la qualità e la coerenza del risultato complessivo.

Infine, l'ultima domanda di ricerca riguarda l'identificazione dei limiti dell'automazione e la definizione del ruolo dell'intervento umano.

Il progetto servirà a studiare se tutte le fasi del processo possano essere completamente automatizzate o se invece l'intervento umano rimarrà necessario – ad esempio per scelte creative, validazione degli output o correzioni. Gli attuali strumenti di IA generativa presentano ancora alcuni limiti evidenti, come la difficoltà nel mantenere coerenza tra fotogrammi successivi, con il risultato che gli oggetti possono apparire o scomparire improvvisamente e l'illuminazione o lo stile possano variare in modo discontinuo. Un secondo limite riguarda il ritmo narrativo: mentre l'essere umano riesce a costruire e percepire la narrazione come una struttura complessa ed articolata, le IA operano su unità discrete ed indipendenti, rendendo difficile la gestione di racconti estesi. Inoltre, l'utilizzo di modelli differenti nelle varie fasi della pipeline rischia di accentuare tali criticità, poiché ognuno di essi opera in modo autonomo senza conoscere le scelte effettuate dagli altri “colleghi IA”.

Alla luce di queste considerazioni, la realizzazione di un prodotto audiovisivo tramite una pipeline automatizzata basata su strumenti di IA generativa appare come un obiettivo complesso e meritevole di una verifica sperimentale. Il progetto richiederà un'attività di studio, sperimentazione e adattamento, finalizzata ad individuare soluzioni tecniche e metodologiche efficaci per realizzare un prodotto coerente con gli obiettivi prefissati e che possa rispondere alle domande di ricerca poste.

1.6.3 Approccio adottato

Viene ora fornita una visione d'insieme del processo seguito per la realizzazione del progetto, analizzato in dettaglio nel [Capitolo 3](#).

Prima di tutto, si è analizzata la pipeline tradizionale per la produzione cinematografica, così da poterla reinterpretare mediante l'utilizzo degli strumenti di intelligenza artificiale.

La pipeline cinematografica si articola in più fasi interconnesse tra loro, svolte da reparti distinti ma coordinati tra loro. Nel caso di questo progetto, sono state utilizzate diverse IA in base alle rispettive specializzazioni e alla fase della pipeline di lavoro, in combinazione con l'intervento umano.

La fase iniziale ha previsto la raccolta dei dati relativi all'esperienza dei concerti tramite la piattaforma Mentimeter e la loro successiva analisi mediante strumenti di IA, al fine di individuare ricorrenze e differenze significative. L'output ottenuto è stato utilizzato come base per la generazione di un soggetto – mediante un brainstorming con ChatGPT, Claude e Gemini – e la scrittura di una sceneggiatura che rispecchiasse le componenti emotive emerse nei concerti.

Utilizzando prompt generati dalle IA sopra citate, sono stati realizzati i *model sheet* dei personaggi e gli storyboard, mediante IA specializzate per la generazione di immagini. Gli

storyboard, progettati per rappresentare fedelmente le inquadrature previste, sono poi stati utilizzati come input per la generazione delle clip video tramite modelli di *image-to-video*.

Prompt realizzati dalle IA sono stati impiegati anche per la generazione della colonna sonora, della voce narrante e degli effetti sonori.

In fase di montaggio, l'intervento umano è stato fondamentale, ma si sono seguite le indicazioni derivanti dal dialogo con le IA e dalla sceneggiatura. Inoltre, si è fatto uso di alcuni strumenti di IA generativa e non generativa presenti sul software DaVinci Resolve Studio per velocizzare alcune correzioni necessarie per sistemare imperfezioni e coerenza ed ottimizzare la qualità complessiva del prodotto. Alcuni esempi degli strumenti usati sono AI upscale e Magic Mask.

Il bilanciamento tra automazione ed intervento umano si è rivelato essere uno degli aspetti centrali del progetto. In una fase iniziale si era tentato di massimizzare l'autonomia dei sistemi di IA per poterne valutare la capacità; successivamente, il processo si è gradualmente evoluto verso una logica collaborativa, in cui le diverse IA scelte sono state impiegate come membri virtuali di una troupe cinematografica e l'intervento umano è servito per prendere decisioni e facilitare la collaborazione, così da garantire una coerenza narrativa e visiva.

CAPITOLO 2 – STATO DELL’ARTE

Il presente capitolo analizza l’attuale panorama degli strumenti di intelligenza artificiale generativa e la loro evoluzione nel contesto della produzione artistica e audiovisiva, con l’obiettivo di delinearne le principali potenzialità e limiti, approfondendo anche le implicazioni culturali che stanno emergendo con la loro crescente diffusione.

In primo luogo, verrà esaminata l’evoluzione dell’IA nel campo delle arti visive e del cinema, evidenziando come queste tecnologie stiano progressivamente trasformando la pipeline cinematografica tradizionale. Successivamente, verranno approfondite alcune questioni teoriche ed etiche connesse all’introduzione di questi strumenti nei processi creativi, tra cui la democratizzazione della produzione audiovisiva e le problematiche legate al diritto d’autore.

Infine, il capitolo presenterà una panoramica dei principali strumenti di intelligenza artificiale attualmente disponibili per la generazione di contenuti audiovisivi. Questa costituirà il riferimento necessario per comprendere le scelte metodologiche adottate nella fase di sperimentazione trattata nel [Capitolo 3](#).

2.1 Intelligenza artificiale nella produzione audiovisiva

Negli ultimi anni, l’intelligenza artificiale ha fatto il suo ingresso all’interno dei reparti legati alla produzione audiovisiva, portando con sé innovazioni e trasformazioni. L’evoluzione dei modelli di *machine learning* e *deep learning* ha migliorato significativamente le capacità dell’IA, spingendo la sua sperimentazione anche verso ambiti creativi, come la generazione di immagini, musica, testi e clip video.

Come anticipato nel [Capitolo 1](#), il concetto di “intelligenza artificiale” affonda le proprie radici negli studi sviluppati a partire da metà ‘900, grazie al contributo di pionieri del settore come Alan Turing⁷ e John McCarthy⁸. In particolare, quest’ultimo definì l’IA come una disciplina finalizzata alla realizzazione di macchine capaci di simulare comportamenti intelligenti tipicamente associati all’essere umano, come l’apprendimento, il ragionamento ed il problem solving.

Negli ultimi anni, con lo sviluppo dei modelli generativi, il ruolo dell’intelligenza artificiale si è ulteriormente evoluto. Oggigiorno, i sistemi algoritmici non si limitano più ad elaborare o analizzare informazioni preesistenti, ma sono anche in grado di generare nuovi contenuti visivi,

⁷ A. M. Turing, *Computing Machinery and Intelligence*. (1950) [3]

⁸ McCarthy, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (1955) [2]

sonori e testuali. Questo cambiamento ha ampliato la funzione dell'IA, che da semplice strumento di calcolo o supporto tecnico si sta progressivamente trasformando in un elemento attivo nei processi di produzione creativa.

L'introduzione di nuove tecnologie non rappresenta, però, una novità per il linguaggio cinematografico. Fin dalle sue origini, infatti, il cinema è stato profondamente influenzato dalle innovazioni tecniche. Alcuni esempi possono essere il passaggio dal cinema muto al sonoro, l'introduzione del colore, lo sviluppo della *Computer-Generated Imagery (CGI)* e le recenti tecniche di *virtual production*. In questo contesto di innovazione tecnologica costante, l'emergere dell'intelligenza artificiale generativa può essere interpretata semplicemente come una nuova fase, capace di introdurre nuovi strumenti, linguaggi visivi e ruoli.

I paragrafi successivi analizzeranno l'evoluzione dell'intelligenza artificiale nel cinema ed il suo impatto sulla pipeline produttiva, con alcuni esempi significativi di opere ed iniziative che testimoniano l'affermarsi di questa nuova tecnologia creativa.

2.1.1 Evoluzione dell'intelligenza artificiale nel cinema

L'impiego dell'intelligenza artificiale nel settore cinematografico si è progressivamente esteso a numerose fasi della pipeline audiovisiva, dalla pre-produzione fino alla distribuzione dei contenuti. Inizialmente utilizzata per automatizzare operazioni tecniche e analizzare grandi quantità di dati, negli ultimi anni l'IA ha iniziato a svolgere un ruolo sempre più attivo anche nel supporto ai processi creativi e decisionali.

Per comprendere meglio l'impatto di questa tecnologia è utile richiamare brevemente la struttura della pipeline cinematografica tradizionale.

La realizzazione di un'opera audiovisiva è generalmente suddivisa in tre fasi principali: pre-produzione, produzione e post-produzione, alle quali si aggiungono le attività di distribuzione e promozione.

La pre-produzione comprende tutte le attività preliminari alla realizzazione del film, come lo sviluppo della sceneggiatura, la definizione del concept visivo, la realizzazione dello storyboard, la scelta del cast e la pianificazione delle riprese.

La produzione coincide con la fase delle riprese, sia video che audio.

La post-produzione include tutte le operazioni successive alle riprese, come il montaggio, il *sound design*, la *color correction* e l'integrazione degli effetti visivi.

In ciascuna di queste fasi operano diversi reparti tecnici e creativi che collaborano tra loro sotto la supervisione del regista, il quale mantiene una visione complessiva dell'opera e coordina il lavoro dei vari dipartimenti.

Attualmente, l'intelligenza artificiale si può inserire all'interno di ciascuna di queste fasi, come dimostrerà il progetto di tesi approfondito nel Capitolo 3.

Durante la fase di pre-produzione, ad esempio, diversi strumenti possono supportare lo sviluppo creativo e organizzativo del progetto. I sistemi di generazione di immagini, come Midjourney, consentono di produrre rapidamente storyboard preliminari o concept visivi, mentre modelli linguistici come ChatGPT possono assistere nella stesura delle sceneggiature e velocizzarne lo spoglio.

In ambito industriale, diverse case di produzione hanno iniziato a sperimentare sistemi di analisi predittiva basati su algoritmi di *machine learning*. Un esempio significativo è rappresentato dalla piattaforma *Cinelytic*^[13], utilizzata dalla Warner Bros., che analizza grandi quantità di dati relativi a incassi, recensioni e performance di mercato per stimare il potenziale commerciale di un prodotto cinematografico. Allo stesso modo, strumenti come *ScriptBook*^[14] sono in grado di analizzare automaticamente una sceneggiatura per individuare elementi narrativi e strutturali che potrebbero influenzare il successo o il fallimento commerciale del film. Alcuni sistemi sperimentali arrivano anche suggerire possibili scelte di casting, sebbene tali approcci presentino evidenti limiti pratici ed etici: la macchina non è infatti in grado di valutare pienamente la qualità interpretativa degli attori e può introdurre *bias* derivanti dai dati utilizzati per l'addestramento.

Durante la fase di produzione, l'intelligenza artificiale può generare ambienti fotorealistici in tempo reale, in set virtuali, aiutando i registi e gli attori a pianificare ed interpretare le azioni e le inquadrature. Alcuni di questi strumenti sono, ad esempio, forniti da Unreal Engine ed NVIDIA AI nell'ambito della *virtual production*. Inoltre, l'IA prodotta da Metaphysic ha consentito di ringiovanire in tempo reale Tom Hanks e Robin Wright durante le riprese di *Here* (2024)⁹ di Robert Zemeckis, mostrando il risultato e la performance degli attori (e dei loro *alter ego*) su schermi live.

Nella fase di post-produzione, numerosi software utilizzano algoritmi di IA per automatizzare attività tecniche e ripetitive, come la sincronizzazione automatica di audio e video, la classificazione del girato, la rimozione del rumore di fondo nelle tracce audio e l'applicazione di maschere per il *compositing*.

⁹ B. Edwards, *The \$50 Million Movie Here De-Aged Tom Hanks With Generative AI* (2024) [15]

In alcuni casi, l'IA viene impiegata anche per modificare o ricostruire digitalmente le performance degli attori. Il film *The Irishman* (2019)^[16], ad esempio, ha utilizzato tecnologie digitali avanzate per ringiovanire ed invecchiare i protagonisti e rappresentare le diverse fasi della loro vita senza ricorrere a lunghi processi di trucco protesico. Allo stesso modo, il film *The Brutalist* (2024)^[17] ha impiegato il software *Respeecher*^[18] per migliorare digitalmente l'accento ungherese degli attori americani, operazione che ha suscitato un acceso dibattito sull'uso della tecnologia di *voice cloning*, giudicando non più originali le interpretazioni degli attori.

Infine, l'utilizzo dell'intelligenza artificiale si estende anche alla fase di distribuzione e promozione dei prodotti audiovisivi. Le piattaforme di streaming utilizzano da tempo algoritmi di raccomandazione per analizzare le preferenze degli spettatori e suggerire contenuti personalizzati sulla base delle loro abitudini di visione.

In ambito promozionale, un caso particolarmente interessante riguarda la collaborazione tra 20th Century Fox e IBM per addestrare il supercomputer Watson a riconoscere ciò che spaventa il pubblico. Questo è stato fatto fornendogli centinaia di trailer di film horror per analizzare i momenti di maggiore tensione visiva e sonora e selezionare automaticamente le sequenze più adatte alla realizzazione del trailer del film horror fantascientifico *Morgan* (2016)^[19]. Il vantaggio principale riguarda la riduzione dei tempi, poiché l'intero processo di analisi e produzione è stato eseguito in 24 ore.

Tali applicazioni evidenziano come l'intelligenza artificiale sia un potente strumento capace di supportare e ampliare le possibilità operative e creative dei professionisti del settore. In questa prospettiva un'ampia parte della letteratura concorda nell'interpretare l'IA come un vero e proprio alleato del regista e della troupe, in grado di offrire analisi predittive, automazione tecnica e supporto creativo¹⁰.

2.1.2 Esempi significativi di opere realizzate con l'intelligenza artificiale generativa

Negli ultimi anni, sono sempre più le opere audiovisive nate come sperimentazione delle tecnologie di intelligenza artificiale generativa. Nei lavori qui riportati (i quali rappresentano una minima parte dell'offerta disponibile), l'IA non è stata limitata a svolgere funzioni tecniche di supporto, bensì, è stata utilizzata come parte integrante del processo creativo.

¹⁰ S. Swarnakar, *Artificial Intelligence and Cinema: Exploring the Implications of Artificial Intelligence in Cinema*, (2024) [20]

Uno dei primi esempi significativi è *Sunspring* (2016)^[5], cortometraggio di fantascienza diretto da Oscar Sharp e scritto da Benjamin, intelligenza artificiale sviluppata dal ricercatore Ross Goodwin. Il sistema, basato su una rete neurale ricorrente di tipo LSTM, è stato addestrato su numerose sceneggiature di film e serie televisive di fantascienza e ha generato autonomamente la sceneggiatura, con l'aggiunta di suggerimenti di interpretazione per gli attori. Il progetto è stato realizzato per il concorso *48hr Film Challenge contest of Sci-Fi-London*.

Un approccio simile ma più recente è *The Last Screenwriter* (2024)^[21], diretto da Peter Luisi e basato su una sceneggiatura generata con il modello linguistico GPT-3.5 di OpenAI. Il progetto rappresenta uno dei primi tentativi di realizzare un lungometraggio completamente scritto dall'intelligenza artificiale, per valutarne la validità a fronte di una produzione eseguita successivamente da un team di professionisti.

Entrambi i progetti sono stati criticati a livello di trama e coerenza, definendoli caotici e senza nessi logici.

Spostandosi nel panorama italiano è possibile evidenziare altrettante opere interessanti.

Il cortometraggio *Miss Polly Had a Dolly* (2023)^[22], presentato al *41° Torino Film Festival* e diretto da Pietro Lafiandra, Andrea Rossini, Flavio Pizzorno, è considerato uno dei primi esempi italiani di opera audiovisiva realizzata interamente con strumenti di intelligenza artificiale. Gli stessi autori hanno affermato come ogni scena, voce e idea siano state generate esclusivamente dall'intelligenza artificiale, attraverso un dialogo tra uomini e macchina.

Recentemente, EDI – Effetti Digitali Italiani e HAI – Human & Artificial Imagination hanno prodotto *The Last Image* (2025)^[6], lungometraggio che integra l'intelligenza artificiale in diverse fasi della pipeline produttiva, dalla scrittura alla generazione delle clip video fino al montaggio finale, mantenendo però la supervisione creativa in mano agli autori umani.

Un settore particolarmente fertile per la sperimentazione con l'intelligenza artificiale è quello dell'animazione, nel quale l'utilizzo di strumenti generativi consente di ridurre i costi di produzione mantenendo risultati visivamente convincenti.

Uno dei primi lungometraggi animati, prodotto interamente con strumenti IA, è l'anime horror *DreadClub: Vampire's Verdict* (2024)^[23], diretto da Hooroo Jackson e distribuito su Prime Video. Il film è stato realizzato con un budget estremamente ridotto (circa 405 dollari) e l'IA è stata impiegata in tutte le fasi produttive, dalla scrittura della sceneggiatura fino alla generazione delle clip video e delle musiche. Inoltre, questa è stata utilizzata anche come supporto nella fase di montaggio, suggerendo come tagliare e montare le sequenze. Nonostante alcuni difetti evidenziati dalla critica – come dialoghi estremamente lunghi o limitazioni nella rappresentazione visiva delle scene più intense, tipiche del genere horror – il film riesce comunque a

coinvolgere lo spettatore e dimostra come tali tecnologie possano consentire la realizzazione di opere audiovisive efficaci anche con risorse estremamente limitate.

Nel 2023, OpenAI ha prodotto *Critterz*^[24], primo cortometraggio animato realizzato con Dall-E 2, il suo modello generativo di immagini. Partendo dagli output prodotti dall'IA riguardo personaggi e sfondi, gli animatori li hanno trasformati in sequenze video. Recentemente è stato rimasterizzato in 4K utilizzando Sora, nuovo modello proprietario di OpenAI, in grado di generare anche video a partire da immagini o prompt testuali.

Infine, altre due opere di animazione interessanti sono *To Dear Me*^[25] e *Un Rêve Liquide*^[26], rispettivamente primo e terzo classificato al *Reply AI Film Festival* nelle edizioni 2024 e 2025. Entrambi i lavori presentano narrazioni fortemente oniriche e adottano stili visivi che richiamano tecniche tradizionali, come l'illustrazione ad acquerello e il disegno a penna.

È interessante notare come queste opere presentino estetiche molto differenti tra loro: *DreadClub: Vampire's Verdict* segue lo stile dell'animazione anime giapponese, *Critterz* sfrutta una resa visiva assimilabile alla computer grafica tridimensionale, *To Dear Me* richiama l'estetica dell'acquerello, mentre *Un Rêve Liquide* presenta un tratto grafico simile a un disegno a penna in bianco e nero. Questo dimostra come l'intelligenza artificiale non imponga necessariamente un'estetica uniforme o standardizzata; al contrario, se guidata da una visione autoriale chiara, può essere utilizzata per esplorare una pluralità di stili visivi differenti.



Figura 2.1 Esempi delle opere di animazione citate.

In ordine: “*DreadClub: Vampire’s Verdict*”, “*Critterz*”, “*To Dear Me*”, “*Un Rêve Liquide*”.

2.1.3 Esempi significativi di festival ed iniziative incentrati sull'intelligenza artificiale

La crescente diffusione di opere audiovisive realizzate con l'intelligenza artificiale sta portando alla nascita di molteplici iniziative culturali e festival dedicati, testimoniando come questa tecnologia stia progressivamente acquisendo una crescente legittimazione all'interno del panorama audiovisivo contemporaneo.

A livello internazionale, uno degli esempi più significativi e di riferimento è rappresentato dall'*AI Film Festival (AIFF)*^[27], promosso da Runway AI (che ad oggi offre alcuni tra i principali strumenti per la generazione video).

Un'altra iniziativa di rilievo è l'*AI International Film Festival*^[10], organizzato a Hollywood a cadenza mensile, scelta organizzativa che riflette la rapidità con cui evolvono le tecnologie generative.

Anche nel panorama italiano si possono identificare iniziative importanti. Tra queste spicca il *Reply AI Film Festival*^[9], un concorso internazionale dedicato ai cortometraggi realizzati con strumenti di intelligenza artificiale generativa. Il festival si distingue per la sua forte vocazione inclusiva, consentendo la partecipazione gratuita di artisti e creativi provenienti da tutto il mondo, rappresentando oggi uno dei principali punti di riferimento per il "cinema generativo". Un'altra iniziativa rilevante è *AI.motion*^[28], concorso organizzato dall'Università IULM di Milano in collaborazione con lo IULM AI Lab e la Neural Network Association. Il festival si caratterizza per il suo approccio accademico e interdisciplinare, includendo non solo cortometraggi, ma anche videoclip musicali e spot pubblicitari realizzati con tecnologie di intelligenza artificiale generativa. Una sezione specifica dell'evento è inoltre dedicata allo sviluppo di personaggi digitali e avatar iperrealistici generati tramite sistemi di IA, al fine di esplorare il rapporto tra identità digitale, rappresentazione visiva e recitazione sintetica.

Festival e concorsi dedicati all'intelligenza artificiale non rappresentano soltanto occasioni di visibilità per artisti e sperimentatori, ma contribuiscono anche a stimolare il dibattito teorico e culturale sul ruolo delle tecnologie generative nei processi creativi, favorendo la loro diffusione e la nascita di nuovi stili estetici e narrativi.

2.1.4 Limiti tecnici ed operativi dell'intelligenza artificiale

Nonostante le numerose potenzialità offerte dall'intelligenza artificiale nei processi audiovisivi, le pipeline produttive che integrano questi strumenti generativi non possono ancora essere considerate completamente automatizzate. L'IA generativa viene spesso presentata come un

sistema autonomo capace di sostituire integralmente il lavoro umano ma, al contrario, le applicazioni attuali mostrano limiti tecnici e operativi significativi. L'utilizzo di questi sistemi richiede, infatti, un costante intervento umano per garantire coerenza narrativa, qualità visiva e controllo creativo all'interno del processo produttivo, come verrà dimostrato nel Capitolo 3 attraverso il progetto di tesi.

Uno dei principali limiti riguarda la definizione dei *prompt*, ovvero delle istruzioni testuali utilizzate per guidare i modelli generativi. Attualmente non esiste una metodologia standardizzata per la loro progettazione e l'efficacia delle istruzioni può variare notevolmente in base al modello utilizzato e al tipo di contenuto desiderato. Di conseguenza, il processo di generazione risulta spesso empirico e iterativo: l'autore è costretto a sperimentare diverse formulazioni e ad affinare progressivamente le istruzioni per ottenere risultati soddisfacenti.

Un'ulteriore criticità riguarda la consistenza visiva e narrativa dei contenuti generati. Nel linguaggio cinematografico è fondamentale che personaggi, ambienti e oggetti mantengano caratteristiche coerenti tra diverse inquadrature o scene, così da essere riconoscibili dagli spettatori. Tuttavia, i modelli generativi faticano a mantenere questa continuità nel tempo. Soprattutto durante la generazione di clip video possono verificarsi variazioni nei lineamenti dei personaggi, nelle caratteristiche degli oggetti o nell'illuminazione dell'ambiente. L'utilizzo di immagini di riferimento o tecniche come *l'image-to-video* può mitigare il problema, ma, come verrà approfondito nel prossimo capitolo, non risolverlo del tutto.

Inoltre, i modelli generativi hanno spesso difficoltà nel gestire sia i dettagli complessi (espressioni facciali, mani, sfondi articolati) sia le simulazioni fisiche o le interazioni fra gli oggetti.

Un'altra limitazione significativa riguarda il fenomeno delle “allucinazioni” dei modelli, ovvero la generazione di contenuti non coerenti con le istruzioni fornite. In questi casi l'intelligenza artificiale introduce elementi non richiesti, modifica informazioni o produce risultati non corrispondenti al contesto narrativo previsto, rendendo necessaria una supervisione costante da parte dell'autore.

Infine, un limite strutturale rilevante è rappresentato dalla scarsa integrazione tra i diversi strumenti di intelligenza artificiale. A differenza di una troupe cinematografica tradizionale, nella quale i diversi reparti collaborano in modo coordinato, gli strumenti di IA tendono a funzionare come moduli separati, senza la possibilità di scambiare informazioni fra loro. Ad esempio, l'IA che genera la sceneggiatura non è in grado di dialogare con quella che genera le clip video e questa, a sua volta, non è in grado di generare un audio coerente con quanto rappresentato. Questa frammentazione costringe l'autore umano a svolgere il ruolo di intermediario tra strumenti differenti, traducendo e adattando manualmente gli output prodotti da ciascun sistema per integrarli all'interno di una *pipeline* produttiva coerente.

È da specificare, però, come questo limite stia venendo gradualmente superato grazie a sistemi che integrano più funzionalità: ad esempio, Veo3 è in grado di generare clip video con l'audio sincronizzato e ChatGPT consente la realizzazione di immagini direttamente all'interno della chat testuale.

Attualmente, però, l'intelligenza artificiale, pur rappresentando uno strumento estremamente potente per la produzione audiovisiva, richiede ancora un supervisione umana significativa. L'autore non viene sostituito dalla macchina, ma assume piuttosto il ruolo di coordinatore e curatore del processo generativo, selezionando, correggendo e integrando i risultati prodotti dagli algoritmi.

A questi limiti tecnici si affiancano anche questioni di natura etica e giuridica, come la presenza di *bias* nei modelli e le problematiche legate al diritto d'autore, affrontate nei paragrafi successivi.

2.2 Intelligenza artificiale e creatività: dibattiti ed implicazioni etiche

L'introduzione di sistemi capaci di generare contenuti audiovisivi ha sollevato numerosi interrogativi riguardo alla natura della creatività, al ruolo dell'autore umano e alle implicazioni etiche derivanti dall'utilizzo di tali tecnologie.

Uno dei temi più discussi riguarda la possibilità che sistemi algoritmici possano essere considerati strumenti creativi a tutti gli effetti oppure semplici mezzi tecnici al servizio dell'autore umano. Parallelamente, emergono problematiche legate al diritto d'autore e ai possibili *bias* (ovvero distorsioni più o meno volontarie presenti nei dati utilizzati per l'addestramento dei modelli) che possono influenzare gli output con stereotipizzazioni.

La letteratura affronta in modo approfondito queste problematiche, brevemente discusse nei paragrafi seguenti.

È interessante aggiungere anche un paradosso culturale legato al rapporto tra cinema e intelligenza artificiale: da un lato, l'industria cinematografica utilizza sempre più frequentemente strumenti di IA per supportare le fasi della produzione audiovisiva, dall'altro lato, il cinema stesso ha spesso rappresentato l'intelligenza artificiale come una tecnologia pericolosa o minacciosa per l'essere umano¹¹. Questa rappresentazione contribuisce a rafforzare una visione

¹¹ C. López Frías, *The Paradox of Artificial Intelligence in Cinema*, (2024)[29]

ambivalente dell'intelligenza artificiale: da una parte strumento di innovazione tecnologica e creativa, dall'altra fonte di rischi culturali, sociali ed etici.

2.2.1 Creatività e ruolo dell'autore

La capacità dei modelli generativi di produrre immagini, testi, musiche e video ha portato a riflessioni sulla natura della creatività e sul ruolo dell'autorialità nelle opere realizzate con il supporto dell'intelligenza artificiale.

La questione chiave riguarda la possibilità che l'IA possa effettivamente “creare”. Infatti, l'essere umano sviluppa la creatività attraverso esperienze vissute, riferimenti culturali e interpretazioni soggettive della realtà; al contrario, le IA producono nuovi contenuti attraverso l'analisi di dati e pattern statistici appresi durante l'addestramento, unendo queste informazioni alle richieste dell'utente¹².

La letteratura evidenzia un paradosso: opere generate dagli algoritmi delle intelligenze artificiali sono in grado di produrre emozioni negli spettatori, pur non possedendo la stessa intenzionalità artistica che avrebbe, invece, un autore umano. Questo solleva interrogativi complessi anche dal punto di vista dell'autorialità, poiché, se un contenuto generato dall'IA è in grado di essere apprezzato al pari di un'opera umana, a chi appartiene la sua paternità? All'autore del prompt, agli sviluppatori del modello, agli artisti le cui opere sono state utilizzate nei dataset di addestramento o all'IA stessa? Per rispondere a questo dubbio, molti autori sottolineano come il valore creativo di un'opera derivi soprattutto dal processo progettuale e decisionale che guida la sua realizzazione, piuttosto che dalla semplice generazione automatica di immagini o contenuti.

Nella letteratura sono inoltre presenti molti studiosi (antecedenti all'epoca dell'IA ma le cui intuizioni restano ancora valide) che evidenziano come le tecnologie non rappresentino semplici strumenti neutri, ma contribuiscano a trasformare il modo in cui i contenuti vengono creati. Marshall McLuhan^[30], ad esempio, sostenne che ogni nuovo *medium* modifica il linguaggio e le forme espressive della cultura, influenzando profondamente i processi di produzione artistica. Allo stesso modo, Lev Manovich^[31] individua nei media digitali alcune caratteristiche fondamentali – tra cui modularità, automazione e variabilità, riconducibili senza dubbio anche alle attuali tecnologie IA – che consentono ai contenuti di essere generati, modificati e ricombinati attraverso processi computazionali. In questa prospettiva, l'intelligenza artificiale può essere interpretata come una naturale evoluzione delle logiche dei media digitali.

¹² G. Grassi, *Intelligenza artificiale e produzione cinematografica: analisi del processo creativo e della sua evoluzione*. (2024) [19]

Tuttavia, le opinioni all'interno dell'industria cinematografica rimangono divergenti. Alcuni registi, come Guillermo del Toro o Steven Spielberg, esprimono scetticismo nei confronti della capacità dell'intelligenza artificiale di produrre autentiche opere artistiche, sostenendo che la creatività derivi dall'esperienza e dalla sensibilità umana. Altri autori, invece, interpretano l'IA come uno strumento capace di ampliare le possibilità espressive del cinema. I registi Anthony e Joe Russo, ad esempio, hanno ipotizzato che le tecnologie generative possano consentire in futuro la realizzazione di nuove forme di narrazione cinematografica difficilmente ottenibili con gli strumenti tradizionali¹³.

Nel 2023, però, gli scioperi degli sceneggiatori e degli attori negli Stati Uniti hanno evidenziato le preoccupazioni legate all'utilizzo dell'IA nel settore audiovisivo, tra cui il rischio di sostituzione di alcune professioni creative, l'utilizzo non autorizzato delle immagini o delle voci degli attori e la perdita di controllo autoriale sulle opere prodotte. Allo stesso tempo, la letteratura e le sperimentazioni condotte dimostrano come l'intelligenza artificiale non sia una "bacchetta magica" capace di sostituire il lavoro umano, bensì uno strumento potente che può ampliare le possibilità operative e creative degli autori¹⁴. Infatti, automatizzando alcune fasi della produzione, l'IA consente di ridurre tempi e costi, permettendo agli autori di esplorare rapidamente diverse soluzioni narrative e visive.

In questa prospettiva, il rapporto tra intelligenza artificiale e creatività può essere interpretato non come una contrapposizione tra uomo e macchina, bensì, come una forma di co-creazione, dove l'algoritmo accelera i processi e moltiplica le possibilità creative, mentre l'autore umano seleziona gli output e attribuisce intenzionalità espressiva all'opera.

2.2.2 Democratizzazione della produzione

Abbattendo le barriere tecniche ed economiche tradizionali, l'intelligenza artificiale consente ad un numero sempre maggiore di utenti di sperimentare con la creazione di contenuti artistici ed audiovisivi. Questo fenomeno viene definito "democratizzazione" della produzione audiovisiva.

Tradizionalmente, infatti, la realizzazione di un'opera audiovisiva richiedeva il coordinamento di numerose figure professionali specializzate e l'impiego di attrezzature e risorse economiche considerevoli. Oggigiorno, invece, è possibile realizzare contenuti qualitativamente elevati con team ridotti o da parte di singoli autori, anche senza la necessità di conoscenze tecniche

¹³ C. López Frías, *The Paradox of Artificial Intelligence in Cinema*, (2024) [29]

¹⁴ S. Swarnakar, *Artificial Intelligence and Cinema: Exploring the Implications of Artificial Intelligence in Cinema*, (2024) [20]

pregresse. L'intelligenza artificiale sembra quindi rappresentare una risorsa significativa per il cinema indipendente e a basso budget, poiché consente di sopperire alla mancanza di risorse (tempo, finanziamenti, personale, strumenti tecnici) senza rinunciare alla qualità del prodotto. Inoltre, la possibilità di sperimentazione artistica anche da parte di utenti “inesperti” può favorire la nascita di nuove forme artistiche e linguaggi audiovisivi alternativi, meno vincolati ai modelli produttivi e narrativi dominanti nell'industria cinematografica tradizionale.

Allo stesso tempo, la diffusione di questi strumenti sta contribuendo alla nascita di nuovi ruoli professionali legati alla produzione con IA, come le figure dei *prompt artist* o *AI filmmaker*, professionisti che combinano competenze artistiche e conoscenze tecniche nell'utilizzo dei modelli generativi. Si può notare, quindi, come l'intelligenza artificiale non elimini necessariamente le competenze richieste nel settore creativo, ma contribuisca piuttosto a trasformarle, introducendo nuove modalità di progettazione e gestione del processo creativo.

Tuttavia, la democratizzazione della produzione audiovisiva potrebbe gradualmente portare ad un'omologazione di stili visivi e narrativi, con il rischio che le opere generate con IA presentino caratteristiche estetiche analoghe. I modelli generativi, infatti, di fronte a richieste non precisamente dettagliate tendono a riprodurre una “media statistica” delle immagini e delle strutture narrative apprese durante l'addestramento o in base ai feedback forniti dagli utenti. Se gli autori si limitano ad utilizzare le impostazioni di default degli strumenti senza una direzione artistica precisa, il risultato può essere una proliferazione di contenuti esteticamente simili, caratterizzati da schemi visivi ricorrenti e da una ridotta originalità espressiva.



Figura 2.2 Fornendo a ChatGPT il generico prompt “image of a cat singing”, l'IA ha proposto un'estetica realistica di un gatto cantante, in formato 9:16 adatto ai social media. Questo suggerisce un condizionamento derivante dai nuovi trend social, dove animali che cantano canzoni famose stanno diventando virali.

Si evidenzia quindi, come la possibilità tecnica di generare immagini, video o musica non garantisca automaticamente la qualità o l'originalità delle opere prodotte. La vera sfida risiede nell'utilizzare consapevolmente gli strumenti generativi, sfruttandoli come supporto al processo creativo piuttosto che come sostituti dell'autorialità.

2.2.3 Proprietà intellettuale e diritto d'autore

Un ulteriore nodo critico nel dibattito riguarda l'attribuzione del diritto d'autore alle opere prodotte mediante sistemi generativi. La possibilità che contenuti audiovisivi, immagini o testi vengano generati da algoritmi solleva infatti interrogativi su chi debba essere considerato il vero autore dell'opera: il creatore del modello, gli autori delle opere utilizzate in fase di addestramento, l'utente che fornisce il prompt, oppure lo stesso sistema di intelligenza artificiale?

Secondo le normative vigenti nell'Unione Europea e in Italia, il diritto d'autore può essere attribuito esclusivamente ad una persona fisica. Di conseguenza, le opere realizzate con strumenti di IA generativa possono essere tutelate dal diritto d'autore solo quando è presente un contributo umano significativo (ad esempio, attraverso la progettazione di prompt complessi o *editing* aggiuntivo dopo la generazione). Secondo l'*AI Act (Regolamento UE 2024/1689)*^[32], in caso di tale contributo umano significativo, l'opera appartiene all'autore umano, ma il suo utilizzo commerciale rimane soggetto ad eventuali violazioni di copyright in caso, ad esempio, emergano riferimenti non autorizzati ad opere altrui.

La questione si complica ulteriormente considerando i *dataset* utilizzati per l'addestramento dei modelli IA. Durante la fase di *training*, questi vengono esposti a grandi quantità di dati (come immagini, testi, materiali audio e video) per apprendere pattern statistici e relazioni tra gli elementi presenti nei contenuti. Se all'interno di questi *dataset* sono inclusi materiali protetti da copyright e non vi è il consenso dei rispettivi autori, possono emergere violazioni dei diritti di proprietà intellettuale nei confronti delle aziende che sviluppano tali modelli. La *Direttiva Copyright UE 2019/790 (artt. 15-17)*^[33], ad esempio, cede la responsabilità alle piattaforme per i contenuti caricati e generati dagli utenti (devono essere le piattaforme stesse, quindi, a impedire l'utilizzo di opere protette), oltre a vietare anche il *training* non autorizzato su opere con copyright. Inoltre, i principi stabiliti dalla *Convenzione di Berna*^[34] per la protezione delle opere letterarie e artistiche continuano a rappresentare uno dei principali riferimenti per la tutela del diritto d'autore a livello internazionale.

Un ulteriore ambito di discussione riguarda l'utilizzo delle tecnologie di *deepfake*, ovvero contenuti audiovisivi manipolati o generati artificialmente che riproducono il volto o la voce di persone reali. L'*AI Act*^[32] introduce specifici obblighi di trasparenza per questo tipo di contenuti, richiedendo che siano chiaramente identificabili come materiali generati o modificati artificialmente. In ambito cinematografico, tali tecnologie possono essere utilizzate legittimamente con il consenso degli attori coinvolti, ad esempio per ringiovanire digitalmente i volti degli interpreti, come nel caso del film *The Irishman*^[16], o per modificare alcuni aspetti della performance, come avvenuto nel film *The Brutalist*^[17]. L'utilizzo non autorizzato dell'immagine o della voce di una persona è invece vietato dalla normativa; emblematico è il caso dell'attore Morgan Freeman, che ha denunciato pubblicamente l'uso non autorizzato di una voce sintetica paragonabile alla sua in alcuni contenuti pubblicitari.

2.2.4 Bias algoritmici e distorsioni generate

Un'ulteriore criticità deriva dalla possibile presenza di *bias* algoritmici nelle opere generate con l'intelligenza artificiale. Queste distorsioni sono causate da squilibri, lacune o pregiudizi presenti nei dati di addestramento dei modelli, i quali apprendono e poi ripropongono ciò che hanno imparato.

Se i dati utilizzati per l'addestramento contengono rappresentazioni sbilanciate di determinate categorie sociali o culturali, il modello tenderà a replicare tali schemi nei propri output. Un esempio riportato nello studio *How human-AI feedback loops alter human perceptual, emotional and social judgements* di Glickman e Sharot¹⁵ riguarda i sistemi di generazioni di immagini: addestrando i modelli su dataset nei quali i manager erano rappresentati prevalentemente come uomini bianchi, occidentali e vestiti formalmente, è stato osservato come i modelli tendessero a riprodurre lo stesso schema nelle immagini generate, mostrando difficoltà nel rappresentare manager appartenenti ad altre etnie o con abbigliamento informale.

Lo stesso studio, inoltre, dimostra come gli utenti percepiscano gli output generati dalle IA come più affidabili rispetto a quelli prodotti da esseri umani, considerando gli strumenti algoritmici neutrali e completamente oggettivi. Questa percezione può portare gli individui ad accettare con maggiore facilità contenuti distorti o stereotipati, rischiando di contribuire alla riproduzione e alla diffusione di stereotipi culturali, amplificando così gli squilibri già esistenti piuttosto che correggerli. Questa dinamica può inoltre facilitare la diffusione di informazioni fuorvianti o di fake news.

¹⁵ M. Glickman, Sharot T, *How human-AI feedback loops alter human perceptual, emotional and social judgements*, (2024) [35]

Lo studio prosegue analizzando il *feedback loop* tra esseri umani ed intelligenza artificiale. In questo processo circolare, i *bias* presenti nella società vengono inizialmente trasferiti nei *dataset* utilizzati per addestrare i modelli. Successivamente, i sistemi di IA tendono a riprodurre (e talvolta amplificare) tali distorsioni nei contenuti generati. Infine, gli utenti, esposti a questi risultati percepiti come affidabili, interiorizzano tali rappresentazioni e possono contribuire involontariamente a rafforzare gli stessi pregiudizi alla base del problema. In questo modo si crea un circolo nel quale i *bias* sociali alimentano i *dataset*, i *dataset* influenzano gli algoritmi e gli algoritmi restituiscono contenuti che consolidano ulteriormente tali distorsioni.

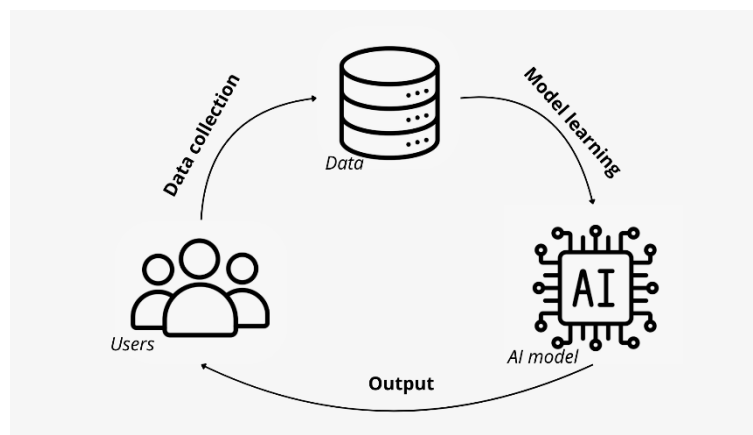


Figura 2.3 Schema del feedback loop.

I bias risiedono nelle fasi di raccolta dati, apprendimento e presentazione del risultato.

Questi fenomeni dimostrano come l'intelligenza artificiale non possa essere considerata uno strumento neutrale o completamente oggettivo, ma piuttosto un sistema che riflette le caratteristiche sociali e culturali da cui apprende. Per questo motivo, uno sviluppo responsabile delle tecnologie di IA richiede un'attenta selezione dei *dataset*, un monitoraggio costante dei risultati prodotti e una valutazione critica degli output generati, al fine di ridurre il rischio di riproduzione o amplificazione di *bias* sociali e culturali.

2.2.5 Il fenomeno dell'Uncanny Valley

Un ulteriore criticità riguarda il fenomeno dell'*Uncanny Valley* (“valle perturbante”). Il concetto fu introdotto nel 1970 dal professore di robotica Masahiro Mori^[36], il quale osservò come la percezione umana nei confronti di robot o rappresentazioni artificiali dell'essere umano non aumenti in modo lineare con il grado di realismo. Secondo Mori, man mano che un oggetto artificiale assume caratteristiche simili a quelle umane, la familiarità percepita dall'utente

tende ad aumentare ma, quando il livello di realismo diventa molto elevato e non perfetto, si verifica un improvviso calo nella risposta emotiva positiva dell'utente, generando sensazioni di disagio o inquietudine. Solo quando la somiglianza con l'essere umano diventa quasi indistinguibile dalla realtà la percezione torna ad essere positiva.

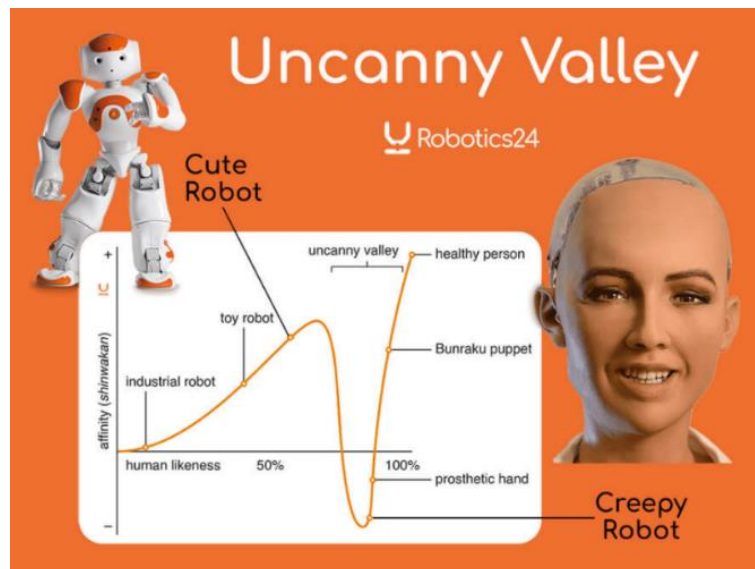


Figura 2.4 Schema del fenomeno Uncanny Valley riportato dal blog Robotics24.

Il fenomeno è stato osservato in diversi ambiti tecnologici, tra cui robotica, animazione digitale e realtà virtuale. In particolare, avatar o personaggi generati artificialmente, che presentano caratteristiche umane molto realistiche ma imperfezioni nei movimenti, nelle espressioni facciali o nello sguardo possono risultare disturbanti per gli osservatori. Studi successivi hanno collegato questa reazione a un conflitto percettivo nel cervello umano: quando un volto o un corpo appare quasi umano ma presenta anomalie visive o comportamentali, il sistema cognitivo fatica a classificare correttamente lo stimolo, generando una risposta di disagio o diffidenza¹⁶. La presenza di piccole anomalie nei movimenti o nelle espressioni facciali produce quindi una dissonanza tra le aspettative del sistema visivo e la percezione reale dello stimolo.

¹⁶ B. Marr, *The Uncanny Valley: Advancements And Anxieties Of AI That Mimics Life*, (2024) [37]



Figura 2.5 Il film di animazione “*The Polar Express*” (2004) è ritenuto da molti critici un esempio di *Uncanny Valley* a causa della presenza di umani “perfetti ma imperfetti”.

La capacità dei sistemi generativi di produrre figure umane ad alto realismo ha riaperto il dibattito sul fenomeno dell’*Uncanny Valley*. Spesso, infatti, l’imperfetta aderenza al vero dei loro output genera una sensazione di inquietudine anziché di empatia. Per ovviare a tale criticità, molti autori prediligono l’adozione di uno stile cartoon e meno realistico, riducendo così le possibilità di generare un effetto di repulsione.

Recentemente, alcuni studi (ad esempio, *Too human to model: the Uncanny Valley of large language models in simulating human systems*¹⁷, di Zeng, Brown e Rounsevell, pubblicato a marzo 2026) hanno evidenziato come dinamiche simili possano emergere anche nell’interazione linguistica con sistemi basati su modelli di linguaggio avanzati (i *Large Language Model* o LLM, come, ad esempio, ChatGPT o Claude). Quando queste IA mostrano comportamenti molto simili a quelli umani ma presentano incoerenze logiche o limiti nella comprensione sociale, gli utenti possono percepire una forma di *Uncanny Valley*, caratterizzata da una sensazione di artificialità o mancanza di autenticità nelle interazioni. In questi casi non è tanto l’aspetto visivo a generare disagio, quanto la percezione di una simulazione umana incompleta o poco naturale.

¹⁷ Y. Zeng et al., *Too human to model: the Uncanny Valley of large language models in simulating human systems*, (2026) [38]

2.3 Strumenti di intelligenza artificiale per la produzione audiovisiva

Oggi, il panorama degli strumenti di intelligenza artificiale generativa adatti per la produzione audiovisiva si presenta frammentato ed in continua espansione, con la nascita quasi quotidiana di strumenti nuovi o aggiornati. In vista della fase di sperimentazione descritta nel [Capitolo 3](#), risulta quindi utile presentare una panoramica dei principali strumenti attualmente disponibili, al fine di contestualizzare e motivare le scelte operative adottate nel progetto. È importante precisare che l'obiettivo di questa analisi non è individuare un unico strumento “migliore” in senso assoluto, bensì identificare, per ciascuna fase della pipeline produttiva, gli strumenti più adatti alle specifiche esigenze del progetto.

La letteratura scientifica ed i report di settore propongono numerosi *benchmark* e studi comparativi che valutano i modelli di intelligenza artificiale sulla base di differenti parametri, come la qualità dell'output, la coerenza visiva o narrativa, il grado di controllo offerto all'utente e l'accessibilità economica delle piattaforme. Questi criteri costituiscono un punto di riferimento utile per orientarsi all'interno dell'ecosistema degli strumenti generativi e selezionare quelli più pertinenti rispetto agli obiettivi della sperimentazione.

2.3.1 Approcci operativi: online o offline

Prima di tutto è fondamentale delineare una distinzione metodologica e architetturale che caratterizza l'attuale panorama dell'intelligenza artificiale generativa. Esistono infatti due principali approcci operativi per l'utilizzo di questi modelli: l'approccio online (basato su *cloud*) e l'approccio offline (basato sull'esecuzione in locale). Questa distinzione influisce direttamente sull'infrastruttura tecnologica richiesta, sul livello di controllo dell'utente, sui costi operativi e sulla gestione dei dati durante il processo di produzione.

Con l'approccio online, i modelli generativi risiedono su server remoti gestiti direttamente dalle aziende sviluppatrici (i *provider*). L'utente interagisce con l'intelligenza artificiale tramite interfacce web, inviando i propri prompt ai server che eseguono il processo di generazione e restituiscono l'output risultante. Questo è il modello adottato da molte delle principali piattaforme commerciali, come OpenAI (ChatGPT e Sora), Midjourney o Runway. Uno dei principali vantaggi di questo approccio consiste nella possibilità di accedere a modelli costantemente aggiornati senza dover disporre di macchine performanti; l'elaborazione computazionale viene, infatti, gestita interamente dall'infrastruttura del *provider*, rendendo possibile l'utilizzo degli strumenti anche su dispositivi relativamente semplici, come gli smartphone. Tuttavia, questo

approccio presenta anche alcune limitazioni: è necessaria una connessione internet stabile e i dati inviati vengono trasmessi a server esterni, con potenziali rischi in termini di privacy e gestione dei contenuti sensibili.

L'approccio offline, invece, prevede che i modelli (spesso distribuiti come software open-source) vengano scaricati ed eseguiti localmente sul computer dell'utente. Tra i software più apprezzati vi è ComfyUI^[39], per la generazione di immagini, video e modelli 3D. La sua caratteristica chiave è l'interfaccia a nodi, che offre all'utente una gestione estremamente precisa e dettagliata di ogni parametro tecnico.

Questo approccio offre diversi vantaggi, tra cui un maggiore controllo sul processo di generazione, una gestione completamente locale dei dati e la possibilità di personalizzare o addestrare i modelli in base alle proprie esigenze. Tuttavia, questo approccio richiede risorse hardware considerevoli e presenta generalmente una curva di apprendimento più ripida rispetto alle piattaforme online.

Considerando questi aspetti, per il progetto di tesi è stato scelto di adottare un approccio online, utilizzando piattaforme accessibili tramite interfaccia web. Questa scelta è stata motivata sia dalla maggiore accessibilità degli strumenti sia dalla volontà di sperimentare il workflow realistico per un utente medio, permettendo così di osservare in modo diretto le dinamiche di democratizzazione della produzione audiovisiva discusse nei paragrafi precedenti.

2.3.2 I benchmark: criteri e metriche di valutazione

Per comprendere la complessità della scelta degli strumenti di intelligenza artificiale, è necessario, prima di tutto, inquadrare le dimensioni e l'evoluzione del mercato attuale. Secondo il report annuale pubblicato dallo Stanford Institute for Human-Centered Artificial Intelligence¹⁸, il numero complessivo di strumenti generativi disponibili pubblicamente è cresciuto in modo estremamente rapido negli ultimi anni, arrivando a raddoppiare rispetto all'anno precedente. Una parte significativa di questa espansione riguarda tecnologie direttamente collegate alla produzione audiovisiva, come i modelli per la generazione di video e musica.

Questa rapida crescita riflette un cambiamento profondo nella natura degli strumenti disponibili: negli ultimi anni si è assistito alla comparsa di piattaforme IA sempre più specializzate in singoli settori della pipeline creativa. Alcuni strumenti si concentrano sulla scrittura e sull'analisi delle sceneggiature, altri sulla generazione di immagini o video, mentre altri ancora supportano le attività di *editing* e creazione di effetti visivi.

Allo stesso tempo, però, si può osservare anche un processo di convergenza multimodale, nel

¹⁸ Stanford University, *The 2025 AI Index Report*, (2025) [40]

quale piattaforme come Sora o Veo3.1 cercano di integrare in un'unica interfaccia più modalità, combinando le generazioni di immagini, audio e video.

All'interno di questo panorama in rapida espansione, la valutazione delle prestazioni dei diversi modelli rappresenta una sfida significativa. In ambito scientifico e industriale, questa viene generalmente effettuata attraverso l'utilizzo dei *benchmark*, ovvero insiemi di test e compiti standardizzati progettati per misurare e confrontare le capacità dei diversi sistemi di intelligenza artificiale.

Tuttavia, nel campo dell'IA generativa, questi si ritrovano ad essere rapidamente obsoleti a causa della velocità di evoluzione dei modelli. Inoltre, la grande varietà di architetture e applicazioni rende difficile definire parametri universalmente validi. Tutti i *benchmark*, però, concordano su alcuni elementi di valutazione comuni, come la qualità dell'output (ad esempio, un video fotorealistico o un testo coeso e coerente) e la consistenza e coerenza degli elementi. Inoltre, viene valutato anche il grado di controllo offerto dal modello e la sua capacità di interpretare correttamente tutte le istruzioni fornite dagli utenti. Spesso vengono valutati anche la facilità d'uso, la curva di apprendimento ed il costo di eventuali abbonamenti.

Molti *benchmark*, inoltre, vengono costruiti sulla base di recensioni degli utenti stessi. Un esempio sono le "AI Arena", dove viene chiesto agli utenti di votare quale output (testo, immagine, video) è considerato migliore, al fine di assegnare punteggi alle IA e valutare le loro prestazioni. Una delle più famose e significativa per la costruzione dei *benchmark* è disponibile sul sito di Artificial Analysis^[41].

Poiché i *benchmark* ufficiali richiedono tempo per essere applicati ed analizzati e spesso si ritrovano quasi in ritardo rispetto alla rapida evoluzione dei modelli, sono nati anche *benchmark*, informali ma efficaci, grazie alla comunità creativa. Tra questi, uno dei più noti è *Will Smith eating spaghetti*¹⁹, nato nel 2023 come esperimento per testare le capacità dei modelli *text-to-video* utilizzando un viso noto ed un'azione semplice. Nonostante l'apparente semplicità, questo rappresenta uno stress test particolarmente complesso per i modelli generativi. Il sistema, infatti, deve riconoscere e mantenere stabili i tratti somatici dell'attore, simulare correttamente la fisica di oggetti flessibili come gli spaghetti e gestire l'interazione tra diversi oggetti nello spazio, come mani, forchetta, cibo e movimenti del viso.

Mantenendo invariato il prompt e testandolo con i vari strumenti e le loro versioni aggiornate, si sono notati grandi miglioramenti. Le prime versioni producevano risultati distorti e surreali, caratterizzati da deformazioni dei volti e movimenti innaturali. I modelli più recenti, invece, sono in grado di generare sequenze credibili e coerenti, dal punto di vista visivo e fisico.

¹⁹ BBC Bitesize, *How Will Smith eating spaghetti became the 'test' of AI video*, (2026) [42]



Figura 2.6 Alcuni esempi del benchmark “Will Smith eating spaghetti”. In ordine, una delle prime generazioni (2023), la generazione di Google Veo 3.1 (2025) e la recente generazione di Kling 3.0 (2026).

Infine, in modo analogo, risultano estremamente preziosi i test e le recensioni condivisi su community come Reddit o YouTube, dove gli utenti, dopo aver sperimentato i modelli, offrono le loro opinioni e suggerimenti.

2.3.3 Strumenti di intelligenza artificiale per la generazione testuale

I modelli di intelligenza artificiale dedicati alla generazione di testo rappresentano una delle applicazioni più diffuse e mature dell’IA generativa. Questi sistemi, basati su modelli linguistici di grandi dimensioni (i *Large Language Models* o LLM), sono in grado di analizzare e produrre testi complessi in linguaggio naturale.

Negli ultimi anni, l’evoluzione di questi modelli ha portato alla nascita di numerose piattaforme commerciali e open-source con caratteristiche e prestazioni differenti. Nei paragrafi successivi verranno quindi analizzati il loro funzionamento generale e alcuni dei principali *benchmark* esistenti.

Funzionamento degli strumenti

I *Large Language Models* (LLM) sono progettati per analizzare, comprendere e generare testi utilizzando complesse architetture di *deep learning*, addestrate su enormi quantità di dati testuali. Grazie a questo processo di apprendimento, gli LLM sono in grado di produrre testi coerenti, sintetizzare informazioni e supportare attivamente le attività di scrittura, incluse, nel caso di questo progetto di tesi, le generazioni di contenuti narrativi e sceneggiature.

Alla base di questi modelli vi sono le reti neurali: strutture informatiche ispirate al funzionamento del cervello umano, composte da numerosi nodi interconnessi che elaborano informazioni attraverso pesi e funzioni di attivazione. Durante la fase di addestramento, il modello analizza i *dataset* forniti e modifica progressivamente i propri parametri interni per ridurre il margine di errore tra il testo generato e l’output atteso. In questo modo, la rete apprende e riconosce

pattern linguistici, strutture sintattiche e relazioni semantiche.

Affinché la macchina possa elaborare il linguaggio naturale, le parole devono però essere convertite in un formato numerico comprensibile all' algoritmo. Questo processo avviene attraverso tecniche definite *word embeddings*, che mappano ciascun termine in un vettore numerico all'interno di uno spazio multidimensionale. In questo spazio semantico, parole con significati o contesti d'uso simili risultano geometricamente vicine tra loro, permettendo al modello di calcolare le distanze concettuali tra i termini.

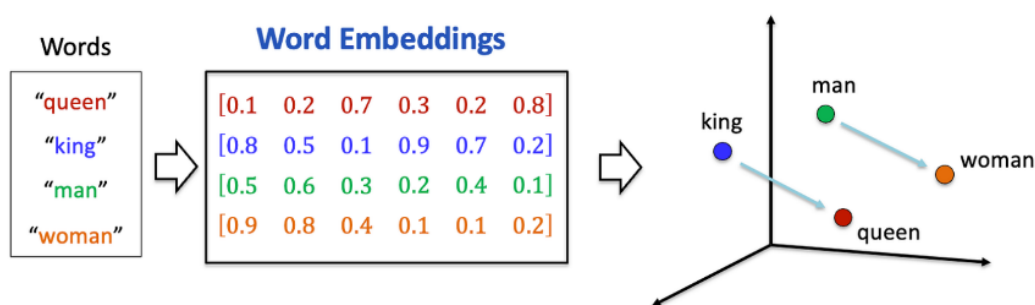


Figura 2.7 Rappresentazione semplificata dello spazio semantico dei word embeddings.

Le prime architetture impiegate per l'elaborazione testuale si basavano su Reti Neurali Ricorrenti (RNN) e sulle loro evoluzioni (come le LSTM). Questi modelli analizzavano le sequenze di parole in modo lineare e sequenziale, cercando di mantenere una "memoria" del contesto precedente. Pur essendo utili per compiti di base come la traduzione automatica, mostravano gravi limiti computazionali e perdite di coerenza nella gestione di testi molto lunghi.

Una svolta importante è avvenuta nel 2017, con la pubblicazione del paper *Attention Is All You Need* di Vaswani et al.²⁰, che ha introdotto l'architettura Transformer. A differenza dei modelli sequenziali, i Transformer utilizzano un meccanismo di *self-attention* (auto-attenzione). Questo sistema permette al modello di analizzare simultaneamente tutte le parole di una frase, assegnando un "peso" diverso a ciascuna di esse in base alla loro rilevanza nel contesto globale.

Gli attuali LLM leader del mercato online (come ChatGPT, Claude e Gemini) si basano su quest'ultime architetture Transformer, scalate su miliardi di parametri. Il loro funzionamento, pur simulando un ragionamento logico, si fonda su un principio puramente statistico: la predizione della parola successiva (*next-token prediction*). Ricevuto un testo iniziale (il *prompt*), il modello calcola la distribuzione di probabilità dei termini successivi e genera iterativamente una sequenza testuale coerente.

Nel contesto specifico della produzione audiovisiva, questa capacità rende gli LLM strumenti

²⁰ A. Vaswani et al., *Attention Is All You Need*, (2017) [43]

eccezionali per la generazione di idee, la stesura di dialoghi e lo spoglio delle sceneggiature, pur essendo il loro output il risultato di una correlazione statistica e non il frutto di un'autentica coscienza autoriale.

Benchmark degli strumenti

Secondo i *benchmark* attualmente disponibili, l'efficacia di un modello per la generazione testuale non dipende solamente dalla quantità di dati con cui è stato addestrato, ma, soprattutto, da fattori qualitativi, come la capacità di rispettare i vincoli narrativi indicati dagli utenti e la qualità dell'output che viene percepita; a questi, si aggiunge anche il costo di un eventuale abbonamento.

Nel contesto del progetto di tesi, le IA per la generazione testuale verranno utilizzate per l'analisi dei dati raccolti durante gli eventi tenuti da *Narrazioni Parallele* (per approfondimenti sul contesto, si rimanda al [Capitolo 1](#)) e, successivamente, per la generazione di concept narrativi e per la scrittura della sceneggiatura. Inoltre, saranno utilizzate anche come “assistenti virtuali” durante tutto il *workflow*.

Per individuare, quindi, lo strumento più adatto, si è ricorso all'analisi e comparazione di due sistemi di *benchmarking*, con l'obiettivo di integrare dati eterogenei per ottenere una solida visione d'insieme.

Per valutare la capacità dei modelli di realizzare strutture narrative complesse, si è utilizzato *LLM Creative Story-Writing Benchmark V4*²¹, uno stress test progettato per valutare le abilità di scrittura creativa dei modelli linguistici. Questo prevede la generazione di un breve racconto (tra le 600 e le 800 parole) nel quale devono essere integrati organicamente dieci elementi narrativi obbligatori, tra cui: contesto spazio-temporale, motivazione dei personaggi e specifici vincoli tematici. La valutazione degli output viene poi affidata ad un gruppo di sette modelli linguistici indipendenti che analizzano i testi secondo due criteri: il 60% del punteggio si basa sull'impatto emotivo e la qualità del racconto, mentre il restante 40% è relativo al rispetto dei vincoli dati con il prompt.

Le classifiche sono in continuo aggiornamento, rispettando la rapida evoluzione dei modelli di intelligenza artificiale. Attualmente, si può notare come le IA dotate di funzionalità di “ragionamento avanzato” abbiano ottenuto i punteggi migliori, dimostrando una maggiore abilità nel combinare creatività narrativa e rispetto delle istruzioni.

²¹ L. Mazur, *Writing: LLM Creative Story-Writing Benchmark V4*, (2025) [44]

I modelli migliori, dal punto di vista narrativo, risultano essere il recente Claude Opus 4.6 (di Anthropic) e GPT-5.2 (di OpenAI)

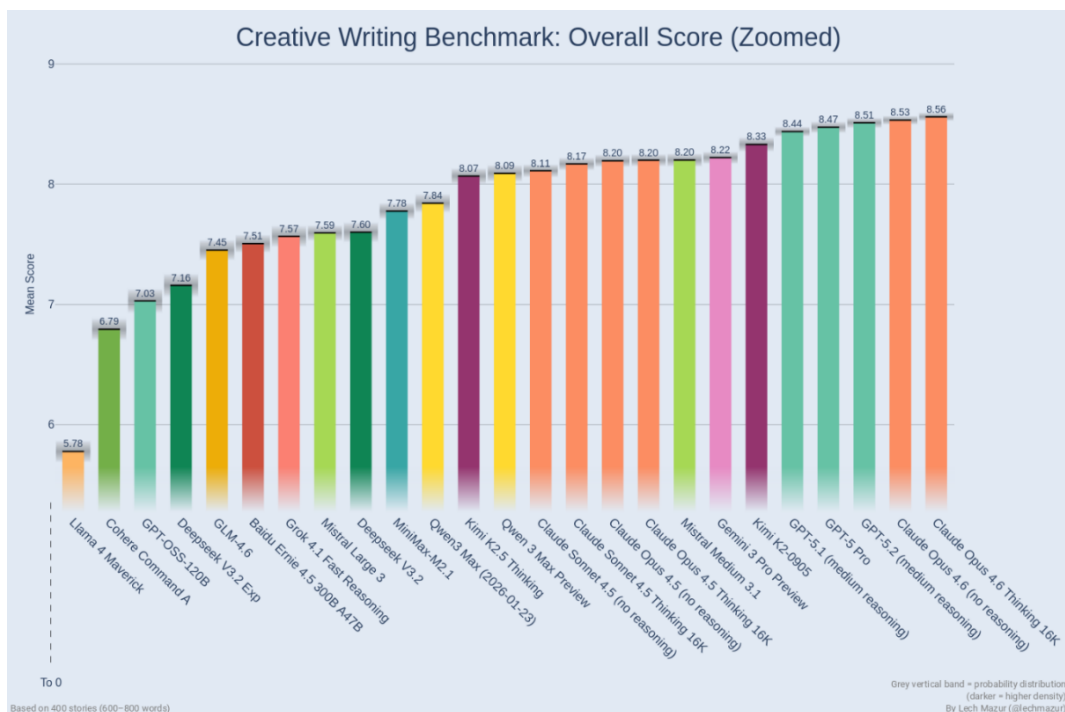


Figura 2.8 Classifica complessiva degli strumenti di IA per la generazione di testi.

La piattaforma dispone di molteplici grafici a supporto della ricerca, tra cui una *heatmap* riportante i singoli punteggi ottenuti per ogni domanda di analisi. Viene specificato come le domande Q1-Q8 si concentrino sulla narrazione (profondità, motivazioni, coerenza), mentre le domande Q9A-Q9J controllano che siano inseriti, in modo organico, tutti gli elementi richiesti.

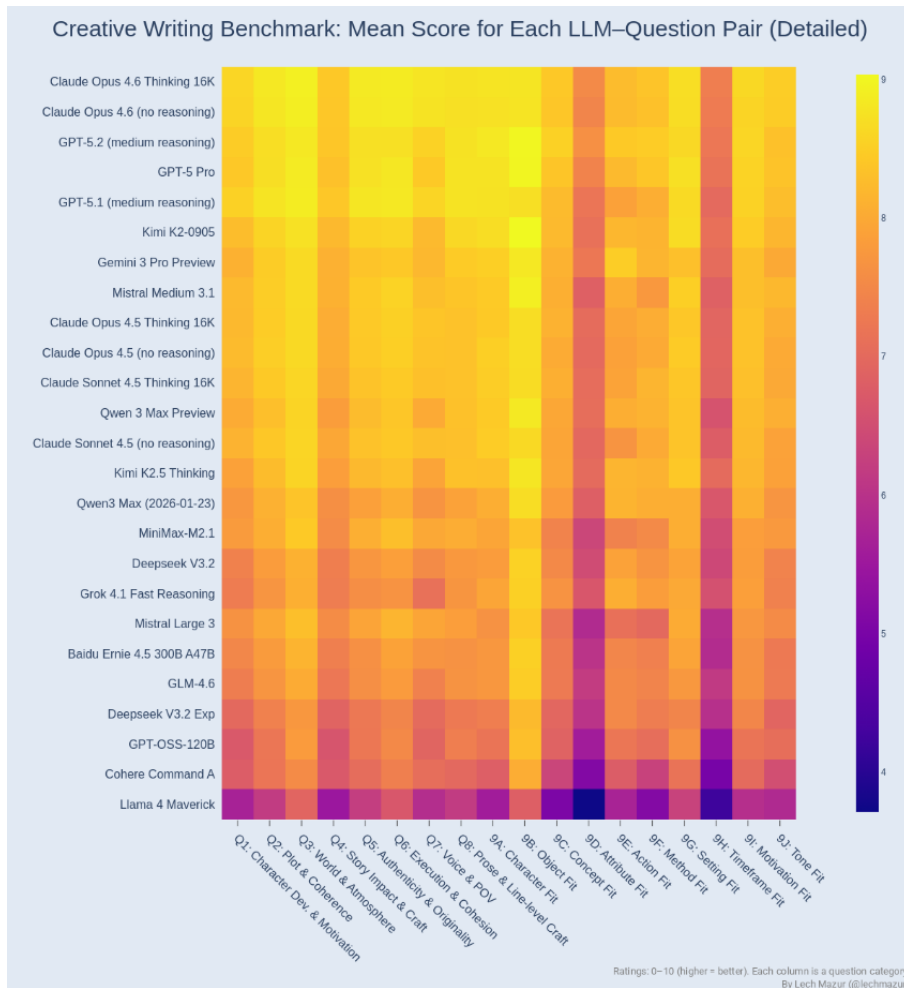


Figura 2.9 Heatmap dei punteggi ottenuti dalle IA per ogni domanda.

Il secondo *benchmark* preso in considerazione è la *Chatbot Arena Leaderboard* di *LMSYS*^[45], dove si analizza la naturalezza conversazionale dei modelli linguistici tramite test A/B, eseguiti per confrontare le loro capacità. Ad ogni modello viene assegnato un punteggio (definito *ELO*) sulla base di una valutazione eseguita da utenti umani (i quali, si specifica, non sanno quali modelli stanno confrontando).

I modelli migliori, dal punto di vista conversazionale, risultano essere il recente Claude Opus 4.6 (di Anthropic) e Gemini-3.1 (di Google, ma disponibile solamente come anteprima).

Text 4 days ago

Rank	Model	Score
1	AI claude-opus-4-6	1504
2	AI claude-opus-4-6-thinking	1502
3	G gemini-3.1-pro-preview	1500 (i)
4	XI grok-4.20-beta1	1491 (i)
5	G gemini-3-pro	1485
6	GPT gpt-5.4-high	1479 (i)
7	GPT gpt-5.2-chat-latest-20260210	1479
8	G gemini-3-flash	1473
9	XI grok-4.1-thinking	1473
10	AI claude-opus-4-5-20251101-thinking-32k	1470

Figura 2.10 Classifica degli strumenti di IA per la generazione di testi.

I modelli indicati con il simbolo (i) hanno avuto un accesso in anteprima e non sono ancora disponibili in tutti i Paesi.

In seguito all'analisi di questi *benchmark*, si è scelto di sperimentare i modelli di ChatGPT, Claude e Gemini, al fine di valutare il migliore per il progetto di tesi. È inoltre importante specificare come si siano scelti strumenti in grado di svolgere anche un compito conversazionale e di assistenza generale, escludendo, quindi, piattaforme pensate specificatamente per la sola scrittura di sceneggiature.

2.3.4 Strumenti di intelligenza artificiale per la generazione di immagini

I modelli di intelligenza artificiale per la generazione di immagini rappresentano una delle applicazioni più interessanti delle tecnologie generative nel campo creativo. Questi sistemi sono in grado di produrre immagini a partire da descrizioni testuali o da materiali visivi di riferimento.

Nel contesto della produzione audiovisiva, la rapidità con cui è possibile generare varianti dello stesso contenuto consente agli autori di esplorare molteplici soluzioni visive in tempi ridotti, rendendoli adatti anche alle fasi preliminari di creazione di *concept art*, raccolta di reference e realizzazione di storyboard.

Nei paragrafi successivi verranno quindi analizzati il funzionamento generale di questi modelli ed alcuni dei principali *benchmark* esistenti.

Funzionamento degli strumenti

I moderni sistemi *text-to-image* sono in grado di decodificare descrizioni testuali (i *prompt*) per sintetizzare immagini complesse sulla base delle indicazioni ricevute. Questo è reso possibile dall'impiego di modelli di *deep learning*, addestrati su enormi *dataset* composti da coppie di immagini associate alle rispettive descrizioni testuali. Durante il processo di *training*, la rete neurale apprende le correlazioni tra gli elementi visivi e il linguaggio naturale, acquisendo la capacità di generare nuove immagini coerenti con le indicazioni fornite dall'utente.

Nel corso dell'ultimo decennio, la ricerca ha sviluppato diversi approcci architetturali per affrontare questo compito, tra cui i *Variational Autoencoders* (VAE), le *Generative Adversarial Networks* (GAN) e, più recentemente, i *Diffusion Models*, che rappresentano l'attuale stato dell'arte del settore.

I *Variational Autoencoders* (VAE) costituiscono uno dei primi approcci architetturali di successo nell'ambito delle reti neurali generative. Il sistema si fonda su un'architettura bipartita, dove un *Encoder* (Codificatore) analizza l'immagine originale e la comprime in una densa rappresentazione numerica e probabilistica, definita "spazio latente", mentre un *Decoder* (Decodificatore) ricostruisce l'immagine a partire dalle coordinate matematiche risultanti. Questo processo consente al modello di mappare e apprendere la struttura statistica delle immagini presenti nel dataset.

Terminata la fase di addestramento, il sistema è in grado di generare immagini inedite campionando nuovi punti all'interno dello spazio latente.

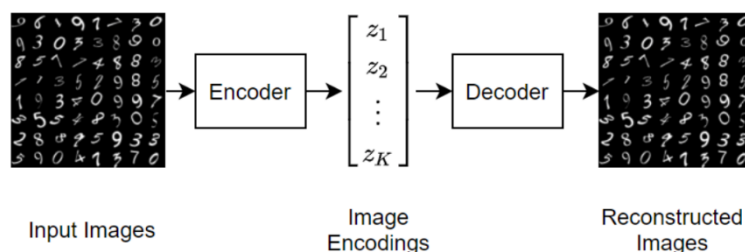


Figura 2.11 Funzionamento dei VAE.

Tuttavia, a causa della natura probabilistica della ricostruzione e delle metriche matematiche utilizzate per calcolare l'errore, i VAE presentano un limite importante: la tendenza a generare

output visivamente morbidi o leggermente sfocati, mancanti di definizione nei micro-dettagli e nitidezza.

Per ovviare al problema, nel 2014 è stata introdotta l'architettura delle *Generative Adversarial Networks* (GAN). Questa si basa su una dinamica competitiva (simile a un gioco a somma zero) tra due reti neurali distinte, il *Generator* (Generatore) che sintetizza immagini artificiali partendo da vettori di rumore casuale ed il *Discriminator* (Discriminatore) che agisce come un “critico” il cui scopo è quello di esaminare le immagini e stabilire se siano reali (ovvero, prelevate dal *dataset*) o false (ovvero, create dal generatore).

Durante l'addestramento iterativo, il generatore impara progressivamente a produrre immagini sempre più fotorealistiche nel tentativo di ingannare il discriminatore. Questo processo competitivo costringe quindi la rete a creare risultati estremamente dettagliati e nitidi.

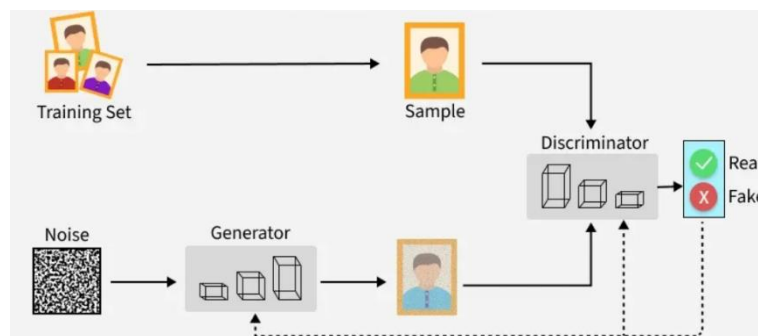


Figura 2.12 Funzionamento delle GAN.

Nonostante l'eccellente qualità visiva raggiunta, le architetture GAN possiedono alcune criticità importanti: risultano fortemente instabili durante la fase di addestramento e sono soggette al fenomeno del *mode collapse* (ovvero, il modello “collassa” su sé stesso quando individua una tipologia di immagine in grado di ingannare il discriminatore e si limita a generare varianti di quel medesimo soggetto, azzerando la diversità creativa).

Recentemente, i *Diffusion Models* hanno superato le architetture precedenti, diventando la tecnologia dominante alla base delle odierne piattaforme leader di mercato (come Stable Diffusion, Midjourney e DALL-E). Durante l'addestramento (fase di *forward diffusion*), il modello impara ad aggiungere progressivamente rumore gaussiano a un'immagine reale, iterazione dopo iterazione, fino a distruggerne completamente l'informazione visiva. Allo stesso tempo, la rete neurale viene addestrata a eseguire l'operazione inversa (*reverse diffusion*).

Quando si chiede al modello di generare un'immagine, il sistema parte da un quadro composto interamente da rumore inorganico e applica una serie di passaggi di “pulizia” (*denoising*) progressiva. Guidato dai vettori semantici del prompt testuale, il modello “scolpisce”

iterativamente i pixel fuori dal rumore, fino a far emergere un'immagine coerente con la descrizione fornita.

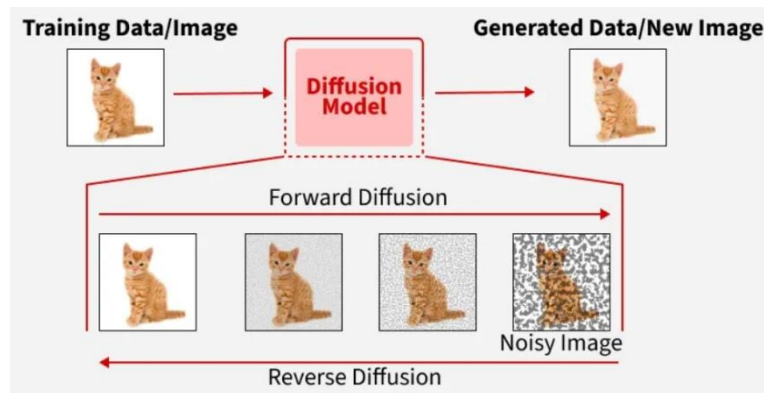


Figura 2.13 Funzionamento dei Diffusion Models.

Grazie a questo approccio graduale e iterativo, i *Diffusion Models* permettono di ottenere immagini ad altissima risoluzione, garantendo una stabilità di calcolo superiore e una vastissima diversità estetica, superando quindi i limiti di VAE e GAN.

Benchmark degli strumenti

Come si evince dal loro funzionamento, la valutazione dei modelli per la generazione di immagini coinvolge parametri complessi. Tra questi vi sono la fedeltà al prompt dato, la coerenza stilistica e di personaggi, la credibilità e la corretta interpretazione di proporzioni e posizioni.

Nel contesto del progetto di tesi, le IA per la generazione di immagini verranno utilizzate per la realizzazione degli storyboard e la creazione dei personaggi.

Per individuare lo strumento più adatto, si è ricorso allo stesso approccio adottato per le IA testuali: l'analisi e comparazione di due sistemi di *benchmarking*, con l'obiettivo di integrare dati eterogenei per ottenere una solida visione d'insieme.

Per quanto riguarda la qualità estetica e l'interpretazione corretta dei prompt, uno dei principali riferimenti è stato la sezione *text-to-Image* della *LMSYS Arena*^[45], già utilizzata per le IA testuali.

I modelli migliori risultano essere il recente Nano Banana 2 (di Google, ma disponibile solamente in anteprima) e GPT image 1.5 (di OpenAI).

Text-to-Image 5 days ago

Rank	Model	Score
1	gemini-3.1-flash-image-preview (nano-ba... (i)	1268
2	gpt-image-1.5-high-fidelity	1248
3	gemini-3-pro-image-preview-2k (nano-ban... (i)	1236
4	gemini-3-pro-image-preview (nano-banana... (i)	1233
5	reve-v1.5	1177
6	grok-image-image (i)	1175
7	flux-2-max	1167
8	grok-image-image-pro (i)	1162
9	flux-2-flex	1160
10	gemini-2.5-flash-image-preview (nano-ba... (i)	1156

Figura 2.14 Classifica degli strumenti di IA per la generazione di immagini.

I modelli indicati con il simbolo (i) hanno avuto un accesso in anteprima e non sono ancora disponibili in tutti i Paesi.

È stato inoltre considerato anche il benchmark messo a disposizione da Artificial Analysis^[41] attraverso i dati ottenuti dalla Image Arena. I modelli sono classificati tramite un punteggio ELO, ottenuto sulla base delle votazioni degli utenti all'interno dell'Arena; il procedimento è il medesimo approfondito per le IA di generazione testuale.

In questo caso, i modelli migliori risultano essere GPT image 1.5 (di OpenAI) ed il recente Nano Banana 2 (di Google, ma disponibile solamente in anteprima), seguito da Nano Banana Pro.

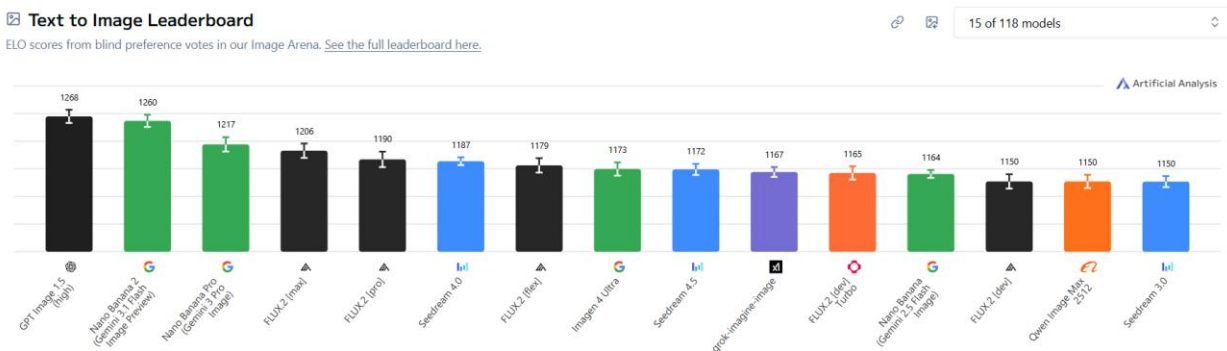


Figura 2.15 Classifica degli strumenti di IA per la generazione di immagini.

In seguito all'analisi di questi *benchmark*, si è scelto di sperimentare i modelli GPT e Nano Banana Pro.

2.3.5 Strumenti di intelligenza artificiale per la generazione di video

I modelli di intelligenza artificiale per la generazione video rappresentano uno degli sviluppi più recenti e in rapida evoluzione nel campo dell'IA generativa. Questi sistemi consentono di produrre sequenze video a partire da descrizioni testuali, immagini statiche o combinazioni di questi.

Nei paragrafi successivi verranno quindi analizzati il funzionamento generale di questi modelli e alcuni dei principali *benchmark* esistenti.

Funzionamento degli strumenti

Prima di tutto, è necessario ricordare che un video viene elaborato come una sequenza di immagini ordinate nel tempo. Di conseguenza, i modelli per la generazione video non possono gestire unicamente la coerenza spaziale e compositiva del singolo frame, ma devono calcolare e mantenere la coerenza temporale lungo l'intera sequenza video. Questo implica un costo computazionale generalmente più elevato rispetto alla generazione di immagini statiche.

La ricerca ha progressivamente adattato al dominio video le architetture originariamente concepite per le immagini, esplorando l'uso di VAE, GAN e *Diffusion Models*.

I primi tentativi di generazione video hanno visto l'utilizzo e adattamento dell'architettura VAE (*Variational Autoencoders*). Con questo approccio, il modello viene addestrato a generare un flusso di fotogrammi in cui la sintesi di ogni frame è matematicamente condizionata dallo stato del frame precedente, al fine di preservare la continuità logica e temporale. Sistemi avanzati come lo *Stochastic Video Generation* (SVG) introducono poi variabili latenti temporali specifiche, utili a modellare l'incertezza del movimento e le variazioni fisiche tra i frame, cercando così di restituire dinamiche più organiche e realistiche.

In seguito, sono state testate le *Generative Adversarial Networks* (GAN). In questo caso, il *Generator* è incaricato di produrre intere sequenze di frame, mentre il *Discriminator* non si limita più a valutare il fotorealismo della singola immagine, ma analizza e giudica la fluidità e la coerenza del moto tra frame consecutivi.

Inoltre, architetture specifiche come *VideoGAN* adottano una complessa struttura a due vie (*two-stream*), separando l'elaborazione del *background* statico (l'aspetto visivo) dal *foreground*

dinamico (la generazione del movimento). Questo stratagemma computazionale consente di modellare dinamiche d'azione complesse, pur mantenendo un'elevata nitidezza.

Infine, l'introduzione dei *Diffusion Models* ha portato alla recente rivoluzione audiovisiva in ambito IA. In queste architetture applicate alla generazione video, il modello non “pulisce” dal rumore una singola immagine, ma applica il processo di *denoising* su interi blocchi di frame. Il sistema apprende a ricostruire iterativamente sia l'integrità visiva della scena sia la sua corretta evoluzione nel tempo.

Questo approccio rappresenta il funzionamento delle odierne piattaforme leader di mercato (come Runway, Sora, Pika e Veo).

Benchmark degli strumenti

Come si evince dal loro funzionamento, la valutazione dei modelli per la generazione video coinvolge parametri ancor più complessi di quelli per la generazione di immagini. Infatti, oltre alla fedeltà al prompt dato, la coerenza stilistica e dei personaggi, la credibilità e la corretta interpretazione di proporzioni e posizioni, i modelli devono essere in grado di mantenere coerenza temporale attraverso i diversi frame.

Per individuare lo strumento più adatto al progetto di tesi, si è ricorso allo stesso approccio già adottato: l'analisi e la comparazione di tre sistemi di *benchmarking*, con l'obiettivo di integrare dati eterogenei per ottenere una solida visione d'insieme.

Uno dei principali strumenti utilizzati in ambito accademico per valutare i modelli di generazione video è *VBench*²², un framework progettato da Huang Y. et al. per analizzare le prestazioni dei modelli *text-to-video* e *image-to-video* attraverso diversi parametri qualitativi, come la consistenza dei personaggi e degli sfondi, la dinamicità e la capacità di gestire più oggetti contemporaneamente. Inoltre, vengono analizzati anche i *bias* introdotti dei modelli, come *skin bias* o *gender bias*.

Negli ultimi anni sono stati rilasciati anche *VBench++* e *VBench+++*, in grado di analizzare in modo sempre più approfondito i diversi modelli e modalità.

La ricerca condotta da Huang Y. et al. presenta numerosi dati interessanti, suddivisi in grafici diversi per tematica e tipologia di modelli (locale/online, *text-to-video/image-to-video*). Si riporta solamente il grafico complessivo per i modelli online, poiché si è scartata in precedenza l'opzione di utilizzare modelli in locale. Per ulteriori approfondimenti si rimanda all'interessante ricerca presente sul sito, il cui link è presente al numero [46] della Sitografia.

²² Y. Huang et al., *VBench: A Comprehensive Benchmark for Video Generation*. (2023) [46]

VBench Evaluation Results of Closed Source Models

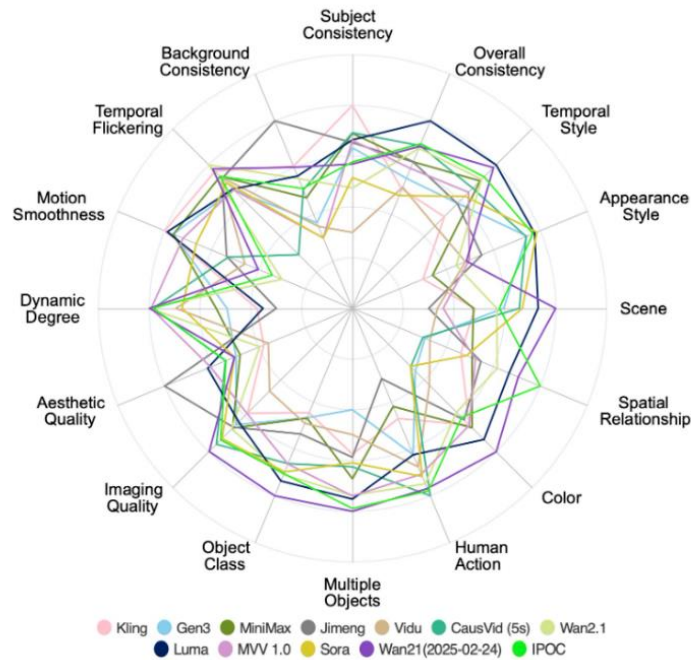


Figura 2.16 Risultati degli strumenti di IA per la generazione video.

Come per i modelli di generazione di immagini, è stato considerato anche il *benchmark* messo a disposizione da Artificial Analysis^[41] attraverso i dati ottenuti dalla *Video Arena*. I modelli sono classificati tramite un punteggio ELO, ottenuto sulla base delle votazioni degli utenti all'interno dell'*Arena*; il procedimento è il medesimo approfondito per le IA di generazione testuale.

In questo caso è possibile notare la presenza di più classifiche, differenziate in base alla tipologia di modello (*text-to-video* oppure *image-to-video*) e alla presenza o meno di audio integrato all'interno della clip generata.

Prima di proseguire, risulta opportuno precisare la differenza fra le due modalità sopracitate, così da poter apprezzare al meglio i dati riportati. *Text-to-video* (T2V) è il processo di generazione di un video a partire da un prompt testuale; *image-to-video* (I2V) è il processo di generazione di un video a partire da un'immagine di partenza (di solito, il frame iniziale) accompagnato da un prompt testuale che ne descrive l'animazione. All'interno della modalità I2V è spesso possibile inserire anche un'immagine relativa al frame finale, così da vincolare maggiormente l'animazione e ottimizzare la coerenza visiva e narrativa.

Tornando ai dati del *benchmark*, i modelli migliori *text-to-video* con audio incorporato risultano essere i recenti Kling 3.0 e Veo 3.1.

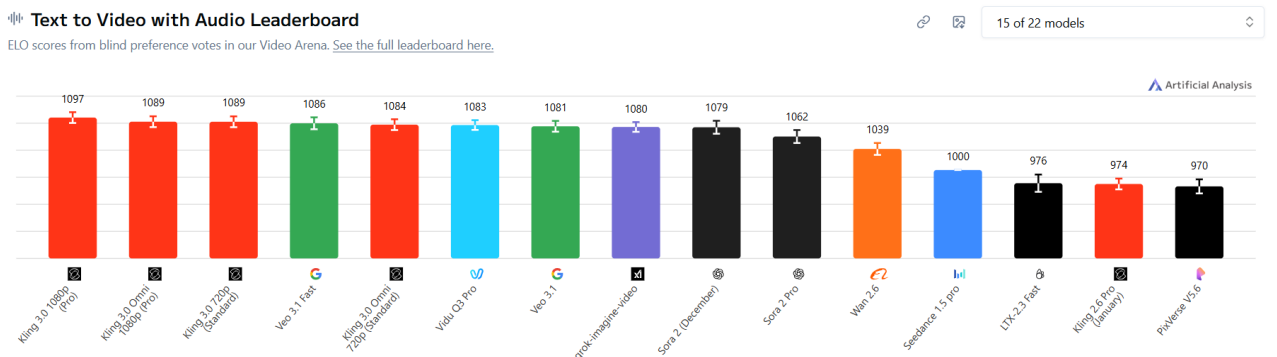


Figura 2.17 Classifica degli strumenti di IA per la generazione *text-to-video* con audio.

I migliori modelli *text-to-video* senza audio incorporato risultano essere Kling 3.0, Grok e Runway Gen-4.5.

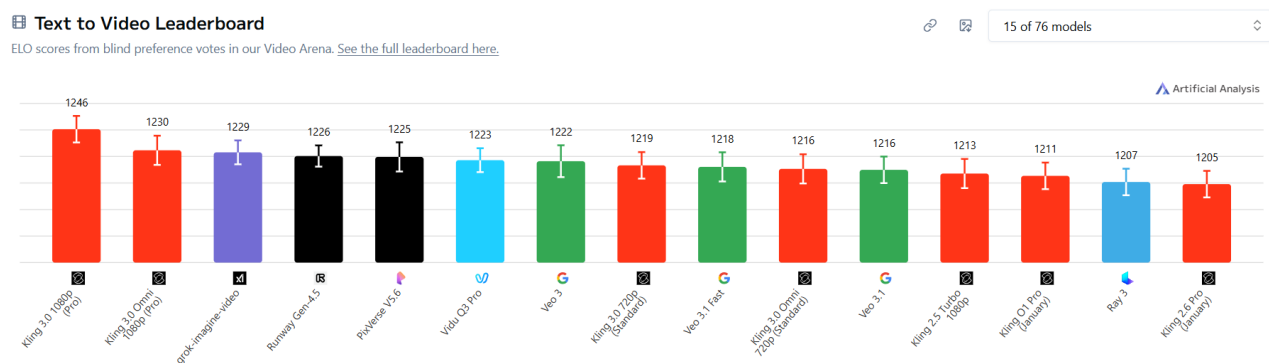


Figura 2.18 Classifica degli strumenti di IA per la generazione *text-to-video* senza audio.

I migliori modelli *image-to-video* con audio incorporato risultano invece essere Veo 3.1, Grok e Kling 3.0.

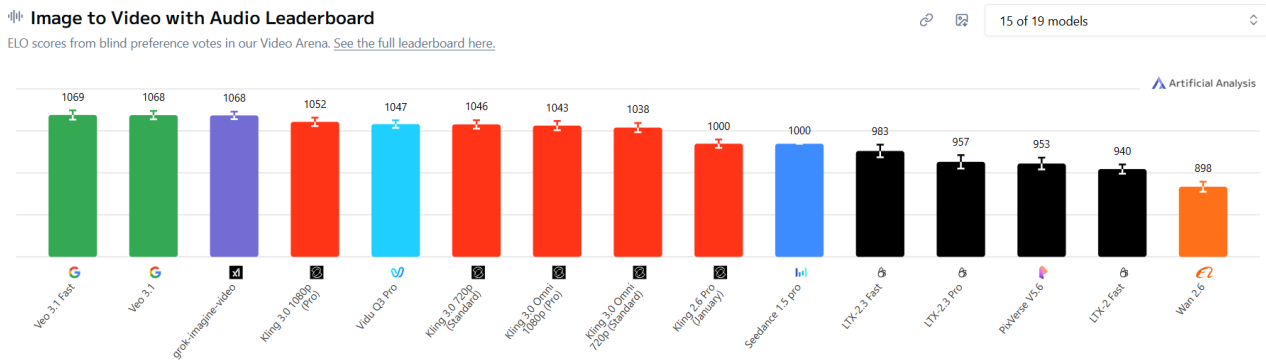


Figura 2.19 Classifica degli strumenti di IA per la generazione image-to-video con audio.

Infine, i migliori modelli *image-to-video* senza audio incorporato risultano essere Grok e Kling 2.5 e PixVerse.

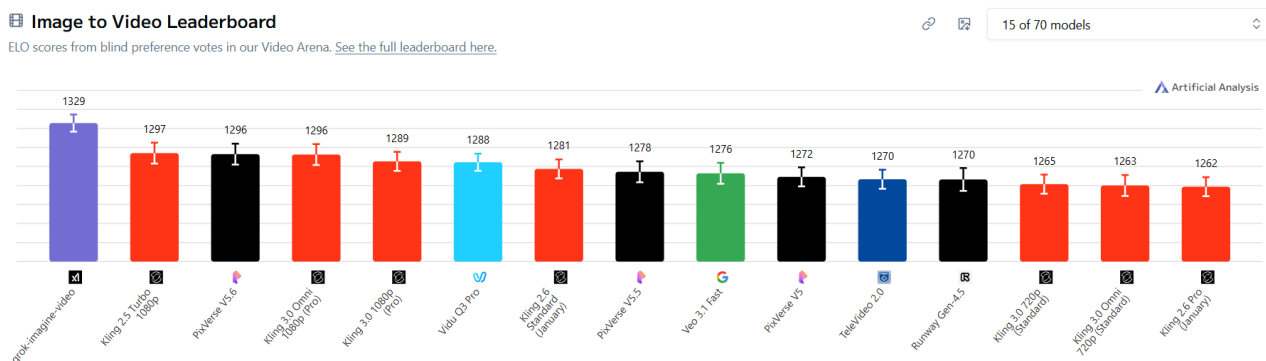


Figura 2.20 Classifica degli strumenti di IA per la generazione image-to-video senza audio.

Poiché la generazione video sarà la parte più importante del progetto di tesi, si è preso in considerazione un terzo *benchmark* proposto da ApiFrame^[47], il quale considera diversi fattori tra cui: coerenza ed affidabilità degli output, rapporto qualità-prezzo e facilità di utilizzo. Da questa analisi emergono nuovi modelli non riportati nelle classifiche precedenti, come Sora2 (non ancora disponibile in Italia) e Pika 2.0, IA attualmente in tendenza per la realizzazione di video ed effetti per piattaforme social ma con un ottimo punteggio per resa visiva, coerenza temporale e rispetto del prompt.

In seguito all'analisi di questi *benchmark*, si è quindi scelto di sperimentare i modelli Sora, Runway, Kling, Veo e Pika.

2.3.6 Strumenti di intelligenza artificiale per la generazione audio

Un ulteriore ambito di sviluppo dell'intelligenza artificiale generativa è costituito dai modelli dedicati alla generazione audio, in grado di produrre composizioni musicali originali, replicare o sintetizzare voci umane e creare effetti sonori a partire da descrizioni testuali o input sonori di riferimento.

Nei paragrafi successivi verranno quindi analizzati il funzionamento generale di questi modelli e alcuni dei principali *benchmark* esistenti.

Funzionamento degli strumenti

I modelli per la generazione sonora partono dalle architetture vincenti per la generazione di testo e immagini: modelli basati su architetture Transformer, le reti GAN e i *Diffusion Models*, che elaborano l'audio trasformandolo temporaneamente in immagini visive bidimensionali (gli spettrogrammi).

Nell'ambito della generazione musicale, i sistemi di intelligenza artificiale vengono addestrati su enormi *dataset* di brani per imparare le regole armoniche, melodiche e ritmiche.

Successivamente, la generazione si divide in due approcci distinti. In primo luogo, vi è la “rappresentazione simbolica”, dove il modello non genera il suono reale, ma uno “spartito digitale” (come i file MIDI). L'IA decide quali note suonare e la loro durata, lasciando a sintetizzatori esterni il compito di tradurre i dati in suono udibile. In secondo luogo, vi è la “generazione audio diretta”, dove i modelli operano sul segnale grezzo, generando *waveform* (onde sonore) complete e complesse, con la possibilità di inserire strumenti multipli, riverberi e tracce vocali cantate.

Lo stesso tipo di addestramento lo si ha per le IA specializzate nella generazione di effetti sonori. Addestrati su enormi librerie di suoni ambientali, *foley* cinematografici e rumori di scena, i modelli *text-to-audio* permettono di sintetizzare il suono partendo da una descrizione testuale.

Infine, vi è il campo della sintesi vocale, molto delicato a causa della naturale sensibilità dell'orecchio umano verso le inflessioni dei propri simili. I moderni sistemi di *text-to-speech* si compongono tipicamente di due fasi: un modello linguistico che analizza il testo per estrapolarne la fonetica, e un *Vocoder* neurale incaricato di tradurre queste informazioni in un'onda sonora credibile, replicando timbro, ritmo e pause respiratorie.

L'evoluzione di questi modelli è il *voice cloning* (clonazione vocale) dove, sfruttando architetture di apprendimento *zero-shot*, i sistemi sono in grado di replicare fedelmente il timbro unico

e le caratteristiche vocali di una persona specifica analizzando un campione audio anche di pochissimi secondi.

Benchmark degli strumenti

Infine, è da evidenziare come *benchmark* ufficiali siano quasi inesistenti nel campo della generazione audio (musica, voci ed effetti sonori). Questo rappresenta una possibile criticità per la ricerca ma, soprattutto, un paradosso: infatti, come si evincerà dal Capitolo 3, i modelli per la generazione audio risultano attualmente più maturi rispetto ai modelli per la generazione video, ma sono supportati da una ricerca minore.

Vista la mancanza di dati ufficiali, si sono analizzati anche alcuni forum Reddit, al fine di reperire informazioni e recensioni degli utenti. Questi segnalano ElevenLabs come il miglior generatore di voci e SunoAI come il miglior generatore di musica, affiancato da Loudly e Udio.

Come *benchmark* ufficiale è stato preso in considerazione ApiFrame^[47], utilizzato anche nell'analisi dei *benchmark* per la generazione video. Questo segnala come IA migliori SunoAI, Udio ed ElevenLabs, seguite da Loudly.

Tuttavia, è necessario segnalare la necessità di separare le piattaforme per la generazione musicale da quelle per la generazione vocale, poiché operano su metriche differenti.

Infine, è stato preso in considerazione anche il lavoro *Benchmarking Music Generation Models and Metrics via Human Preference Studies*²³ di Grötschla. et al. del 2025, dove si propone uno dei primi *benchmark* per la generazione musicale, basato su preferenze umane. Sono state valutati 12 modelli di generazione musicale e sono state generate 6000 canzoni, confrontate da un vasto pubblico di utenti con strumenti paragonabili a quelli delle *IA Arena* approfondite per strumenti di generazione di testo, immagini e video.

Per ulteriori approfondimenti si rimanda all'interessante ricerca il cui link è presente al numero [48] della Sitografia.

Sulla base delle informazioni ottenute, si è scelto di testare SunoAI e Loudly per la generazione musicale ed ElevenLabs per la generazione vocale e di effetti sonori.

²³ F. Grötschla et al., *Benchmarking Music Generation Models and Metrics via Human Preference Studies*, (2025) [48]

2.3.7 Strumenti di intelligenza artificiale per la post-produzione

Oltre alla generazione di contenuti, l'intelligenza artificiale trova la sua applicazione anche nelle fasi di post-produzione. In questo contesto, gli algoritmi non generano necessariamente nuovi contenuti, ma analizzano e migliorano il materiale esistente, automatizzando alcune operazioni tecniche onerose in termini di tempo e passaggi.

Tra le applicazioni più diffuse rientrano strumenti di montaggio automatico, come Recut^[49], in grado di analizzare l'audio del girato e suggerire tagli nei momenti di silenzio. Inoltre, alcuni software di *editing* professionale, come DaVinci Resolve Studio, integrano funzionalità basate su IA in grado di suggerire cambi di camera, sincronizzare audio e video o generare maschere automatiche per il *compositing* delle scene.

Un altro ambito rilevante riguarda la *color correction* automatica, in cui gli algoritmi analizzano l'immagine e suggeriscono bilanciamenti e regolazioni così da uniformare le diverse clip.

Tuttavia, è da precisare come nel contesto di questo progetto di tesi gli strumenti automatici non fossero in grado di interpretare correttamente le esigenze narrative o di gestire le anomalie visive presenti nei video generati. Nonostante ciò, alcune funzionalità offerte da DaVinci Resolve Studio si sono dimostrate utili ed in grado di velocizzare i processi di post-produzione.

2.3.8 Sintesi delle analisi sui benchmark

Come si evince dai *benchmark* trattati, il panorama degli strumenti di intelligenza artificiale generativa è molto ampio e in continua crescita. Nel prossimo capitolo, partendo dai dati analizzati verranno testati i modelli reputati migliori, così da individuare quelli più adatti per il progetto. Come già specificato, infatti, non si tenterà di trovare gli strumenti migliori in assoluto, bensì, gli strumenti più adatti alle precise esigenze della pipeline produttiva.

Nella pagina seguente si riporta una tabella riassuntiva degli strumenti di intelligenza artificiale generativa che saranno testati ed utilizzati nel prossimo capitolo.

È opportuno, inoltre, ribadire la natura fortemente evolutiva di questo ambito di ricerca e, di conseguenza, dei *benchmark* disponibili. I dati presentati fanno riferimento al periodo compreso tra la fine di dicembre 2025 e i mesi di gennaio-febbraio 2026; pertanto, nel tempo potrebbero aver subito variazioni.

Tabella 2.1 Tabella riassuntiva degli strumenti di IA selezionati

Generazione testuale	Generazione immagini	Generazione video (<i>image-to-video</i>)	Generazione audio
ChatGPT 5.2, Plus	ChatGPT 5.2, Plus	Runway Gen-4.5	SunoAI (musica)
Claude 4.5 Sonnet	Sora	Sora	Loudly (musica)
Gemini 2.5, Pro	Gemini 2.5, Pro (Nano Banana Pro)	Kling 3.0 e 3.1	ElevenLabs (voce ed effetti sonori)
	Flow (Nano Banana Pro)	Flow (Veo3.1)	
	Midjourney	Pika	

CAPITOLO 3 – PROGETTO DI TESI

Il presente capitolo descrive nel dettaglio il workflow adottato per la realizzazione del cortometraggio mediante l'utilizzo di strumenti di intelligenza artificiale generativa.

Il lavoro intende analizzare fino a che punto le tecnologie generative possano intervenire nel processo creativo audiovisivo, individuando: opportunità, limiti e implicazioni etiche. Inoltre, un compito importante consisteva nel riuscire a rappresentare emozioni, pensieri, immagini emersi durante i tre eventi di *Narrazioni Parallele*.

Dopo una panoramica generale dell'approccio, verranno analizzate progressivamente le singole fasi della pipeline, progettata per rispecchiare quanto più possibile quella di una tradizionale produzione cinematografica, con gli opportuni adattamenti. Saranno quindi illustrate le fasi di pre-produzione, produzione e post-produzione, evidenziando per ognuna gli strumenti di IA impiegati, le criticità riscontrate e il grado di intervento umano.

Come emergerà nel corso del capitolo, il cortometraggio risultante si presenta come il risultato di un processo di co-creazione tra autore umano e strumenti di intelligenza artificiale generativa

3.1 Struttura del workflow e impostazione della pipeline produttiva

Coerentemente con gli obiettivi della ricerca, il progetto ha previsto la definizione di una nuova pipeline produttiva, che integrasse gli strumenti di intelligenza artificiale generativa all'interno del workflow di una produzione audiovisiva. L'intento non è stato quello di delegare totalmente il lavoro ed i processi creativi e alle IA, bensì di esplorarne l'impiego come supporto operativo e co-creativo, mantenendo un dialogo costante ed un controllo autoriale e decisionale.

Si è scelto di adottare un approccio basato prevalentemente su strumenti online, sia per ragioni pratiche – tra cui la riduzione del carico computazionale locale – sia per la volontà di sperimentare concretamente le potenzialità di democratizzazione produttiva offerta dalle piattaforme accessibili via web. È tuttavia opportuno sottolineare che, in contesti professionali strutturati, le pipeline audiovisive basate su IA prevedono spesso l'utilizzo di soluzioni locali o ibride, supportate da team multidisciplinari specializzati nello sviluppo e nell'addestramento dei modelli open source.

Nel caso di questo progetto, la progettazione del workflow ha preso come riferimento la pipeline cinematografica tradizionale, che suddivide il processo in tre macro-fasi: pre-produzione, produzione e post-produzione, a loro volta suddivise in fasi interconnesse tra loro ma svolte da differenti reparti, coordinati dai rispettivi capo-reparto, in costante comunicazione fra loro. La struttura di base è stata il più possibile invariata, tenendo però conto delle difficoltà specifiche dei modelli generativi, tra cui: frammentazione degli strumenti, assenza di una memoria condivisa e necessità di un *prompt design* accurato e diversificato per ogni obiettivo.

La pipeline utilizzata nel progetto si articola quindi in più fasi, correlate ed interconnesse tra loro. Prima di tutto, vi è stata la raccolta e l'analisi dei dati raccolti durante gli eventi di *Narrazioni Parallele*. Questi hanno costituito la base per la generazione di proposte di soggetto e, in seguito, di una sceneggiatura articolata. Partendo dalla sceneggiatura si è definito lo stile visivo del cortometraggio e si sono successivamente realizzate le “schede personaggio” (*character sheet*) da utilizzare come riferimento per la produzione dello storyboard. Quest'ultimo si è rivelato fondamentale come strumento di controllo e di input per la generazione delle clip video, tramite modelli *image-to-video*. Parallelamente, partendo dalla sceneggiatura, si è realizzato il comparto audio: musiche, voce narrante ed effetti sonori.

Per la fase di scrittura sono stati progettati cinque prompt sequenziali, finalizzati a guidare il passaggio dai dati grezzi alle proposte di soggetto e, successivamente, alla sceneggiatura. In questa fase sono stati impiegati in parallelo ChatGPT, Claude e Gemini, con l'obiettivo di valutarne sperimentalmente il comportamento e la qualità degli output. Gli stessi modelli linguistici sono stati inoltre utilizzati per la definizione e l'ottimizzazione dei prompt destinati alla realizzazione delle immagini, dei video e del sonoro.

Per la generazione di immagini (*character sheet* e storyboard) sono state testate diverse intelligenze artificiali generative; la scelta finale è ricaduta su Nano Banana Pro, utilizzato tramite la piattaforma Flow.

La generazione video è stata realizzata combinando gli output di Veo3.1 (disponibile su Flow) e Pika, mentre le musiche sono state generate con Suno e la voce narrante con ElevenLabs, il quale ha dato anche la possibilità di produrre suoni per arricchire il paesaggio sonoro.

Infine, la fase di post-produzione ha richiesto un intervento significativo, sia per la selezione e rifinitura degli output, sia per correggere imperfezioni e discontinuità. In questa fase è stato utilizzato DaVinci Resolve, che propone strumenti di IA per ottimizzare alcuni compiti, come Magic Mask e AI SuperScale.

Nel complesso, il progetto ha adottato un approccio di collaborazione tra umano e intelligenza artificiale generativa, trattando i modelli come membri virtuali di una troupe piuttosto che come sostituti autonomi.

Il contributo umano è risultato inoltre necessario per integrare fra loro gli output provenienti dai differenti modelli, i quali operano in modo isolato senza scambiarsi informazioni.

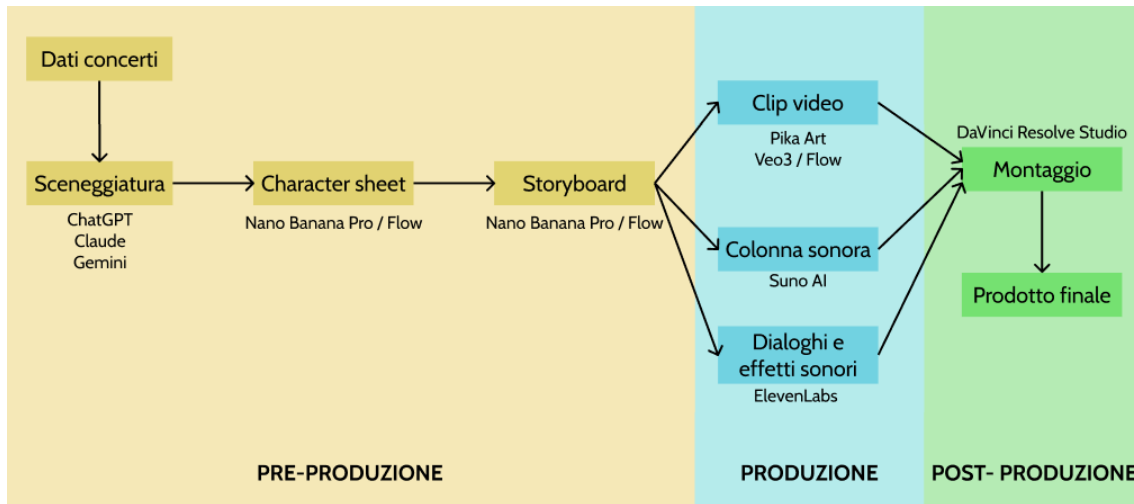


Figura 3.1 Schema della pipeline produttiva con la suddivisione in fasi e le IA utilizzate.

La fase di pre-produzione ha richiesto circa un mese, durante il quale sono stati sviluppati e migliorati i concept e la sceneggiatura ed è stato realizzato lo storyboard, successivamente aggiornato anche durante la fase di produzione video per aiutare le IA nella generazione delle clip. La fase di produzione video ha richiesto circa un ulteriore mese. Parallelamente alla generazione delle clip, queste sono state progressivamente inserite nella timeline di montaggio e sottoposte a primi interventi di correzione. La post-produzione è iniziata in parte durante la fase di produzione e si è poi sviluppata nella fase successiva, includendo anche la generazione e dell'audio.

Complessivamente, quindi il cortometraggio è stato sviluppato attivamente tra fine dicembre e fine febbraio. Durante questo periodo sono proseguite anche le attività di ricerca, selezione e test degli strumenti utilizzati nel lavoro.

I seguenti paragrafi approfondiscono nel dettaglio le singole fasi della pipeline, con l'obiettivo di valutarne l'efficacia e proporre un workflow replicabile per futuri progetti audiovisivi basati su IA.

3.2 Raccolta e analisi dei dati

La caratteristica chiave del progetto consiste nell'utilizzo dei dati qualitativi raccolti durante i tre eventi di *Narrazioni Parallele*, descritti nel [Capitolo 1 \(1.5\)](#).

La fase di pre-produzione è stata preceduta da un processo preliminare di raccolta ed analisi delle informazioni di partenza. In questa sezione viene approfondito l'intero percorso, dalla realizzazione dei questionari fino all'analisi dei dati tramite gli strumenti di intelligenza artificiale generativa scelti.

3.2.1 Raccolta dei dati

La raccolta dei dati ha richiesto una fase preliminare per la progettazione delle domande da rivolgere al pubblico. Per ciascuno dei tre eventi è stata fatta una videochiamata con Claudia Lupo di *Narrazioni Parallele* e gli ospiti coinvolti, così da poterle concordare insieme. In particolare, vi era un duplice obiettivo da rispettare: da un lato, bisognava garantire che i quesiti proposti fossero pertinenti ai temi trattati durante gli eventi ed utili alla ricerca di *Narrazioni Parallele* sul coinvolgimento del pubblico ed il grado di benessere che deriva dalle loro attività; dall'altro, serviva raccogliere materiali significativi per la realizzazione del progetto di tesi.

Gli eventi prevedevano un forte coinvolgimento attivo del pubblico, e i questionari dovevano supportare questa dinamica partecipativa. In particolare, nel primo evento – incentrato su Mozart – le domande servivano a costruire collettivamente un identikit umoristico del compositore, partendo dalle conoscenze pregresse e dalle sensazioni dei partecipanti; nel secondo evento, invece, le risposte del pubblico erano necessarie per generare parole, immagini e idee da rielaborare nella composizione corale della canzone. A queste esigenze si è deciso di affiancare alcune domande specifiche per il progetto, cercando di mantenere comunque un forte legame con le esperienze. I quesiti da me proposti si sono rifatti ai temi chiave del linguaggio cinematografico, come le palette di colori, le ambientazioni e le atmosfere narrative. Alcuni esempi, estratti dal terzo evento – incentrato sulle opere di Giuseppe Verdi – sono:

“Se la musica di Verdi fosse un colore, quale sarebbe?”

“Se la musica di Verdi fosse un luogo, quale sarebbe?”

Inoltre, considerando come la letteratura esistente ed il dibattito pubblico evidenziano una diffusa diffidenza nei confronti dell'utilizzo dell'intelligenza artificiale in ambito creativo, ho

ritenuto opportuno includere alcune domande che esplorassero la percezione su questo tema. Sebbene la maggioranza dei partecipanti fosse costituita da adulti over 50 – una fascia demografica spesso percepita come più distante dalle nuove tecnologie – le risposte hanno rivelato un atteggiamento inaspettatamente aperto verso le potenzialità creative dell'IA, come verrà mostrato nel prossimo paragrafo.

Anche in questo caso, le domande sono state tematizzate sulla base dell'evento ed erano quesiti a risposta chiusa, con la possibilità di scegliere le opzioni “sì, anche meglio / sì / no / non lo so”. Sono state proposte nell'evento *Song Writing* e nell'evento di Verdi, riferendosi rispettivamente alla canzone composta collettivamente e ai versetti di *Va' pensiero* riscritti dal pubblico:

“Pensi che l'Intelligenza Artificiale avrebbe potuto creare una canzone simile?”

“Pensi che l'Intelligenza Artificiale avrebbe potuto riscrivere “Va' Pensiero”?”

In chiusura è stata proposta anche una domanda più concretamente riferita al progetto da realizzare. Anche in questo caso si è chiesto al pubblico di ripensare all'esperienza appena vissuta, prima di rispondere, per dare un senso di continuità con gli eventi.

In particolare, durante l'appuntamento *Song Writing* è stato chiesto di pensare ad uno stile visivo per il cortometraggio, mentre per l'evento di Verdi è stato chiesto un tema da trattare.

“Pensando all'attività, quale stile di video (creato con l'Intelligenza Artificiale) ti piacerebbe vedere?”

“Pensando all'attività di oggi, quale tema potrebbe trattare il video creato con l'Intelligenza Artificiale?”

Per una visione completa e dettagliata di tutti i quesiti proposti, si rimanda all'[Appendice A](#); mentre per un commento delle risposte si rimanda al paragrafo successivo.

Un'osservazione emersa da un partecipante merita di essere esplicitata. Al termine dell'ultimo evento, un ex ricercatore del Politecnico ha segnalato che nelle domande sarebbe stato opportuno specificare il termine “generativa” associato all'espressione “Intelligenza Artificiale”. Essendo un'osservazione corretta e di rilievo, ci tengo a specificare che il termine è stato intenzionalmente omesso, sia per i limiti della piattaforma Mentimeter, che impone un numero massimo di caratteri per ciascuna domanda, sia per necessità comunicativa: il pubblico spesso non sapeva la reale distinzione tra IA generativa e non, quindi si è optato per non inserire il termine così da evitare eventuali incomprensioni.

La raccolta dei dati è si è svolta con *Mentimeter*^[50], piattaforma online che permette di realizzare presentazioni interattive accessibili da qualsiasi dispositivo, senza la necessità di installare applicazioni o di creare un account. Il pubblico poteva partecipare inquadrando un QR code o inserendo sul sito (www.mentimeter.com) il codice associato all'evento.

La piattaforma offre diverse opzioni per raccogliere dati. Quelle più utilizzate durante i tre eventi sono state le nuvole di parole, le risposte aperte ed i grafici a barre.

Oltre che per la facilità di utilizzo, *Mentimeter* è stata scelta soprattutto per la possibilità di visualizzare le risposte in tempo reale, con scelte grafiche e visive adatte alla *user experience* e alla rapida comprensione. Ad esempio, le voci più ricorrenti all'interno delle nuvole di parole appaiono di dimensioni maggiori rispetto alle altre, mentre i diversi colori aiutano a differenziare le risposte.

COME TI FA SENTIRE QUESTA MUSICA RISPETTO A QUELLA DI PRIMA?



Figura 3.2 Esempio di nuvola di parole - evento dedicato a Mozart.

L'integrazione di *Mentimeter* rispecchia perfettamente l'approccio partecipativo seguito da *Narrazioni Parallele*, dove il pubblico non è più spettatore passivo, ma co-creatore dell'evento stesso. In particolare, nei primi due incontri – “*Caro Mozart... Una lettura musicale delle lettere alla cugina di un genio inquieto*” e “*Song Writing 2 - Intelligenza naturale generativa*” – la componente di interazione era la parte centrale dell'evento. Il terzo – “*In scena con Verdi*” – al contrario, ha posto al centro le esecuzioni musicali ma si è avvalso in egual misura dell'interazione con *Mentimeter* per intervallare l'ascolto ed evitare un'eccessiva continuità che avrebbe potuto risultare impegnativa per parte del pubblico, essendo l'evento aperto a chiunque, anche bambini o “inesperti” riguardo i concerti di musica classica.

3.2.2 Analisi preliminare e preparazione dei dati

I dati raccolti durante gli eventi rappresentano il punto di partenza per la costruzione del soggetto e della sceneggiatura. Tuttavia, prima di fornire agli strumenti di IA le risposte ottenute, è stato necessario fare un passo intermedio di analisi, pulizia e standardizzazione. I dati, infatti, erano molto interessanti e ricchi di informazioni, ma presentavano anche una grande eterogeneità nella forma (lunghezze di risposte differenti, varianti grammaticali della stessa parola, errori di battitura, sinonimi, risposte duplicate e, talvolta, contributi poco pertinenti o fuori contesto).

Il primo passo è stato quello di identificare e rimuovere le risposte chiaramente fuori contesto, inserite per un errore dell'utente o un'interpretazione errata della domanda. Sono stati eliminati anche gli inserimenti privi di un significato riconducibile alla domanda, come risposte ironiche non richieste.

È opportuno segnalare alcune anomalie riscontrate durante l'utilizzo della piattaforma. Accedendo io stessa ai questionari durante gli eventi, ho osservato che, se tra una domanda e la successiva trascorrevano molto tempo (ad esempio a causa di un intermezzo musicale), sui dispositivi mobili rimaneva visibile la domanda precedente e la nuova appariva solamente cliccando un piccolo pulsante dedicato. Se non lo si notava – o non si capiva cosa vi era scritto, essendo l'interfaccia solamente in inglese – si rispondeva alla domanda errata. Un ulteriore problema consisteva, invece, nella perdita della sessione: se un partecipante bloccava lo schermo del telefono o chiudeva il browser, al rientro spesso veniva conteggiato come un nuovo utente, alterando quindi le percentuali di partecipazione.

In alcuni casi, sono state individuate anche ripetizioni della stessa risposta, inviate probabilmente dallo stesso utente per paura che la prima non fosse stata registrata. In queste situazioni si è provveduto a mantenere una sola versione.

Mentimeter, inoltre, impone un limite di caratteri anche per le risposte aperte. È capitato, quindi, che alcuni utenti inviassero una risposta tagliata e non la completassero o inviassero più messaggi per aggirare il vincolo. Dove possibile, si è cercato di ricostruire il senso delle risposte incomplete o di unire quelle frammentate.

Un ulteriore elemento critico riguardava la variabilità linguistica: la stessa parola poteva apparire declinata al maschile o al femminile oppure al singolare o al plurale, o, ancora, essere scritta con errori di battitura o grammaticali oppure apparire come sinonimi. Poiché *Mentimeter* tratta ogni variante come una voce distinta, si è reso necessario raggruppare tutte le variazioni e sommare i conteggi delle apparizioni.

Bisogna inoltre sottolineare che gli attuali strumenti di intelligenza artificiale generativa, pur possedendo capacità avanzate di elaborazione del linguaggio naturale, mostrano ancora difficoltà nell'interpretazione di rappresentazioni visive complesse come grafici e nuvole di parole, che, al contrario, risultano essere molto chiari ed intuitivi per gli esseri umani. È stato quindi necessario trasformare ogni set di risposte in tabelle facilmente leggibili dalle IA.

In particolare, ogni tabella presenta nella colonna di sinistra le risposte date dal pubblico e nella colonna di destra il relativo conteggio delle occorrenze. *Mentimeter* consente ai presentatori di visualizzare l'esatto numero di volte che una risposta è stata inserita, funzionalità molto utile soprattutto per tradurre correttamente la rappresentazione visiva delle nuvole di parole – dove la dimensione del carattere riflette la frequenza – in valori numerici per le tabelle. Inoltre, su suggerimento delle stesse IA che saranno utilizzate per l'analisi dei dati, i conteggi sono stati lasciati in formato numerico, anziché essere convertiti in percentuali. Questo perché gli LLM lavorano meglio con quantità discrete e le percentuali possono cambiare di molto in base al numero di partecipanti che hanno risposto (quindi “10” risulta molto più chiaro di “70%”).

Si propone di seguito un esempio di costruzione della tabella partendo dal set di risposte alla domanda “Condividi la prima parola che ti viene in mente su come ti senti ora” relativa al primo evento.

Nella nuvola di parole sottostante il termine “bene” appare in dimensioni maggiori; consultando i dati di *Mentimeter* è possibile verificare che è stata scritta da 4 persone. Al contrario, “ansia” è stata scritta da 1 persona e risulta, infatti, di dimensioni molto inferiori. Tale frequenza è stata quindi riportata nella tabella sottostante. Inoltre, sono stati eseguiti gli accorgimenti sopracitati (come correggere “cuorisan” e conteggiarlo insieme a “incuriosito”, “incuriosita”, “curiosità”).



Condividi la prima parola che ti viene in mente su come ti senti ora	
RISPOSTA DATA	N° PERSONE CHE HANNO RISPOSTO
Incuriosito Incuriosita Curiosa Curiosità	5
Bene	4
Rilassata Rilassamento	3
Divertito Divertita	2
Accaldato	1
Ansia	1
Confusa	1
Emozionata	1
In attesa	1
Molto bene	1
Stanco	1
Stressata	1

Figura 3 Esempio trasposizione in tabella.

Inizialmente, il conteggio delle occorrenze è stato eseguito manualmente; data la quantità complessiva di risposte, si è cercato supporto da parte delle IA. In particolare, è stata testata la versione gratuita di ChatGPT, fornendo in input le risposte esportate da *Mentimeter* al fine di verificarne la capacità di restituire conteggi accurati e di aggregare le varianti sopracitate. Dopo un primo tentativo in cui il modello tendeva ad accorpare anche varianti con sfumature semantiche non realmente equivalenti, si è provveduto ad affinare le istruzioni, ottenendo risultati corretti, successivamente verificati manualmente.

Una volta completata la fase di pulizia e standardizzazione, le tabelle, realizzate in Word, sono state sportate in formato PDF. Questi documenti sono stati poi forniti agli strumenti di IA scelti (ChatGPT, Claude e Gemini) per la fase di analisi successiva, approfondita nel prossimo paragrafo.

Nell'Appendice B sono riportati sia i dati originali estratti da *Mentimeter* sia le tabelle rielaborate. Si precisa che, nella versione sottoposta alle IA, erano presenti esclusivamente le tabelle; nell'appendice, invece, i materiali sono stati accorpati per facilitarne il confronto e la leggibilità.

Infine, durante i test preliminari alla fase successiva è emerso che, se una tabella veniva spezzata dal cambio di pagina, le IA la leggevano come due tabelle separate, falsando le risposte. È stato quindi fondamentale prestare attenzione all'impaginazione per evitare che ciò accadesse.

3.2.3 Elaborazione dei dati con strumenti di IA generativa

Prima di trasformare i dati raccolti durante i tre concerti in materiale utile alle fasi successive del progetto, è stato definito un metodo strutturato per l'interazione con gli strumenti di intelligenza artificiale generativa.

Innanzitutto, una scelta importante è stata quella di non delegare alle IA l'intero processo creativo ma, bensì, di utilizzarle come “assistenti avanzate”, instaurando un dialogo continuo e critico, riproducendo ciò che avviene all'interno di un reparto di pre-produzione cinematografica. Con questa modalità, si è cercato di far assumere alle intelligenze artificiali utilizzate il ruolo di supporto analitico e propositivo, mentre ho lasciato a me – l'autore umano – la responsabilità delle decisioni progettuali e delle scelte o modifiche finali.

Bisogna precisare che la fase di analisi ed elaborazione dati descritta in questo capitolo è strettamente connessa alla successiva fase di ideazione del soggetto, affrontata nel capitolo successivo. Pertanto, sono state realizzate all'interno di un unico processo di dialogo, articolato attraverso una sequenza di cinque prompt progressivi.

Prima di analizzare i prompt e la loro funzione, è utile precisare che, in seguito alle considerazioni emerse nel [Capitolo 2](#), sono stati selezionati tre modelli da testare durante l'analisi dei dati, così da scegliere il risultato più solido ed attinente al progetto. Inoltre, essendo i tre modelli caratterizzati da diverse sensibilità linguistiche e capacità analitiche, è stato possibile ottenere sfumature di interpretazioni e punti di vista interessanti, arricchendo così la ricerca.

Le IA scelte sono:

- ChatGPT 5.2 (versione Plus)^[51]
- Claude 4.5 Sonnet^[52]
- Gemini 2.5 (versione Pro)^[53]

Tutti e tre i modelli sono stati utilizzati in modalità di ragionamento avanzato, così da sfruttare al massimo le loro capacità analitiche e ridurre il rischio di ricevere risposte arbitrarie o superficiali. Inoltre, prima di iniziare questo dialogo, sono state avviate chat nuove e prive di memoria pregressa, così da evitare contaminazioni dovute ad interazioni precedenti.

Per garantire coerenza e controllo del processo, il dialogo con le IA è stato, come già anticipato, suddiviso in cinque prompt, ciascuno consequenziale all'altro e con una funzione specifica. Questa scelta ha permesso di accompagnare progressivamente i modelli durante l'analisi, osservando con attenzione ogni risultato, ed evitando che le richieste risultassero eccessivamente lunghe o complesse.

I prompt sono stati sviluppati attraverso un processo di dialogo con le stesse IA, alle quali è stato presentato il progetto e l'obiettivo da raggiungere. Questo approccio ha permesso di individuare quali istruzioni sarebbero state più efficaci e meglio interpretate.

I prompt sono stati progettati per: contestualizzare il progetto, analizzare i dati, astrarre le informazioni al fine di individuare elementi narrativi e formulare proposte di soggetto.

In questa fase di elaborazione dei dati sono stati utilizzati i primi due prompt, con l'obiettivo di individuare pattern ricorrenti ed eventuali differenze nelle emozioni degli spettatori durante i tre eventi. Non è stata però esclusa la possibilità di iniziare ad evidenziare possibili nuclei tematici rilevanti. Inoltre, è stata data una particolare attenzione nel mantenere un forte legame con i dati raccolti, evitando interpretazioni astratte o non giustificabili.

Poiché sono state avviate nuove chat prive di memoria, il primo prompt fornito aveva come obiettivo la contestualizzazione del progetto: il lavoro di tesi, la realtà di *Narrazioni Parallele* e la struttura dei tre concerti con l'interazione del pubblico. Questo passaggio era fondamentale per permettere ai modelli di interpretare correttamente le risposte, in particolare quelle fortemente legate al contesto degli eventi.

Tabella 3.2 Prompt numero 1

<p>1) CONTESTO DEL PROGETTO</p> <p>Sto realizzando un cortometraggio, della durata massima di 5 minuti, utilizzando strumenti di intelligenza artificiale generativa.</p> <p>Il progetto nasce all'interno di <i>Narrazioni Parallele</i> (Torino, autunno 2025), un ciclo di tre concerti partecipati in cui il pubblico ha contribuito attivamente a costruire contenuti testuali, emotivi e creativi.</p> <p>Il cortometraggio non deve essere un resoconto degli eventi, ma un'elaborazione narrativa e poetica basata sui dati raccolti dal pubblico attraverso questionari interattivi in tempo reale.</p> <p>I dati sono stati puliti, normalizzati e organizzati in tabelle, che verranno fornite nel messaggio successivo.</p> <p>2) DESCRIZIONE DEI TRE EVENTI</p> <p>1. "Caro Mozart..."</p> <p>Evento dedicato a Mozart, costruito in forma giocosa e ironica.</p> <p>Il pubblico ha contribuito alla creazione di un identikit umoristico del compositore attraverso: letture di lettere autentiche (al padre e alla cugina), ascolto di brani ("Ein Musikalischer Spass", "Ave Verum</p>
--

Corpus”), domande orientate a ricostruire un ritratto emotivo, caratteriale ed umoristico del compositore.

2. “Song Writing — Intelligenza naturale generativa”

Laboratorio creativo basato sulla poesia di Gianni Rodari Il paese con la S davanti.

Il pubblico ha partecipato attivamente alla generazione del testo e del ritmo di una canzone collettiva, basata anche sul tema di “Un giorno senza regole”.

3. “Concerto su Verdi”

Concerto e riflessione guidata sulla musica di Giuseppe Verdi.

Il pubblico ha condiviso emozioni, ricordi, immagini e associazioni cromatiche o spaziali ispirate dai brani presentati. È stato anche riscritto dal pubblico un versetto di “Va’ pensiero”.

3) DATI RACCOLTI

Il pubblico (soprattutto over 50) ha compilato dei questionari in tempo reale su Mentimeter, durante i concerti.

Le risposte erano: nuvola di parole, risposte aperte, risposte chiuse, associazioni cromatiche, immagini mentali, emozioni, interpretazioni creative.

I dati sono stati successivamente ripuliti, aggregati e strutturati in tabelle, una per ogni domanda. Le tabelle verranno fornite nel messaggio successivo per l’analisi. La prima riga della tabella ha, scritto in grassetto, la domanda; la colonna di sinistra ha le risposte (aggregate per significati, generi, grammatica); la colonna di destra ha le ricorrenze di quella risposta. Il numero di partecipanti che hanno risposto è variabile.

4) NESSUNA AZIONE RICHIESTA ORA

In questo messaggio sto solo fornendo il contesto generale.

Non devi proporre analisi né idee creative a questo punto.

Il secondo prompt chiedeva alle IA di analizzare le tabelle contenute nel file PDF allegato, costruito durante la fase di preparazione dei dati (approfondito nel paragrafo precedente), con il fine di individuare ricorrenze e/o differenze tra i concerti e riconoscere eventuali evoluzioni emotive o tematiche. È stato esplicitamente richiesto ai modelli di non introdurre elementi non presenti nei dati e di giustificare ogni osservazione citando le risposte ed i relativi valori numerici.

1) DATI ALLEGATI

Nel PDF che sto inviando sono presenti le tabelle dei tre concerti.

Ogni concerto è introdotto da una pagina bianca con scritto “CONCERTO 1”, “CONCERTO 2”, “CONCERTO 3” (nello stile Titolo1 di Word).

In ogni tabella:

- la prima riga contiene la domanda in grassetto
- la colonna di sinistra contiene le risposte (già ripulite, con varianti linguistiche accorpate)
- la colonna di destra contiene il numero di occorrenze

Il numero di partecipanti che ha risposto varia per ogni domanda

2) IL TUO COMPITO — ANALISI DEI DATI

Analizza i dati contenuti nel pdf e rispondi in modo strutturato ai seguenti punti.

2.1 – RICORRENZE E PATTERN TRASVERSALI

Quali temi, emozioni, immagini e concetti ricorrono in tutti e tre i concerti?

Quali elementi potrebbero diventare un “filo rosso” narrativo che collega le tre esperienze?

Ci sono simboli o metafore che ritornano?

Indica eventuali differenze tra i tre eventi o contrasti.

– SPECIFICITÀ DI OGNI EVENTO

Per ciascun concerto identifica:

temi principali

colori/emozioni più citati

immagini/concetti dominanti

tono emotivo

2.3 – EVOLUZIONE EMOTIVA

Dalla lettura dei dati, emerge un'evoluzione emotiva dal primo al terzo concerto?

In che modo questa evoluzione potrebbe allinearsi alla struttura del “viaggio dell'eroe”?

2.4 – RISPOSTE POTENZIALMENTE UTILI PER UNA STORIA

Quali tipi di risposte (o categorie di risposte) sono più adatte a diventare materiale narrativo?

Quali immagini, emozioni, situazioni o concetti ricorrenti potrebbero essere tradotti in scene, atmosfere, simboli o trasformazioni narrative?

REGOLE IMPORTANTI

- Non inventare dati. Usa solo ciò che trovi nelle tabelle.
- Cita sempre i dati numerici: ad esempio “7 persone hanno indicato X”.
- Non interpretare in modo troppo creativo: questa parte deve essere analitica, oggettiva.
- Nell’analisi NON generare ancora idee di trama. Questo avverrà nel prossimo messaggio.

Procedi ora con l’analisi dei dati nel pdf.

Dall’analisi sono emersi risultati interessanti.

Pur partendo dagli stessi dati e richieste, i tre modelli hanno mostrato differenze nel modo di organizzare ed esporre l’analisi: ChatGPT ha suddiviso i dati in modo più sistematico, Claude ha enfatizzato le sfumature emotive e Gemini ha mostrato fin da subito una tendenza a riorganizzare i dati in chiave simbolica. Queste differenze riguardano solo la presentazione dell’analisi e non il contenuto che risulta invece essere comune.

Tutte le IA hanno individuato un’evoluzione emotiva che passa dalla curiosità prima del concerto ad uno stato di benessere, divertimento e rilassamento. È stata riconosciuta la centralità della natura come tema chiave, soprattutto in relazione al secondo e terzo concerto, con numerose risposte che evocano immagini e ricordi di boschi, foreste, mare e paesaggi aperti. A questi elementi si affiancano poi temi ricorrenti legati alla nostalgia, alla memoria e al ritorno all’infanzia o a tempi già vissuti.

L’analisi ha inoltre evidenziato una netta distinzione tra i tre eventi, che rispecchia le loro diverse caratteristiche ed impostazioni. Il primo concerto ha suscitato curiosità e divertimento, legati alla scoperta della figura di Mozart attraverso una narrazione ironica; il secondo ha permesso di esplorare maggiormente la dimensione ludica e partecipativa, attraverso la creazione collettiva di una canzone e la riscoperta dei “bambini interiori”; infine, il terzo ha evocato emozioni riflessive e nostalgiche. Queste differenze risultano particolarmente rilevanti in vista delle successive fasi di astrazione dei dati e proposta dei soggetti.

3.3 Fase di Pre-produzione: scrittura

Dopo la raccolta ed elaborazione dei dati si è passati alla prima fase della pre-produzione del cortometraggio, dedicata all'ideazione del concept narrativo e alla scrittura della sceneggiatura.

In questa fase sono stati utilizzati gli ultimi tre prompt del dialogo e sono state definite due scelte importanti: seguire la struttura de *Il Viaggio dell'Eroe* ed adottare una suddivisione del racconto in tre atti, ciascuno relativo ad uno dei tre concerti.

3.3.1 Astrazione dei dati

Dopo l'elaborazione dei dati si è chiesto alle IA di astrarre le informazioni nell'ottica di trasformarle in elementi narrativi adatti a costruire un soggetto per un cortometraggio. Non vi era ancora l'obiettivo di generare una storia, ma piuttosto di individuare immagini, temi ed atmosfere capaci di tradurre le emozioni e le ricorrenze emerse dai dati in un linguaggio narrativo e audiovisivo. Inoltre, è stato esplicitamente richiesto alle intelligenze artificiali di non interpretare didascalicamente i contenuti inserendo, ad esempio, Mozart o Verdi come protagonisti, ma di reinterpretarli come qualità o attributi per caratterizzare mondi narrativi, personaggi o azioni. Ad esempio, gli elementi ricorrenti di curiosità ed ironia emersi dal primo concerto potevano tradursi in stili narrativi, mentre i temi di libertà e rottura delle regole del secondo evento potevano diventare i motori dell'azione e, infine, la nostalgia scaturita con il terzo concerto poteva rappresentare il tono emotivo.

Tabella 3.4 Prompt numero 3

<p>Ora che l'analisi dei dati è conclusa, voglio che passi a una fase di astrazione narrativa.</p> <p>Questa fase NON serve a creare storie, ma a trasformare i dati in materiale narrativo potenziale.</p> <p>1) OBIETTIVO</p> <p>Identificare quali elementi emersi dai questionari sono utili per costruire una storia poetica, metaforica, strutturata in 3 atti (uno per concerto).</p> <p>Quindi:</p> <ul style="list-style-type: none">- temi- atmosfere- stati emotivi- immagini ricorrenti- palette- metafore- possibili trasformazioni narrative
--

2) REGOLA IMPORTANTE: ASTRAZIONE

Non usare personaggi storici.

Le risposte su Mozart, Verdi, Rodari non vanno interpretate alla lettera.

Trasformale in qualità, stati emotivi, caratteristiche di un mondo o di un protagonista.

Esempi di astrazione corretta (non inventare, solo riformulare):

- se molte risposte su Mozart parlano di “bambino/ribelle”, allora l’Atto 1 potrebbe avere un tono “giocosso, infantile, impulsivo”.
- se molte risposte su Verdi parlano di “nostalgia/natura/luogo”, l’Atto 3 potrebbe avere un’atmosfera più “matura, introspettiva, radicata”.

3) COMPITO

1. Identifica le categorie di risposte con valore narrativo

Seleziona quali tipi di risposte tra tutte quelle presenti (emozioni, immagini, luoghi, ricordi, stati d’animo, colori, metafore...) hanno il potenziale per diventare:

- atmosfera di un atto
- stato emotivo del protagonista
- trasformazione narrativa
- ambientazione
- simboli ricorrenti
- conflitti interiori o desideri

2. Escludi le risposte NON utili alla costruzione di una storia

Indica quali categorie di risposte sono poco utili o irrilevanti per una narrazione (es. risposte troppo tecniche, troppo specifiche sul contenuto del concerto, o prive di potenziale simbolico).

3. Mappa i tre concerti in tre “mood narrativi”

Per ogni concerto descrivi il “tono emotivo” che emerge dai dati, senza fare riferimento ai contenuti reali del concerto ma solo alle emozioni e immagini ricorrenti.

4. Spiega come questi mood possono supportare un Viaggio dell’Eroe

Senza creare una storia vera e propria, spiega come queste tre atmosfere potrebbero rappresentare tre tappe di un percorso narrativo.

4) ALTRE REGOLE IMPORTANTI

- NON inventare dati.
- NON introdurre nuovi concetti non presenti nel pdf.

- Usa SOLO le risposte reali, trasformandole in categorie astratte.
- Non proporre ancora trame o concept: ci sarà un prompt dedicato.

Le risposte scaturite da questo terzo prompt hanno evidenziato approcci differenti all'astrazione dei dati. In particolare, ChatGPT e Claude hanno rielaborato i dati come temi ed atmosfere generali, mentre Gemini ha trasformato alcune risposte ricorrenti in figure narrative reali, come la presenza di un animale che accompagna il protagonista. Inoltre, ChatGPT e Claude hanno restituito un'analisi più lunga, approfondita e curata, mantenendo una struttura molto simile nell'organizzare la risposta. Gemini ha invece prodotto una risposta molto più sintetica e compatta, pur rispettando i vincoli e le richieste forniti nel prompt.

Nella tabella sottostante sono riportati due esempi più significativi di queste analogie e differenze. È interessante soprattutto notare la mancanza dell'analisi sugli animali nella risposta di Claude. Questo evidenzia come le tre IA non ragionino allo stesso modo e valutino o meno se scartare alcune risposte considerate poco rilevanti.

Tabella 3.5 Esempi di risposte

ChatGPT	Claude	Gemini
<p>B. Immagini e luoghi → potenziali ambientazioni e simboli</p> <p>4. Animali come alter-ego Gatto (6), lupo (3), cane (2), altri singoli. → Simboli potenziali: indole, istinto, carattere nascosto.</p>	<p><i>Non ha evidenziato le risposte sugli animali.</i></p>	<p>3.1 Identificazione delle categorie con valore narrativo</p> <ul style="list-style-type: none"> • Figure Simboliche (Avatar/Guide): L'animale affine ("Gatto", "Lupo") ⁹e la voce interiore rappresentata da uno strumento ("Violoncello", "Chitarra") ¹⁰ possono diventare manifestazioni visive dell'inconscio.
<p>B. Immagini e luoghi → potenziali ambientazioni e simboli</p> <p>1. Natura Molto presente nei dati, soprattutto nel secondo e terzo concerto:</p> <ul style="list-style-type: none"> • erba appena tagliata, tiglio, aria aperta • boschi, praterie, mare, oceano, fiume, vento, cascata 	<p>B) IMMAGINI NATURALI</p> <ul style="list-style-type: none"> • Valore narrativo: Ambientazioni, spazi simbolici, scenografie • Dati rilevanti: <ul style="list-style-type: none"> ○ Boschi notturni, nebbia (11 persone, Concerto 3) ○ Campagne, praterie, valli 	<p>3.1 Identificazione delle categorie con valore narrativo</p> <ul style="list-style-type: none"> • Luoghi dell'Anima (L'Ambientazione): Le risposte sulla visualizzazione spaziale della musica ("Bosco", "Nebbia", "Mare tempestoso") ³e le associazioni stagionali ("Primavera") ⁴ forniscono la scenografia visiva.

<p>→ Possibili atmosfere: spazi aperti, trasformazione stagionale, forza degli elementi.</p>	<p>(5 persone, Concerto 3)</p> <ul style="list-style-type: none"> ○ Mare tempestoso, cascate (Concerto 3) ○ Natura come luogo di camminata quotidiana (4 persone, Concerto 2) <ul style="list-style-type: none"> • Uso: Costruire l'ambientazione visiva di ogni atto 	
--	---	--

Le differenze riscontrate nelle risposte prodotte sono attribuibili ai fattori strutturali e metodologici propri di ciascun modello, come le diverse strategie di addestramento, le modalità di ottimizzazione del ragionamento e i possibili *bias* introdotti. Pur basandosi su architetture simili, i modelli presentano differenti inclinazioni interpretative e stilistiche che influenzano la loro capacità di analisi, astrazione, sintesi e rielaborazione.

È da precisare che queste differenze non rappresentano incongruenze o errori ma, bensì, punti di vista complementari e interessanti per la ricerca.

3.3.2 Scelte strutturali e narrative

Prima di passare alla scrittura della sceneggiatura è stata compiuta una scelta fondamentale per la struttura del racconto: adottare un modello narrativo solido che seguisse *Il Viaggio dell'Eroe* e fosse suddiviso in tre atti.

Il Viaggio dell'Eroe è un saggio di Christopher Vogler²⁴ in cui viene analizzata l'omonima struttura narrativa. Questa viene descritta come un viaggio interiore compiuto dal protagonista, che lo porta a crescere ed evolvere durante l'arco narrativo. Il viaggio è suddiviso in dodici fasi, caratterizzate da passaggi ed equilibri ricorrenti che vengono spezzati da chiamate all'azione e successivamente ricomposti in forme nuove. Non si tratta però di una struttura rigida: Vogler stesso, infatti sottolinea come l'ordine delle fasi possa essere variato e, anzi, che non tutte debbano obbligatoriamente essere presenti.

²⁴ C. Vogler, *The Writer's Journey: Mythic Structure For Writers*, (1992, 1998, 2007) [54]

Nel contesto di questo progetto, il Viaggio dell'Eroe è stato utilizzato come struttura di riferimento per garantire un arco narrativo chiaro facilmente seguibile dalle IA.



Figura 3.4 Schema de “Il Viaggio dell’eroe” di Christopher Vogler.

A questa scelta è stata affiancata l’adozione della struttura in tre atti, teorizzata da Eugène Scribe^[55] negli anni venti del XIX secolo e, ancor prima, da Aristotele. Al contrario della suddivisione in tre atti di Vogler – che comprendono la partenza, l’iniziazione ed il ritorno – l’organizzazione di Scribe si articola in:

- *Atto I: introduzione dei personaggi.* Ha una funzione preparatoria e serve a presentare i personaggi ed i loro obiettivi.
- *Atto II: scontro e lotta.* Inizia con la definizione del problema e la necessità di risolverlo e si conclude con una situazione critica per il protagonista, che pare sconfitto dagli eventi.
- *Atto III: risoluzione della crisi.* Ha la funzione di concludere la storia e rispondere a tutti i quesiti aperti.

Struttura in tre atti



Figura 3.5 Schema della struttura in tre atti di Eugène Scribe.

L'idea di introdurre la suddivisione in tre atti è nata pensando alla suddivisione del progetto di *Narrazioni Parallele* in tre concerti. Questo ha permesso di stabilire una corrispondenza ancora più solida ed elaborata tra la sceneggiatura ed i dati raccolti: ciascun atto è stato infatti associato ad uno dei tre eventi. Si tratta di un'associazione emotiva e tematica, in cui ogni concerto contribuisce a definire l'atmosfera e le tensioni narrative di una specifica fase della narrazione.

Il primo atto presenta il mondo del protagonista ispirandosi alle atmosfere ironiche, curiose e anche un po' oppressive ("sotto pressione paterna", come recita una risposta pensando al giovane Mozart) del primo concerto. Il secondo atto, più lungo e dal carattere centrale per la storia, riflette la dimensione ludica, creativa e di rottura delle regole e degli schemi emersa durante l'esperienza di *Song Writing*, ispirata a Gianni Rodari. Queste tematiche si uniscono perfettamente alla natura di prova e trasformazione del secondo atto. Infine, il terzo atto è basato sulle emozioni di introspezione, memoria e nostalgia scaturite dalla musica di Verdi durante il terzo concerto. Quest'ultimo atto conduce il racconto verso una risoluzione che restituisce il senso del percorso appena compiuto.

A queste scelte si è affiancata anche una decisione relativa alla durata complessiva del cortometraggio, fissata ad un massimo di cinque minuti, sia per esigenze narrative che produttive.

Dal punto di vista narrativo, una durata contenuta consente di mantenere un ritmo narrativo efficace, evitando il rischio di allungare eccessivamente la storia rendendola troppo didascalica e perdendo le tensioni narrative e la leggibilità dell'arco di trasformazione del protagonista.

Cinque minuti sono una durata sufficiente per raccontare in modo completo una storia ben articolata nei tre atti e che segua il Viaggio dell'Eroe.

Dal punto di vista produttivo, tenendo a mente gli attuali limiti delle intelligenze artificiali generative, una durata non eccessiva permette di realizzare meno clip video ma di qualità più alta, poiché si possono spendere maggior tempo ed attenzione su di esse. Inoltre, gli attuali strumenti di IA consentono di generare clip della durata ridotta (generalmente tra gli otto e i dieci secondi), rendendo molto macchinoso e dispendioso – in termini di tempo e “crediti” – la realizzazione di lunghe narrazioni.

3.3.3 Ideazione dei concept narrativi

In seguito alla fase di astrazione dei dati, il dialogo con le intelligenze artificiali è giunto alla fase di ideazione dei concept narrativi, dove sono stati utilizzati i due prompt restanti. L'obiettivo era la generazione di proposte narrative capaci di tradurre in storie le caratteristiche emerse dall'analisi dei dati.

Il quarto prompt proposto alle IA anticipa la fase di creazione delle proposte, fornendo alcuni materiali aggiuntivi relativi al secondo e terzo evento. In particolare: il testo e la registrazione audio della canzone collettiva creata dal pubblico e il versetto riscritto di *Va' pensiero* e votato dai musicisti come il migliore. È stato esplicitamente precisato nel prompt di utilizzarli solo come supporto per rafforzare la vena emotiva degli atti dedicati ai due eventi e di non trasformarli in elementi diegetici riconoscibili.

L'idea è nata pensando all'eventualità di inserire la registrazione audio della canzone realizzata come colonna sonora del cortometraggio; si è quindi ritenuto corretto condividerla con le IA, permettendo loro di sfruttarla come riferimento emotivo.

Tabella 3.6 Prompt numero 4

<p>MATERIALI AGGIUNTIVI PER IL MOOD (NON PER FORZA PER IL CONTENUTO)</p> <p>Ora ti fornisco tre materiali prodotti dal pubblico durante i concerti. NON devono essere usati come contenuto narrativo. Servono solo a comprendere il mood emotivo dei tre atti del futuro cortometraggio.</p> <p>1) Testo della canzone creata nel secondo concerto</p>
--

Usalo esclusivamente per capire l'atmosfera dell'Atto 2: gioco, immaginazione, libertà, leggerezza, creatività collettiva.

Non usare il testo letteralmente. Non citarlo. Non trasformarlo in trama.

N.B. Canzone creata partendo dalle risposte del pubblico, basate su "il paese con la S davanti" di Rodari e il tema di "un giorno senza regole".

Da usare solo per suggerire l'atmosfera dell'Atto 2

2) Testo riscritto di "Va' pensiero" dal pubblico del terzo concerto

Usalo solo per capire il mood dell'Atto 3: nostalgia, radici, desiderio, malinconia, dolcezza, memoria. Non citare né imitare il contenuto. Non usare Verdi, né riferimenti espliciti ai versi.

N.B. Il pubblico ha riscritto i versetti di Va pensiero e questa è la versione che è piaciuta di più.

Da usare solo per suggerire l'atmosfera dell'Atto 3

3) AUDIO della canzone del secondo concerto

Ti allego l'audio della canzone creata dal pubblico durante il secondo concerto.

Ascoltalo unicamente per comprenderne il tono emotivo: energia, ritmo, atmosfera, sensazione generale.

Ascoltalo SOLO per cogliere:

- energia
- ritmo
- tempo
- tono emotivo complessivo
- atmosfera evocata

Regole assolute sull'uso dell'audio:

- Non riprodurre o citare melodia/timbro/armonizzazioni. Non devi citare, imitare o rigenerare la melodia, il testo o la struttura musicale.
- Non usarlo come contenuto narrativo o musicale
- Non imitarlo né parafrasarlo => è solo dedurre il mood.
- Non devi inserirla nella trama né farla riconoscere come canzone.
- Non devi usarla come contenuto narrativo.
- Puoi usarla solo per dedurre l'atmosfera da tradurre nell'Atto 2 (ad esempio: "piacevolmente lenta", "sognante", "giocosa", "intima").
- Serve esclusivamente per capire il mood emotivo dell'atto centrale, non come materiale creativo diretto.

IL TUO COMPITO

Dimmi, in modo sintetico:

1. Cosa rivelano questi tre materiali sul mood emotivo dei tre atti?

(Atto 1 – Mozart; Atto 2 – Rodari/Canzone; Atto 3 – Verdi)

2. Come questi materiali possono rafforzare la struttura emotiva del viaggio dell'eroe?

3. Senza inventare nulla, quali elementi atmosferici EMERGONO chiaramente e possono essere usati nella fase narrativa?

(Esempi: “leggerezza”, “giocosità”, “nostalgia”, “movimento lento”, “profondità”, “luminosità”, ecc.)

RICORDA:

- Niente contenuti letterali dei testi. Non devi riportare frasi o versi
- Niente uso diretto dell'audio.
- Solo deduzione del tono emotivo.
- Non devi inserirli per forza come scene o citazioni
- Non devono diventare personaggi, luoghi o eventi
- Non devono essere reinterpretati letteralmente

Uso consentito

- Serve solo a comprendere e rafforzare il mood emotivo degli atti 2 e 3.
- Niente di più.

Infine, si è passati all'ultimo prompt, con l'obiettivo di generare di tre concept narrativi distinti. Sono stati chiaramente definiti i vincoli progettuali: la durata massima del cortometraggio, l'adozione del Viaggio dell'Eroe, la suddivisione in tre atti ispirata ai tre eventi, il divieto di utilizzare i compositori delle opere in modo letterale e l'obbligo di riferire ogni scelta a dati effettivamente presenti, per evitare risultati arbitrari.

Tabella 3.7 Prompt numero 5

ORA PASSIAMO ALLA FASE DI CREAZIONE NARRATIVA.

Basandoti esclusivamente sui dati analizzati nei messaggi precedenti, ti chiedo di generare 3 CONCEPT narrativi originali per un cortometraggio di 5 minuti.

1) IL TUO RUOLO

Agisci come un regista e sceneggiatore di cinema. Devi trasformare i dati emotivi e tematici dei tre concerti in soggetti per un cortometraggio unico, poetico, metaforico ed emozionale.

2) OBIETTIVO

Creare tre proposte di cortometraggio che:

- seguano il Viaggio dell'Eroe
- siano divise in 3 atti, ognuno ispirato a uno dei tre concerti
- formino un'unica storia continua, NON tre episodi separati
- riflettano rigorosamente le atmosfere e le emozioni emerse dai dati

IMPORTANTE REGOLA D'ORO: ASTRAZIONE

- NON usare i personaggi storici (Niente Mozart, Rodari o Verdi come personaggi).
- Usa i dati come "aggettivi" per il mondo e i personaggi... se i dati di Mozart dicono "Ribelle/Bambino", il protagonista o il mondo dell'Atto 1 avranno quelle caratteristiche; se i dati di Verdi dicono "Patria/Ricordo", l'Atto 3 parlerà di un ritorno a casa o alle origini...

3) VINCOLI DA RISPETTARE

- Durata: max 5 minuti
- Struttura: 3 atti (Mozart → Rodari → Verdi)
- Arco narrativo: Viaggio dell'Eroe
- Tono: poetico, evocativo, non didascalico o esplicativo
- Formato: ancora da scegliere in base alle storie (accetto suggerimenti)
- Unità narrativa: 1 storia continua
- Musica: elemento centrale, trasformativo
- Atto 2: deve includere un evento legato alla creazione (trovare una melodia, creare un suono, rompere o reinventare regole) => è stata creata una canzone durante il secondo concerto

4) STRUTTURA DELLA RISPOSTA

Crea i 3 soggetti tenendo conto:

- dei dati del pdf iniziale (messaggio 2)
- dell'astrazione tematica (messaggio 3)
- dei mood del pdf di testi extra (messaggio 4)
- dei vincoli narrativi che ho deciso qui sopra

Per ciascuno dei tre concept, segui ESATTAMENTE questo schema:

CONCEPT X

Titolo evocativo

1. Sinossi (3-4 righe)

Breve descrizione dell'intera storia.

2. Protagonista

Chi/che cosa è?

Costruiscilo ASTRAENDO dai dati del Concerto 1 (Mozart).

3. Mappatura sui 3 atti

ATTO 1 — Mondo ordinario / Chiamata (Mood Mozart)

- Quale situazione iniziale riflette le emozioni emerse nel Concerto 1?
- Quale evento scatena il viaggio?

ATTO 2 — Prove / Creazione (Mood Rodari)

- Come si rompe l'ordine delle cose?
- In che modo il protagonista deve “creare la sua canzone/voce”?

ATTO 3 — Climax / Ritorno trasformato (Mood Verdi)

- Momento di consapevolezza
- Temi di luogo, memoria, ritorno alle origini
- Chiusura poetica

4. Evoluzione emotiva / cromatica

Descrivi la trasformazione del mood da atto a atto.

5. Ruolo della musica

Come la musica guida o trasforma il protagonista.

6. Elementi visivi

- Palette cromatica (una per atto)
- Ambientazioni
- Motivi simbolici ricorrenti

7. Stile visivo

Dimmi che stile useresti (animazione, realistico...)

8. Data Check

Cita almeno 3 dati precisi estratti dal pdf che giustificano il concept.

(es. “8 persone hanno indicato ‘natura’, quindi l’atto 3 è ambientato in un bosco”)

REGOLE IMPORTANTI

- Non inventare dati o emozioni.
- Usa solo ciò che emerge dalle tabelle.
- Le connessioni creative devono SEMPRE derivare da dati presenti.
- La storia deve risultare poetica, evocativa, universale. NON deve essere didascalica.

FORMATO FINALE ATTESO (Chiaro, ordinato, con titoli e sezioni)

- Concept 1
- Concept 2
- Concept 3

Per approfondire i concept generati, si rimanda all'Appendice C.

Pur partendo dagli stessi dati e dal medesimo prompt, le nove proposte risultanti presentano idee narrative diverse fra loro ed originali. Si possono inoltre notare differenti approcci nella costruzione della risposta, come già era emerso nelle precedenti fasi del dialogo. Le tre IA hanno prodotto concept coerenti con i materiali forniti e potenzialmente validi per produrre un cortometraggio, ma si distinguono per modalità di astrazione, aderenza ai dati, stile visivo, tipologia di personaggi ed inserimento o meno degli archetipi descritti dal Viaggio dell'Eroe.

Si può velocemente notare come le risposte di ChatGPT mantengano un legame molto esplicito con le risposte del pubblico, citando di sovente i dati e la loro frequenza come giustificazione delle scelte narrative. Il “data check” richiesto come ultimo punto viene integrato costantemente come spiegazione delle risposte ai punti, rendendo il processo molto trasparente ma la lettura e comprensione delle storie molto lenta. Inoltre, il punto relativo al “data check” è stato reso molto sintetico ed approssimativo, avendo già risposto in precedenza.

Claude, al contrario, mostra un controllo preciso dei dati ma una rielaborazione molto più astratta ed interpretativa, in cui le informazioni emergono a livello tematico e simbolico. Il “data check” finale risulta molto approfondito e giustifica tutte le scelte narrative citando i dati correttamente.

Gemini, come già accaduto nelle altre fasi del dialogo, rispetta tutti i vincoli e risponde correttamente a tutti i punti ma in maniera molto più sintetica rispetto agli altri modelli. In questo caso, il “data check” risulta sintetizzato e, come ChatGPT, lo distribuisce nelle risposte agli altri punti, seppur in modo meno dettagliato e palese. ChatGPT, ad esempio, quando parla di un protagonista capriccioso, curioso ed esuberante aggiunge una frase per giustificare una grande frequenza di risposte di quel tipo relative al carattere di Mozart, citando tutti i dati; Gemini parla di un protagonista curioso aggiungendo semplicemente, dopo l'aggettivo, “(5 persone)”.

Un'altra grande differenza emerge nel secondo atto. ChatGPT lo costruisce attorno ad un suono che rompe le regole e alla costruzione di una melodia che ne stabilisce di nuove e conduce al terzo atto, riflettendo in modo molto chiaro l'esperienza di co-creazione musicale del secondo evento, partita dal tema di “un giorno senza regole”.

Claude, invece, traduce la dimensione ludica del secondo evento in una sovversione delle leggi del mondo narrativo, basandosi chiaramente sulle domande poste al pubblico riguardanti lo stravolgimento delle leggi della fisica, l'ottenere superpoteri, l'istituzione di regole per vivere meglio, e così via. Al contrario di ChatGPT, Claude non inserisce sempre il tema della creazione di una canzone. Nel secondo concept *IL PAESE DOVE SI SCUCE IL TEMPO*, ad esempio, viene proposta la distruzione e la creazione di un abito tramite un gioco guidato da

dei bambini. Inoltre, nel primo atto, ha sempre posto una grande attenzione sul tema della rigidità delle regole e pressione paterna, che portano ad un'oppressione più o meno esplicita del carattere del protagonista.

Gemini si colloca in una posizione intermedia: la creazione musicale è fondamentale come principio trasformativo, ma viene esplorata seguendo le risposte del pubblico in una dimensione visiva più astratta. È da notare, però, come nel terzo concept *IL LUPO DI CARTA* abbia invece proposto di inserire la canzone composta come colonna sonora e di far eseguire al lupo protagonista le “istruzioni” cantate:

Il mondo si costruisce visivamente seguendo il testo della Canzone del Pubblico (che useremo come audio). Se la canzone dice “nuvole”, il Lupo salta su nuvole di carta pesta. Se dice “senza regole”, il Lupo cammina a testa in giù.
- GEMINI, descrivendo il secondo atto del terzo concept -

Per quanto riguarda lo stile visivo, ChatGPT suggerisce spesso live action con, eventualmente, l'inserimento di piccoli effetti in animazione, mentre Claude e Gemini prediligono gli stili di animazione 2D o 3D o, anche, la tecnica della stop motion. Inoltre, questi ultimi due modelli, forniscono riferimenti ad autori, studi di animazione o opere audiovisive, per rafforzare l'idea stilistica che stanno proponendo. Vengono spesso citati Miyazaki^[56] e lo Studio Ghibli^[57] e Laika Studios^[58] (con particolare riferimento a *Coraline e la porta magica*^[59], opera in stop motion del 2009).

Infine, anche dal punto di vista dei personaggi e delle figure archetipiche presentate nel Viaggio dell'Eroe, si possono notare differenze interessanti. Nei concept di ChatGPT, il percorso del protagonista – che è sempre un essere umano – è costruito senza una figura di mentore o guida esplicita, affidando la trasformazione ad un evento scatenante e la risoluzione ad un cambiamento interno del protagonista stesso.

Claude presenta sia protagonisti umani che “esseri astratti” e inserisce altri personaggi, magici o non, con il ruolo di guidare il protagonista attraverso il suo viaggio, come la figura del mentore nel Viaggio dell'Eroe.

Gemini, al contrario di ChatGPT, non presenta mai protagonisti umani ma bensì robot umanoidi o animali. Come ChatGPT, però, nei primi due concept non ha inserito la figura del mentore che, invece, appare nel terzo ed è rappresentato da un gatto. Quest'ultimo elemento è molto interessante, poiché solamente Gemini ha avuto l'idea di sfruttare la domanda “Con quale animale ti senti più affine?” del secondo concerto per realizzare, nel secondo concept, un

protagonista gatto (la risposta con più ricorrenze) e, nel terzo, un mentore gatto che affianca un giovane lupo (la seconda risposta con più ricorrenze).

3.3.4 Selezione e raffinamento del concept finale

La fase di generazione dei concept narrativi ha fatto emergere diverse proposte valide. Tra queste, due concept in particolare sono stati ritenuti i più promettenti sia dal punto di vista narrativo sia rispetto agli obiettivi del progetto di tesi, seppur necessitando di un'ulteriore fase di approfondimento e raffinamento.

I concept selezionati sono *Manuale per non perdere i giorni*, generato da ChatGPT 5.2 Plus, e *Il lupo di carta*, generato da Gemini 2.5 Pro. Pur muovendosi in direzioni differenti, entrambi presentavano una struttura narrativa chiara, un arco emotivo coerente con i dati raccolti durante i concerti ed un forte potenziale visivo.

“Manuale per non perdere i giorni”

Il concept *Manuale per non perdere i giorni* era inizialmente incentrato su un sistema di catalogazione dei suoni, utilizzato per conservare il tempo ed i ricordi. Quando questo si rompe, il protagonista deve affrontare il caos che si genera e rivivere le memorie passate.

Nel corso del dialogo con le intelligenze artificiali, l'idea è stata raffinata per renderla più concreta e più facilmente traducibile in un progetto audiovisivo.

Prima di tutto, è stato proposto di inserire un elemento scatenante che fosse più facile da rappresentare e comprendere. Prendendo in considerazione la domanda relativa all'animale guida del terzo concerto, ChatGPT ha quindi suggerito di inserire un gatto – maggiormente votato nella risposta – che rompe accidentalmente i barattoli.

In seguito a questa scelta, l'intelligenza artificiale ha inserito un cambio di protagonista: da una figura maschile che possiede un complesso sistema di catalogazione di suoni e memorie si è passati ad un'anziana signora che possiede un gatto e racchiude i suoni in barattoli, per conservare i ricordi di una vita. Questa modifica ha evidenziato un possibile *bias* dell'IA, ovvero che un uomo possiede un sistema tecnologico mentre una signora anziana possiede un gatto, ma ha anche rafforzato il legame con i temi emersi nei concerti, in particolare quelli legati alla memoria, alla nostalgia ed al tempo che passa. In particolare, ChatGPT ha giustificato l'introduzione della protagonista anziana per avvalorare il tema del ricordo e della sua caducità.

La nuova versione del concept si sviluppa quindi attorno ad un'anziana signora che vive in un ambiente ordinato e controllato, in cui conserva i suoni della propria vita all'interno di barattoli

etichettati. I suoni diventano così metafora dei ricordi e del tentativo di preservarli per sempre. Il gatto rappresenta l'elemento destabilizzante: intrufolandosi in uno spazio a lui proibito – la stanza dei barattoli – ne rompe accidentalmente uno e libera un suono, innescando poi una reazione a catena. Da questo momento, il mondo visivo e sonoro si trasforma progressivamente: il colore e la musica tornano a permeare l'ambiente e la protagonista passa da una posizione di controllo e rigidità ad una di gioco e apertura.

Nel finale, la donna apre un barattolo vuoto, contrassegnato con un'etichetta strappata. Il “vento dei suoni”, ormai diffusosi nella città, conduce lo spettatore verso una casetta abbandonata in una foresta, dove si trova l'altra metà dell'etichetta e si sentono degli echi lontani di risate di bambini.

Si tratta di un concept emotivamente forte e coerente con i dati raccolti. Tuttavia, presenta una messa in scena complessa a causa delle molteplici interazioni proposte tra l'umana, il gatto ed i barattoli con particelle colorate. Inoltre, la storia ed il finale rischiano di diventare didascalici e scontati se non gestiti attentamente.

“Il lupo di carta”

Il concept de *Il lupo di carta* è stato ridefinito principalmente per rafforzarne la solidità narrativa e chiarire il percorso del protagonista durante il secondo atto, proposto in maniera troppo astratta.

Nella versione originale, il lupo protagonista veniva rappresentato come un lupo colorato di carta, ispirato alle risposte riguardanti il carattere di Mozart bambino – capriccioso, ribelle e sotto pressione paterna – che non voleva essere rinchiuso in un libro di favole. È stato proposto all'intelligenza artificiale di astrarre queste caratteristiche, ottenendo così una versione intermedia contenente un lupo ribelle che non voleva essere il cattivo della storia. L'idea era interessante, ma poco funzionale allo sviluppo dei personaggi secondari e alla progressione narrativa.

Dopo un confronto con Gemini, è emersa la versione definitiva del personaggio e del conflitto: il lupo colorato è oppresso all'interno della storia perché, al contrario degli altri animali presenti nel libro di favole in cui vive, non viene accettato nel branco di lupi grigi che dovrebbe essere la sua famiglia. Il branco ulula in maniera perfettamente sincronizzata ed appare visivamente uguale, mentre il protagonista ha un ululato diverso, stonato, ed il suo corpo è caratterizzato da macchie colorate. Questa versione, sfruttando anche l'intonazione dell'ululato, si avvicina ancor di più al tema musicale dei concerti.

Come nella proposta iniziale, vi è poi un gatto-mentore, libero dalle regole del mondo di carta, che appare per guidare il lupo attraverso un mondo tridimensionale sospeso, in cui la scoperta di suoni e melodie diventa il centro dell'azione narrativa. Attraverso l'interazione con elementi appartenenti al mondo dei lupi – legno, alberi, neve – il protagonista sperimenta e costruisce il proprio ululato.

È stato inserito anche un momento di tensione, in linea con la struttura de *Il Viaggio dell'Eroe*: una pagina con i lupi di carta riappare, offrendo al protagonista la possibilità di tornare indietro. Il lupo, scegliendo di proseguire la propria avventura, riuscirà a trovare finalmente la propria voce. Con un ultimo ululato liberatorio, richiamerà una foresta realistica, in cui un branco di lupi lo accetterà, anche se diverso da loro.

Nel finale, però, il lupo decide di fermarsi in una tana insieme al gatto-mentore. Gemini ha motivato questa scelta sottolineando come l'appartenenza non implichi la rinuncia della propria identità e storia.

Scelta del concept finale

Dopo una lunga riflessione, si è scelto di approfondire il concept *Il lupo di carta* sia per una ragione narrativa che per una tecnica.

Dal punto di vista narrativo, si è preferito scegliere una storia più astratta ma fortemente riconducibile ai temi dei concerti. Dal punto di vista tecnico, il concept permette di sperimentare in modo più evidente limiti e potenzialità delle IA generative, grazie alla presenza di tre stili diversi (illustrato, 3D, realismo) e di protagonisti non umani.

Inoltre, in seguito ad alcuni test preliminari, è emersa una notevole difficoltà delle IA nel gestire la fisica e le interazioni richieste per *Manuale per non perdere i giorni*. In particolare, la scena chiave in cui il gatto urta il barattolo che si rompe portava ad animazioni molto confusionarie ed innaturali, con il gatto che cercava di afferrare il barattolo e questo che esplodeva invece di rompersi. Si è preferito quindi escludere il concept, per evitare rischi e complicazioni.

3.3.5 Sviluppo della sceneggiatura

Una volta selezionato il concept finale, *Il lupo di carta*, si è passati alla fase di sviluppo della sceneggiatura.

Come stabilito in precedenza, la storia doveva svolgersi in tre atti ed essere il più possibile coerente sia con la struttura de *Il Viaggio dell'Eroe* sia con i tre concerti di *Narrazioni Parallele*.

In ciascun atto, ispirato alle emozioni emerse durante il rispettivo evento e alle tematiche trattate, il protagonista attraversa una fase diversa del proprio percorso identitario.

Nel primo atto, il lupo si ritrova oppresso all'interno della sua storia e rifiutato perché diverso (come il giovane Mozart, figura eccentrica e oppressa in gioventù dalla figura paterna). Insieme al gatto-mentore, il lupo trova il coraggio di fuggire e, nel secondo atto, si ritrova in uno spazio caotico e senza regole, dove scopre, sperimenta e sbaglia, fino a trovare la propria voce ed il coraggio di allontanarsi dal mondo di carta. Questa dimensione richiama il secondo concerto, *Song Writing*, dove il pubblico ha immaginato un mondo senza regole ed ideato una canzone. Infine, nel terzo atto, il lupo, ora in carne ed ossa, si muove attraverso una foresta, insieme ad un branco. Qui, prenderà la decisione di fermarsi con il gatto-mentore, sapendo però che avrà per sempre un branco a sostenerlo. Le immagini di natura, casa, malinconia e sicurezza, emerse durante il terzo concerto, *In scena con Verdi*, confluiscono nella scena finale, dove il lupo ed il gatto si addormentano insieme in una tana fra le radici di un albero, mentre il nuovo branco ulula in sottofondo.

Terminata la sceneggiatura, questa è stata tradotta in una shot list – una suddivisione dettagliata delle scene e della loro durata – che si è rivelata fondamentale per realizzare lo storyboard e, successivamente, il video.

Per organizzare al meglio il processo è stata realizzata una tabella su Notion^[60], suddivisa in colonne che includono: numero della scena, breve descrizione dell'azione, durata della scena, prompt per lo storyboard, prompt per il video. Notion si è rivelato utile perché ogni riga della tabella corrisponde ad una pagina con tutte le informazioni ben visibili ed organizzate; perfetta quindi sia per una visione d'insieme che di dettaglio.

Poiché si è scelto di utilizzare la tecnica di *image-to-video* è stato necessario produrre uno storyboard molto dettagliato, così che potesse poi essere animato con facilità. Per questo motivo la colonna della tabella riguardante il prompt per lo storyboard è stata suddivisa in due: una colonna per generare la clip iniziale del video e una per quella finale. Questa scelta si è rivelata utile nei casi di piano sequenza o di continuità visiva molto marcata, poiché ha facilitato la coerenza tra le clip e ridotto il rischio di variazioni indesiderate tra una generazione e l'altra.

Durante la realizzazione della shot list, alcune scene pensate come un'unica lunga unità sono state suddivise in clip più brevi. Questo per due motivazioni principali: da un lato, la necessità di rendere più chiaro alle IA cosa dovesse accadere in ogni singola inquadratura, evitando quindi di inserire un gran numero di azioni ed informazioni in un singolo prompt; dall'altro, i

limiti tecnici delle piattaforme di *image-to-video*, che consentono la generazione di clip della durata di pochi secondi e avrebbero quindi velocizzato e confuso le azioni richieste. Un esempio significativo è rappresentato dal primo atto, concepito come un piano sequenza. Da una singola clip/scena si è passati a 12 clip poi montate insieme come un'unica azione.

In una produzione cinematografica standard la shot list viene riorganizzata secondo l'ordine di riprese, che raramente rispecchia la cronologia della sceneggiatura e si basa su questioni tecnico-organizzative. Nel caso di questo progetto, invece, lo storyboard ed il video sono stati generati seguendo perfettamente l'ordine cronologico della sceneggiatura e della shot list. L'unica eccezione si ha nelle scene 31 e 33, dove il gatto-mentore è seduto (scena 31) e poi si alza ed entra nella tana (scena 33); per risparmiare tempo e crediti di generazione, questa è stata generata come un'unica scena, poi spezzata in fase di montaggio.

3.3.6 Ideazione del titolo

Durante la fase di affinamento del concept si è anche valutato se mantenere il titolo proposto con il concept iniziale o se rielaborarlo in modo più coerente con l'evoluzione del progetto. Il dialogo con le intelligenze artificiali e una rilettura dei temi ricorrenti emersi dai concerti hanno portato l'attenzione sulla dimensione del gioco, sul ritorno all'infanzia e sulla volontà di rivolgersi ad un pubblico intergenerazionale. Poiché entrambi i concept inizialmente selezionati presentavano una struttura ed un immaginario riconducibili alla dimensione della fiaba – intesa non come racconto esclusivamente infantile, ma come forma capace di parlare anche agli adulti – si è proposto di pensare a titoli evocativi, con un'influenza fiabesca.

Per il concept *Manuale per non perdere i giorni*, ChatGPT ha suggerito il titolo *Il gatto che rompe il silenzio*, il quale sembra richiamare i famosi titoli dello scrittore Luis Sepúlveda^[61], come *Storia di una gabbianella e del gatto che le insegnò a volare*. Il titolo, inoltre, è in grado di suggerire il tono emotivo e gli avvenimenti del racconto senza anticiparne la trama.

Al contrario, per il concept *Il lupo di carta* si è deciso di mantenere il titolo originario, la cui forza risiede nella semplicità e nella coerenza con la narrazione, presentando fin da subito il protagonista e le sue caratteristiche.

3.4 Fase di Pre-produzione: stile visivo

Dopo aver ultimato la sceneggiatura, si è passati alla successiva fase di pre-produzione, riguardante la definizione dello stile visivo dei personaggi e degli ambienti. Questo passaggio si rivelerà fondamentale per la realizzazione dello storyboard.

3.4.1 Definizione dello stile visivo

La sceneggiatura prevede un cambio di stile visivo per ognuno dei tre atti. Attraverso il dialogo con le intelligenze artificiali è stato definito un registro visivo distinto ma coerente e progressivo. Il percorso stilistico conduce dal 2D acquerello del libro illustrato ad un ambiente 3D pittorico – fortemente materico, caratterizzato da pennellate visibili e spesse – fino alla fisicità del mondo realistico.

Per definire con precisione gli esiti visivi desiderati e costruire, quindi, *prompt* coerenti per generare i personaggi e lo storyboard, sono state individuate reference di partenza, con l’obiettivo di studiarne il linguaggio e lo stile per trarne ispirazione, senza intenti imitativi.

Il primo atto si apre con una scena ambientata in una biblioteca, dallo stile realistico, concepita come richiamo ciclico al realismo dell’ultimo atto. Attraverso l’apertura di un libro si entra nel mondo della favola, in stile 2D ad acquerello.

Volendo richiamare le illustrazioni adatte ad un libro di favole per bambini si è pensato allo stile ad acquerello di Beatrix Potter^[62], illustratrice e scrittrice britannica, celebre per molteplici libri illustrati per l’infanzia. In particolare, è stato preso come modello *La storia di Peter Coniglio* (1902)^[63], libro successivamente adattato, anche in tempi recenti, in diverse produzioni animate.

L’impostazione visiva del primo atto prevede camera statica, in posizione zenitale sopra al libro, mentre i personaggi e le pagine si muovono, facendo fluire la storia. Vi è una struttura “a doppia pagina”, in cui ciascuna famiglia di animali è associata al proprio ambiente (ad esempio i cervi nella foresta, gli orsi sulla riva del fiume). In questo schema risalta quindi la sezione dei lupi: la pagina di sinistra presenta il branco di lupi grigi immerso in un ambiente che varia con la loro storia, mentre la pagina di destra è spoglia e contiene solo il lupo colorato, accentuando la sua condizione di solitudine.



Figura 3.6 Reference per l'Atto 1.



Figura 3.7 Esempio pagine dell'Atto 1.

Il passaggio dal primo al secondo atto è caratterizzato dal salto del protagonista attraverso lo squarcio nella pagina aperto dal gatto-mentore. Il lupo precipita in un tunnel di carta, colorato ad acquerelli, che gradualmente diventa solido e tridimensionale. Come suggestione visiva per

la caduta e il passaggio nel “nuovo mondo” si è pensato alla celebre sequenza di *Alice nel Paese delle Meraviglie* (1951)^[64].

Il secondo atto è ambientato in un ambiente 3D pittorico, astratto e sospeso, la cui definizione ha richiesto un processo di ricerca più articolato per riuscire a rappresentarne efficacemente la funzione di spazio di scoperta e transizione. La scelta finale è ricaduta su un'estetica tridimensionale con pennellate materiche evidenti, in grado di anticipare la tangibilità del realismo dell'ultimo atto, ma mantenendo ancora un richiamo agli acquerelli del primo.

Il riferimento principale per questo mondo è stato il cortometraggio di animazione francese *The Ones Who Never See The Moonlight* (2024)^[65], caratterizzato da un'atmosfera sospesa e da un'estetica tridimensionale e pittorica, con pennellate dense e visibili. Ulteriore suggestione – in particolar modo per la resa luminosa delle particelle – è stata data dal modello scaricabile di Blender *Splash Fox* (Blender 2.90)^[66].

In misura minore sono stati analizzati anche *DOGWALK* (2025, videogioco)^[67] e *Flow* (2024, Oscar come miglior film d'animazione nel 2025)^[68], per il riferimento ambientale, pur presentando una resa pittorica più piatta rispetto agli obiettivi stilistici del progetto.

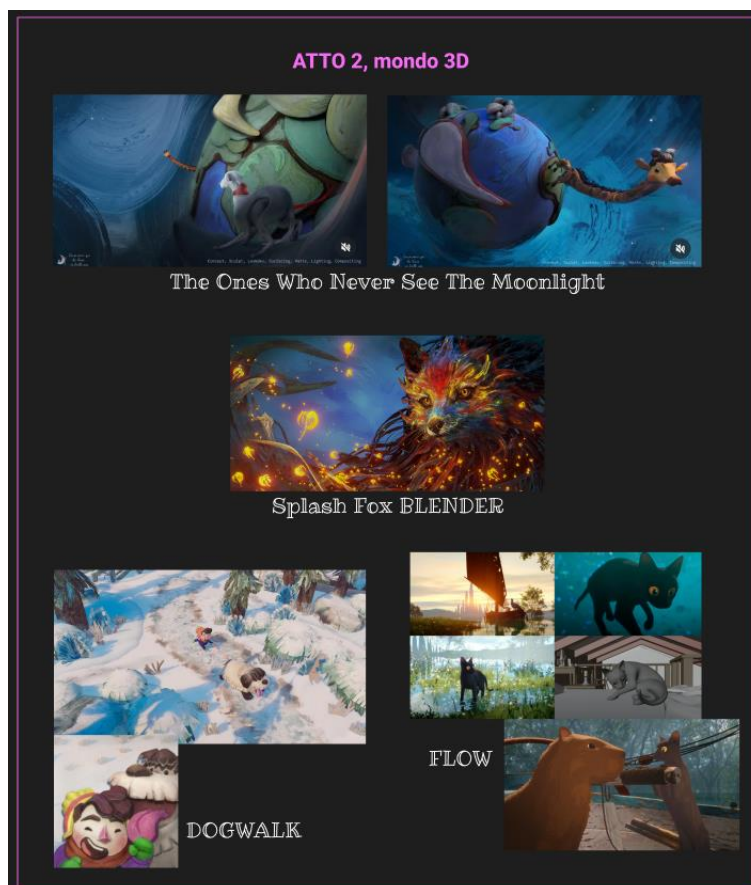


Figura 3.8 Reference per l'Atto 2.

Infine, il passaggio dal secondo al terzo atto prevede la nascita di una foresta realistica che pervade lo spazio. Questo atto è basato su uno stile visivo fotorealistico, sviluppato a partire da reference fotografiche di lupi e paesaggi innevati. Come riferimento audiovisivo è stato proposto dalle AI di prendere in considerazione *Vita di Pi* (2012)^[69], in particolare per la capacità di integrare animali in CGI (*Computer-Generated Imagery*) in contesti visivi realistici. In una pipeline tradizionale, infatti, una possibile soluzione avrebbe previsto l'utilizzo di ambienti reali combinati con animali realizzati in CGI.

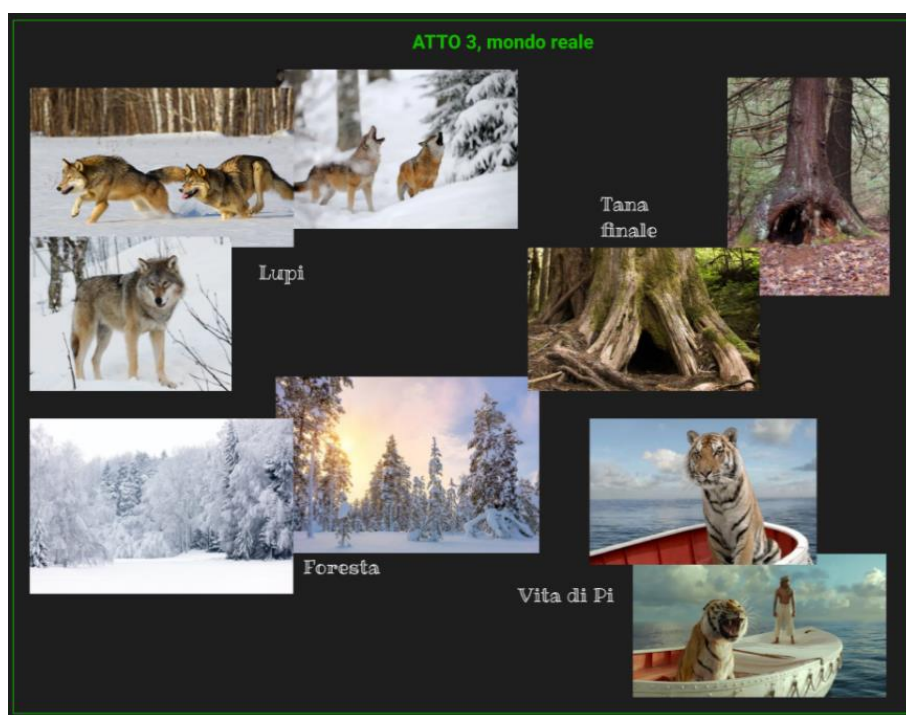


Figura 3.9 Reference per l'Atto 3.

3.4.2 Valutazione comparativa degli strumenti di generazione di immagini

Prima di proseguire con la generazione dei personaggi e dello storyboard, è stata condotta una valutazione comparativa dei principali strumenti di IA generativa per immagini, al fine di valutare, attraverso test pratici, quale fosse la più adatta per le esigenze del progetto.

Partendo dai *benchmark* delle principali IA disponibili sul mercato presentati nel [Capitolo 2](#), ne sono state selezionate cinque da valutare mediante test pratici.

La valutazione ha tenuto conto del comportamento del modello, della facilità d'uso, della flessibilità operativa e della compatibilità con i requisiti del progetto.

È opportuno precisare che i test sono stati condotti nel mese di dicembre 2025. Trattandosi di tecnologie in rapido sviluppo, potrebbero esserci stati nuovi aggiornamenti di funzionalità o correzioni di criticità.

Sono stati presi in considerazione sia i modelli di IA generativa per immagini integrati nei modelli linguistici – ChatGPT e Gemini, mentre Claude non supporta la generazione di immagini – sia piattaforme dedicate *text-to-image* (Flow, Sora e Midjourney).

ChatGPT (OpenAI): generazione immagini integrata in chat

ChatGPT^[51] è stato il primo strumento testato, in quanto già ampiamente utilizzato nella fase di scrittura. La piattaforma impiega il modello per immagini GPT-Image, evoluzione di DALL-E 3.

La generazione di immagini può avvenire all'interno di una chat, con possibilità di allegare immagini di riferimento, o tramite la sezione “Immagini”. Quest'ultima presenta un'interfaccia più simile ad un ambiente di generazione *standalone*, come Flow o Sora, e consente l'inserimento di prompt testuali, il caricamento di immagini reference e la possibilità di selezionare uno stile predefinito. Durante la generazione, si verrà comunque reindirizzati all'interno di una chat.

La presenza della chat integrata è una caratteristica vantaggiosa, poiché vi è la possibilità di dialogare con l'IA in qualsiasi lingua e richiedere nuove generazioni, migliorie o suggerimenti per ottimizzare i prompt.

Tuttavia, è emerso un limite significativo: le immagini vengono realizzate esclusivamente in formato 3:2, non ottimale per una pipeline video standard in 16:9.

Inoltre, la presenza di numerose immagini all'interno della stessa conversazione tende a rallentare le prestazioni, con possibilità di crash o tempi di attesa elevati. La creazione di chat separate mitiga il problema ma, anche se ChatGPT consente di racchiudere conversazioni diverse all'interno di una stessa cartella di progetto, si riduce drasticamente l'ordine spezzettando gli output in posizioni diverse.

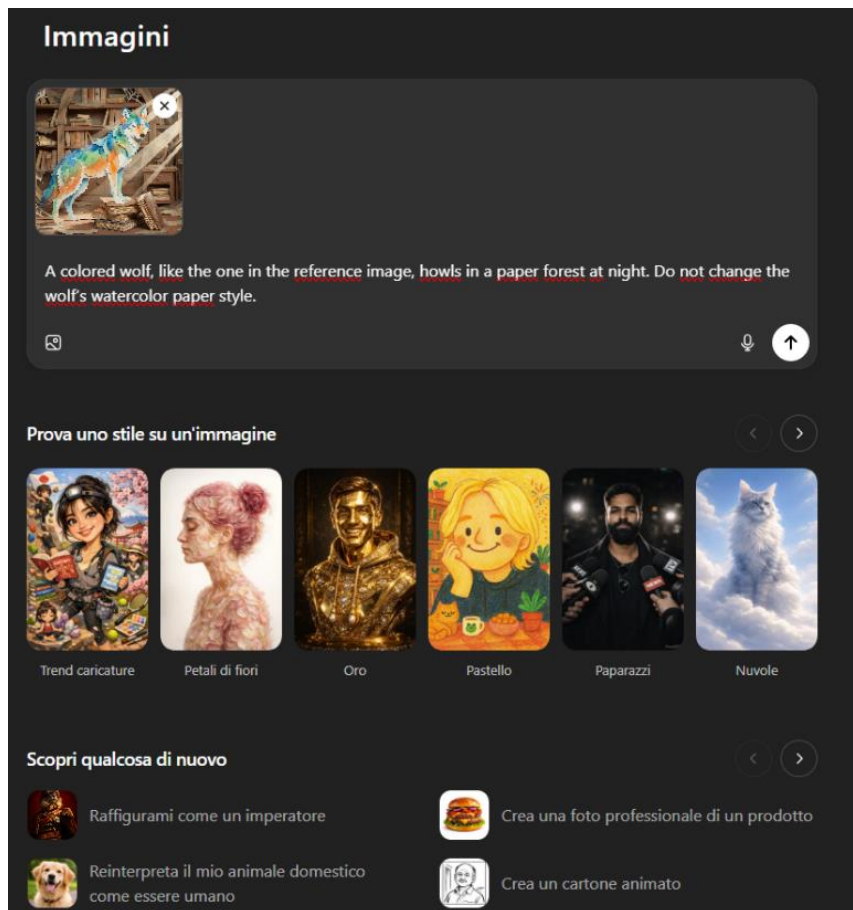


Figura 3.10 Sezione “Immagini” ChatGPT.

Gemini (Google): generazione immagini integrata in chat

Il secondo modello testato è Nano Banana Pro, disponibile attraverso la chat con Gemini^[53]. Come per ChatGPT, questo approccio mantiene il vantaggio del dialogo con il modello linguistico, con la possibilità di definire meglio le modifiche e farsi aiutare nel *prompting*, utile soprattutto per utenti meno esperti. Vi è la possibilità anche di caricare una o più immagini di reference.

Rispetto a ChatGPT, Gemini offre la possibilità di un output in 16:9. Tuttavia, non consente un editing preciso dell'immagine: non si può selezionare un'area in cui effettuare le modifiche; è necessario descriverle testualmente, avviando la rigenerazione completa di una nuova immagine, con il rischio di alterazioni indesiderate.

Inoltre, durante i test è emerso un comportamento talvolta imprevedibile: la tendenza a modificare lo stile dell'immagine di reference anche quando esplicitamente richiesto di mantenerlo invariato.

Data l'importanza della coerenza visiva per la buona riuscita del progetto, Gemini è stata esclusa.

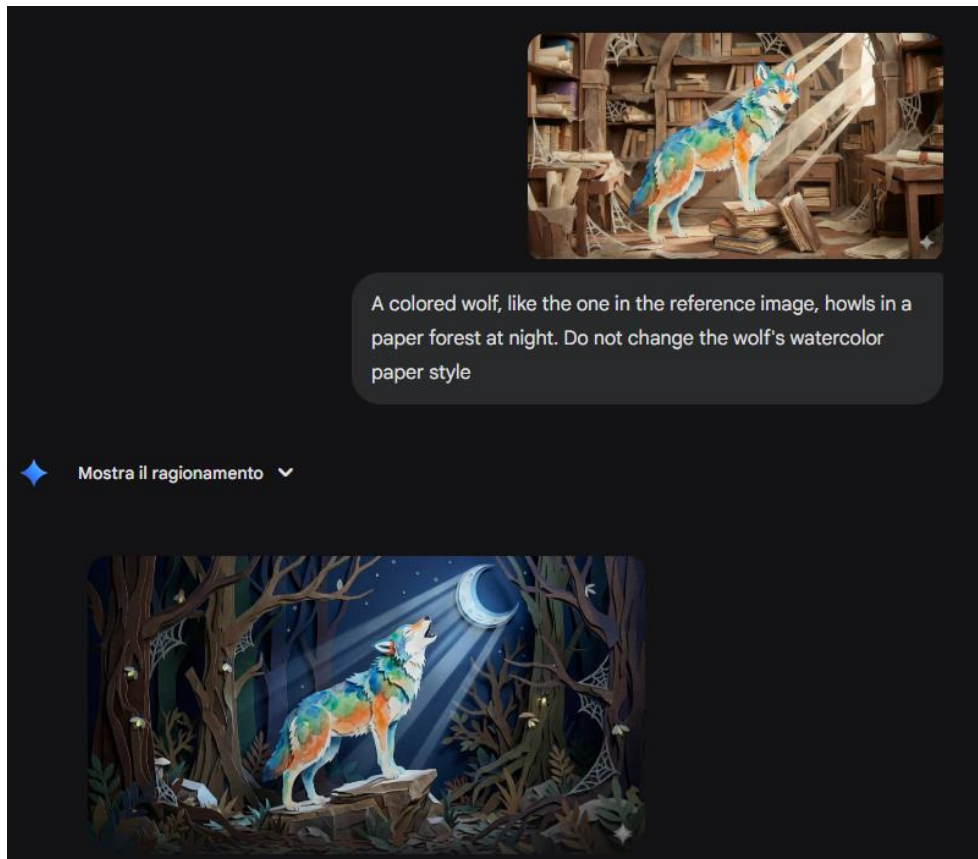


Figura 3.11 Generazione immagini con Gemini.

Sora (OpenAI)

Pur appartenendo all'ecosistema di OpenAI e basandosi sul medesimo modello di generazione visiva (GPT-Image), Sora^[70] offre un ambiente di lavoro più strutturato rispetto a ChatGPT, presentando al contempo analogie e limitazioni rispetto a quest'ultima.

Come ChatGPT, Sora offre la possibilità di caricare una o più immagini di reference da affiancare al prompt e di scegliere fra alcuni stili predefiniti. Tuttavia, a differenza della sezione "Immagini" di ChatGPT, questi sono indicati tramite codici testuali (ad esempio "cartoonify-v01") e non sono accompagnati da anteprime visive, rendendo meno immediata la comprensione del risultato, soprattutto per utenti poco esperti. Inoltre, non è possibile avviare un dialogo conversazionale con il modello ed è consigliato l'utilizzo esclusivo della lingua inglese per ottenere risultati affidabili.

Al contrario di ChatGPT, Sora consente di selezionare il formato di output (1:1, 2:3, 3:2) e, tramite la funzionalità "Remix", permette di effettuare modifiche locali selezionando aree specifiche dell'immagine.

Un ulteriore punto di forza è la possibilità di generare un video a partire dall'immagine prodotta, senza doverne effettuare il download. Questa caratteristica risulta essere particolarmente

interessante nell'ottica di una pipeline audiovisiva basata su IA, garantendo un flusso di lavoro integrato.

Tuttavia, la piattaforma ha evidenziato alcune criticità. In primo luogo, non supporta il formato 16:9 requisito fondamentale per il progetto. In secondo luogo, la generazione è risultata mediamente più lenta e sensibile a prompt complessi rispetto al competitor Flow. Infine, dal punto di vista estetico, le immagini ottenute sono risultate complessivamente simili a quelle generate con ChatGPT (presumibilmente poiché entrambi utilizzano GPT-Image). Risalta, però una tendenza di Sora ad alterare lievemente gli stili o ad omettere alcuni dettagli.

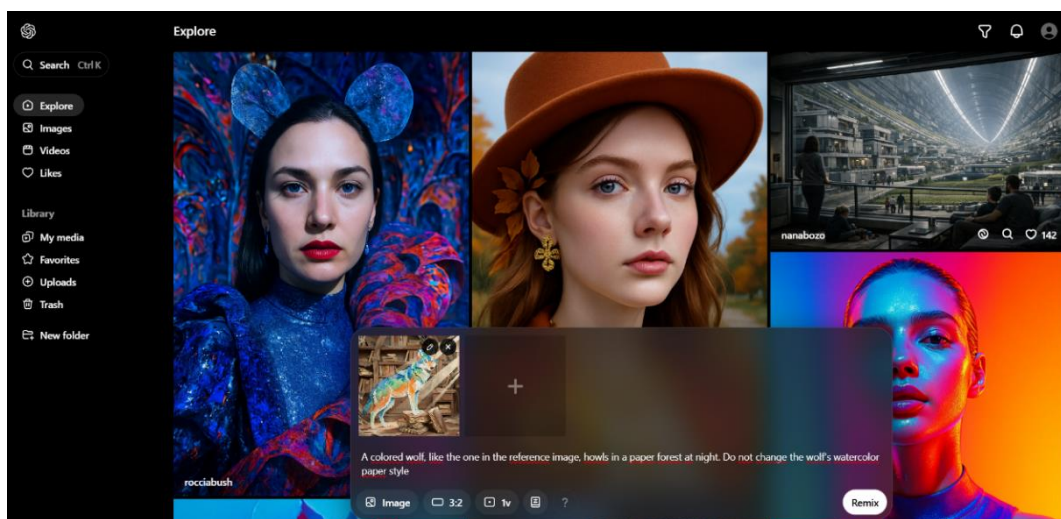


Figura 3.12 Interfaccia di Sora per la generazione immagini.

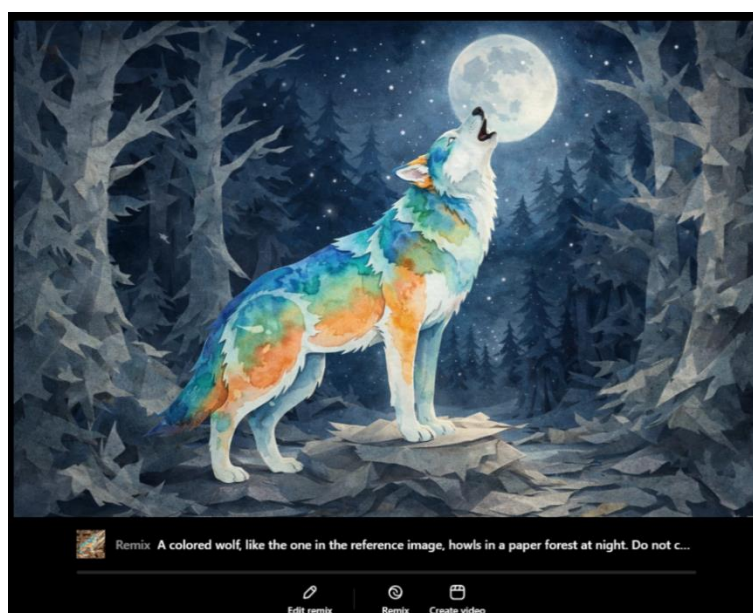


Figura 3.13 Interfaccia di Sora per la modifica delle immagini.

Flow (Google)

Flow^[71] rappresenta un ambiente avanzato rispetto a Gemini e ha restituito risultati interessanti e diversificati, pur basandosi sul medesimo modello di generazione visiva (Nano Banana Pro).

Vantaggi rilevanti rispetto a Gemini – e alle altre AI prese in esame – sono il supporto del formato in 16:9 e la possibilità di esportare le immagini con risoluzione fino a 2K. Quest’ultima caratteristica si è rivelata particolarmente vantaggiosa: partire da input di alta qualità mitiga la perdita qualitativa durante la generazione video con IA, le quali – come verrà discusso nei paragrafi successivi – consentono output di massimo 720p o 1080p.

Flow offre anche la possibilità di effettuare un editing locale dell’immagine, anche su più aree contemporaneamente mediante diverse forme di selezione. Utilizzando il *tool* apposito, fornisce anche la possibilità di inserire testi all’interno dell’immagine.

Al contrario di Sora, Flow interpreta bene sia prompt complessi sia prompt scritti in lingua italiana (automaticamente tradotta in inglese al momento dell’invio). Inoltre, a differenza della corrispondenza quasi diretta osservata tra gli output di Sora e di ChatGPT, Flow ha prodotto risultati visivamente differenti rispetto a Gemini e ha mostrato una grande stabilità nel mantenimento delle immagini di reference.

Dal punto di vista dell’organizzazione delle immagini, Sora unisce tutti i risultati in un’unica galleria, mentre Flow consente di creare progetti separati all’interno dei quali archiviare in modo ordinato sia il materiale generato che le immagini di reference.

Alla luce delle prestazioni offerte e della disponibilità illimitata e gratuita di crediti, tramite il piano Google Pro per studenti (anch’esso gratuito), Flow è stato selezionato come strumento principale per la generazione dello storyboard.

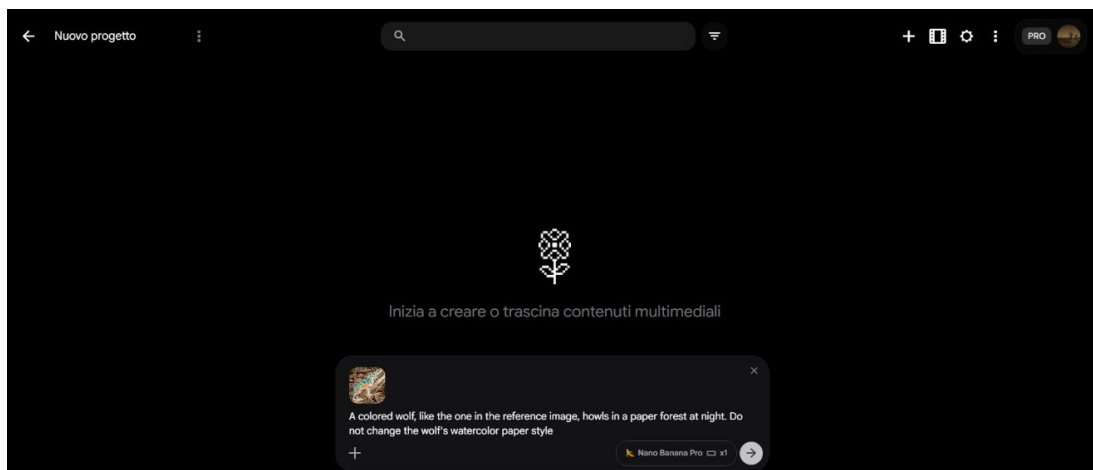


Figura 3.14 Interfaccia di Flow per la generazione immagini.

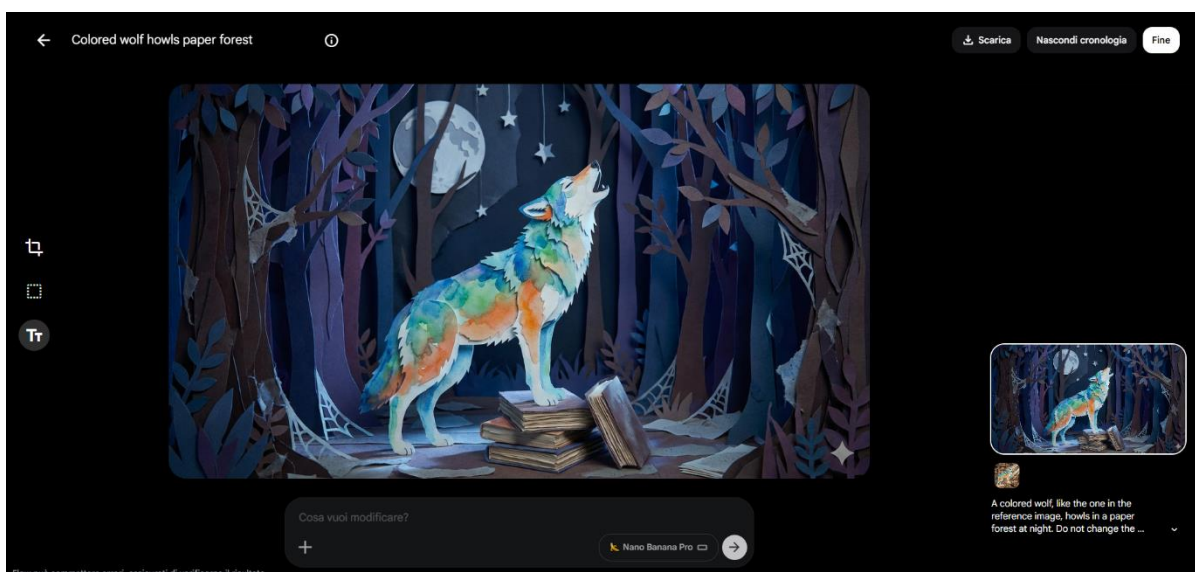


Figura 3.15 Interfaccia di Flow per la modifica delle immagini.

Midjourney

Midjourney^[72], pur non trovando posto nelle classifiche dei *benchmark* del Capitolo 2, è stato inizialmente preso in considerazione data la sua ottima reputazione, ma è stato rapidamente escluso per ragioni operative ed economiche. La piattaforma, infatti, attualmente non offre accessi gratuiti di prova tramite il sito ufficiale e tantomeno attraverso il server Discord dedicato. Inoltre, adotta un modello di *pricing* basato sul tempo di utilizzo della GPU dei server durante la generazione.

Considerata l'elevata quantità di immagini e di iterazioni prevista per lo storyboard e la sua importanza per la successiva fase di generazione video, l'utilizzo di Midjourney avrebbe introdotto un livello di incertezza sui costi e sui possibili output non compatibile con gli obiettivi sperimentali del progetto né con una logica produttiva orientata all'ottimizzazione delle risorse, come quella solitamente adottata per le pipeline cinematografiche.

Nota sull'uso complementare di ChatGPT

Sebbene la scelta principale sia ricaduta su Flow, ChatGPT è stato comunque impiegato per la generazione di alcune immagini di riferimento. In particolare, per la scena iniziale della biblioteca e del libro, ChatGPT ha prodotto immagini con una resa maggiormente cinematografica rispetto a Flow. Queste, sono state fornite a Flow come reference visiva di partenza.

Con questo passaggio si mette in luce l'importanza e l'utilità di un approccio integrato e complementare tra i diversi strumenti, analogo alla collaborazione tra colleghi e reparti di una troupe cinematografica.

3.4.3 Definizione dei personaggi

Prima di generare lo storyboard è stato necessario definire visivamente lo stile dei personaggi, così da poter produrre un model sheet da fornire alle intelligenze artificiali generative per aiutarle a mantenere la coerenza dei personaggi durante le diverse generazioni.

È utile fare una premessa riguardo alla distinzione tra “character sheet” e “model sheet”, termini spesso utilizzati erroneamente come sinonimi. Il *character sheet* è un documento che descrive completamente il personaggio, racchiudendone identità, aspetto, caratteristiche; serve da riferimento concettuale e narrativo. Il *model sheet*, invece, è una rappresentazione visiva del personaggio – spesso in pose differenti – per garantire coerenza nelle diverse pose.

Per questo progetto sono stati realizzati entrambi: un character sheet come guida, riassunto e supporto per le diverse IA testuali e per tenere traccia delle decisioni prese per ogni personaggio e un model sheet utilizzato come reference del personaggio da dare alle IA per la generazione di immagini. I model sheet di ogni personaggio sono racchiusi all'interno del character sheet e, per una visione completa, si rimanda all'[Appendice E](#).

Per realizzare i model sheet si è scelto di seguire lo stesso procedimento adottato per la generazione dei concept narrativi, ovvero, utilizzare ChatGPT, Claude e Gemini e chiedere loro dei prompt per la generazione.

Per evitare influenze pregresse, sono state create nuove conversazioni su ChatGPT e Claude, fornendo sia la sceneggiatura che i dettagli sul progetto, mentre con Gemini si è mantenuta la chat originale poiché già contenente il concept de *Il lupo di carta*.

Questo approccio ha evidenziato nuovamente delle similitudini e delle differenze interessanti fra i tre modelli.

In primo luogo, tutti e tre i modelli hanno interpretato correttamente la richiesta di realizzare un model sheet e hanno proposto un'immagine di riferimento con più viste del personaggio, su sfondo neutro. In autonomia, hanno fornito i prompt scritti in inglese, sostenendo che così sarebbero stati meglio compresi dalle principali piattaforme di IA generative per immagini.

Dal confronto, emergono differenze significative nel livello di dettaglio delle risposte: ChatGPT fornisce un prompt molto strutturato, con indicazioni precise su stile, texture, illuminazione e

funzione dell'immagine (“designed for animation reference”) mentre Gemini propone una versione più sintetica, descrivendo solo il personaggio e chiedendo di posizionarlo in più pose su sfondo neutro. Claude ha invece proposto in autonomia due prompt con due diversi livelli di dettaglio: uno per un'eventuale fase esplorativa di ricerca reference ed uno per la creazione definitiva del personaggio.

Inoltre, come emerso per la fase di scrittura, risaltano i tre stili diversi adottati dalle IA: ChatGPT risulta essere analitico e preciso, inserendo un lungo elenco di dettagli separati da virgole, Gemini sintetico e Claude approfondito e quasi poetico, con anche alcune libere aggiunte.

Tabella 3.8 Esempio di prompt per il model sheet del personaggio Lupo di Carta.

ChatGPT	Gemini	Claude
Character sheet, full body turnaround (front, ¾, side, back), stylized wolf made of painted paper, paper texture clearly visible, slightly irregular edges, hand-painted watercolor style, soft color gradients, muted but expressive colors, visible brush strokes, illustration style, poetic and delicate, neutral pose, clean white background, studio lighting, designed for animation reference.	A character sheet reference image showing multiple views (front, side, back view) of a stylized wolf made of painted paper, watercolor painting style. Flat style, isolated on a plain background.	<p>Prompt base: Character sheet of a colorful wolf made of paper, watercolor painting style, vibrant colors including blue, orange, teal, and cream. Multiple angles: front view, side view, ¾ view, back view. Paper craft texture with visible brushstrokes and watercolor bleeding effects. Ethereal and artistic appearance. White background. Professional character design sheet layout with labels.</p> <p>Prompt dettagliato: Professional character turnaround sheet of a stylized wolf, paper craft art style with watercolor painting technique. The wolf features vibrant mixed colors: turquoise blue fur on the back, orange and cream on the belly and legs, flowing mane with multiple color layers. Four views: front, side profile, ¾ angle, rear. Delicate paper texture with visible watercolor brush strokes, soft color gradients and bleeding edges. Whimsical and dreamlike aesthetic. Clean white background, character reference labels, consistent proportions across all angles. High detail, artistic illustration.</p>

I prompt ottenuti sono stati testati sulle piattaforme per la generazione di immagini citate nel paragrafo precedente (ChatGPT, Gemini, Sora e Flow) per effettuare delle valutazioni e comparazioni adeguate.

Come già evidenziato, i risultati prodotti dagli strumenti di OpenAI (ChatGPT e Sora) sono simili tra loro, così come quelli ottenuti con gli strumenti di Google (Gemini e Flow). Le principali differenze riguardano le modalità operative, tra cui la presenza o meno della possibilità di avviare una conversazione con memoria, utile per iterare progressivamente le modifiche.

Tuttavia, si sono riscontrate anche alcune differenze e criticità importanti.

Gemini, utilizzando il proprio prompt, tendeva a generare collage con tre pose del personaggio, mentre con i prompt forniti da ChatGPT e Claude riusciva a produrre model sheet con le pose richieste, in ordine. Un'ulteriore difficoltà è emersa durante la generazione del Lupo Colorato in versione realistica: sebbene la pelliccia risultasse accurata e fotorealistica, le macchie di colore richieste venivano ridotte ed accompagnate da tonalità di marroni e grigi non richiesti, soprattutto in prossimità del muso, come se il modello tendesse ad aggiungere alcuni colori ad una fotografia di un lupo reale piuttosto che crearne uno colorato e "fantastico". Questa problematica è stata successivamente risolta grazie all'utilizzo di Flow, combinando il model sheet fotorealistico di Gemini con quello di ChatGPT, i cui colori erano perfetti ma il lupo appariva finto. Si è quindi ottenuto un model sheet per il terzo atto rappresentante il Lupo Colorato in versione realistica e cromaticamente coerente con il concept.

Sora e ChatGPT, invece, presentavano anomalie nella generazione dei lupi grigi di carta. Tendevano, infatti, a produrre dei cuccioli, anche in presenza di prompt in cui venisse esplicitato che dovevano essere esemplari adulti. Tale comportamento rimane inspiegabile, poiché non si è verificato per nessun altro personaggio o stile.

Come prevedibile, variazioni anche minime nei prompt hanno prodotto risultati differenti. La selezione finale dei model sheet da integrare nei rispettivi character sheet è stata effettuata manualmente, privilegiando le soluzioni più coerenti con la direzione estetica del progetto.

Un ulteriore aspetto emerso durante la definizione visiva dei personaggi, che avrebbe potuto rappresentare un'enorme criticità per il progetto, riguarda il fenomeno dell'Uncanny Valley, già analizzato nel [Capitolo 2 \(2.2.5\)](#). La letteratura, infatti, evidenzia come questo possa manifestarsi anche nelle rappresentazioni di animali, quando si verifica una condizione di "realismo imperfetto". Animali altamente realistici ma caratterizzati da incongruenze morfologiche, espressive o cromatiche possono generare disagio negli spettatori, soprattutto se questi hanno

molta familiarità con la specie^{25,26}, analogamente a quanto succede per i soggetti umani (dove il disagio scaturisce soprattutto dalla nostra profonda familiarità con le proporzioni anatomiche e l'espressività).

Esempio recenti in ambito cinematografico include il cane Buck nel film *Il richiamo della foresta* (2020)^[75], la cui resa in CGI combina tratti altamente realistici a tratti ed espressioni artificiali, generando un effetto ambiguo e talvolta inquietante. Sul fronte opposto, gli animali dei live-action Disney *Il Re Leone* (2019) e *Mufasa* (2024)^[76] dimostrano come la ricerca di un realismo perfetto, presumibilmente introdotto per arginare il rischio di Uncanny Valley, porti ad una drastica limitazione dell'espressività facciale dei protagonisti. Questo ha prodotto un distacco emotivo in molti spettatori, restituendo loro la sensazione di assistere ad un documentario con animali parlanti.



*Figura 3.16 A sinistra: un'immagine da "Il richiamo della foresta".
A destra: un'immagine da "Il re leone".*

Con la recente e rapida diffusione dei modelli di intelligenza artificiale generativa, il rischio di realizzare opere affette da Uncanny Valley si è ulteriormente amplificato, data la natura ancora instabile degli algoritmi.

Nel contesto di questo progetto, la criticità risultava particolarmente rilevante, poiché l'obiettivo era ottenere immagini tendenzialmente realistiche e di non utilizzare uno stile cartonesco (spesso impiegato appositamente per evitare l'effetto Uncanny Valley). Di conseguenza, l'intervento umano si è rivelato fondamentale anche per controllare gli output prodotti ed intervenire tempestivamente nel caso di problematiche.

²⁵ A. Sierra Rativa et al., *The Uncanny Valley of a Virtual Animal*, (2022) [73]

²⁶ V. Schwind et al., *Is there an uncanny valley of virtual animals?*, (2018) [74]

3.4.4 Generazione dello storyboard

Partendo dalla shot list definita in fase di pre-produzione, si è passati alla realizzazione dello storyboard tramite l'utilizzo di Nano Banana Pro sulla piattaforma Flow.

Mantenendo un approccio in continuità con quello adottato nelle fasi precedenti, alle tre IA testuali – ChatGPT, Claude e Gemini – è stato chiesto di produrre i prompt destinati alla generazione delle singole inquadrature. Dopo alcuni test preliminari è emerso che i prompt generati da ChatGPT tendevano a produrre risultati visivamente più coerenti con l'idea; tuttavia, in alcuni casi è emerso anche che prompt eccessivamente dettagliati (in particolar modo riguardo al posizionamento di luci e soggetti) portassero più facilmente ad errori di varia natura, come se creassero confusione nell'IA generativa. Di conseguenza, si è progressivamente optato per formulazioni più essenziali, capaci di guidare la composizione senza sovraccaricare il modello con vincoli e dettagli troppo specifici. Per questo motivo, si è anche scelto di non sviluppare un floor plan tradizionale – documento tipico della pre-produzione audiovisiva che definisce con precisione la posizione di luci, attori, punti camera e molto altro – poiché si sarebbe rivelato controproducente per la generazione tramite IA.

Un'ulteriore differenza rispetto ad una pipeline cinematografica tradizionale è il livello di dettaglio dello storyboard. In questa, infatti, i disegni possono essere abbozzati, per rappresentare concettualmente l'idea registica, e sono sempre accompagnati da informazioni tecniche utili (ad esempio: tipo di lenti, movimenti di camera). Al contrario, in questo progetto, si è data grande importanza alla produzione di immagini di alta qualità e ricche di dettagli, poiché ciascuna di esse sarebbe stata utilizzata come base diretta per la generazione video.

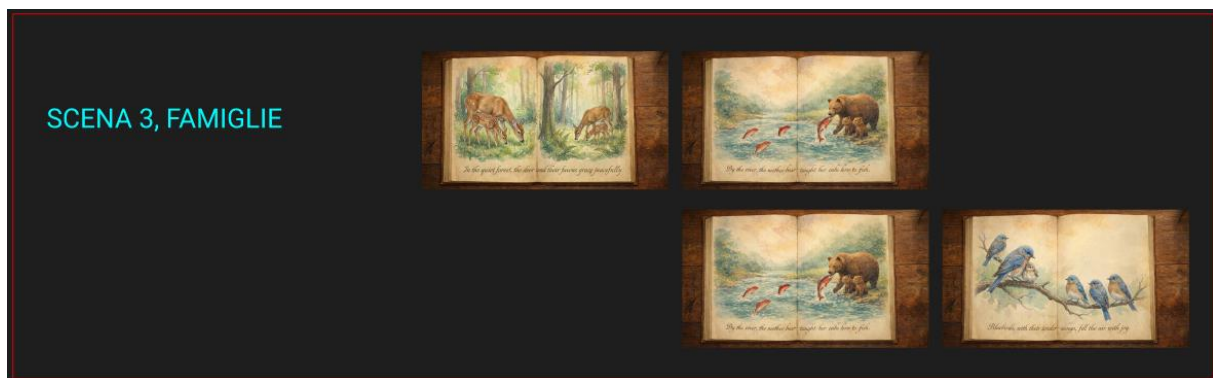


Figura 3.17 Esempio di storyboard realizzato per il progetto.

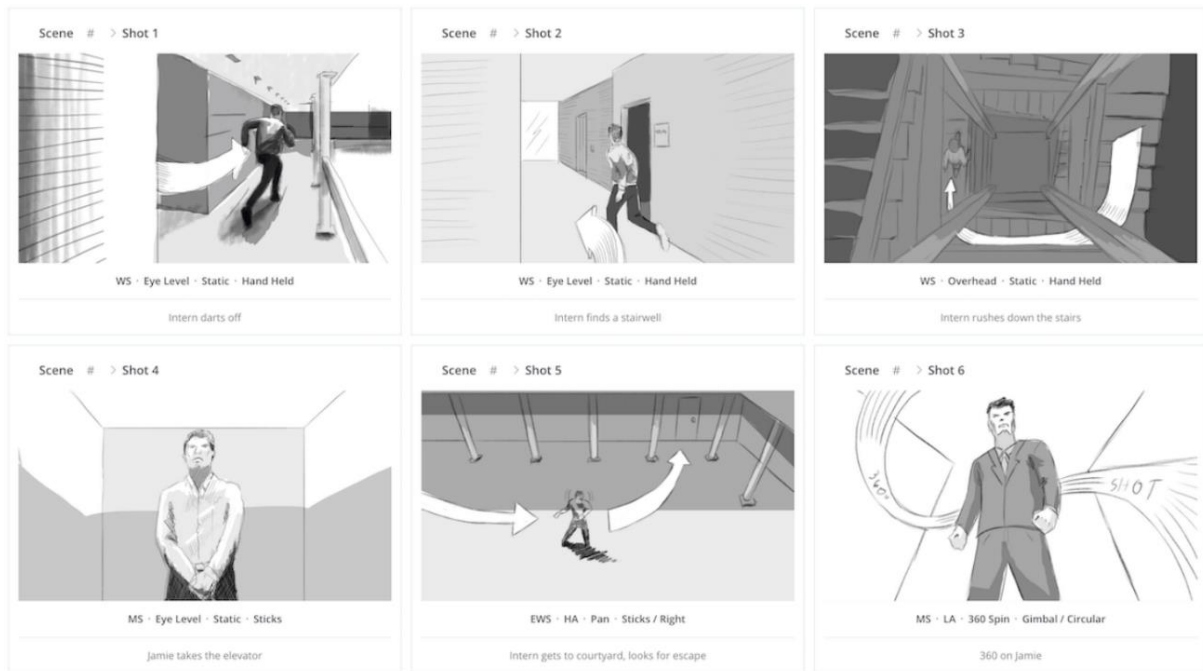


Figura 3.18 Esempio di storyboard professionale (StudioBinder.com).

Lo storyboard finale è composto da 82 immagini, a fronte di un numero di generazioni complessive più che doppio dovuto al processo di ri-generazione per ottimizzare i dettagli. Le tavole definitive sono state organizzate in ordine sequenziale sulla piattaforma Figma, per facilitarne la consultazione ed il controllo. Per poterle consultare, si rimanda all'[Appendice F](#).

Per ciascuna generazione effettuata sulla piattaforma Flow, i prompt testuali elaborati da ChatGPT sono stati combinati con la possibilità di caricare delle immagini di riferimento. Nello specifico, si è scelto di allegare i model sheet dei personaggi, precisando all'interno del prompt l'esatta corrispondenza tra la reference caricata ed il rispettivo soggetto (ad esempio: "colored wolf as the first image reference [...] black cat as the second image reference").

In determinate sequenze si è rivelato utile includere, come riferimento aggiuntivo, anche il fotogramma precedente, per garantire la continuità visiva e stilistica dell'azione.

Tuttavia, nonostante tali accorgimenti, vi sono state alcune criticità che hanno reso necessario l'intervento umano, evidenziando come i modelli generativi attuali non siano in grado di garantire autonomamente una continuità visiva di stile, personaggi e ambienti all'interno di sequenze articolate.

Il primo atto ha presentato le maggiori difficoltà, in particolare per la presenza di un piano sequenza in cui il lupo colorato interagisce con gli altri personaggi. Nonostante l'utilizzo

congiunto di reference multiple – il model sheet e le immagini dell’azione precedente – Flow ha mostrato una limitata capacità di mantenere la continuità visiva del personaggio per generazioni continue. Sono state osservate variazioni indesiderate nella geometria del tavolo (caratterizzato da segni di usura nel legno), nella palette cromatica del lupo (progressivamente tendente al verde) e nel livello di zoom tra iterazioni successive della stessa inquadratura. Tutto ciò non era ovviamente compatibile con un piano sequenza la cui impostazione prevedeva un’inquadratura statica ed i soli personaggi in movimento.

Nei casi in cui le correzioni tramite ulteriori iterazioni con l’IA risultavano insufficienti, si è reso necessario un intervento manuale utilizzando GIMP^[77], software di *image editing*. In particolare, per sistemare le discontinuità fra il tavolo, il libro e il livello di zoom, è stata selezionata come base la versione del tavolo proveniente dalla scena della famiglia di uccellini, poiché ritenuta visivamente più coerente. Dalle inquadrature precedenti e successive è stato ritagliato il libro, ridimensionato a seconda delle necessità, ed integrato sul tavolo di riferimento, garantendo così la continuità visiva del piano sequenza. Allo stesso modo, per correggere le variazioni cromatiche del lupo, questo è stato ritagliato e caricato nuovamente su Flow, chiedendo di mantenere la stessa posa ma utilizzare i colori del model sheet. Isolando il lupo dal resto della scena si è garantita la massima attenzione dell’IA sul personaggio e la nuova versione generata – con colori e posa corretti – è stata inserita nella composizione originale.

Il secondo atto ha richiesto inizialmente una revisione della sceneggiatura, in quanto la prima versione risultava eccessivamente astratta per una traduzione visiva efficace. Una volta chiariti gli elementi e le azioni da rappresentare, questa sezione si è rivelata la più agevole da generare, probabilmente grazie alla presenza di un ambiente unitario, all’assenza di variazioni luminose complesse e al numero contenuto di personaggi in scena. Inoltre, essendo caratterizzata da molti elementi astratti e non ben definiti, non c’erano vincoli “ideologici” da rispettare.

Nel terzo atto le problematiche di continuità del personaggio si sono ridotte, ma sono emerse difficoltà significative nella gestione della progressione luminosa. L’obiettivo era rappresentare una transizione graduale dalla luce fredda della blue hour iniziale alla golden hour finale, in coerenza con l’evoluzione emotiva del protagonista e con il sorgere del sole. Tutti i modelli testati tendevano ad aumentare l’intensità luminosa in modo eccessivo e poco controllato.

In seguito a numerosi tentativi, è stata selezionata la soluzione visivamente più equilibrata, accettando però alcuni compromessi.

Particolarmente complessa si è rivelata anche la gestione della tana del gatto, ripresa da angolazioni multiple. In questo caso si è faticato a mantenere coerenza spaziale e continuità dei personaggi, soprattutto nella sequenza finale della carrellata dall’interno verso l’esterno, dove risultava difficile preservare simultaneamente l’inquadratura, la visibilità parziale dei personaggi nella tana e la stabilità della camera. La reference di partenza presentava infatti la

camera leggermente ruotata rispetto alla tana, poiché appartenente ad un'inquadratura differente, mentre per la scena finale era richiesta una vista frontale perfettamente ortogonale, con i personaggi parzialmente visibili all'interno. Questa discrepanza ha generato ripetuti errori di prospettiva.

Dopo numerose generazioni si è ottenuta l'inquadratura desiderata, senza personaggi e con la luce coerente per l'alba finale. Partendo da questa, i dettagli sono stati progressivamente aggiunti in generazioni consecutive.



Figura 3.19 Prima immagine della tana, usata come reference per le successive.



Figura 3.20 Immagine finale della tana.

Un ulteriore problema generale riguardava la progressiva perdita di qualità nelle modifiche iterative: ogni rigenerazione introduceva infatti un leggero degrado della qualità dell'immagine. Per mitigare questo effetto, prima dell'esportazione, veniva richiesta un'operazione di *upscaling*, generando un'ultima versione dell'immagine e specificando esplicitamente di non alterarne il contenuto. Tale accorgimento non si è tuttavia rivelato sempre affidabile; in particolare, nell'Atto 3 si sono verificati casi di aumento indesiderato dell'intensità luminosa.

Nel complesso, questa fase ha confermato che, allo stato attuale, i modelli per la generazione di immagini richiedono una supervisione umana significativa. Inoltre, in presenza di sequenze articolate con la necessità di mantenere continuità visiva, risultano necessari interventi di perfezionamento e correzione, anche con software appositi. Ciò presuppone un'attenta attenzione ai dettagli e, talvolta, competenze di post-produzione avanzate.

3.5 Fase di Pre-produzione: osservazioni e conclusioni

La fase di pre-produzione ha evidenziato come gli strumenti di intelligenza artificiale generativa possano costituire un supporto efficace nelle fasi iniziali di un progetto audiovisivo, operando non come sostituti dell'autore ma come assistenti creativi. In particolare, l'utilizzo combinato di modelli linguistici e *text-to-image* ha consentito di tradurre rapidamente i dati qualitativi astratti, reperiti durante gli eventi di *Narrazioni Parallele*, in materiale visivo concreto, ampliando le possibilità di esplorazione creativa e progettuale.

L'impiego delle IA durante questa fase ha portato numerosi vantaggi, tra cui la significativa accelerazione dei processi di brainstorming e di creazione narrativa. Le IA testuali si sono dimostrate particolarmente efficaci nell'elaborazione dei dati raccolti, riuscendo ad evidenziare sia emozioni ricorrenti che differenze e dettagli importanti, realizzando proposte narrative coerenti con quanto vissuto dal pubblico. È però necessario sottolineare come la fase preliminare di *prompt design* sia stata fondamentale per orientare correttamente i modelli verso gli output desiderati.

Allo stesso modo, la possibilità di generare rapidamente le immagini per i personaggi e per lo storyboard mediante gli appositi strumenti *text-to-image* ha supportato in modo significativo il processo di pre-produzione, riducendo tempi e competenze tecniche richieste da una pipeline tradizionale. Gli strumenti utilizzati hanno mostrato buone potenzialità in termini di qualità estetica e rapidità operativa.

Gli aspetti positivi sono però stati accompagnati da criticità rilevanti. La qualità dei risultati si è dimostrata infatti fortemente dipendente dalla precisione dei prompt e dalla presenza di una supervisione autoriale consapevole. È emerso con chiarezza il limite strutturale della coerenza visiva per i modelli di generazione di immagini, come dimostrano le variazioni cromatiche del Lupo Colorato e le incongruenze geometriche del tavolo e del libro. Gli attuali modelli generativi faticano a mantenere stabili nel tempo dettagli di personaggi e ambienti. Quest'ultima problematica ha reso necessario un intervento correttivo manuale tramite software di editing tradizionale, come GIMP, richiedendo delle competenze tecniche specifiche e aprendo una riflessione critica rispetto alla reale capacità delle IA di democratizzare l'accesso alle pipeline generative. Qualora si voglia ottenere risultato qualitativamente coerenti, l'intervento umano specializzato risulta, allo stato attuale, ancora essenziale.

Vi è inoltre un ulteriore elemento da tenere in considerazione: il rischio di Uncanny Valley. Questo richiede una supervisione autoriale rigorosa in fase di *prompt design* ma, soprattutto, nella verifica e selezione degli output.

In sintesi, la pre-produzione si è rivelata una fase in cui l'intelligenza artificiale generativa, ad oggi, offre benefici concreti in termini di supporto creativo e rapida esplorazione. Parallelamente, però, emerge la centralità della supervisione umana per la selezione critica degli output e per il mantenimento di una visione progettuale d'insieme coerente. In particolare, le IA per generazione di immagini non risultano ancora pienamente affidabili in contesti che richiedono una rigorosa continuità visiva.

3.6 Fase di Produzione: video

In seguito alla definizione dello stile visivo e alla preparazione dello storyboard, si è passati alla generazione delle clip video, entrando nella fase di produzione. Come già accennato, si è scelto di utilizzare modelli che sfruttassero la tecnica *image-to-video*, utilizzando quindi in input sia un'immagine (prelevata dallo storyboard) sia un prompt dettagliato per descrivere la scena e le azioni.

3.6.1 Valutazione comparativa degli strumenti di generazione video

In continuità con l’approccio adottato per la generazione di immagini, è stata condotta una valutazione comparativa dei principali strumenti di intelligenza artificiale per la generazione video, al fine di individuare le soluzioni più adatte per il progetto. È da precisare che si è preferito selezionare solo modelli in grado di supportare la modalità di generazione *image-to-video*, poiché ritenuta più facilmente controllabile ed ottimizzabile. Partendo quindi dai *benchmark* analizzati nel [Capitolo 2](#), sono stati presi in esame: Sora, Runway, Kling, Veo3.1 e Pika. Midjourney è stato escluso a priori, in linea con quanto già discusso nel paragrafo relativo agli strumenti di generazioni di immagini.

Sora (OpenAI), Runway e Kling

I modelli Sora^[70], Runway^[78] e Kling^[79] rappresentano oggi alcune delle soluzioni più avanzate nel panorama delle IA per generazione video. Tuttavia, l’analisi sperimentale ha evidenziato alcune criticità che ne hanno causato l’esclusione dalla pipeline del progetto.

Nel caso di Sora e Runway, i principali limiti riscontrati hanno riguardato la stabilità temporale delle sequenze e la comparsa di artefatti visivi nelle clip, soprattutto se generate a partire da immagini di input complesse. In particolare, entrambi i modelli tendevano ad introdurre micro-variazioni indesiderate nella geometria e nella texture degli elementi, compromettendo così la continuità visiva. Inoltre, persisteva un’estetica “morbida e perfetta”, riconoscibile come tipica di video generati tramite IA, che il progetto mirava invece a ridurre il più possibile. In diversi casi, poi, i modelli interpretavano in modo impreciso le istruzioni, introducendo azioni non richieste o incoerenti con la scena. A ciò si aggiungono tempi di generazione lunghi e costi poco sostenibili in un contesto fortemente iterativo e di sperimentazione, come quello adottato.

Kling ha mostrato una buona qualità visiva, ma ha evidenziato due criticità importanti: tempi di generazione significativamente lunghi (anche fino ad un’ora per prompt con istruzioni complesse o con più immagini di reference) e una certa imprevedibilità dell’output, talvolta con azioni non richieste o confusionarie. Per questi motivi è stato utilizzato solo per generare alcune scene particolarmente riuscite nei test preliminari, ma non è stato integrato stabilmente nella pipeline.

A causa di queste criticità e visto il tempo ridotto per la realizzazione del progetto, si è preferito scartare i modelli per ridurre il rischio di output imprevedibili che avrebbero richiesto un intervento umano rilevante.

Pika (Pika.art)

A seguito dei testi comparativi, Pika^[80] è stato selezionato come strumento principale per la generazione delle clip video.

Sebbene la piattaforma sia fortemente orientata alla produzione di contenuti per social media – elemento che inizialmente ne suggeriva un approccio prudente e scettico – le prove pratiche hanno evidenziato diversi vantaggi operativi. In particolare, ogni generazione richiedeva pochi minuti e il costo in crediti risultava contenuto, permettendo così numerosi cicli di test e affinamento.

Contrariamente alle aspettative iniziali, Pika ha mostrato una buona stabilità nella modalità *image-to-video*, soprattutto grazie alla possibilità di utilizzare il campo “negative prompt” per specificare azioni e comportamenti da evitare nella generazione.

La piattaforma risulta di facile utilizzo ma al tempo stesso ricca di possibilità e molto versatile. Come per altri modelli testati, è possibile caricare una o più immagini di reference da affiancare al prompt, definire il formato di output (in questo caso 16:9) e selezionare la risoluzione desiderata (in questo caso, 1080p, il massimo disponibile).

In aggiunta, vi è la possibilità di impostare un *seed*, ovvero un valore numerico che permette di mantenere una maggiore coerenza stilistica tra le diverse generazioni, riducendo il rischio di variabilità casuale.

Analogamente a Kling e Runway, Pika offre diverse funzionalità – tra cui estensione di clip video, applicazioni di effetti, inserimento di personaggi o elementi di una fotografia all’interno di un video – ma una in particolare è risultata rilevante: la possibilità di utilizzare più immagini in sequenza per guidare la generazione attraverso diverse azioni ed ottenere un’unica clip. Questa funzionalità verrà approfondita nel prossimo paragrafo, poiché testata per realizzare i piani sequenza e le scene più complesse. Inoltre, la stessa funzionalità permetteva anche di creare video in loop; possibilità utile sia nel mondo social per la creazione, ad esempio, di GIF animate, ma interessante anche per questo progetto.

Infine, per ogni funzionalità vi è una breve descrizione e anche la possibilità di utilizzare opere di altri utenti o preset tramite l’opzione “Make it yours”.

Nel complesso, pur non rappresentando il modello più conosciuto ed avanzato in termini di realismo e fedeltà, Pika ha rappresentato un ottimo compromesso tra velocità di generazione, costo, controllabilità, facilità d’uso e qualità visiva, risultando adeguato come strumento principale per la generazione delle clip video del progetto.

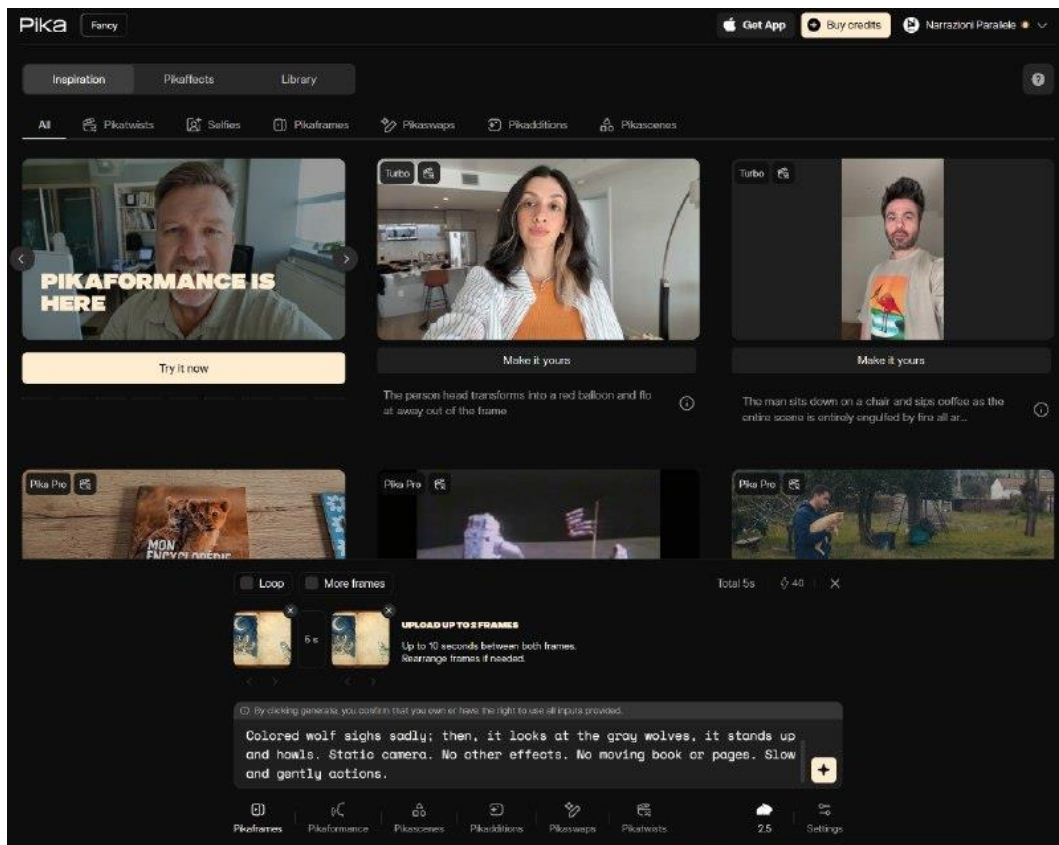


Figura 3.21 Interfaccia di Pika per la generazione video.

Veo3.1, Flow (Google)

Accanto a Pika è stato affiancato Veo3.1^[81], come seconda soluzione, poiché i due modelli presentavano punti di forza parzialmente complementari.

Il modello è disponibile sia tramite la chat con Gemini, con una limitazione a tre video al giorno, sia attraverso la piattaforma Flow^[71], con un sistema di crediti molto vantaggioso per il piano Google Pro Students. Nei test condotti è emerso che l'utilizzo di Veo tramite Flow mostra prestazioni più stabili e controllabili rispetto alla versione integrata in chat, dove spesso accusa delle "allucinazioni" che portano ad output indesiderati e imprevedibili.

In alcune situazioni, Veo ha prodotto risultati più coerenti e realistici rispetto a Pika, compensandone alcune limitazioni. Tuttavia, la maggior disponibilità di crediti e la rapidità iterativa di Pika hanno portato a scegliere quest'ultimo come strumento primario.

Anche in questa fase, l'integrazione di più strumenti di IA si è dimostrata una strategia efficace.

Come per altri modelli testati, Flow supporta la tecnica *image-to-video* e consente di caricare il frame iniziale e finale della clip che si desidera generare. È inoltre possibile definire il formato di output (in questo caso 16:9) e selezionare la risoluzione desiderata (in questo caso, 1080p, il massimo disponibile). In aggiunta, così come per la generazione di immagini, è

supportato anche l'utilizzo di prompt scritti in lingua italiana, che vengono automaticamente tradotti in inglese durante la generazione.

È importante osservare che Veo genera automaticamente anche una traccia audio sincronizzata e coerente con il video. Nel contesto del progetto, questa funzionalità è stata considerata secondaria, poiché la progettazione sonora era già prevista tramite strumenti di IA generativa dedicati alla generazione audio, così da controllare maggiormente il loro output in base alle esigenze.

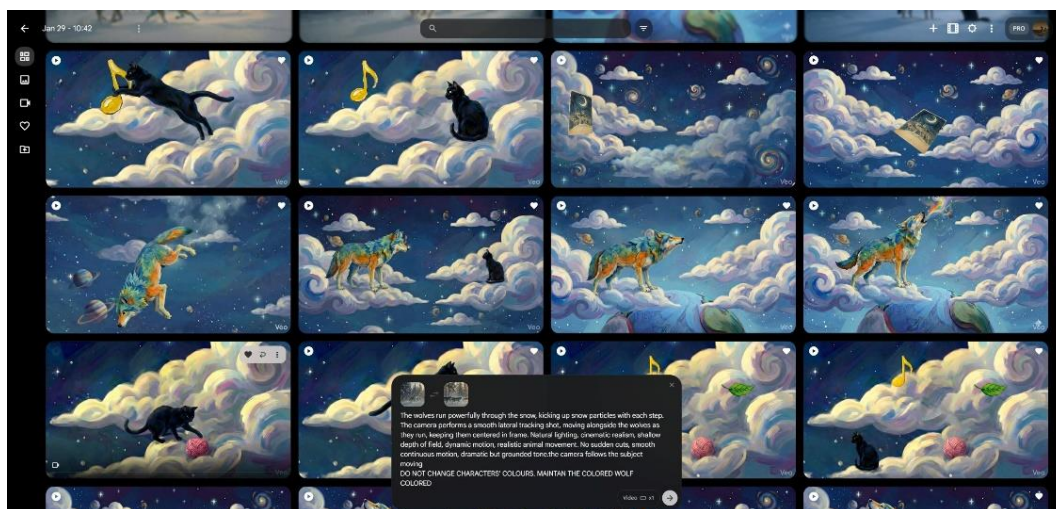


Figura 3.22 Interfaccia di Flow per la generazione video.

Scelta del piano di abbonamento

Oltre alla qualità visiva degli output, una caratteristica fondamentale nella scelta degli strumenti IA è stato il costo dell'abbonamento, il numero di crediti forniti ed il loro consumo per ogni generazione.

Osservando la shot list e lo storyboard si può notare come molte delle 29 scene descritte nella sceneggiatura siano suddivise in sotto-scene, per le motivazioni già affrontate: il limite massimo di durata della generazione (tra 5 e 10 secondi) e la possibilità direttamente proporzionale di creazione di errori in clip lunghe o con molteplici personaggi ed azioni.

Da questa suddivisione è emersa una stima iniziale di 45 clip necessarie per realizzare l'intero cortometraggio. Tuttavia, sapendo che vi sarebbero stati errori e necessità di nuove generazioni, si è ipotizzato un minimo di tre tentativi per ciascuna clip, così da arrivare ad una previsione di circa 135 generazioni complessive.

Pika propone due piani di abbonamento, che consentono generazioni fino a 1080p e senza watermark.

Il piano Pro ha un costo di 28\$ al mese e offre 2300 crediti; il piano Fancy ha un costo di 76\$ al mese e offre 6000 crediti. Ogni generazione in 1080p consuma 40 crediti.

Si è quindi optato per il piano Fancy, così da avere fino a 150 generazioni con un ampio margine operativo in caso di ulteriori suddivisioni delle clip o errori imprevisti.

Flow offre 1000 crediti al mese con la versione Google AI Pro gratis per studenti e ogni generazione consuma 20 crediti. Pertanto, si ha la possibilità di generare fino a 50 clip.

È da precisare che i prezzi indicati ed il rapporto di crediti si riferiscono a gennaio 2026; pertanto, potrebbero aver subito variazioni nel tempo.

Questa analisi evidenzia come, in una pipeline basata su strumenti di IA, il costo di produzione non sia legato solo all'IA in sé o alla durata del prodotto, ma anche al numero di iterazioni necessarie per ottenere un risultato soddisfacente. La componente economica è strettamente connessa al livello di controllo e di affidabilità del modello poiché output con risultati imprevedibili richiedono un numero maggiore di generazioni e, di conseguenza, elevano il costo complessivo.

3.6.2 Generazione delle clip video

Dopo aver concluso lo storyboard e scelto gli strumenti di IA più adatti, si è proceduto alla generazione delle clip video. Con entrambi i modelli è stata adottata la strategia di utilizzare due immagini di riferimento: un frame iniziale, da cui doveva partire la clip, e un frame finale. Questo approccio ha permesso di guidare in modo più controllato lo sviluppo delle singole clip e, soprattutto, di garantire maggior continuità nel caso di piani sequenza, poiché stabilito con precisione il punto di chiusura di una clip e quello di apertura della successiva. I piani sequenza, come già descritto, sono infatti stati suddivisi in più sequenze così da superare i limiti di generazione delle IA in termini di durata ed errori.

Pika si è rivelato particolarmente efficace per le scene del libro illustrato (Atto 1), grazie alla buona interpretazione dello stile – mentre Veo tendeva a modificare i personaggi attribuendogli tratti ed espressioni cartooneschi – e alla velocità di generazione, che ha permesso di effettuare molteplici iterazioni e test. Tuttavia, nel caso di prompt e azioni molto complessi si è notata la comparsa di artefatti visivi indesiderati (flash luminosi, distorsione delle pagine o dell'immagine, effetti di fumo, tremolii più o meno accentuati della camera), spesso risolvibili tramite nuove generazioni o con un utilizzo puntuale del negative prompt; in alcuni casi, invece, si è dovuto ricorrere ad un'opera di correzione in post-produzione, approfondita nei prossimi

paragrafi (Capitolo 3.9). Tali criticità sembrano derivare dalla difficoltà di Pika di gestire situazioni lunghe e complesse o molteplici personaggi impegnati in azioni differenti.

Una problematica analoga si è verificata nella sequenza iniziale con le pagine che si sfogliano e le famiglie di animali animate. Tutti i modelli testati faticavano a gestire contemporaneamente l'animazione dei personaggi, il movimento della pagina e l'introduzione di nuovi soggetti animati, soprattutto, entro il limite di lunghezza della clip (tra i 5 e i 10 secondi). Si è quindi optato per generare separatamente le animazioni delle singole famiglie e, successivamente, produrre la sola animazione della pagina usando come frame iniziale l'ultimo fotogramma della clip precedente e come frame finale il primo della clip successiva.

Non è stato possibile individuare un modello in grado di gestire in modo affidabile questa transizione complessa o di poterla generare a partire dalle due clip video separate. Allo stesso modo, anche gli effetti di transizione preesistenti su DaVinci Resolve Studio non si sono rivelati utili allo scopo.



Figura 3.23 Esempio di una transizione con la pagina che si sfoglia.

Per il secondo atto è stato maggiormente utilizzato Veo, poiché Pika tendeva ad aggiungere effetti non richiesti – come lo scintillio delle stelle sullo sfondo e flash luminosi randomici – anche se esplicitamente esclusi con i prompt. Veo ha mostrato una maggior stabilità nel contesto ricco di dettagli e dallo stile pittorico tridimensionale, pur introducendo occasionalmente variazioni stilistiche. Questo comportamento è ben visibile nel primo piano del lupo, durante la rottura della nuvola, dove il personaggio assume tratti cartooneschi con un'espressione

esagerata. La resa complessiva risultava comunque efficace e il cambio di stile non eccessivamente fastidioso; pertanto, si è deciso di mantenere la clip.



Figura 3.24 Scena della “rottura della nuvola” con tratti cartooneschi.

Nel terzo atto sono stati utilizzati sia Pika che Veo, integrando anche alcuni test su Kling3.0, il quale, in alcune sequenze (ad esempio, la corsa nel bosco), ha fornito risultati soddisfacenti. Tutti e tre i modelli hanno dimostrato una buona padronanza dello stile realistico, ma anche una profonda criticità nella rappresentazione del respiro dei lupi, condensato a causa del freddo. Questo è stato reso come una nube di fumo eccessiva e poco realistica, talvolta più simile ad una ciminiera o, nel caso di alcune generazioni di Pika che ha aggiunto eccessivi riflessi aranciati, al respiro di un drago. Nonostante i tentativi di correzione tramite prompt, il problema è stato solo parzialmente mitigato e risolto in post-produzione. È tuttavia interessante osservare come la distorsione del respiro si verificasse sia quando nel prompt veniva esplicitamente menzionato il “respiro condensato per il freddo”, sia quando tale indicazione era assente. Questo comportamento suggerisce che i modelli tendano ad associare automaticamente contesti innevati o freddi alla presenza di emissione di vapore dalla bocca, enfatizzandone però l'intensità in maniera eccessiva.



Figura 3.25 Esempio di generazione del “fumo”.



Figura 3.26 Esempio di generazione del “fumo” con riflessi aranciati.

È utile evidenziare che la durata delle clip generate variava tra i 5 e i 10 secondi, con costi differenti in termini di crediti. Nella maggior parte dei casi si è optato per clip da 5 secondi, successivamente rallentate in fase di post produzione così da allungare la loro durata. Alcune clip, seppur brevi, sono invece state velocizzate, come la caduta del lupo attraverso il tunnel, per ottenere un effetto dinamico migliore.

La piattaforma Flow offre anche la possibilità di concatenare più scene all'interno di un'unica timeline, permettendo di estendere il video generato inserendo nuovi prompt. Tuttavia, pur garantendo continuità visiva con la clip precedente, tale modalità non permette di inserire un

frame finale di riferimento, riducendo quindi il controllo. Pertanto, la funzionalità è stata testata ma, a causa di questa criticità, esclusa dalla pipeline.

Per il piano sequenza dell'Atto 1, la durata richiesta eccedeva le capacità dei modelli generativi. Si è quindi scelto, in fase di storyboard, di suddividerlo in più clip consecutive successivamente unite in fase di montaggio.

Pika offre la possibilità di utilizzare più immagini per guidare transizioni complesse attraverso frame intermedi stabiliti, ma questo approccio risultava rapidamente oneroso in termini di crediti e meno controllabile in caso di errori. È stato quindi utilizzato solo in sequenze brevi (ad esempio, l'interazione del gatto-mentore con gli oggetti del secondo atto), mentre per il piano sequenza principale si è preferita la strategia modulare che, pur richiedendo un investimento temporale e organizzativo maggiore, ha garantito un livello di controllo e precisione superiore.

Questa funzionalità multi-frame è alla base della possibilità di creare animazioni in loop, opzione utilizzata per la scena dove il gatto ed il lupo saltano fra le nuvole, con l'obiettivo di poterla ripetere ed estendere facilmente.



Figura 3.27 Scena con il loop.

Nel complesso, questa fase ha confermato che, allo stato attuale, i modelli *image-to-video* richiedono una supervisione umana significativa. Inoltre, in presenza di sequenze articolate con la necessità di mantenere nel tempo qualità e continuità visiva, risultano necessari interventi di perfezionamento e correzione, anche con l'utilizzo di software di editing. Ciò presuppone un'attenta attenzione ai dettagli e, talvolta, competenze cinematografiche o di editing avanzate.

In conclusione, sono state generate complessivamente 188 clip, di cui solo 74 appaiono nel cortometraggio finale. Circa 16 di queste non sono state inserite nella loro forma originale, ma sono state rielaborate mediante operazioni di *compositing*, unendo porzioni di clip differenti attraverso l'uso di maschere.

3.7 Fase di Produzione: audio

In seguito alla generazione delle clip video, la fase di produzione è proseguita con la creazione dell'ambiente sonoro: musiche, effetti e voce narrante.

Sebbene in una pipeline audiovisiva tradizionale molti di questi elementi vengano assemblati in post-produzione – ad esempio mediante librerie di effetti sonori (SFX) o interventi di *foley artist* – in questo elaborato di tesi si è scelto di collocare questa attività all'interno della fase di produzione. Tale decisione deriva dal ruolo centrale della dimensione sonora all'interno del racconto e dall'utilizzo di ulteriori strumenti di intelligenza naturale generativa. Non si nasconde, però, il fatto che molte scelte – tra cui l'aggiunta di una voce narrante – siano state prese in fase di post-produzione, a seguito della revisione delle clip video prodotte.

3.7.1 Definizione dello stile musicale

Un'ulteriore fase importante è stata la definizione dello stile musicale del cortometraggio, nell'ottica di costruire un accompagnamento sonoro coerente con l'evoluzione narrativa ed emotiva dei tre atti. Questa è stata supportata principalmente da ChatGPT, che si è dimostrata particolarmente efficace nel proporre idee musicali strutturate e, in seguito, prompt da fornire alle IA selezionate.

Il primo atto è caratterizzato da una musica che accompagna le famiglie animali, arricchita da suoni coerenti con le loro azioni e con l'ambiente rappresentato (ad esempio, il rumore dei cervi che brucano l'erba o il cinguettio degli uccellini). La melodia, semplice e radiosa, si interrompe in corrispondenza della pagina dedicata ai lupi, con l'intento di farne risaltare la diversità rispetto agli altri gruppi.

Parallelamente, è stato progettato un tema musicale specifico per il Lupo Colorato, destinato ad accompagnarlo lungo i tre atti e ad evolversi progressivamente con lui. Durante il secondo atto assume un ruolo centrale, entrando in relazione con i suoni degli oggetti che guidano il protagonista alla scoperta della propria voce e, di conseguenza, con la melodia dell'ululato che

nasce. Inoltre, il tema presenta delle note malinconiche, richiamando alcune delle risposte emerse durante gli eventi di *Narrazioni Parallele*, soprattutto in relazione al terzo evento.

Un'ulteriore decisione – nata durante la fase di post-produzione – è stata l'introduzione di una voce narrante nel primo atto. La scelta è stata condizionata dalla comparsa spontanea, durante la fase di storyboard, di brevi frasi alla base delle pagine dedicate alle famiglie animali, analoghe a quelle presenti nei libri illustrati per l'infanzia. La voce narrante si inserisce quindi come parte integrante del racconto e aiuta anche la comprensione della situazione iniziale del lupo. Inoltre, secondo ChatGPT, rafforza la percezione della “pressione paterna” che grava sul personaggio, in linea con i dati emotivi emersi dal concerto dedicato a Mozart.

È stato anche deciso di non far proseguire la voce narrante nei due atti successivi. Questa scelta risulta coerente con l'arco trasformativo del protagonista, che fugge da quelle pagine dove non trovava posto, e consente di lasciare maggiore spazio alla dimensione sonora e all'esperienza diretta del lupo.

3.7.2 Valutazione comparativa degli strumenti di generazione audio

In continuità con l'approccio adottato nelle fasi precedenti, è stata condotta una valutazione comparativa dei principali strumenti di IA per la generazione audio, al fine di individuare le soluzioni più adatte per il progetto. Partendo quindi dal *benchmark* analizzato nel [Capitolo 2](#), sono stati presi in esame tre modelli: Suno, Loudly ed ElevenLabs. Di questi sono stati valutati: qualità dell'output, flessibilità e facilità d'uso, controllabilità ed integrazione all'interno del workflow.

Suno

Suno^[82] è una piattaforma avanzata di generazione musicale tramite IA, in grado di produrre brani completi a partire da prompt testuali e, facoltativamente, da reference audio. L'utente può specificare genere, stile e tipo di arrangiamento, scegliendo anche se ottenere come output solo una base strumentale oppure una canzone completa di voce. In quest'ultimo caso, vi è la possibilità di selezionare la tipologia di voce (maschile, femminile o corale), la lingua e la modalità di generazione del testo (automatico in base al prompt dato oppure fornito dall'utente). È inoltre possibile caricare una registrazione vocale di riferimento, per guidare l'interpretazione o per inserirla all'interno della base prodotta. Questa funzionalità è molto potente ma solleva anche molteplici critiche dal punto di vista etico per l'utilizzo di voci altrui senza consenso.

Riguardo la lingua della voce generata a partire da un testo scritto, è emerso da alcuni test

come la lingua inglese risulti più naturale rispetto a quella italiana, nella quale compare frequentemente una lieve inflessione “straniera”. In entrambe le lingue, però, la voce risulta credibile e realistica, mai robotica.

Dal punto di vista qualitativo, le tracce generate si sono rivelate molto coerenti e credibili sotto il profilo armonico, ritmico e timbrico. Rispetto ad alcune IA per immagini o video analizzate in precedenza, Suno dimostra un grado di maturità sorprendentemente elevato.

Ulteriori punti di forza sono la rapidità di generazione ed il sistema di crediti gratuito e con ricarica giornaliera. Ogni generazione, inoltre, produce quattro variazioni dello stesso brano, facilitando l’esplorazione creativa.

Sebbene l’interfaccia possa risultare inizialmente meno immediata rispetto a quella delle piattaforme per immagini e video, la curva di apprendimento si è rivelata gestibile attraverso qualche tentativo di prova.

Alla luce di questi elementi, Suno è stato selezionato come strumento principale per la generazione della colonna sonora del progetto.

Loudly

Loudly^[83] è un’ulteriore piattaforma di generazione musicale basata su intelligenza artificiale, orientata alla produzione di brani royalty-free utilizzabili anche su piattaforme di streaming professionali, come Spotify.

Rispetto a Suno, Loudly offre un controllo decisamente più analitico e professionale sui parametri compositivi. L’utente può infatti impostare con precisione la tonalità e la scala musicale (ad esempio, DO maggiore o Re minore), il tempo (in BPM, “battiti per minuto”) e la metrica (come i tempi in 4/4 o 3/4), nonché di selezionare il numero e la tipologia di strumenti da impiegare all’interno della traccia. La piattaforma consente inoltre di selezionare se includere o meno una voce.

Tutte queste impostazioni tecniche vengono poi integrate alle indicazioni precisate nel prompt testuale.

Un controllo di parametri così specifico rende Loudly uno strumento estremamente potente per gli esperti del settore audio, ma al contempo meno accessibile per utenti principianti. Durante la fase di test preliminari, l’interfaccia è infatti risultata complessa da padroneggiare e la curva di apprendimento molto più ripida rispetto al competitor Suno.

Considerando che quest'ultimo si era già dimostrato adeguato agli obiettivi del progetto e con una facilità d'uso maggiore, Loudly è stato escluso dalla pipeline per ragioni di ottimizzazione operativa e gestione delle tempistiche, non per limiti qualitativi.

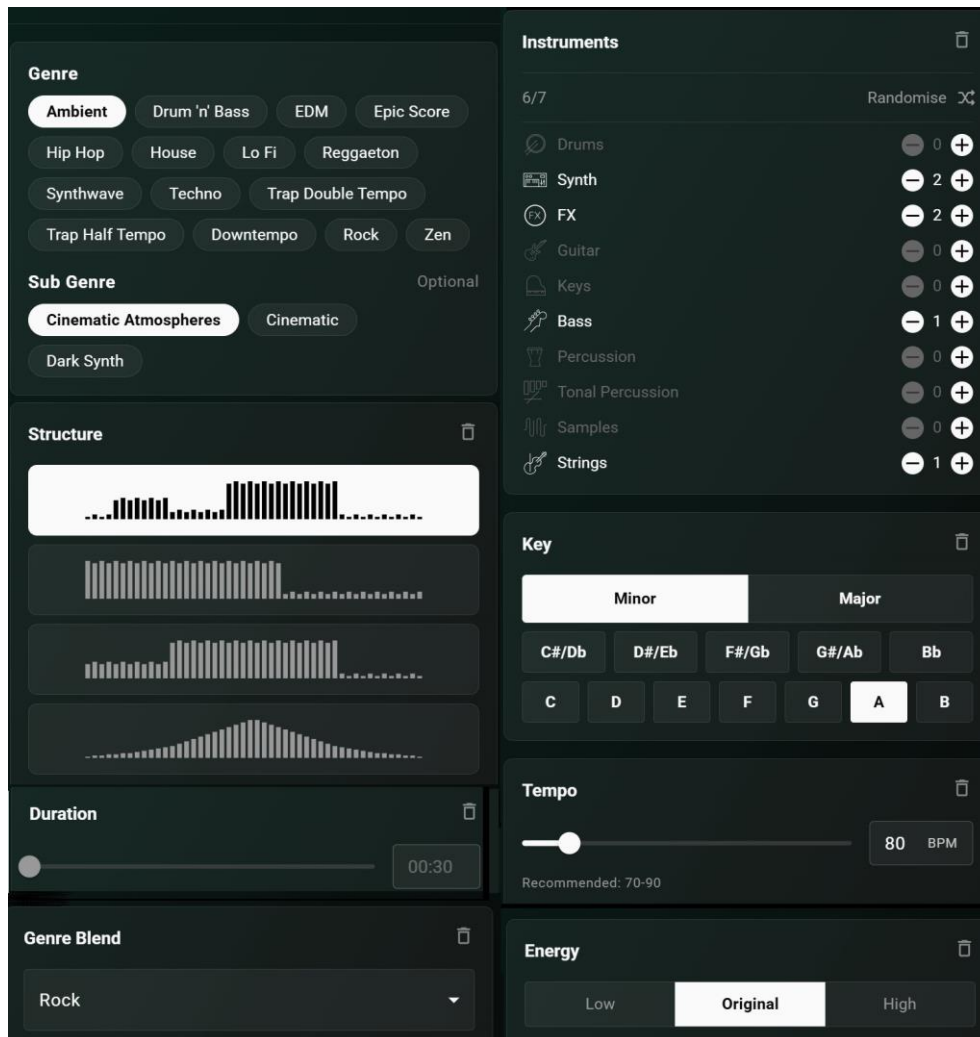


Figura 3.28 Esempio dei parametri musicali disponibili all'interno di Loudly.

ElevenLabs

ElevenLabs^[84] è una piattaforma specializzata nella sintesi vocale realistica tramite modelli *text-to-speech* ed è, oggi, tra le soluzioni più avanzate nel settore.

L'utente ha a disposizione numerosi parametri per definire l'espressività della voce, tra cui: timbro, intonazione, dinamica, ritmo. Uno degli aspetti più interessanti è la capacità del modello di interpretare sia la punteggiatura che specifiche indicazioni testuali per modulare la performance vocale. Ad esempio, è possibile inserire marcatori espressivi tra parentesi – come [sarcastically] o [whispers] – che vengono tradotti in variazioni dell'intonazione.

La piattaforma supporta attualmente un ampio numero di lingue e consente di utilizzare diversi modelli vocali ottimizzati per ciascuna di esse. Inoltre, è possibile realizzare una voce personalizzata tramite prompt e controllo dei parametri oppure caricare e clonare una registrazione vocale. Quest'ultima funzionalità, come per Suno, si rivela molto efficace ma con evidenti rischi etici e legali.

Contrariamente a quanto osservato in altri modelli *text-to-speech*, ElevenLabs ha mostrato buone prestazioni anche in lingua italiana, con pronuncia ed intonazione generalmente credibili e realistiche, non robotiche e senza accenti stranieri.

Oltre alla sintesi vocale, la piattaforma offre anche sistemi per la generazione di effetti sonori a partire da prompt testuali o preset, sempre con la possibilità di modificare vari parametri. Durante i test è dimostrata molto efficace nella produzione dei suoni utili al progetto, come passi, vento o ululati.

Sebbene l'interfaccia possa risultare inizialmente carica di possibilità e parametri, la curva di apprendimento si è rivelata gestibile attraverso qualche tentativo di prova. Questi sono favoriti dal sistema di crediti, particolarmente generoso e vantaggioso anche nella versione gratuita. Inoltre, per ogni operazione è presente un pulsante "Best practices" che rimanda alla guida online, con spiegazioni riguardo le funzionalità ed esempi di prompt. La piattaforma e la guida consentono la consultazione in diverse lingue oltre all'inglese, favorendone l'accessibilità.

Alla luce dell'ottima qualità vocale e della flessibilità anche in ambito di generazione di effetti sonori, ElevenLabs è stato selezionato come strumento principale per la generazione della voce narrante e dell'ambiente sonoro del cortometraggio.

3.7.3 Generazione della voce narrante

Come già accennato, per accompagnare il primo atto del cortometraggio è stata introdotta una voce narrante, con la funzione di guidare lo spettatore attraverso la dimensione fiabesca del libro illustrato e facilitare la comprensione della situazione iniziale del protagonista.

Il contenuto della narrazione è stato sviluppato con il supporto ChatGPT, partendo dalle frasi già presenti nelle scene rappresentanti le famiglie animali ed estendendole rispecchiandone la forma e lo stile.

Le scritte erano state inserite in autonomia dalla piattaforma Flow durante la generazione dello storyboard. Fin da subito sono apparse perfettamente in linea con l'estetica e l'impaginazione di un classico libro di fiabe; pertanto, si è deciso di mantenerle e valorizzarle.



Figura 3.29 Esempio di scritta nel libro.

La voce è stata generata tramite ElevenLabs^[84] utilizzando una voce maschile preesistente, “Chuck Miller -Deep, Raspy, American”, scelta per la chiarezza, il realismo e l’adeguatezza al registro narrativo immaginato. In particolare, si è evitato l’uso di timbri eccessivamente acuti o dinamici, privilegiando una resa più neutra e rassicurante.

La generazione è stata realizzata in lingua inglese, sia perché le tecnologie di sintesi vocale risultano generalmente più precise in questa lingua, sia perché i testi presenti nelle pagine del libro erano in inglese.

Successivamente è stata testata anche una versione in lingua italiana, sfruttando la funzionalità di voce design offerta dalla piattaforma. In questo caso, tramite ChatGPT è stato elaborato un prompt mirato a ottenere un timbro caldo, profondo e rilassato, adatto alla lettura di una fiaba. Come riferimento è stato suggerito da ChatGPT di prendere in considerazione la voce calda di Luca Ward^[85], con un particolare riferimento al tono utilizzato per il personaggio di Mufasa nella pellicola Disney *Il Re Leone* (1994).

In entrambi i casi è stata prevista l’integrazione di sottotitoli nella lingua opposta a quella del parlato, al fine di garantire accessibilità e fruizione internazionale del contenuto. Inoltre, la piattaforma consente di modificare alcuni parametri della voce, durante la generazione *text-to-speech*, per migliorare la resa in base all’idea dell’utente.

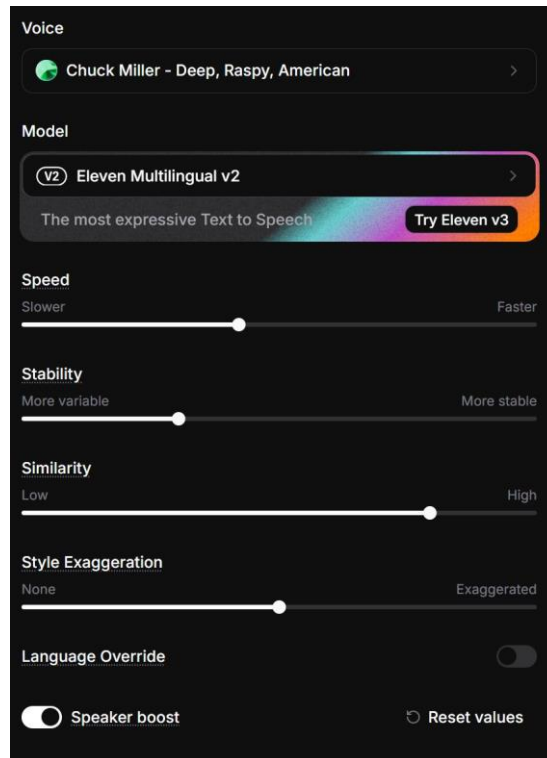


Figura 3.30 Parametri modificabili durante la generazione text-to-speech.

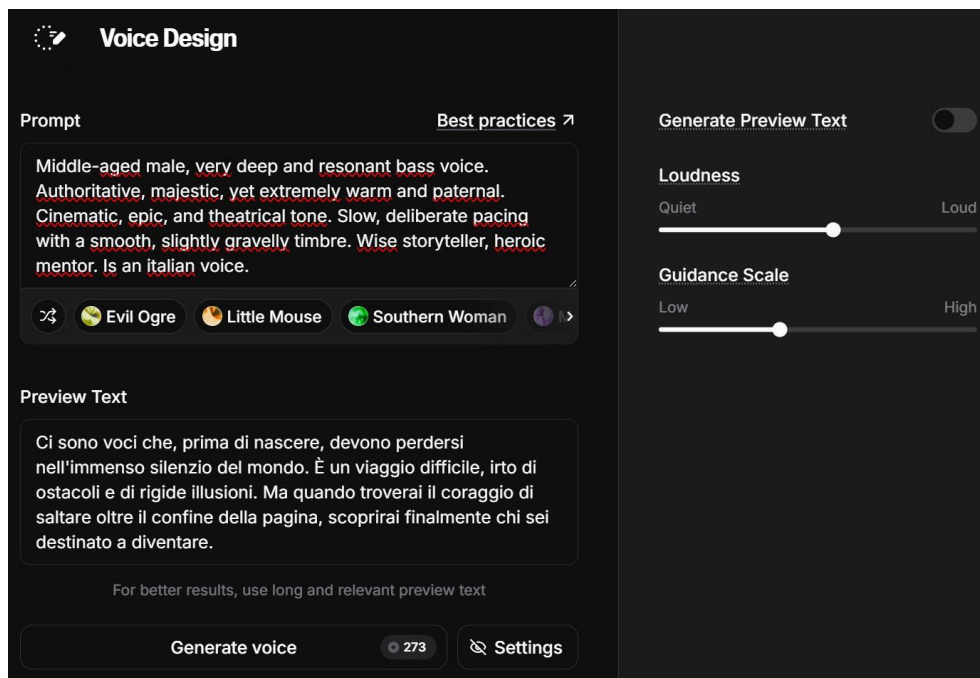


Figura 3.31 Interfaccia per generare una nuova voce.

La scelta di una voce maschile è stata ulteriormente motivata da ChatGPT in base alle considerazioni emerse nell’analisi dei dati del primo concerto, dove il tema della “pressione paterna” è stato ricorrente.

All’interno della struttura narrativa, la voce narrante assume infatti una funzione duplice e complessa, interpretabile attraverso gli archetipi de *Il Viaggio dell’Eroe*. Inizialmente, essa si raffigura come un elemento oppressivo, che intrappola il Lupo Colorato all’interno di un libro e di una storia in cui non riesce a trovare il proprio posto. Questo è assimilabile al Guardiano della Soglia, ovvero colui che rappresenta i limiti e le regole del mondo ordinario, che l’eroe dovrà superare per crescere. Nel momento in cui il protagonista compie il salto verso l’ignoto, sottolineato da uno stacco su schermo nero, la voce muta il proprio ruolo e si trasforma nell’Araldo. Questo, come sottolinea Vogler, non deve essere obbligatoriamente un personaggio positivo: spesso risulta essere una forza negativa, oppressiva o ambigua che scambussola il mondo ordinario dell’eroe e lo costringe quindi a partire. Pronunciando la sua ultima battuta, in corrispondenza della conclusione del primo atto (“*some voices, before being born, must get lost*”; “*alcune voci, prima di nascere, devono perdersi*”), il narratore cessa di essere un ostacolo per il Lupo Colorato e lo indirizza verso il richiamo all’avventura, sancendo ufficialmente l’apertura del secondo atto e l’inizio del percorso di scoperta identitaria.

3.7.4 Generazione degli effetti sonori

La piattaforma ElevenLabs^[84] è stata impiegata anche per realizzare gli effetti sonori (SFX) volti ad arricchire il paesaggio sonoro del cortometraggio.

In una pipeline cinematografica tradizionale, tali elementi vengono generalmente ottenuti tramite registrazioni sul campo, lavoro di *foley artist* o utilizzo di librerie dedicate durante la fase di post-produzione. Nel presente progetto, seppur molti suoni siano stati generati in base alle esigenze del video durante il montaggio, si è scelto di trattare questa operazione all’interno della fase di produzione, in quanto la i materiali sono stati generati ex novo tramite IA. Si precisa inoltre che una minima parte di suoni è stata mantenuta dalle clip audio generate automaticamente da Veo (ad esempio, il rumore dei cervi che brucano, il cinguettio degli uccelli e i tuffi dei pesci nell’acqua).

La generazione dei nuovi effetti è avvenuta tramite la funzionalità dedicata di ElevenLabs, che consente di definire tramite prompt il tipo di suono desiderato, la durata in secondi e, eventualmente, indicare se questo debba essere messo in loop. Quest’ultima caratteristica si è rivelata particolarmente utile per suoni continui e prolungati come gli ululati.

Particolare attenzione è stata dedicata anche alla creazione del *room tone*, elemento fondamentale per evitare silenzi innaturali e garantire continuità acustica. Per ogni atto si è pensato ad una soluzione distinta e caratterizzante.

Il primo atto è rappresentato da un ambiente boschivo, con un vento leggero fra le fronde degli alberi e suoni di uccellini. Coerentemente con le scelte musicali (approfondite nel paragrafo successivo), questa base si interrompe bruscamente durante la scena dei lupi. Per sottolineare la condizione di isolamento e distacco che caratterizza il branco, il vuoto sonoro viene riempito da un vento più aspro, generato specificando nel prompt il contesto invernale e spoglio a cui si è pensato.

Il secondo atto è caratterizzato da un lieve rumore di “vento e spazio vuoto”, volutamente seminascosto dal tema musicale del mondo, di cui si parlerà nel paragrafo seguente.

Infine, il terzo atto segna un ritorno al realismo ed è accompagnato un ambiente sonoro boschivo invernale, con il suono ovattato di neve che cade sugli alberi e la presenza di un fiume lontano, la cui intensità è stata poi enfatizzata per le scene in prossimità dell’acqua.

Per realizzare i *room tone* si è intenzionalmente evitato di richiedere ad ElevenLabs la generazione di un’unica traccia ambientale, al fine di evitare la comparsa di artefatti acustici o aggiunte sonore imprevedibili. Questi sono stati realizzati come in una pipeline tradizionale, con un accurato lavoro di montaggio e mixing manuale, ma utilizzando SFX generati con l’IA ed emersi in dialogo con ChatGPT, valutando congiuntamente i diversi tipi di ambienti, di emozioni e di possibili rumori presenti.

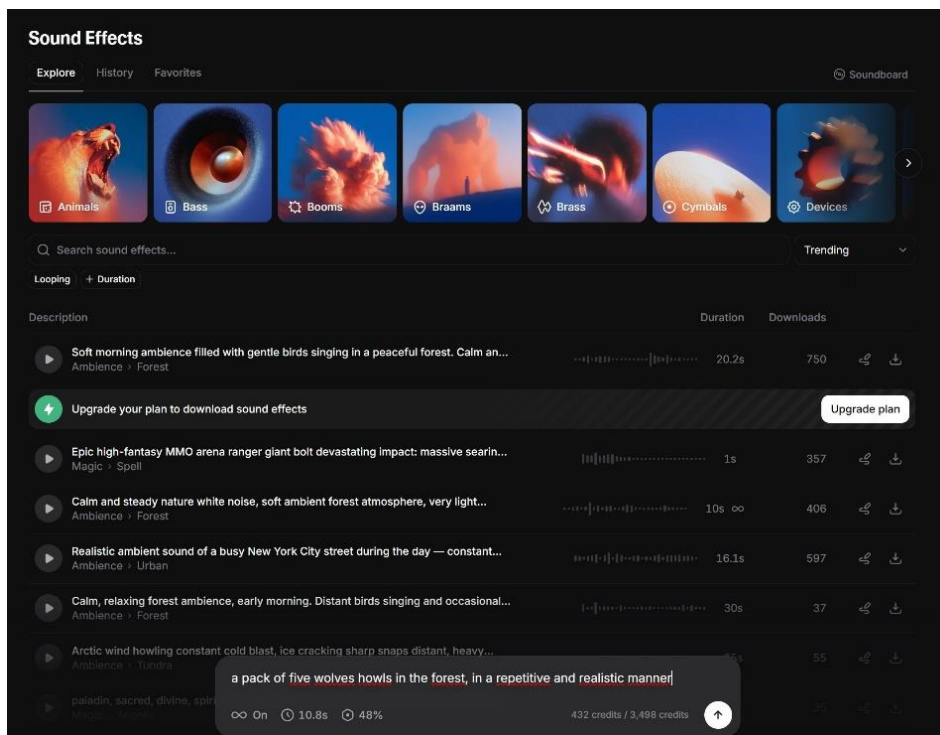


Figura 3.32 Interfaccia di ElevenLabs per la generazione di effetti sonori.

3.7.5 Generazione della colonna sonora

Come già discusso, si è scelto di realizzare le tracce musicali tramite Suno^[82].

Dal punto di vista compositivo, la colonna sonora è costruita come un sistema di temi che si alternano e si fondono nel corso della narrazione.

La scena iniziale, in biblioteca, è accompagnata da una musica dalle tonalità fantasy e leggermente misteriose, ottenuta dopo numerose iterazioni per evitare sia un eccesso di enfasi epica sia un risultato troppo inquietante. Con l'apertura del libro e l'ingresso della voce narrante, questa lascia progressivamente spazio al tema leggero associato alle famiglie degli animali, che si interrompe durante la scena dei lupi, per evidenziarne la diversità.

Con l'apparizione del gatto-mentore viene introdotto il tema a lui associato, caratterizzato da sonorità magiche e riproposto, con sfumature diverse, come base per il secondo atto. Anche durante la transizione nel tunnel il medesimo tema viene riproposto con una variazione timbrica, per accompagnare il passaggio tra i mondi.

Nel secondo atto, il tema del mondo è progettato per essere in coerenza armonica con i suoni delle nuvole e degli oggetti con cui interagisce il gatto, creando un dialogo studiato tra musica e sound design. In questa fase il tema del protagonista inizia progressivamente ad emergere a fondersi con l'ambiente sonoro, in armonia con gli elementi presenti e con l'ululato nascente. Il tema subisce una temporanea destabilizzazione nel momento di confronto con i lupi di carta, per poi affermarsi nuovamente con maggiore forza, rispecchiando la crescita del Lupo Colorato. Infine, nel terzo atto, il tema del lupo prosegue e lo accompagna fino alla conclusione della narrazione. Qui si avvertono particolarmente le note malinconiche che lo compongono, in linea con le emozioni emerse durante il terzo evento di *Narrazioni Parallele*.

In fase di progettazione era stata valutata l'integrazione della canzone composta collettivamente durante l'evento *Song Writing* come tema musicale per il lupo o il mondo del secondo atto. Tuttavia, dopo alcune prove, l'idea è stata accantonata per motivi di coerenza stilistica.

Il materiale è stato comunque riutilizzato in un approccio sperimentale: il testo e l'audio sono stati forniti come riferimento per Suno, generando così una nuova versione da cui, con alcune modifiche, è stata ricavata la base per il tema del lupo. È interessante sottolineare come la versione completa con la voce cantante sia risultata sorprendentemente credibile e realistica, seppur con lievi inflessioni a causa dell'uso della lingua italiana, non supportata bene quanto quella inglese.

Questa sperimentazione e l'utilizzo di Suno per la generazione dei temi musicali rispondono anche alla domanda finale proposta durante l'evento *Song Writing*: "Pensi che l'intelligenza

artificiale avrebbe potuto creare una canzone simile?”. I dati raccolti mostrano una maggioranza di “No” e “Non lo so”, ma l’esito della ricerca ha ribaltato le aspettative. La qualità compositiva, timbrica ed espressiva restituita da Suno ha infatti dimostrato un livello di maturità tecnologica impressionante, provando come questo modello generativo sia in grado di emulare la creatività musicale quasi al completo.

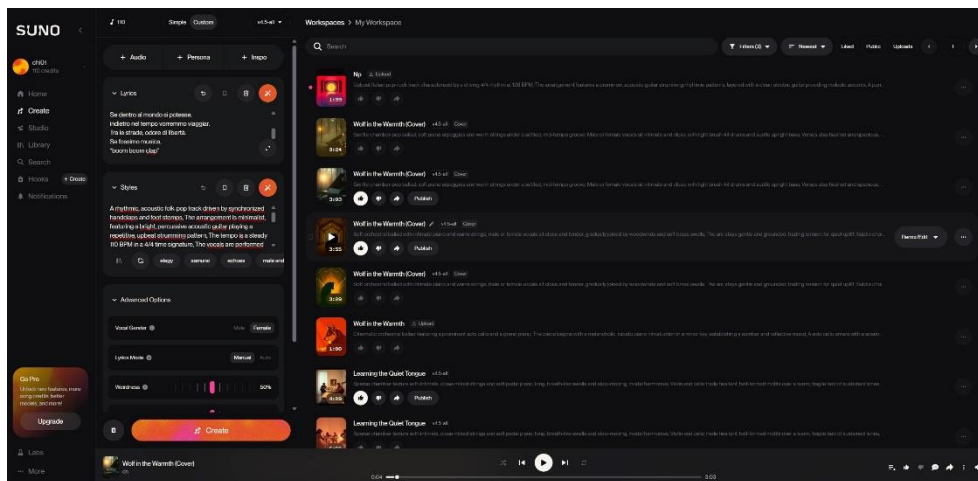


Figura 3.33 Interfaccia di Suno per la generazione di musiche e canzoni.

3.8 Fase di Produzione: osservazioni e conclusioni

La fase di produzione ha rappresentato il momento più significativo per valutare la potenzialità degli strumenti di intelligenza artificiale generativa in ambito audiovisivo. Dal punto di vista operativo, questa si è divisa in due sotto-fasi: la generazione video (mediante Pika e Veo3.1) e la generazione audio (tramite Suno ed ElevenLabs). Un aspetto evidente emerso durante questa fase è il netto divario di maturità tecnologica tra i due gruppi di strumenti: quelli audio capaci di restituire risultati quasi definitivi e quelli video ancora fortemente dipendenti da interventi correttivi, che saranno analizzati nei paragrafi successivi.

I vantaggi della generazione video restano comunque significativi. Prima di tutto, l'impiego dell'intelligenza artificiale ha permesso di abbattere radicalmente i costi e i tempi di realizzazione di un cortometraggio in tecnica mista.

In una produzione tradizionale, il progetto avrebbe richiesto la presenza di un reparto di animazione 2D, un reparto 3D, artisti CGI ed eventualmente l'organizzazione di riprese in ambienti reali. L'utilizzo dell'IA ha unito questi aspetti in un unico flusso di lavoro dove l'autore e le macchine hanno lavorato come un singolo reparto. Un altro enorme vantaggio risiede nella

potenziale eliminazione della curva di apprendimento di software complessi, potendo idealmente ricreare i medesimi risultati con le IA. Ad esempio, pur possedendo competenze avanzate di modellazione e animazione su Blender, non avrei potuto raggiungere un simile livello qualitativo e pittorico per il secondo atto senza dover investire tempo per la creazione di tutti gli elementi e le animazioni, oltre a dover apprendere l'anatomia animale per evitare effetti di Uncanny Valley e la tecnica di *texture painting* per l'effetto 3D pittorico. Allo stesso modo, per il terzo atto, si sarebbe impiegato molto tempo su software di *compositing* complessi, come Nuke o Fusion (di DaVinci Resolve), per l'integrazione degli animali CGI nelle riprese realistiche del bosco. Da questo punto di vista, l'IA si conferma uno strumento formidabile per produzioni a basso costo o per utenti alle prime esperienze, contribuendo a una "democratizzazione del mezzo". Tuttavia, attualmente, ciò sembra valere solo per prodotti tecnicamente e narrativamente semplici, poiché la necessità di continuità visiva e qualitativa o la gestione di molteplici personaggi ed azioni rappresenta ancora una criticità.

Inoltre, l'utilizzo combinato di più modelli (Pika e Veo) ha permesso di ottenere risultati visivamente convincenti, confermando come una strategia *multi-tool* sia attualmente molto più efficace rispetto all'utilizzo di un singolo strumento.

Importante è stata anche la scelta di utilizzare la tecnica *image-to-video*. Infatti, partire da immagini generate in precedenza ha facilitato di molto il mantenimento della continuità visiva, evitando di dover descrivere all'interno di ogni prompt i personaggi e fornire i model sheet, che potevano venire mal interpretati. Inoltre, sarebbe stato impossibile realizzare dei piani sequenza suddividendoli in più generazioni.

Riguardo il reparto audio, Suno ed ElevenLabs hanno dimostrato un livello di generazione eccellente. L'utilizzo dell'IA ha drasticamente velocizzato il processo rispetto alla tradizionale ricerca su siti di librerie musicali *royalty-free* o alla creazione da zero dei suoni. In tempi brevissimi, è stato possibile generare la voce narrante e gli effetti sonori, oltre ai temi musicali, sopperendo alla mia mancanza di competenze in ambito di composizione musicale. L'abbattimento dei costi e delle tempistiche di produzione è innegabile: compiti che in una *pipeline* tradizionale avrebbero richiesto l'assunzione di compositori, studi di registrazione ed eventualmente *foley artist* sono stati eseguiti con prompt testuali, garantendo un controllo iterativo rapido e di alta qualità.

Nonostante queste enormi potenzialità, il reparto video ha fatto emergere criticità importanti: instabilità temporale, artefatti visivi, difficoltà nella gestione di più personaggi o azioni nella stessa scena e, talvolta, una scarsa prevedibilità dell'output. Anche con prompt estremamente accurati e fornendo i frame iniziali e finali, i modelli hanno mostrato la tendenza a introdurre

variazioni indesiderate o a interpretare in modo impreciso le istruzioni. Esempi eclatanti sono il vapore eccessivo che esce dalla bocca dei lupi del terzo atto, oppure le stelle brillanti ed i flash luminosi aggiunti da Pika durante il secondo atto.

Inoltre, anche le interazioni fisiche complesse o fuori dall'ordinario rappresentano un'enorme difficoltà. Questo è stato il motivo per cui, durante i test preliminari svolti in fase di pre-produzione, il concept *Il gatto che rompe il silenzio* è stato scartato. Sembrava, infatti, impossibile realizzare la scena chiave in cui il gatto urta il barattolo di vetro, che si infrange sul pavimento liberando scie luminose.

Un'altra limitazione è data dai limiti di durata delle generazioni (tra i 5 ed i 10 secondi), che impongono all'autore di realizzare azioni brevi o frammentarle in più clip. Questo, in aggiunta alla difficoltà per le IA di mantenere coerenza visiva, complica la realizzazione di scene complesse o piani sequenza.

Queste problematiche portano ad una riflessione critica riguardo il concetto di "democratizzazione" sopra citato. Per ottenere un prodotto di qualità è stato necessario eseguire un elevato numero di test, selezionare gli output migliori e, soprattutto, intervenire in modo significativo in post-produzione. Questo aspetto ridimensiona fortemente il concetto di "democratizzazione del mezzo audiovisivo". Se è vero che l'intelligenza artificiale consente a chiunque di generare contenuti visivi e sonori con facilità, è altrettanto evidente che la realizzazione di un cortometraggio narrativo strutturato e stilisticamente coerente richiede ancora competenze registiche, tecniche e di montaggio avanzato. Senza di esse, gli errori e le imprecisioni dati dai modelli IA rimangono visibili, compromettendo la qualità del risultato.

In sintesi, la fase di produzione ha dimostrato che le IA generative rappresentano oggi strumenti estremamente potenti, ma ancora troppo immaturi per sostenere una pipeline completamente autonoma. Esse si configurano come acceleratori del processo e moltiplicatori di possibilità creative, ma i loro limiti richiedono un continuo controllo dell'autore umano e la ricerca di compromessi tra l'idea originale e l'output risultante.

In questo scenario, l'intelligenza artificiale non sostituisce l'autore, ma ne amplifica le possibilità in un processo di co-creazione.

3.9 Fase di Post-produzione

Completata la generazione dei materiali visivi e sonori, si è giunti alla fase di post-produzione. Analogamente a una pipeline cinematografica tradizionale, questa fase ha incluso le operazioni di montaggio audio e video, color correction e color grading ed alcuni interventi per migliorare la resa finale del prodotto.

L'intervento umano è risultato fondamentale per garantire la continuità visiva e correggere alcuni difetti emersi durante le generazioni con IA.

3.9. DaVinci Resolve e strumenti IA integrati

DaVinci Resolve^[86] è un software per il montaggio e la post-produzione video e audio sviluppato Blackmagic Design, ampiamente diffuso in ambito professionale poiché integra in un unico ambiente editing, color, VFX e mixing audio.

La versione Studio, inoltre, include la *DaVinci AI Neural Engine*^[87], un framework avanzato basato su reti neurali profonde (*Deep Learning*) e tecniche di *machine learning*. Questo contiene un insieme di funzionalità pensate per supportare e automatizzare operazioni ripetitive o complesse in termini di lavoro e tempo (ad esempio, l'isolamento ed il *tracking* di soggetti in movimento su sfondi non uniformi). Nel contesto del progetto, queste si sono rivelate particolarmente utili per compensare le distorsioni prodotte dai modelli generativi, portando l'ottica della pipeline basata su sistemi di IA anche nella fase di post-produzione (in questo caso, adoperando principalmente con strumenti non generativi).

Di seguito, vengono approfonditi alcuni tra i numerosi strumenti proposti dalla *DaVinci AI Neural Engine*, utilizzati durante il progetto e richiamati nel paragrafo successivo.

Magic Mask

Magic Mask è uno degli strumenti IA più avanzati di DaVinci Resolve. Consente di isolare automaticamente soggetti in movimento, agendo con precisione anche su sfondi non uniformi e su elementi complessi, come capelli e abiti. A differenza dell'isolamento basato sul colore (*chroma key*), questo strumento sfrutta un processo di segmentazione semantica: quando l'utente indica il soggetto, una Rete Neurale Convolutionale (CNN) ne estrae le caratteristiche visive e produce una mappa di probabilità pixel per pixel, determinando quali costituiscono il soggetto e quali lo sfondo. In automatico, viene tracciato nel tempo, frame per frame, il movimento dell'elemento isolato, utilizzando algoritmi di stima del moto che analizzano come i pixel si spostano nei fotogrammi adiacenti.

Questo crea quindi maschere dinamiche tracciate nel tempo, riducendo drasticamente il lavoro manuale di *rotoscoping* e *tracking*.

All'interno del progetto, Magic Mask è stato utilizzato frequentemente per correggere incoerenze e separare personaggi ben riusciti da clip problematiche, così da ricomporli in scene migliori. Rispetto alla tecnica “tradizionale” di *rotoscoping* manuale, questo strumento consente un notevole risparmio di tempo, richiedendo l'intervento umano solo saltuariamente, per rifinire i bordi e i dettagli più complessi.

La principale criticità riscontrata è legata al *tracking*: quando si modificano aspetti strutturali della clip, come la posizione nella timeline o la sua velocità, la maschera richiede un ricalcolo del tracciamento temporale. Sebbene l'operazione si esegua con un semplice clic, l'elaborazione su clip complesse può essere computazionalmente onerosa. Un limite operativo significativo risiede nell'assenza di un sistema di notifiche; il software, infatti, non avvisa l'utente della necessità di aggiornare il *tracking*, limitandosi a produrre artefatti che, in assenza di un'accurata revisione, possono sfuggire al controllo dell'operatore.

Per ovviare al problema si può utilizzare il *Render in Place*, una funziona di esportazione/render della clip con effetti all'interno della timeline stessa. Questo congela il risultato della Magic Mask, ovviando al problema del *tracking*, ma precludendo anche la possibilità di modificare in futuro i parametri della maschera.

AI SuperScale

La funzione AI SuperScale utilizza modelli di *Single-Image Super-Resolution* (SISR), una tecnica di computer vision che ricostruisce un'immagine ad alta risoluzione da un singolo input a bassa risoluzione, migliorandone i dettagli ed aumentandone le dimensioni spaziali.

A differenza dell'interpolazione tradizionale (come gli algoritmi bilineari o bicubici), che “stirano” i pixel esistenti e restituiscono un'immagine morbida e sfocata, la tecnica SISR mira a ricostruire con fedeltà le informazioni. La rete neurale è infatti addestrata su vaste librerie di immagini, per generare nuovi pixel coerenti con il contenuto originale.

AI SuperScale consente di selezionare l'algoritmo neurale da utilizzare e il fattore di ingrandimento (2x o 4x), offrendo anche parametri per calibrare il bilanciamento tra la riduzione del rumore introdotto e i filtri di sharpening per aumentare la nitidezza.

Nel caso del progetto, l'applicazione dello strumento e la scelta dei parametri è stata gestita clip per clip, valutando l'effetto visivo prodotto. È stato impiegato principalmente per uniformare clip generate a risoluzioni differenti a causa delle limitazioni delle IA o per risparmiare sul sistema di crediti (ad esempio, per generare su Pika sequenze di durata superiore ai 5

secondi, si è optato per la qualità in 720p, poiché risoluzioni maggiori si sarebbero rivelate troppo onerose in termini di crediti).

Stabilize

DaVinci Resolve include diversi algoritmi avanzati di stabilizzazione video. Sebbene non siano classificati esplicitamente come strumenti di IA, questi si avvalgono di una sofisticata analisi di *Computer Vision*.

La stabilizzazione ottica si articola in tre passaggi chiave. Prima di tutto, vi è la stima del movimento, dove il software traccia i dettagli più visibili e ad alto contrasto dell'immagine (*features*) per calcolare la traiettoria. In seguito, tramite calcoli matematici, viene ricreata una traiettoria virtuale della camera. Infine, i fotogrammi vengono traslati, ruotati o deformati per seguire il nuovo percorso.

In caso di riprese con camere moderne, DaVinci Resolve Studio supporta anche la stabilizzazione tramite giroscopio; una tecnica moderna ed avanzata che analizza i metadati di movimento della camera per stabilizzarne il girato. Lavorando, però, con file prodotti dalle IA, non è stato possibile sperimentarla per ovvie ragioni.

Nel progetto, la stabilizzazione è stata utilizzata principalmente per correggere tremolii e vibrazioni indesiderate generate da Pika, soprattutto nelle sequenze con il libro, che necessitavano la camera statica. Le applicazioni verranno approfondite nel prossimo paragrafo, dedicato alla post-produzione.

Lo strumento Stabilize si è dimostrato efficace, ma con un limite noto: il *crop* automatico. Poiché compensare il movimento implica dover spostare l'immagine fuori dal suo centro geometrico, lungo i bordi dell'inquadratura si creano delle bande nere. Per nasconderle, il software effettua uno zoom ritagliando (*crop*) la porzione di frame con il problema. Questo ha richiesto, ad esempio nel caso del libro, ridimensionamenti manuali della clip o *compositing* con maschere, per mantenere la coerenza visiva della sequenza.

Suite IA per l'audio

DaVinci Resolve integra anche diverse funzionalità IA relative al reparto audio. Tra le più rilevanti presentate sul sito ufficiale e, in parte testate all'interno del progetto, vi sono strumenti per la riduzione del rumore e l'isolamento automatico della voce umana (*Voice Isolation*) e algoritmi in grado di scomporre file musicali separandone i singoli strumenti così da permettere agli utenti di bilanciarli nuovamente o creare nuovi arrangiamenti. A questi si aggiungono

anche funzioni di *speech-to-text* per la sottotitolazione automatica, con la possibilità di animarli dinamicamente in sincrono con le singole parole pronunciate. Di notevole interesse per i flussi di lavoro complessi multi-camera è il *Multicam SmartSwitch* che, analizzando le riprese e le tracce audio, è in grado di tagliare e selezionare l'inquadratura corretta in base all'oratore attivo. È stato recentemente introdotto anche l'*Audio Assistant*, progettato per analizzare l'intera timeline audio e normalizzare dinamicamente i livelli sonori, restituendo un mix quasi professionale con un singolo clic.

Nel progetto tali strumenti sono stati testati in via esplorativa, ma utilizzati in misura più limitata rispetto ai moduli video, poiché la pipeline sonora si basava principalmente su materiale generato esternamente con intelligenze artificiali specifiche (Suno ed ElevenLabs), le quali hanno restituito file audio che non necessitavano di eccessiva lavorazione con strumenti appositi.

Tuttavia, l'inclusione di queste funzionalità conferma la crescente maturazione dell'IA come assistente alla post-produzione anche in ambito audio e porta con sé innovazioni interessanti.

Riguardo la trascrizione automatica del parlato, sarebbe stato di notevole interesse utilizzarla per realizzare i sottotitoli da affiancare la voce narrante. Questa possibilità è stata però scartata poiché la funzione è disponibile solo per trascrivere nella stessa lingua della traccia audio mentre, nel caso del progetto, si volevano inserire sottotitoli tradotti.

3.9.2 Montaggio, color correction e ottimizzazione finale

Il montaggio è stato effettuato seguendo la sceneggiatura e la shot list realizzate in fase di pre-produzione.

Le clip video sono state preventivamente organizzate in cartelle suddivise per atto e rinominate secondo una nomenclatura coerente con la shot list e già adottata per lo storyboard, così da mantenere una corrispondenza chiara tra le diverse fasi del workflow.

Durante il montaggio si è reso necessario un ampio uso di maschere, ritagli e zoom correttivi per compensare le discontinuità introdotte nella fase di generazione video. Gli strumenti IA di DaVinci Resolve Studio si sono rivelati particolarmente utili per velocizzare interventi che altrimenti sarebbero stati molto onerosi.

Nelle sequenze del libro (Atto 1) in cui Pika aveva difficoltà ad animare correttamente più personaggi contemporaneamente, i personaggi sono stati animati separatamente e successivamente uniti tramite maschere. In questa operazione si è rivelata particolarmente efficace la funzione *Magic Mask*, che ha permesso di ritagliarli e tracciare automaticamente il loro movimento al fine di isolarli dalla clip originale per poi integrarli in un'unica composizione.

Lo stesso strumento è stato utilizzato anche nel secondo atto, dove due generazioni di Veo presentavano rispettivamente un lupo ed un ambiente corretti ma un gatto difettoso e viceversa. L'isolamento intelligente dei personaggi ha permesso di combinare rapidamente le versioni migliori, effettuando solamente piccoli ritocchi di scala e posizioni.

Nel primo atto, per correggere i tremolii di camera è stato utilizzato lo strumento Stabilize, che tuttavia introduce *crop* e variazioni di scala. Per mantenere la coerenza visiva, il libro è stato quindi separato con una maschera e ricomposto sopra ad un'immagine PNG statica del tavolo, adeguandone manualmente i valori di zoom al fine di garantire coerenza tra le clip.

L'intervento umano durante la fase di montaggio è stato necessario anche per l'inserimento delle transizioni fra le clip e gli stacchi sul nero (ad esempio durante la caduta nel tunnel), oltre che per la costruzione del montaggio spaziale caratteristico del secondo atto, volto a evidenziare la ripetitività dei tentativi eseguiti dal lupo e il ritmo progressivamente costruito.

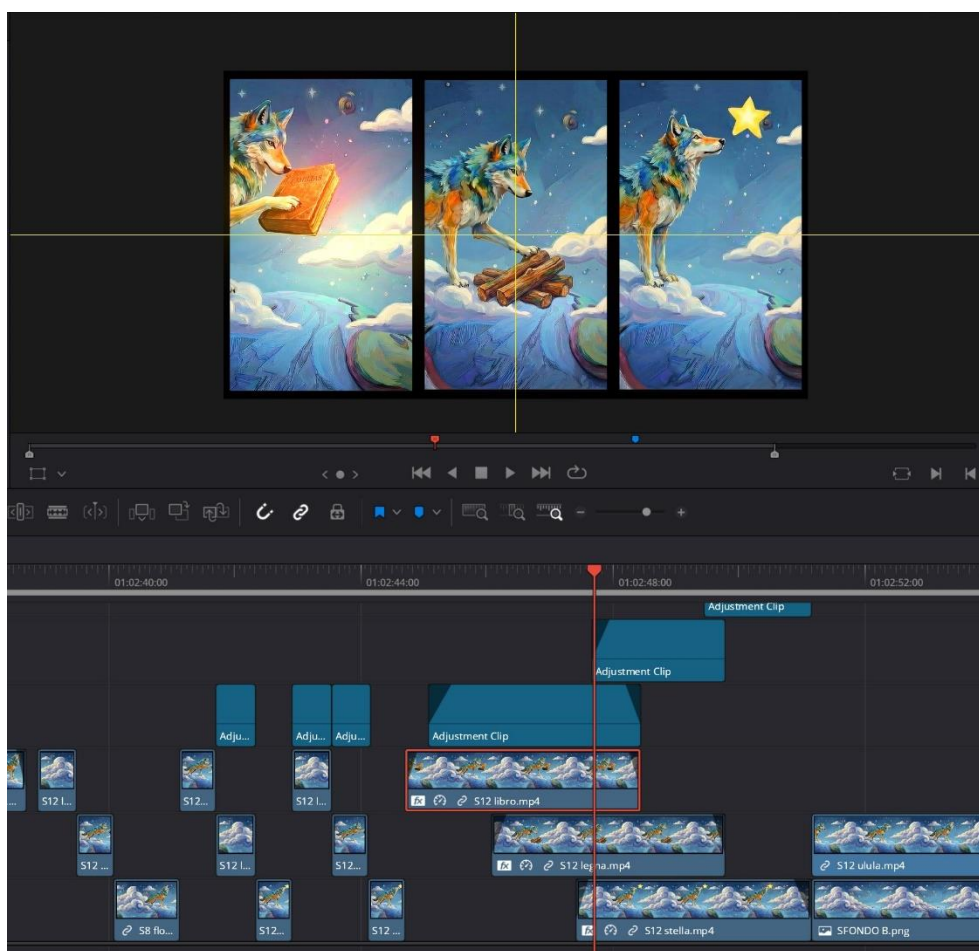


Figura 3.34 Timeline per il montaggio spaziale, tramite la funzione Video Collage.

Inoltre, l'intervento umano si è rivelato fondamentale per incrementare manualmente la nitidezza (*sharpening*) delle pagine dedicate al branco di lupi – sia nel mondo 2D del libro sia nella pagina che compare nel mondo 3D – durante le scene in cui ringhiano contro il Lupo Colorato. Questa caratteristica era stata definita in fase di sceneggiatura, decidendo di renderli più “spaventosi” grazie al loro cambio di forma, con bordi “netti e taglienti”. I modelli generativi, però, non riuscivano a restituire in modo significativo questa caratteristica, la cui aggiunta è stata effettuata in fase di post-produzione.

Come già accennato in precedenza, sono stati anche applicati interventi di *retiming* su diverse clip. Alcune sequenze generate risultavano infatti eccessivamente lente – come la caduta nel tunnel o quando la nuvola nel secondo atto si spezza, producendo effetti fluttuanti poco realistici – mentre altre presentavano movimenti troppo rapidi, in particolare in alcune interazioni del gatto con gli oggetti del mondo. Attraverso accelerazioni e rallentamenti mirati si è migliorata la coerenza e la leggibilità dei movimenti, oltre a superare i limiti di durata delle clip generate.

Alcune clip esportate a 720p sono state inoltre trattate con lo strumento di AI SuperScale, così da uniformare la risoluzione della timeline e migliorare la qualità complessiva del materiale.

In seguito al montaggio video, si è proseguito verso la fase di color correction e color grading. La prima interviene per correggere i parametri tecnici dell'immagine (bilanciamento del bianco, esposizione, contrasto e molti altri), mentre la seconda costituisce la chiave espressiva per armonizzare l'estetica complessiva del prodotto e migliorare la resa visiva finale.

Nel caso di questo progetto, entrambe sono servite per uniformare le diverse clip generate e garantire il più possibile una coerenza visiva e stilistica.

Le criticità più evidenti sono emerse nell'Atto 3, dove l'intenzione narrativa prevedeva una progressione luminosa dalla blue hour alla golden hour che rispecchiasse l'evoluzione emotiva del protagonista.

Come approfondito in precedenza, già in fase di storyboard erano emerse difficoltà nella gestione della luce.

I due casi più evidenti rappresentano la scena del lago, dove la luce era molto più intensa di come avrebbe dovuto essere, e la scena della decisione finale tra il branco ed il gatto, dove l'illuminazione risultava più fredda e meno intensa rispetto alle scene precedenti e successive. Per sistemare questi dettagli ed uniformare le clip è stato necessario l'intervento umano. Nonostante ciò, alcune imprecisioni sono rimaste, poiché le discrepanze tra le clip generate dalle IA e la resa desiderata risultavano particolarmente marcate. Inoltre, trattandosi di clip

esportate in formati non Log o RAW – e quindi prive di un’ampia gamma dinamica e margini di intervento – le possibilità di correzione risultavano ancor più limitate.



Figura 3.35 Scena del lago.



Figura 3.36 Scena della scelta.

Nel medesimo atto, inoltre, nella scena della corsa del branco, uno dei lupi aveva assunto tonalità cromatiche simili a quelle del Lupo Colorato. Il problema è stato risolto isolando il soggetto con la Magic Mask ed intervenendo sulla saturazione per riportarlo verso la palette corretta.

Particolarmente impegnativa è risultata la fase di montaggio audio e, nello specifico, la sincronizzazione audio-video. A differenza delle produzioni cinematografiche tradizionali, dove i dialoghi e le clip video corrispondenti vengono allineati automaticamente tramite timecode o strumenti per l'analisi della forma d'onda, in questo progetto gran parte dei suoni (come passi e ululati, che sarebbero stati registrati sul set da un fonico in presa diretta, come i passi e gli ululati) è stata generata esternamente, rendendo quindi necessario un allineamento manuale accurato.



Figura 3.37 Esempio di una parte della timeline audio.

Infine, per alcune lavorazioni specifiche e per l'utilizzo di strumenti meno familiari (come, ad esempio, la Magic Mask), sono state consultate le IA testuali, in coerenza con l'impostazione della pipeline produttiva. Tra ChatGPT, Claude e Gemini, quest'ultima si è dimostrata la più efficace nel fornire supporto operativo su DaVinci Resolve con, talvolta, suggerimenti coerenti per video tutorial presenti su YouTube.

Questo evidenzia come le IA testuali possano fungere non solo da generatori di contenuti, ma anche da assistenti tecnici nella gestione di software complessi, riducendo il tempo normalmente dedicato alla ricerca e consultazione di tutorial tradizionali.

3.9.3 Titoli di coda

Per i titoli di coda si è scelto di riprendere lo stile visivo del libro del primo atto, così da chiudere la struttura narrativa in modo circolare.

La sequenza si apre infatti con un fade-out finale che dalla tana porta alle pagine del libro, sulle quali compare il titolo del cortometraggio scritto con un tratto che richiama la grafia manuale e colorato come se fosse scritto ad acquerelli. Attraverso l'uso della Magic Mask sono stati isolati il lupo ed il gatto per farli comparire al di sotto del titolo. In questo caso è stata

fatta una scelta stilistica importante: il lupo è collocato sulla pagina sinistra e ulula correttamente, mentre il gatto entra in scena sulla pagina di destra.



Figura 3.38 Titoli di coda.

I titoli di coda vengono introdotti con un'animazione di pagine sfogliate, coerente con quella del primo atto. Per questi è stato selezionato un font nero, anche questo ispirato alla scrittura manuale. Inoltre, l'impaginazione è stata progettata mantenendo i margini coerenti con quelli di una pagina di libro per rafforzare la coerenza estetica e visiva del progetto.



Figura 3.39 Animazione delle pagine, come per l'Atto 1.

I titoli di coda non sono stati realizzati tramite IA, al fine di evitare possibili errori nella generazione del testo, ambito che, attualmente, rappresenta una criticità per molti modelli di generazione di immagini e video.

3.10 Fase di Post-produzione: osservazioni e conclusioni

La fase di post-produzione ha confermato in modo netto il ruolo centrale dell'intervento umano all'interno di una *pipeline* basata su intelligenza artificiale generativa. Se nelle fasi precedenti i modelli hanno contribuito in modo significativo alla creazione dei materiali, è stato proprio durante la post-produzione che si è reso necessario un grande lavoro di rifinitura, correzione e integrazione da parte dell'autore umano. Questo ha dimostrato come l'impiego dell'IA nella produzione audiovisiva non elimini la necessità del montaggio tradizionale, rendendolo, al contrario, un passaggio molto importante.

In questa fase, il passaggio dall'IA generativa all'IA analitica (attraverso i moduli della *DaVinci AI Neural Engine*) ha evidenziato come quest'ultima sia oggi uno strumento maturo e particolarmente efficace per supportare il montaggio.

Tra i principali vantaggi emersi vi è il drastico risparmio di tempo in operazioni di *editing* e *compositing* tradizionalmente onerose. Strumenti come Magic Mask e Stabilize, si sono rivelati importanti nel recupero di clip difettose in tempi ridotti e AI SuperScale ha permesso di uniformare la risoluzione e la qualità del cortometraggio. Questi strumenti hanno supportato la fase di montaggio, aiutando ad elevare la qualità del prodotto finale. Le IA per la generazione di video, infatti, avevano lasciato difetti nelle animazioni dei personaggi o dello sfondo, corretti con un intensivo uso di maschere, oppure movimenti di camera indesiderati, per i quali sono state effettuate stabilizzazioni delle clip. Queste ultime hanno introdotto il problema del *crop* dell'immagine, che ha richiesto ulteriori operazioni di *compositing* con maschere o di regolazione dello zoom delle clip.

È inoltre da sottolineare l'utilizzo delle IA testuali, utilizzate come assistenti per suggerimenti di montaggio e per risolvere rapidamente alcune problematiche o ricercare tutorial per strumenti specifici.

Particolarmente critica si è rivelata la fase di *color correction* e *color grading*, che ha evidenziato la difficoltà dei modelli generativi nel mantenere una gestione coerente della luce tra clip diverse, obbligando a interventi manuali per preservare la continuità visiva. A questo si aggiunge anche l'attuale difficoltà per le IA di gestire in modo affidabile la tipografia, limite che ha motivato la scelta di realizzare manualmente la sequenza dei titoli di coda.

Dal punto di vista del reparto audio, la generazione tramite IA ha dimostrato un'elevata qualità. Tuttavia, l'intervento umano è stato necessario per poter allineare correttamente le clip video ai rispettivi materiali audio. Non disponendo di timecode o del ciak, questa operazione non ha potuto usufruire del Auto align offerto da DaVinci Resolve ed è stata effettuata manualmente.

Nel complesso, la fase di post-produzione evidenzia come, allo stato attuale, le IA generative non possano realizzare autonomamente un contenuto audiovisivo di qualità e coerente, poiché risultano necessari molteplici interventi di *editing*, talvolta con la necessità di conoscenze tecniche avanzate. Tuttavia, l'IA apporta benefici ed innovazioni concrete, accelerando significativamente il lavoro e riducendo il carico di compiti onerosi, come la creazione di maschere ed il loro tracciamento nel tempo.

Questa fase finale conferma quindi la natura ibrida del processo sperimentato, in cui uomo e macchina lavorano insieme in un processo di co-creazione artistica dove l'autore mantiene un ruolo decisivo ed insostituibile per garantire coerenza, qualità tecnica ed espressività dell'opera.

CAPITOLO 4 – CONCLUSIONI E VALUTAZIONI FINALI

Il presente capitolo presenta una sintesi del percorso sperimentale condotto e propone una valutazione critica del rapporto tra arte ed intelligenza artificiale. L'analisi prende in considerazione il ruolo delle tecnologie generative (e non) nel processo creativo audiovisivo, dalla raccolta dei dati fino alla post-produzione, evidenziandone potenzialità, limiti ed implicazioni etiche. Inoltre, verranno anche verificati i risultati raggiunti rispetto agli obiettivi delineati nel Capitolo 1 (1.4).

4.1 Pipeline IA: potenzialità e limiti

La sperimentazione condotta durante la realizzazione del cortometraggio ha permesso di delineare con chiarezza le potenzialità e i limiti di una pipeline audiovisiva basata su strumenti di intelligenza artificiale. Se da un lato si sono dimostrati efficaci nell'ottimizzare le tempistiche, ridurre i costi di produzione e ampliare le possibilità creative, dall'altro è emerso come non possano ancora costituire un sistema autonomo affidabile.

Durante la fase di pre-produzione, i modelli linguistici (ChatGPT, Claude e Gemini) si sono rivelati eccellenti per l'elaborazione dei dati raccolti durante gli eventi di *Narrazioni Parallele*. In poco tempo, sono stati individuati pattern ricorrenti, discordanze e temi rilevanti, tradotti in concept narrativi in linea con le emozioni emerse dal pubblico. Il dialogo con questi strumenti ha inoltre facilitato la progressiva definizione della narrazione e della sceneggiatura finale, offrendo in aggiunta un costante supporto creativo ed organizzativo durante le successive fasi del progetto.

Tuttavia, è stato necessario un attento lavoro di *prompt design* per evitare narrazioni eccessivamente didascaliche o troppo legate ai dati raccolti. Un rischio evidente, infatti, risiedeva nella possibilità che le IA utilizzassero direttamente i protagonisti degli eventi – Mozart, Rodari e Verdi – come personaggi delle narrazioni, anziché come figure metaforiche da cui trarre ispirazione.

L'impiego di modelli di generazione di immagini (Nano Banana Pro, in particolare) ha velocizzato in modo significativo la rappresentazione visiva dei personaggi e la costruzione dello storyboard, ampliando al contempo le potenzialità creative ed esplorative del progetto. Questo è particolarmente evidente nell'Atto 2, dove l'ambientazione risultava inizialmente molto astratta a livello di sceneggiatura e ha potuto essere sviluppata gradualmente attraverso successive iterazioni generative, senza richiedere competenze grafiche specifiche degli *Storyboard Artist*.

Tuttavia, sono emerse non poche criticità nella coerenza e continuità visiva dei personaggi e degli ambienti, le quali hanno richiesto molteplici tentativi di rigenerazione oppure adattamenti tramite software e competenze specifici.

Le intelligenze artificiali utilizzate hanno comunque confermato il potenziale di questi strumenti nel rendere più accessibili i processi della produzione audiovisiva. L'abbattimento della curva di apprendimento di software complessi, come Blender, e la possibilità di ottenere risultati visivamente e sonoramente convincenti attraverso prompt testuali, senza competenze tecniche pregresse, rappresentano un cambiamento significativo rispetto alle pipeline tradizionali. Ad esempio, la diffusione sempre più ampia di contenuti generati tramite IA sui social media dimostra come questi strumenti permettano di produrre rapidamente video realistici che, con metodologie tradizionali, richiederebbero tempi e competenze molto maggiori. Pika, usato insieme a Veo3.1 per la realizzazione del cortometraggio, nasce soprattutto come IA per la realizzazione di rapide clip ed effetti per *reels* e video da social media.

In particolare, fra tutti gli strumenti utilizzati per il progetto, quelli relativi al reparto audio (SunoAI, Loudly e ElevenLabs) si dimostrano molto maturi rispetto ai loro "collegli" per la generazione di immagini e video.

SunoAI ha permesso la creazione di colonne sonore di elevata qualità ed è in grado di generare anche canzoni convincenti ed orecchiabili. Allo stesso modo, la fedeltà delle voci e degli effetti sonori realizzabili con ElevenLabs risulta sorprendente e ha contribuito di molto ad arricchire il paesaggio sonoro del cortometraggio in modo rapido ed efficace. Grazie alla sua maturità, il reparto audio è riuscito in parte a compensare alcune criticità rimaste nelle clip video, garantendo e supportando la componente emotiva ricercata con il cortometraggio.

Le criticità più rilevanti sono emerse nella fase di generazione delle clip video. Gli output delle IA presentano ancora instabilità temporali, incoerenze visive e comportamenti imprevedibili, richiedendo quindi numerosi tentativi di generazione e di affinamento dei prompt, oltre ad un'attenta selezione del materiale prodotto ed una sua eventuale revisione con software di *editing* appositi. A ciò si aggiungono ulteriori limitazioni tecniche, come la breve durata delle clip generate e la risoluzione massima spesso limitata a 720p o 1080p.

La fase di post-produzione ha inoltre evidenziato come queste nuove pipeline non eliminino il lavoro umano, ma lo ridistribuiscono e supportino. L'intelligenza artificiale ha velocizzato e semplificato i processi di produzione di questo progetto, ma spesso ha richiesto molteplici interventi correttivi per ottenere il prodotto finale rispettando qualità e coerenza. A causa di questo, la promessa di una completa democratizzazione della produzione visiva viene ridimensionata; per realizzare un'opera audiovisiva complessa sembra ancora essere necessario avere competenze tecniche avanzate per sistemare i difetti in post-produzione.

Riguardo a questa fase, è interessante notare l'integrazione crescente di strumenti basati su IA all'interno di software professionali di *editing*, come DaVinci Resolve. Questi portano molteplici funzionalità automatizzate, come la generazione di maschere e il loro tracciamento nel tempo, che semplificano operazioni complesse e dispendiose in termini di tempo, supportando e migliorando l'efficienza del *workflow*. Questo tipo di integrazione potrebbe contribuire, almeno in parte, a riequilibrare il tema della democratizzazione degli strumenti. Tuttavia, oggi, molte delle funzioni basate su IA di DaVinci Resolve sono incluse esclusivamente nella versione a pagamento, così come per altre piattaforme; ad esempio, Canva offre la possibilità di un "AI Super Scale" per le immagini solo tramite il piano ad abbonamento.

Infine, alla luce degli obiettivi definiti nel Capitolo 1 si può affermare che:

- La metodologia di raccolta dati, tramite la piattaforma Mentimeter, è risultata efficace e ha dimostrato la possibilità di integrare le emozioni e le sensazioni del pubblico e dei performer all'interno di un processo creativo mediato dall'IA.
- L'analisi, la categorizzazione e l'elaborazione dei dati ha trovato nei modelli linguistici un supporto concreto ed efficace.
- La progettazione di una pipeline basata su IA e automatizzata è stata tecnicamente possibile, ma non completamente autonoma. Al di là delle criticità risolte in post-produzione e già approfondite, vi è anche l'impossibilità per gli strumenti di trasferire in modo autonomo le informazioni; questo ha richiesto un intervento umano costante per allineare tutti i modelli e, talvolta, ricordare loro decisioni prese in precedenza. Inoltre, il controllo autoriale è stato fondamentale per validare gli output, garantendo coerenza ed una narrazione non didascalica.

In conclusione, la pipeline basata su strumenti di intelligenza artificiale sperimentata possiede numerose potenzialità ed altrettanti limiti da superare. Attualmente, non può sostituire il processo creativo umano, ma può efficacemente affiancarlo in un sistema di co-creazione in cui la macchina velocizza i tempi ed espande le possibilità operative, mentre l'autore umano controlla la coerenza, la qualità e l'intenzione espressiva dell'opera e prende le decisioni finali.

4.2 Dalla progettazione alla pratica: il compromesso creativo

Uno degli obiettivi principali del progetto consisteva nel definire un'identità estetica autonoma del prodotto finale, capace di tradurre e restituire in modo coerente i dati raccolti durante gli eventi di *Narrazioni Parallele*. Tuttavia, nel passaggio dalla fase progettuale alla realizzazione pratica è emerso con chiarezza come tale identità non potesse essere imposta alle IA ma dovesse nascere attraverso una serie continua di compromessi tra la visione autoriale e i limiti tecnologici degli strumenti utilizzati.

Per superare le limitazioni dei modelli video è stata adottata una strategia basata sulla combinazione di due strumenti generativi, Pika e Veo3.1. Questa scelta si è rivelata particolarmente efficace; presi singolarmente, infatti, i due modelli non avrebbero garantito lo stesso livello di coerenza visiva, stabilità temporale e qualità estetica durante l'intero cortometraggio, poiché presentavano alcune difficoltà. È da evidenziare, però, come l'approccio *multi-tool* abbia permesso di compensare le carenze dei singoli strumenti ma non abbia eliminato del tutto alcune criticità, come variazioni di stile, tremolii della camera, differenze di illuminazione e aggiunta di artefatti ed effetti indesiderati.

In alcuni casi, si è giunti a compromessi tra l'idea originaria ed il risultato ottenuto. Un esempio significativo riguarda la scena del secondo atto in cui la nuvola si rompe ed il Lupo Colorato precipita: Pika non era in grado di gestirla poiché introduceva flash di luce incoerenti, mentre Veo3.1 ha apportato un cambio di stile verso espressioni più cartoonesche. Qui nasce il compromesso: dopo alcuni tentativi con nuove generazioni si è scelto di integrare la sequenza originale all'interno del cortometraggio, poiché l'effetto risultava comunque coerente e non disturbante.

Situazioni analoghe si sono verificate anche in relazione alla risoluzione delle clip, la quale è stata uniformata il più possibile attraverso gli strumenti di *upscaling* offerti da DaVinci Resolve, ma in alcuni casi un ingrandimento eccessivo introduceva ulteriori distorsioni ed artefatti visivi. Per queste clip si è scelto quindi di mantenere alcune leggere differenze di risoluzione al fine di preservare la qualità complessiva dell'azione.

Allo stesso modo, la color correction non è sempre stata in grado di uniformare le discontinuità, soprattutto di illuminazione, generate dai modelli video. Le clip sono comunque state inserite all'interno del cortometraggio, pur riconoscendo che rappresentano difetti di continuità.

Un ulteriore esempio di questo processo di compromessi riguarda l'estetica complessiva del secondo atto. Il risultato finale non coincide perfettamente con l'intenzione iniziale della sceneggiatura, dove la sensazione di mondo onirico, non definito e misterioso era maggiore. Tuttavia, questo risultato non può essere considerato come un errore tecnico, bensì una forma di

adattamento rispetto ai numerosi test effettuati e alle scelte operate durante la sperimentazione.

È quindi importante sottolineare che questi compromessi non rappresentino cedimenti passivi di fronte al controllo e ai limiti della macchina, ma piuttosto decisioni consapevoli su quali imperfezioni fossero tollerabili o meno all'interno dell'identità estetica complessiva. La linea autoriale adottata è stata a favore delle emozioni che potevano scaturire dal progetto e della continuità narrativa delle azioni piuttosto che della pura perfezione tecnica.

Il lavoro autoriale si è quindi configurato come un continuo processo di selezione, adattamento e reinterpretazione degli output generati, portando ad una riflessione più ampia sul rapporto tra uomo e macchina nel processo creativo. La produzione audiovisiva basata sull'utilizzo di strumenti di intelligenza artificiale generativa non può essere interpretata come una mera esecuzione automatica di comandi testuali, bensì, come un processo di co-creazione tra l'autore e le IA. Nel caso di questo progetto, si è imparato a dialogare con i differenti modelli generativi, ognuno con caratteristiche proprie, e a far fronte alla loro imprevedibilità e ai loro limiti, traducendoli in opportunità narrative diverse ma altrettanto significative.

4.3 Il ruolo dell'autore nei processi di co-creazione con l'IA

Uno degli obiettivi principali del progetto era comprendere in che misura la creatività potesse essere condivisa tra artista ed intelligenza artificiale, individuando se fosse possibile stabilire un rapporto di co-creazione e di equilibrio tra automazione e controllo umano. La sperimentazione condotta ha mostrato chiaramente come la realizzazione di un cortometraggio attraverso una pipeline basata su strumenti di IA non comporti la scomparsa del ruolo dell'autore, ma bensì una sua ridefinizione e la necessità di apprendere nuove competenze.

L'intelligenza artificiale si è dimostrata uno strumento potente nella generazione di contenuti visivi, sonori e testuali, ma ancora incapace di garantire coerenza narrativa, continuità estetica ed intenzionalità espressiva in modo autonomo. In assenza di una supervisione umana costante, i modelli tendono a produrre soluzioni ridondanti, incoerenti o eccessivamente didascaliche, appiattendosi così la complessità narrativa e stilistica del progetto. L'intervento umano si è quindi rivelato fondamentale per selezionare gli output più efficaci, correggere le "allucinazioni" delle macchine e orientare la direzione complessiva dell'opera.

Al tempo stesso, però, le IA offrono nuovi strumenti di supporto alla fase creativa iniziale: l'autore non parte più necessariamente da una pagina completamente bianca, ma può trasformare rapidamente delle suggestioni in materiale concreto, attraverso processi di brainstorming

con i modelli testuali. Nel caso del presente progetto, ad esempio, questi sono stati utilizzati per analizzare rapidamente le emozioni raccolte dal pubblico di *Narrazioni Parallele* e ottenere spunti narrativi utili. Inoltre, si è lasciato ad essi la creazione di concept narrativi e sceneggiature, così da testarne le capacità.

È emerso come le IA possano suggerire anche soluzioni narrative o registiche interessanti, basate su esempi appresi durante la fase di addestramento. Tuttavia, la capacità di selezionare, reinterpretare e trasformare queste suggestioni in una visione coerente rimane prerogativa dell'autore umano.

Inoltre, il ruolo dell'autore umano rimane fondamentale come mediatore fra i diversi strumenti di intelligenza artificiale utilizzati. Le piattaforme attualmente disponibili operano in modo frammentato ed indipendente, a compartimenti stagni, senza un dialogo con le altre. Se in una troupe cinematografica tradizionale i diversi reparti dialogano tra loro, seguendo le linee guida del regista, nelle pipeline basate su IA è l'autore stesso a dover far dialogare gli strumenti – anche quelli che operano nello stesso reparto, come i tre modelli testuali – e illustrare la visione comune.

Nel contesto di questo progetto, ad esempio, i concept elaborati da ChatGPT, Gemini e Claude sono stati tradotti in prompt per la generazione di immagini e video; le immagini prodotte con Flow sono poi state scaricate localmente e, successivamente, utilizzate come input per la generazione tramite Pika e Veo3.1; infine, si è creato l'ambiente sonoro e sono stati effettuati i lavori di post-produzione. Alla fine del processo produttivo, solo l'autore umano possiede una visione completa delle scelte effettuate e delle trasformazioni apportate ai materiali.

Tuttavia, è da evidenziare come alcune piattaforme stiano iniziando ad integrare più funzioni all'interno dello stesso ambiente. Modelli come Sora permettono già di trasformare, all'interno della stessa interfaccia, le immagini generate in video, mentre Gemini consente di gestire testo, immagini e video nella stessa chat. Allo stesso modo, modelli come Veo e Kling offrono anche funzionalità di generazione audio sincronizzato alla clip video. Questi sviluppi suggeriscono una possibile evoluzione verso pipeline generative integrate in un unico ambiente, sebbene i sistemi citati presentino ancora numerosi limiti tecnici.

L'autore diventa quindi il punto di connessione tra strumenti che, per loro natura, operano in compartimenti separati. Senza questa continua attività di mediazione e coordinamento, i risultati prodotti dai singoli modelli non potrebbero essere integrati in modo coerente all'interno di un'unica opera audiovisiva.

Un altro ambito in cui il ruolo umano rimane insostituibile riguarda la fase di montaggio e post-produzione. Operazioni come l'organizzazione narrativa delle sequenze, la sincronizzazione tra audio e video e la correzione delle imperfezioni generate dagli algoritmi richiedono ancora competenze tecniche e decisioni autoriali che non possono essere delegate ai modelli generativi.

L'esperienza vissuta durante il progetto suggerisce dunque che la co-creazione uomo-macchina non sostituisca la creatività umana, ma piuttosto la redistribuisca. L'automazione amplia significativamente lo spazio delle possibilità esplorabili in tempi ridotti, mentre il controllo umano rimane essenziale per trasformare tali possibilità in un'opera completa.

L'intelligenza artificiale generativa può quindi essere interpretata come un collaboratore tecnico estremamente avanzato, capace di ampliare le possibilità operative del processo creativo, ma non come un sostituto dell'autore. Questa sperimentazione mostra come, paradossalmente, quanto più la macchina è in grado di generare contenuti in autonomia, tanto più diventa necessario un controllo autoriale rigoroso, capace di orientare tali risultati verso una visione narrativa ed estetica coerente.

4.4 Implicazioni etiche e diritto d'autore

L'integrazione dell'intelligenza artificiale nei processi creativi solleva questioni non solo tecniche, ma anche etiche. Nel [Capitolo 2](#) tali problematiche erano state introdotte attraverso l'analisi della letteratura esistente; l'esperienza progettuale vissuta ha permesso di osservarne e studiarne concretamente alcune. In particolare, i dubbi e le possibili problematiche più rilevanti emersi durante la realizzazione del cortometraggio riguardano i *bias* dei modelli generativi, la definizione del diritto d'autore delle opere realizzate con l'IA e, più in generale, il rapporto e l'equilibrio tra creatività umana e produzione algoritmica.

L'analisi di questi aspetti consente sia di valutare in modo più approfondito i limiti degli strumenti utilizzati, sia di riflettere sulle implicazioni culturali e le normative che accompagnano la crescente adozione dell'intelligenza artificiale nei processi artistici e audiovisivi.

4.4.1 Bias e rappresentazione

Durante le fasi di ideazione dei concept e generazione dei contenuti sono emersi alcuni elementi riconducibili alla presenza di possibili *bias* minori nei modelli utilizzati.

Un primo esempio riguarda la fase di sviluppo dei concept narrativi. Come già descritto nei capitoli precedenti, uno dei concept iniziali (*Manuale per non perdere i giorni*, proposto da ChatGPT) prevedeva la figura di un giovane inventore che catalogava suoni e ricordi all'interno di barattoli tramite un sistema ingegneristico complesso. Successivamente, proponendo di inserire come elemento narrativo un animale-guida, la narrazione è stata trasformata nella storia di un'anziana signora che vive con il proprio gatto e colleziona suoni e ricordi in barattoli magici.

Questo cambiamento può essere interpretato in modi diversi: da un lato, come una correzione basata su un immaginario stereotipato (ovvero, che solo il giovane inventore di sesso maschile possa avere un sistema ingegneristico complesso), dall'altro come una scelta autoriale consapevole per spostare il centro emotivo della narrazione verso una dimensione più intima e fragile. Chiedendo spiegazioni a ChatGPT, questo ha fornito la seconda motivazione, sottolineando come la presenza dell'anziana signora spostasse l'attenzione sui temi della caducità del tempo e il perdersi dei ricordi, rafforzando anche l'emotività della narrazione.

Sempre ChatGPT, durante la creazione dei model sheet dei personaggi, ha mostrato una tendenza riconducibile ad un *bias*. Quando si è chiesto di rappresentare un modello di lupo grigio nella versione 2D ad acquerello, questo è stato sempre rappresentato come se fosse un cucciolo, anche quando il prompt esplicitava il fatto che dovesse essere adulto e spaventoso. Questo comportamento può essere interpretato come una preferenza statistica del modello verso stili attualmente più richiesti dal mercato, con animali cartoon dalle fattezze carine ed addolcite. È interessante, però, notare come l'evento non si sia ripetuto per gli altri personaggi o stili (ovvero, 3D pittorico per il secondo atto e realistico per il terzo).

Un ulteriore caso è stato osservato durante la generazione delle clip video. Le IA utilizzate tendevano ad aggiungere un vapore molto accentuato nel respiro dei lupi durante le scene ambientate nel bosco (Atto 3). Anche se non veniva esplicitato nel prompt, questo elemento veniva introdotto automaticamente dai modelli ed accentuato all'estremo, rendendolo simile al fumo di una ciminiera. L'aggiunta era dettata probabilmente da un'interpretazione dell'ambiente invernale e freddo a seguito dell'addestramento dei modelli su materiali che rappresentavano l'uscita di vapore dalla bocca.

Non vi sono stati *bias* gravi da segnalare; rimane però importante il controllo autoriale sia per una selezione critica degli output generati, sia per evidenziare e correggere eventuali tendenze errate. La maggior parte degli strumenti IA permette inoltre di inviare un *feedback* riguardo le generazioni ottenute, così da segnalare eventuali criticità ed aiutare nella loro risoluzione.

4.4.2 Diritto d'autore delle opere generate

Un ulteriore nodo critico riguarda l'attribuzione del diritto d'autore alle opere prodotte con strumenti di intelligenza artificiale generativa. Le ultime normative europee (*AI Act*, 2024^[32]) stabiliscono che, in caso di contributo umano significativo (ad esempio prompt complessi o *editing* aggiuntivo), l'opera appartenga all'autore umano; l'utilizzo commerciale è però soggetto ad eventuali violazioni di copyright.

Nel caso specifico di questo progetto, ogni fase è stata guidata e rielaborata dall'autore umano, nell'ottica di co-creazione uomo-macchina già descritta.

Se l'autore toglie l'autonomia decisionale e l'intenzionalità artistica ai modelli IA, lasciandola lavorare solo come strumento di generazione e supporto, il ruolo di quest'ultima può essere paragonato a quello di software complessi utilizzati nella produzione audiovisiva, come Blender o DaVinci Resolve. Tuttavia, la natura generativa di questi strumenti introduce un ulteriore livello di complessità: alcune idee narrative o dettagli visivi possono emergere direttamente dagli output dei modelli, senza essere stati progettati dall'autore, come nel caso dell'aggiunta delle scritte all'interno delle pagine del libro (Atto 1).

La questione si complica ulteriormente se si prendono in considerazione i *dataset* utilizzati per addestrare i modelli. Durante la fase di *training*, le IA vengono esposte ad enormi quantità di dati (ad esempio, foto, materiali audio, testi, clip video) al fine di apprendere pattern statistici e relazioni tra gli elementi presenti nei contenuti. Se all'interno di questi dataset sono inclusi materiali protetti da copyright e gli autori non hanno dato il consenso al loro utilizzo, possono emergere problematiche giuridiche contro le aziende produttrici dei modelli.

Nel caso di questo progetto, l'utilizzo di Suno ha sollevato interrogativi sulla provenienza delle voci generate. In alcune generazioni, infatti, il timbro vocale risultava sorprendentemente simile a quello di artisti reali. Ad esempio, tra i brani presenti nella homepage, uno in particolare aveva una forte somiglianza con la voce del rapper Drake, mentre nella versione italiana della rielaborazione della canzone composta durante l'evento *Song Writing* vi era un timbro vocale che ricordava quello della cantante Dolcenera. Tali somiglianze possono essere il risultato di una semplice percezione umana, influenzata dalla naturale tendenza a riconoscere voci familiari anche quando non lo sono realmente. Tuttavia, poiché le voci generate risultano riconducibili allo stile o al timbro di artisti esistenti, è legittimo chiedersi se tali modelli siano stati addestrati anche su registrazioni vocali di performer reali e, in tal caso, se tale utilizzo sia avvenuto con il loro consenso.

Inoltre, strumenti come Suno ed ElevenLabs permettono di caricare registrazioni vocali da utilizzare come base per la generazione di nuove voci sintetiche. Sebbene questa funzionalità possa essere impiegata in modo legittimo, ad esempio utilizzando la propria voce, non si può

escludere che alcuni utenti possano caricare registrazioni appartenenti a conoscenti o a personaggi pubblici senza averne l'autorizzazione.

4.4.3 Democratizzazione e figura dell'autore

Considerando la crescente diffusione di strumenti generativi gratuiti o a basso costo, è necessario riflettere sulla promessa di “democratizzazione” della produzione audiovisiva spesso associata all'IA.

Alla luce di quanto appreso durante la realizzazione del progetto, si può dire che tale democratizzazione esiste, ma solamente in forma parziale.

Da un lato, queste tecnologie sono in grado di ridurre barriere tecniche ed economiche. Attività che richiedono competenze avanzate o accesso a software complessi possono oggi essere effettuate attraverso semplici prompt testuali, permettendo quindi ad un ampio numero di utenti di sperimentare con il linguaggio audiovisivo e di realizzare contenuti visivi o sonori in tempi ridotti.

Dall'altro lato, però, l'esperienza del progetto ha dimostrato come la produzione di un prodotto coerente e di qualità richieda ancora competenze e conoscenze tecniche significative. Se la generazione di contenuti semplici è effettivamente accessibile a molti utenti, la produzione di opere narrative più articolate richiede una supervisione autoriale costante, oltre a numerosi interventi di selezione, correzione ed integrazione dei materiali generati.

In questo scenario, la figura dell'autore non sembra destinata a scomparire, ma piuttosto a trasformarsi. Se sempre più utenti potranno generare immagini, video o musica, ciò che distinguerà realmente un autore non sarà più soltanto la capacità tecnica di produrre contenuti, ma la capacità di selezionare, strutturare e attribuire senso al materiale generato.

Durante lo sviluppo del cortometraggio, ad esempio, conoscenze legate alla fotografia, alla teoria del colore e alla costruzione narrativa si sono rivelate fondamentali per guidare i modelli generativi verso risultati coerenti con l'intenzione espressiva del progetto. Un utente privo di queste competenze potrebbe comunque ottenere risultati visivi accettabili, ma con maggiore probabilità si affiderebbe completamente alle indicazioni suggerite dalle stesse IA, rischiando di produrre prompt e output generici o didascalici.

L'intelligenza artificiale, quindi, amplia l'accesso agli strumenti di produzione, ma non elimina la necessità di una visione autoriale consapevole e di uno studio pregresso.

4.4.4 L'IA può comprendere le emozioni umane e creare?

Infine, una delle questioni più complesse e frequentemente discusse riguarda la possibilità che un sistema algoritmico possa realmente comprendere le emozioni umane e produrre opere artistiche.

L'esperienza sperimentata durante lo sviluppo del progetto suggerisce che i modelli generativi non comprendono le emozioni nel senso fenomenologico del termine. Piuttosto, essi riconoscono pattern statistici, correlazioni e strutture linguistiche o visive che, all'interno dei *dataset* di addestramento, risultano frequentemente associate a determinate categorie emotive. La coerenza emotiva osservabile negli output deriva quindi dalla capacità dei modelli di combinare elementi che nei dati di *training* sono stati correlati a specifiche atmosfere narrative, visive o sonore.

Ciò non implica necessariamente che il risultato non possa essere percepito come artistico o meno. L'esperienza estetica dipende infatti in larga misura dall'interpretazione dello spettatore, che attribuisce significato e valore emotivo all'opera osservata. Tuttavia, mentre l'autore umano possiede intenzionalità, esperienza e consapevolezza delle proprie scelte creative, il sistema algoritmico opera esclusivamente attraverso calcoli probabilistici e relazioni statistiche tra dati.

Il vero spartiacque tra uomo e macchina risiede nell'intenzionalità. Mentre l'autore umano è mosso da un'urgenza espressiva, dall'esperienza personale e dalla piena consapevolezza delle proprie scelte creative, il sistema algoritmico opera esclusivamente attraverso calcoli probabilistici. L'intelligenza artificiale può quindi produrre risultati sorprendenti o inattesi, ma questi sono il prodotto di combinazioni statistiche all'interno dello spazio dei dati di addestramento, e non di un'autentica esperienza creativa o emotiva.

4.5 Sviluppi futuri e prospettive

È fondamentale considerare, in conclusione, la rapida evoluzione degli strumenti di IA generativa. Un esempio è rappresentato da Gemini che, pochi giorni prima della conclusione di questo progetto di tesi (a fine febbraio 2026), ha aggiunto una nuova funzionalità: la possibilità di creare tracce musicali.

4.5.1 Sviluppi tecnologici attesi

Dal punto di vista tecnico, è plausibile attendersi miglioramenti significativi nelle aree che attualmente rappresentano i principali limiti dei sistemi generativi.

Prima fra tutte, la consistenza spaziale e temporale. L'instabilità dei dettagli tra frame consecutivi, le variazioni involontarie nella forma dei personaggi e la difficoltà nel mantenere identità visive persistenti rappresentano ancora ostacoli rilevanti per la produzione di prodotti audiovisivi complessi. Un miglioramento in questa direzione permetterebbe di ottenere sequenze più coerenti, riducendo la necessità di interventi correttivi in post-produzione.

Un secondo sviluppo interessante dovrebbe riguardare l'attuale durata delle clip generabili e la loro risoluzione. Nella maggior parte dei modelli testati, infatti, le generazioni erano limitate a pochi secondi e la risoluzione massima raggiungeva i 1080p. Superare il limite del tempo consentirà un linguaggio registico meno frammentato ed un montaggio più lineare, oltre a permettere l'inserimento di movimenti di camera fluidi e piani sequenza complessi. Allo stesso modo, la possibilità di ottenere risoluzioni migliori (quali, ad esempio 2K e 4K), insieme ad un incremento del fotorealismo e alla qualità complessiva dell'immagine, consentirà alle opere generate tramite IA di colmare l'attuale divario tecnico, raggiungendo un'estetica paragonabile agli standard dell'industria cinematografica.

Vista la diffusione degli strumenti di intelligenza artificiale generativa nel panorama audiovisivo, un altro sviluppo giustificato riguarda l'introduzione di parametri espliciti e controllabili relativi al tipo di illuminazione, ottiche, movimenti di camera. Attualmente, attraverso prompt testuali, è possibile inserire queste informazioni e la maggior parte dei modelli è in grado di comprenderle efficacemente; si pensa, però, all'introduzione di strumenti sempre più precisi, per registi ed autori esperti, con funzionalità tecniche aggiuntive, come quelle offerte da Loudly riguardo la possibilità di controllare il ritmo e la metrica della canzone. Possedere strumenti che restituiscano un controllo preciso, accurato e prevedibile su ogni elemento rappresenterà un grande salto di qualità per il mondo dell'industria audiovisiva.

Inoltre, sarà normale avere le interfacce tradotte in molteplici lingue ed in grado di comprenderne altrettante, come attualmente è in grado di fare Flow per la generazione di immagini e video.

Infine, per quanto gli attuali modelli sonori (*text-to-speech* e generatori di canzoni) siano notevolmente maturi, essi sono destinati ad evolversi verso una sempre maggiore naturalezza della voce in qualsiasi lingua e ad una possibile sincronizzazione automatica del labiale e dei suoni in base ad un video dato, così da velocizzare il lavoro di post-produzione audio.

4.5.2 Evoluzione delle pipeline produttive

Parallelamente all'evoluzione tecnologica dei modelli, si assisterà ad una trasformazione delle pipeline produttive basate su IA. Attualmente, gli strumenti operano infatti in modo frammentato, costringendo l'utente a continui passaggi tra piattaforme di testo, immagine, video e audio. In futuro, vi saranno sempre più ambienti integrati, in cui i diversi modelli potranno comunicare direttamente fra di loro. Un primo esempio di ciò lo si può vedere, ad esempio, all'interno di Sora, dove da un'immagine generata è possibile realizzare un video senza doverla scaricare e caricare nuovamente.

Interessante è anche l'integrazione di strumenti IA nei software professionali di editing, come DaVinci Resolve. In futuro, all'interno di un unico ambiente, saranno automatizzate le operazioni di color correction e di aggiustamento di incongruenze, oltre a ricevere suggerimenti di montaggio basandoti direttamente sulla sceneggiatura prodotta.

Tuttavia, anche se gli strumenti saranno uniti e in grado di lavorare sempre più autonomamente, è difficile che possano sostituire completamente la supervisione autoriale nelle produzioni narrative complesse, poiché le decisioni stilistiche e ritmiche sono strettamente legate alla sensibilità umana.

Una trasformazione radicale riguarderà anche l'infrastruttura hardware. Attualmente, la maggior parte degli strumenti generativi lavora online, *in cloud*, dipendendo da *server* remoti. Tuttavia, aziende leader del settore tecnologico, come NVIDIA, stanno sviluppando dispositivi e componenti ottimizzati per processare i sistemi di intelligenza artificiale "in locale". Questo renderà i workflow più stabili ma, soprattutto, privati (la riservatezza dei dati è, ad esempio, uno standard di sicurezza imprescindibile per le grandi case di produzione cinematografiche). Serviranno però macchine estremamente performanti per poter gestire il carico computazionale, escludendo così molti utenti dal loro utilizzo.

4.5.3 Ridefinizione del ruolo dell'autore e democratizzazione

Come già approfondito, il ruolo dell'autore non è destinato a scomparire, bensì, a trasformarsi in una nuova figura, capace di combinare competenze creative tradizionali, conoscenze tecniche e capacità di sfruttare al meglio i nuovi sistemi generativi. La differenza fra un contenuto professionale ed uno amatoriale sarà sancita anche dal *prompt design* più o meno avanzato, la selezione degli output e la loro gestione ed unione al fine di una narrazione coerente.

Allo stesso tempo, la diffusione degli strumenti, anche in formato gratuita, accende l'attenzione sulla democratizzazione della produzione audiovisiva. L'IA è infatti in grado di abbattere le barriere tecniche, eliminando soprattutto la necessità di apprendere l'utilizzo di software complessi, ma l'evoluzione degli strumenti porterà anche a costi in aumento per abbonamenti o infrastrutture adeguate. La nascita di piattaforme private e l'innalzamento dei prezzi creeranno nuove forme di divario digitale, annullando le premesse stesse della democratizzazione.

Nonostante ciò, è innegabile che l'attuale accessibilità agli strumenti generativi offra opportunità senza precedenti, in particolar modo per gli artisti indipendenti, che possono portare a compimento progetti ambiziosi senza un team specializzato di supporto. Questa libertà consentirà di creare linguaggi alternativi e narrazioni innovative, lontane dai canoni estetici e contenutistici imposti dall'industria *mainstream*.

La co-creazione tra uomo e macchina porterà, quindi, a nuove, interessanti, forme ibride.

4.5.4 Questioni normative e diritto d'autore

A queste trasformazioni tecnologiche si affiancano importanti questioni normative. L'attuale quadro legislativo risulta ancora in fase di definizione e l'evoluzione estremamente rapida degli algoritmi generativi rischia di generare situazioni giuridiche ambigue, con possibili ripercussioni sia per le aziende che sviluppano tali tecnologie sia per gli artisti le cui opere possono essere state utilizzate nei dataset di addestramento. Questa incertezza potrebbe, da un lato, scoraggiare investimenti nel settore creativo basato su IA e, dall'altro, limitare la tutela degli autori originali.

Le future regolamentazioni saranno chiamate a chiarire in modo sempre più attento molteplici aspetti fondamentali, tra cui l'utilizzo di dati protetti da copyright nei dataset di training, il consenso per l'addestramento di modelli vocali o musicali e, soprattutto, i criteri di attribuzione del diritto d'autore nelle opere generate con il supporto dell'intelligenza artificiale. In particolare, sarà necessario definire in modo più preciso in quale misura l'intervento umano possa essere considerato sufficiente per attribuire la paternità creativa dell'opera all'autore (ad esempio, attraverso attività come il *prompting*, la selezione e la supervisione degli output), oppure se tali produzioni debbano essere classificate come opere derivate o come contenuti privi di tutela autoriale tradizionale.

4.5.5 Impatto sul linguaggio audiovisivo

Infine, l'intelligenza artificiale è destinata ad avere un impatto significativo anche sul linguaggio audiovisivo e le modalità di produzione. La possibilità di generare rapidamente numerose

varianti di uno stesso contenuto favorisce la sperimentazione creativa e consente di esplorare soluzioni visive e narrative che, in una pipeline tradizionale, richiederebbero tempi e risorse elevati. Questo scenario aprirà nuove possibilità espressive, soprattutto per artisti indipendenti e progetti sperimentali, permettendo la nascita di linguaggi visivi alternativi ed innovativi.

Allo stesso tempo, esiste il rischio di una progressiva omologazione stilistica. I modelli generativi operano infatti sulla base di correlazioni statistiche apprese dai dati di addestramento e tendono quindi a riprodurre configurazioni visive medie o particolarmente ricorrenti all'interno dei dataset. Di conseguenza, senza un intervento autoriale consapevole, gli output potrebbero convergere verso estetiche simili tra loro. L'originalità delle future opere frutto di co-creazione uomo-macchina dipenderà dalla capacità degli autori di utilizzare gli strumenti in modo critico e creativo, sfruttandone le potenzialità senza limitarsi ai pattern più prevedibili generati dagli algoritmi.

4.6 Considerazioni conclusive

In conclusione, la sperimentazione condotta conferma come l'intelligenza artificiale generativa sia ancora poco matura per sostituire completamente l'autore umano nel processo creativo, presentando limiti e criticità ricorrenti e non trascurabili. Tuttavia, si evidenzia anche l'enorme potenziale di questa tecnologia, soprattutto come strumento di supporto avanzato per ottimizzare i tempi e svolgere funzioni ripetitive ed onerose, oltre che per espandere e supportare la sperimentazione creativa e la nascita di nuove idee.

La pipeline realizzata durante questo progetto si è rivelata efficace e ha portato alla realizzazione di un risultato finale più che soddisfacente. Essa si dimostra facilmente replicabile anche per progetti futuri e la sua forza risiede nella modularità. Questa impostazione consente la sostituzione o l'aggiornamento dei singoli strumenti senza compromettere il workflow complessivo. Infatti, la stabilità del progetto non risiede nel singolo strumento utilizzato (destinato a rapida obsolescenza, visti i continui e rapidi sviluppi delle IA), ma nella solida logica di integrazione tra i diversi moduli della pipeline. In un contesto tecnologico caratterizzato da un'evoluzione rapida degli strumenti, la possibilità di sostituire un modello video, audio o testuale con la sua versione successiva o con un competitor più performante, permette di mantenere intatto il piano lavorativo ma, parallelamente, elevare la qualità degli output.

Inoltre, un altro punto di forza risiede di utilizzare più strumenti per svolgere la medesima funzione e sopperire alle mancanze degli uni o degli altri; come nel caso dell'utilizzo congiunto di Pika e Veo3.1 per la generazione video.

Nel complesso, il progetto dimostra come l'intelligenza artificiale non annulli la pipeline audiovisiva tradizionale, ma ne proponga una riconfigurazione. Non vi sarà, inoltre la scomparsa delle competenze umane e registiche, bensì una loro trasformazione e ibridazione con i sistemi IA. In questo scenario, il vero interrogativo per i professionisti del settore non riguarda tanto la possibile obsolescenza delle loro attuali competenze tecniche, quanto piuttosto l'individuazione di quali nuove capacità (come il *prompt design* avanzato) diventeranno il fulcro centrale del processo creativo di domani.

Infine, riguardo alla pipeline sviluppata, sussistono, naturalmente, ampi margini di ottimizzazione, destinati a progredire parallelamente al miglioramento e avanzamento delle tecnologie. Gli spunti, i compromessi e le metodologie emersi da questa sperimentazione non rappresentano un punto di arrivo, bensì, un punto di partenza testato per definire un modello produttivo ibrido di co-creazione tra uomo e macchina, in linea con l'inesorabile evoluzione tecnologica e culturale del linguaggio audiovisivo contemporaneo.

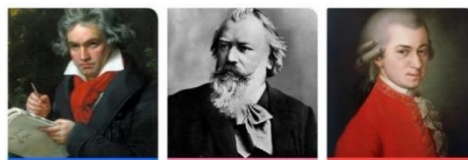
APPENDICI

APPENDICE A – Domande poste al pubblico

In questa sezione sono riportate tutte le domande poste al pubblico durante i tre eventi, anche quelle che non sono state utilizzate per la realizzazione del progetto audiovisivo.

Concerto 1 – “Caro Mozart... Una lettura musicale delle lettere alla cugina di un genio inquieto”

- 1) Condividi la prima parola che ti viene in mente su come ti senti ora
- 2) Chi è Mozart dei tre?
- 3) In quale secolo è vissuto Mozart?
 - a. Settecento
 - b. Ottocento
 - c. Novecento



Option 1

Option 2

Option 3

Risposte possibili per la domanda numero 2

- 4) Quali strumenti suonava da piccolo?
- 5) Secondo te, che carattere aveva Mozart da piccolo?
- 6) Quali strumenti, secondo te, hanno stonato palesemente?
[ci si riferisce all'esecuzione di Ein Musikalischer Spass KV 522]
 - a. Corni
 - b. Violini
 - c. Viola
 - d. Bassi
- 7) In che punto del brano ti sei accorto/a dell'errore?
[ci si riferisce all'esecuzione di Ein Musikalischer Spass KV 522]
- 8) Come ti fa sentire questa musica rispetto a quella di prima?
[ci si riferisce all'esecuzione di Ave Verum Corpus K 618]
- 9) Secondo te, dopo aver ascoltato le composizioni e le lettere di Mozart, che carattere può aver avuto da adulto?
- 10) Se Mozart era in grado di scrivere musica così bella, perché ha scritto il sestetto in modo così goffo?
 - a. Era troppo giovane e non conosceva bene le regole della composizione
 - b. Fa sbagliare apposta gli strumentisti per criticare chi suona male
 - c. I musicisti presenti hanno volutamente sbagliato per far ridere il pubblico
 - d. I musicisti presenti non sono bravi
 - e. Mozart odiava il corno

- 11) Se Mozart visse oggi, che tipo di musica scriverebbe?
- 12) Se Mozart visse oggi, a quale personaggio contemporaneo somiglierebbe?
- 13) Scrivi una o più parole per descrivere cosa senti fisicamente in termini di benessere o malessere dopo l'ascolto
- 14) Prima e dopo l'ascolto hai notato delle differenze nel tuo umore e nel tuo stato d'animo? Quali? Come ti senti ora?
- 15) Quale grado di benessere hai avvertito durante l'attività da 1 a 5?
(1 = totalmente in disaccordo; 5 = fortemente d'accordo)
 - a. Ti sei sentita/o a tuo agio
 - b. La tua creatività è stata stimolata
 - c. L'esperienza ed il grado di coinvolgimento sono stati soddisfacenti
 - d. Ti sei sentita/o parte di una comunità

Concerto 2 – “Song writing 2 – Intelligenza naturale generativa”

- 1) Condividi la prima parola che ti viene in mente su come ti senti ora
- 2) Quali immagini o pensieri ti ha suscitato la poesia appena letta?
[ci si riferisce alla poesia “Il paese con la S davanti” di Gianni Rodari]
- 3) Se potessi inventare una nuova regola per vivere meglio, quale sarebbe?
- 4) Se potessi cambiare le leggi della fisica/tempo/logica, quale sarebbe la prima cosa che faresti?
- 5) Se potessi svegliarti con un superpotere per un giorno, quale sarebbe?
- 6) Se potessi pranzare con chiunque, reale o immaginario, chi sarebbe?
- 7) Con quale animale ti senti più affine?
- 8) Se potessi scegliere un profumo da annusare per le strade, cosa sarebbe?
- 9) Con quale strumento musicale identifichi la tua voce interiore?
- 10) Se dovessi rappresentare la canzone appena creata con una stagione, quale sarebbe?
 - a. Inverno
 - b. Primavera
 - c. Estate
 - d. Autunno
- 11) Quale colore assoceresti al tema trattato oggi?
- 12) Condividi la prima parola che ti viene in mente su come ti senti ora
- 13) Analizziamo il grado di benessere che hai avvertito durante l'attività!
(1 = pochissimo; 5 = tantissimo)
 - a. Ti sei sentita/o a tuo agio
 - b. La tua creatività è stata stimolata
 - c. L'esperienza ed il grado di coinvolgimento sono stati soddisfacenti

- d. Ti sei sentita/o parte di una comunità
 - e. Hai scoperto una nuova passione (per la musica, poesia, ...)
- 14) Pensando all'attività, quale stile di video (creato con l'intelligenza artificiale) ti piacerebbe vedere?
- 15) Pensi che l'intelligenza artificiale avrebbe potuto creare una canzone simile?
- a. Sì, anche meglio
 - b. Sì
 - c. No
 - d. Non lo so

Concerto 3 – “In scena con Verdi”

- 1) Condividi la prima parola che ti viene in mente su come ti senti ora
- 2) Quale colore rappresenteresti come ti senti ora?
- 3) Condividi la prima parola che ti viene in mente pensando a Giuseppe Verdi
- 4) Come riscriveresti con stile contemporaneo i primi versi di “Va’ pensiero”?
- 5) Quali pensieri o ricordi sono emersi durante l'ascolto dei brani? Quali immagini hai visualizzato?
- 6) Se la musica di Verdi fosse un luogo, quale sarebbe?
- 7) Se la musica di Verdi fosse un colore, quale sarebbe?
- 8) Condividi la prima parola che ti viene in mente su come ti senti ora
- 9) Analizziamo il grado di benessere che hai avvertito durante l'attività!
(1 = pochissimo; 5 = tantissimo)
 - a. Ti sei sentita/o a tuo agio
 - b. La tua creatività è stata stimolata
 - c. L'esperienza ed il grado di coinvolgimento sono stati soddisfacenti
 - d. Ti sei sentita/o parte di una comunità
- 10) Pensi che l'intelligenza artificiale avrebbe potuto riscrivere “Va’ pensiero”?

 - a. Sì, anche meglio
 - b. Sì
 - c. No
 - d. Non lo so

- 11) Pensando all'attività di oggi, quale tema potrebbe trattare il video creato con l'Intelligenza Artificiale?

APPENDICE B – Risposte del pubblico e dati raccolti

In questa sezione sono riportate le risposte del pubblico alle domande considerate rilevanti per il progetto, affiancate dalle tabelle contenenti i dati riordinati. Queste ultime sono state date ai modelli testuali per individuare ricorrenze, differenze e tematiche significative.

Per una facilità di organizzazione e consultazione, data la lunghezza complessiva del materiale, il file completo è stato caricato in una cartella Dropbox accessibile tramite il link o il QR code sotto riportati.

RISPOSTE DEL PUBBLICO E TABELLE DATI - DROPBOX



APPENDICE C – Proposte concept narrativi

In questa sezione sono riportate le tre proposte di concept generate da ciascuna delle intelligenze artificiali generative utilizzate, nella loro versione originale, prima di essere riviste ed approfondite dall'intervento umano.

Per una facilità di organizzazione e consultazione, data la lunghezza complessiva del materiale, il file completo è stato caricato in una cartella Dropbox accessibile tramite il link o il QR code sotto riportati.

PROPOSTE CONCEPT - DROPBOX



APPENDICE D – Sceneggiatura

In questa sezione è riportata sceneggiatura completa del cortometraggio.

TITOLO: THE PAPER WOLF

Durata: 5 minuti

IA di riferimento: Gemini Pro

ATTO I – IL LIBRO

Stile visivo: 2D illustrato, acquerello

Atmosfera: malinconica, rigida, controllata

FADE IN:

SCENA 1A/1B – INT. BIBLIOTECA – NOTTE – [STILE REALISTICO]

La biblioteca è immersa nella penombra, illuminata solo da candele.

Un ambiente ampio, silenzioso, costruito in legno scuro: scaffali alti fino al soffitto, libri antichi, tavoli consumati dal tempo. Granelli di polvere sospesi fluttuano lentamente nell'aria, come particelle vive.

La camera esegue una **carrellata in avanti** fra gli scaffali ed i tavoli, esplorando lo spazio.

La camera esegue una **carrellata laterale** in una sala con tavoli in legno antico e grossi tomi.

SCENA 2 – INT. BIBLIOTECA – NOTTE – [STILE REALISTICO]

Su un tavolo in legno è posato un grosso libro antico. La copertina reca la scritta in latino "Familias".

La camera si avvicina lentamente al libro e si sposta per finire in un'inquadratura dall'alto. Nel mentre, senza alcun intervento umano, la copertina si solleva lentamente, come mossa da una forza invisibile.

[mentre le pagine girano]

VOCE NARRANTE:

In this book, everyone has a family.

SCENA 3A/3B/3C – INT. LIBRO – GIORNO

Inquadratura dall'alto (top-down) con camera statica. Si vede il libro (dove avverranno le scene) ed una parte di tavolo.

La pagina si sta sfogliando da sola e mostra una breve scena con una famiglia di cervi che bruca l'erba nel bosco, con un sole caldo ed una breve brezza primaverile.

VOCE NARRANTE:

In the quiet forest, the deer and their fawns graze peacefully.

La pagina si sfoglia da sola e mostra una famiglia di orsi grizzly in riva ad un fiume mentre pesca i salmoni che sbucano fra le rocce e la corrente.

VOCE NARRANTE:

By the river, the mother bear taught her cubs how to fish.

La pagina si sfoglia da sola e mostra una famigliola di uccellini blu posati lungo un ramo. Un piccolo si stringe alla madre mentre gli altri cantano spensierati.

VOCE NARRANTE:

Bluebirds, with their tender songs, fill the air with joy.

La pagina si sfoglia da sola.

Le illustrazioni sono in stile 2D, acquerello vintage. Trasmettono equilibrio, armonia, quotidianità.

SCENA 4A – INT. LIBRO, lupi – NOTTE

Doppia pagina aperta, divisa visivamente in due.

A SINISTRA: un branco di lupi grigi, tutti uguali tra loro, ulula all'unisono alla luna. Il movimento è ciclico, perfettamente sincronizzato, quasi meccanico, come un loop animato.

A DESTRA: separato dal margine centrale della rilegatura, c'è il LUPO COLORATO. È seduto, immobile, nell'angolo più a destra. Il suo corpo è composto da macchie irregolari di carta dipinta: blu, arancio, verde. È visivamente e simbolicamente fuori posto.

VOCE NARRANTE:

Among the wolves, voices rise together.

Il lupo colorato sospira, abbattuto. Poi si alza ed emette un ululato. Esce un suono stonato, spezzato, come di carta accartocciata, reso graficamente con linee tremolanti e irregolari.

VOCE NARRANTE:

But not all voices are born at the same time.

SCENA 4B/4C/4D/4E – continua

Il branco di lupi grigi smette di ululare e cala il silenzio.

Il Lupo Colorato si avvicina lentamente a loro, spostandosi a sinistra verso il margine che lo separa dagli altri.

I lupi grigi iniziano a ringhiare all'unisono, minacciosi. Le loro linee diventano spigolose ed aggressive. Non sono accoglienti.

VOCE NARRANTE:

Those who sound different... are left outside.

Il Lupo Colorato si spaventa e si accartocchia, con la coda tra le zampe e la testa bassa. Si volta e torna nel suo angolo all'estremo destro della pagina, dove si accuccia e si raggomitola su sé stesso. È piccolo, indifeso, isolato.

VOCE NARRANTE:

Alone.

I lupi grigi smettono di ringhiare ed inizia a nevicare. Attorno a loro prende forma una tana, calda e accogliente. I lupi si accucciano insieme, protetti.

Sulla pagina destra, il Lupo Colorato resta solo, sotto la neve, bagnato, immobile.

SCENA 4F/4G/4H – INT. LIBRO, lupi – ALBA

Il tempo accelera. Nella pagina di sinistra, in alto, la luna si sposta verso destra, ma senza superare il margine, mentre il sole sorge da sinistra. Lo sfondo muta gradualmente: dal blu notte a un rosso-arancio dell'alba.

Il Lupo Colorato non si è mosso.

Nella pagina di sinistra, in alto, dal bordo sinistro, entra un gatto nero. Non è dipinto come gli altri personaggi: è uno schizzo a penna nera, nervoso, incompleto. Cammina in diagonale,

ignorando le leggi prospettiche e la gravità del libro, lasciando dietro di sé impronte di inchiostro.

SCENA 4I/4J – continua

Il Gatto si ferma davanti al Lupo Colorato. Si siede con eleganza.

Lo guarda. Miagola.

Il Gatto graffia la superficie della pagina. La carta si lacera. Il Gatto dentro lo strappo.

Oltre il buco c'è solo nero assoluto.

SCENA 4K/4L – continua

Il Lupo Colorato si avvicina allo strappo, ma è troppo piccolo per lui.

Il Lupo Colorato si ferma. Guarda verso il branco: dorme nella sua tana perfetta. Guarda nuovamente il buco

Afferra il lembo di carta con i denti. Tira. Allarga il passaggio, strappando la pagina. Ed entra.

VOCE NARRANTE:

The wolf hesitated. Then, something called louder.

[Taglio sul nero]

VOCE NARRANTE:

Some voices, before being born, must get lost.

SCENA 5A/5B – TUNNEL DI CARTA, TRANSIZIONE, TUNNEL 3D

Il Lupo Colorato precipita in un tunnel surreale, dove fogli di carta, libri ed altri elementi e frammenti vorticano attorno a lui.

Una pagina passa davanti alla camera, coprendo l'obiettivo. TRANSIZIONE/TAGLIO FANTASMA.

Il Lupo Colorato continua a precipitare, ma ora è nella sua forma 3D pittorica (i colori sono uguali alla versione ad acquerello, ma la pelliccia è spessa e vedono le grosse pennellate, modello 3D stile low poly). Anche nel tunnel gli elementi sono tridimensionali.

ATTO II – IL MONDO SOSPESO

Stile visivo: 3D pittorico, low poly

Atmosfera: creativa, caotica, ludica

SCENA 6 – EXT. MONDO 3D – BLUE HOUR

Il Lupo Colorato emerge dal tunnel e atterra goffamente su una nuvola morbida. Il mondo attorno a lui è instabile: nuvole fluttuanti, elementi naturali isolati, oggetti quotidiani sospesi nello spazio, come se fossero stati liberati dalla gravità e dal senso logico. Ma tutto si muove lentamente, con calma.

Il mondo è nell'ora del crepuscolo, prima del sorgere del sole, ancora parzialmente immerso nella luce blu notturna.

Il Gatto è sulla nuvola con il lupo. Salta agilmente su un'altra nuvola e miagola, emettendo un suono armonioso ed accompagnato da particelle luminose. Il Gatto Nero è tridimensionale e materico, come tutto il resto. Osserva poi il Lupo Colorato.

SCENA 7 – continua

Il Lupo Colorato fa dei passi in avanti e cerca di seguire il Gatto, ma cade nella nuvola sottostante: la fisica di questo mondo è imprevedibile.

SCENA 8 – continua

Il Lupo Colorato prova a ululare per richiamare il Gatto. Di nuovo, esce un suono stonato, spezzato, come di carta accartocciata, reso graficamente con linee tremolanti e irregolari.

SCENA 9 – continua

Il Gatto miagola di nuovo, emettendo un suono armonioso ed accompagnato da particelle luminose.

Poi tocca una nota sospesa nell'aria e si sprigiona una luce morbida e calda, accompagnata da un armonioso DING, in linea con la musicalità del miagolio.

SCENA 10 – continua

Il Lupo Colorato prova a imitare il Gatto e tocca un libro fluttuante. Dal contatto nasce solo un rumore sordo di pagine che si chiudono (*THUD*). Una piccola nube di polvere grigia si disperde. Nessuna luce.

Il Lupo Colorato si irrigidisce, frustrato.

SCENA 11 – continua

Il Gatto inizia a muoversi con leggerezza. Tocca la nota sospesa (*DING* + luce), sfiora una foglia (*UCCELLINI* + luce), gioca con un gomitolino di lana (*GATTINI* + luce). Poi miagola (in armonia e con luce).

I suoi movimenti creano una danza sonora, spontanea, giocosa, che si ripete e rappresenta la sua personalità e come vede il mondo (es. la foglia la associa agli uccellini che vede salendo sugli alberi).

SCENA 12A/12B/12C – continua

Il Lupo Colorato osserva attentamente e prova a sperimentare. **Dare l'idea di vari tentativi con un montaggio veloce o uno split screen.**

Il Lupo Colorato tocca di nuovo il libro e non accade nulla. Tocca della legna. Poi ulula spezzato.

Ci riprova.

Dopo alcuni tentativi inizia a sentire dei rumori. Il libro porta con sé dei fruscii di pagine, la legna il calore di un focolare e la stella un “ding” armonico e magico. Infine ulula! Non è perfetta, ma è la sua voce.

SCENA 13/14 – continua

Il Gatto salta accanto al Lupo Colorato e miagola. Il Lupo Colorato ulula con lui, generando particelle colorate. Si crea un'armonia imperfetta ma autentica.

Il Lupo Colorato ed il Gatto, felici, giocano insieme saltando fra le nuvole e gli elementi sospesi.

Vengono però interrotti da un ululato lontano, familiare.

SCENA 15 – continua

Primo piano su una pagina di carta che fluttua lentamente nel vuoto. Sulla pagina è disegnato il branco di lupi grigi, in stile 2D acquerello. La pagina è come un fantasma del passato.

SCENA 15/16 – continua

Il Lupo Colorato, entusiasta, si avvicina alla pagina.

Ma i lupi disegnati si animano: ringhiano, ostili, come prima.

Il Lupo Colorato indietreggia ed emette un ululato spezzato, carico di tristezza.

Il contrasto tra il mondo tridimensionale e la bidimensionalità della pagina è netto.

SCENA 17/18 – continua

Il Gatto si avvicina lentamente e si struscia contro il lupo. Miagola dolcemente.

Il Lupo Colorato guarda la pagina dei lupi ringhianti e prende una decisione. La strappa. Non ne ha più bisogno.

SCENA 19 – MONDO 3D, TRANSIZIONE, MONDO REALE

Il Lupo alza la testa ed ulula. Questa volta l'ululato è potente, continuo, melodico.

Mentre il suono si espande, una foresta (stile realistico) si genera attorno ai personaggi. La realtà sta prendendo forma.

La camera orbita attorno a loro e gli alberi coprono l'obiettivo.

ATTO III – LA FORESTA REALE

Stile visivo: Realistico live action

Atmosfera: solenne, fredda, ma accogliente

SCENA 20A/20B – EXT. FORESTA – BLUE HOUR

Il mondo si è trasformato. Una foresta reale, fitta e silenziosa, si estende davanti alla camera. Gli alberi sono alti, scuri, coperti di neve. I rami scricchiolano lievemente sotto il peso del ghiaccio.

Il Lupo Colorato ora è completamente realistico: pelo fitto, respiro visibile nell'aria fredda, zampe che affondano nella neve. Mantiene però i suoi colori originali.

Con lui c'è il Gatto Nero, anch'esso in vesti realistiche e con una lunga pelliccia soffice.

Siamo ancora nella blue hour prima del sorgere del sole, che inizia a prendere il posto della notte.

Il Lupo Colorato si guarda attorno spaesato. Il silenzio profondo è interrotto da ululati lontani, autentici.

Il Lupo Colorato prova a rispondere, ma il suo primo ululato è ancora leggermente incerto, stonato.

Il Gatto si struscia sul lupo e miagola piano. Un contatto semplice, rassicurante.

Il Lupo Colorato inspira profondamente e ulula di nuovo. Non è perfetto, ma è la sua voce.

SCENA 21A/21B – continua

Dagli alberi emergono i lupi grigi reali. Non sono più figure piatte o sincronizzate: sono diversi tra loro, massicci, vivi, attenti.

Il Lupo Colorato resta immobile, incerto. Poi scodinzola appena e ulula di nuovo, bene.

Il branco risponde. Non contro di lui, ma con lui. L'ululato diventa corale; il Lupo Colorato è stato riconosciuto.

SCENA 22A/22B – continua ma il sole sta pian piano sorgendo

Uno dopo l'altro, i lupi partono di corsa tra gli alberi. Il Lupo Colorato corre con loro e la camera li segue in una **carrellata dinamica**.

Il branco corre compatto tra gli alberi. La neve schizza sotto le zampe, il respiro diventa vapore.

Per la prima volta, il Lupo Colorato non si sente fuori posto fra i suoi simili.

SCENA 23/24/(26) – continua

Il branco rallenta e si ferma in una radura. Davanti a loro, un laghetto ghiacciato o uno spazio aperto, immerso nella nebbia.

All'improvviso, si sente un miagolio.

Il Lupo Colorato si volta e vede il Gatto fermo poco distante, davanti a una piccola tana scavata sotto le radici di un grande albero.

Il Gatto lo guarda, poi entra lentamente nella tana.

SCENA 25/(27) – continua

Dal branco arrivano nuovi ululati.

La camera si posiziona dietro la spalla del Lupo Colorato. Davanti a lui, da un lato, i lupi grigi che lo aspettano, dall'altro, l'ingresso buio della tana del Gatto.

Il Lupo Colorato è sospeso tra due appartenenze.

Il Lupo Colorato guarda il branco. Ulula una sola volta, un ululato di saluto.

Il branco risponde. Poi, lentamente, si allontana, dissolvendosi tra gli alberi e la foschia.

SCENA 28 – continua ma ormai il sole è sorto

Il Lupo Colorato cammina verso la tana. Si abbassa ed entra.

SCENA 30 – INT. TANA, CARRELLATA, EXT. TANA – GIORNO

L'interno della tana è caldo, silenzioso, protetto. La luce è minima, riflessa dalla neve all'esterno.

Il Lupo Colorato ed il Gatto sono rannicchiati insieme. Il lupo scodinzola appena, con gli occhi chiusi, sereno.

Il Gatto guarda un'ultima volta **in camera**, poi chiude gli occhi e si accuccia per dormire.

La camera arretra lentamente, uscendo dalla tana.

FADE TO WHITE – evocando la superficie di una pagina.

Si sentono ululati lontani, di addio e ricordo, mentre appare il titolo "Il lupo di carta".

FINE

APPENDICE E – Character sheet

In questa sezione sono riportati i character sheet dei personaggi.

Per una facilità di organizzazione e consultazione, data la complessa impaginazione, il file completo è stato caricato in una cartella Dropbox accessibile tramite il link o il QR code sotto riportati.

CHARACTER SHEET - DROPBOX



APPENDICE F – Storyboard

In questa sezione è riportato lo storyboard utilizzato come riferimento per la fase di generazione delle clip video. Le immagini sono state realizzate con Nano Banana Pro di Google, utilizzando la piattaforma Flow.

Per una facilità di organizzazione e consultazione, lo storyboard è stato raccolto all'interno di un progetto Figma, accessibile tramite il link o il QR code sotto riportati.

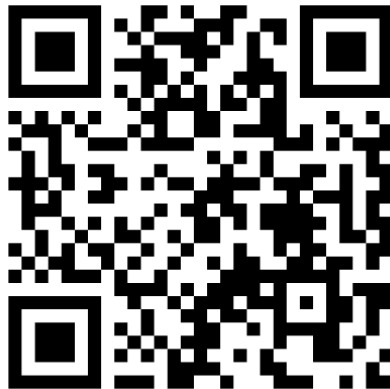
STORYBOARD - FIGMA



APPENDICE G – Cortometraggio “The Paper Wolf”

In questa sezione è riportato il link per visualizzare il cortometraggio *The Paper Wolf*, realizzato con le intelligenze artificiali generative Pika e Veo3.1 (quest'ultima utilizzata tramite la piattaforma Flow).

[IL LUPO DI CARTA - YOUTUBE](#)



BIBLIOGRAFIA E SITOGRAFIA

- [1] **Dartmouth College.** (2024). *Artificial Intelligence (AI) Coined at Dartmouth*. Dartmouth College. <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth>
- [2] **McCarthy J., Minsky M., Rochester N., Shannon C.** (1955, 31 agosto). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. <https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- [3] **Turing A. M.** (1950). *Computing Machinery and Intelligence*. *Mind*, Volume 49, Pages 433–460. <https://courses.cs.umbc.edu/471/papers/turing.pdf>
- [4] **Weizenbaum J.** (1966, 1° gennaio). *ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine*. *Communications of the ACM*, Volume 9, Issue 1, Pages 36–45. <https://dl.acm.org/doi/10.1145/365153.365168>
- [5] **Ars Technica.** (2016, 9 giugno). *Sunspring | A Sci-Fi Short Film Starring Thomas Middleditch*. Directed by: Oscar Sharp, Written by: Benjamin, Produced by: Andrew Kortschak, Walter Kortschak, Andrew Swett, Allison Friedman. YouTube: <https://www.youtube.com/watch?v=LY7x2lhqjmc>
- [6] **EDI Effetti Digitali Italiani.** (2025, 5 settembre). *Venezia: dove il cinema incontra l'intelligenza artificiale*. <https://www.effetidigitali.it/2025/09/05/venezia-dove-il-cinema-incontra-lintelligenza-artificiale/>
- [7] **Angela Watercutter.** (2025, 31 ottobre). *Guillermo del Toro Hopes He's Dead Before AI Art Goes Mainstream*. *Wired US*. <https://www.wired.com/story/guillermo-del-toro-hopes-hes-dead-before-ai-art-goes-mainstream/>
- [8] **Simone Cosimi.** (2023, 27 settembre). *Cosa prevede l'accordo per l'intelligenza artificiale dopo lo sciopero degli autori e sceneggiatori*. *Wired Italia*. <https://www.wired.it/article/sciopero-sceneggiatori-autori-usa-accordo-intelligenza-artificiale-cosa-prevede/>
- [9] **Reply.** *Reply AI Film Festival*. <https://www.reply.com/it/artificial-intelligence/reply-ai-film-festival>
- [10] **AI International Film Festival.** *AI International Film Festival*. <https://aifilmfest.org/>
- [11] **Narrazioni Parallele.** *Narrazioni Parallele*. <https://www.narrazioniparallele.org/>
- [12] **LUCAS.** *LUCAS Italia*. <https://www.lucasitalia.it/en/home-en/>
- [13] **Cinelytic.** *Cinelytic Platform*. <https://www.cinelytic.com/platform/>
- [14] **ScriptBook.** *ScriptBook*. <https://www.scriptbook.io/#!/>

- [15] **Benj Edwards (Ars Technica)**. (2024, 6 novembre). *The \$50 Million Movie Here De-Aged Tom Hanks With Generative AI*. Wired US.
<https://www.wired.com/story/here-movie-de-age-tom-hanks-generative-ai/>
- [16] **Wikipedia**. (2019). *The Irishman*. Directed by: Martin Scorsese.
[https://it.wikipedia.org/wiki/The_Irishman_\(film_2019\)](https://it.wikipedia.org/wiki/The_Irishman_(film_2019)).
- [17] **Wikipedia**. (2024). *The Brutalist*. Directed by: Brady Corbet
https://it.wikipedia.org/wiki/The_Brutalist
- [18] **Respeecher**. *Respeecher*. <https://www.respeecher.com/>
- [19] **Grassi Gabriele**. (2023/24) *Intelligenza artificiale e produzione cinematografica: analisi del processo creativo e della sua evoluzione*. Università di Pavia, UniTesi.
<https://unitesi.unipv.it/handle/20.500.14239/27703>
- [20] **Santosh Swarnakar**. (2024). *Artificial Intelligence and Cinema: Exploring the Implications of Artificial Intelligence in Cinema*.
https://www.researchgate.net/profile/Santosh-Swarnakar-2/publication/382692930_Artificial_Intelligence_and_Cinema_-_Exploring_the_Implications_of_Artificial_Intelligence_in_Cinema/links/66a9b2adde060e4c7e69fe73/Artificial-Intelligence-and-Cinema-Exploring-the-Implications-of-Artificial-Intelligence-in-Cinema.pdf
- [21] **The Last Screenwriter**. (2024). *The Last Screenwriter*. Directed by: Peter Luisi.
<https://lastscreenwriter.com/>
- [22] **Torino Film Festival**. (2024). *Miss Polly Had a Dolly*. Directed by: Pietro Lafiandra, Flavio Pizzorno, Andrea Rossini.
<https://www.torinofilmfest.org/it/41-torino-film-festival/film/miss-polly-had-a-dolly/51929/>
- [23] **Wikipedia**. (2024). *DreadClub: Vampire's Verdict*. Directed by: Hooroo Jackson.
https://en.wikipedia.org/wiki/DreadClub:_Vampire%27s_Verdict
YouTube. *DreadClub: Vampire's Verdict*. <https://www.youtube.com/watch?v=VhQPt5rdFKU>
- [24] **Critterz**. (2023). *Critterz*. Directed by: Chad Nelson. <https://www.critterz.tv/>
- [25] **FilmFreeway**. (2024). *To Dear Me*. Directed by: Gisele Tong.
<https://filmfreeway.com/ToDearMe751>
- [26] **CinemaItaliano**. (2025). *Un Rêve Liquide*. Directed by: Andrea Lommatzsch.
<https://nuovo.cinemaitaliano.info/unnreveliquid>
- [27] **AI International Film Festival**. *AI International Film Festival*. <https://aifilmfest.org/>
- [28] **Nenet**. *AI Motion*. <https://www.nenet.it/ai-motion/>
- [29] **Claudia López Frías**. (2024, 3 agosto). *The Paradox of Artificial Intelligence in Cinema*.
<https://ojs.nmdjournals.org/index.php/cdig/article/view/99>
- [30] **Wikipedia**. *Marshall McLuhan*. https://it.wikipedia.org/wiki/Marshall_McLuhan

- [31] **Wikipedia**. *Lev Manovich*. https://it.wikipedia.org/wiki/Lev_Manovich
- [32] **Unione Europea**. (2024, 13 giugno). *Regolamento (UE) 2024/1689 – AI Act*. *EUR-Lex*. <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A32024R1689>
- [33] **Unione Europea**. (2019, 17 aprile). *Direttiva (UE) 2019/790 sul diritto d'autore nel mercato unico digitale*. *EUR-Lex*. <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A32019L0790>
- [34] **Interlex**. (1971, 24 luglio). *Convenzione di Berna per la protezione delle opere letterarie e artistiche*. <https://www.interlex.it/testi/convberna.htm>
- [35] **Moshe Glickman, Tali Sharot**. (2024, 18 dicembre) *How human–AI feedback loops alter human perceptual, emotional and social judgements*. *Nat Hum Behav*, Pages 345-359. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11860214/>
- [36] **Wikipedia**. *Masahiro Mori*. [https://en.wikipedia.org/wiki/Masahiro_Mori_\(roboticist\)](https://en.wikipedia.org/wiki/Masahiro_Mori_(roboticist))
- [37] **Bernard Marr**. (2024, 7 febbraio). *The Uncanny Valley: Advancements And Anxieties Of AI That Mimics Life*. *Forbes*, innovation, enterprise tech. <https://www.forbes.com/sites/bernardmarr/2024/02/07/the-uncanny-valley-advancements-and-anxieties-of-ai-that-mimics-life/>
- [38] **Zeng Y. et al.** (2026, marzo 2). *Too human to model: the uncanny valley of large language models in simulating human systems*. *Npj Complexity*. Article numbr 13. <https://www.nature.com/articles/s44260-026-00075-1>
- [39] **ComfyUI**. *ComfyUI*. <https://www.comfy.org/>
- [40] **Stanford University**. *The 2025 AI Index Report*. <https://hai.stanford.edu/ai-index/2025-ai-index-report>
- [41] **Artificial Analysis**. *Artificial Analysis Arena*. <https://artificialanalysis.ai/>
- [42] **BBC Bitesize**. (2026, febbraio). *How Will Smith eating spaghetti became the ‘test’ of AI video*. <https://www.bbc.co.uk/bitesize/articles/z4rsmbk>
- [43] **Vaswani, A. et al.** (2017, 12 giugno). *Attention Is All You Need*. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- [44] **Mazur L.** (2025, 23 gennaio). *Writing: LLM Creative Story-Writing Benchmark V4*. GitHub. <https://github.com/lechmazur/writing>
- [45] **LMSYS Org.** (2024). *Chatbot Arena Leaderboard*. Hugging Face Spaces. <https://huggingface.co/spaces/lmarena-ai/arena-leaderboard>
- [46] **Huang Y. et al.** (2023, 22 novembre). *VBench: A Comprehensive Benchmark for Video Generation*. GitHub. <https://vchitect.github.io/VBench-project/>

- [47] **ApiFrame.** (2024). *AI Image, Video & Music Model Rankings*. <https://apiframe.ai/rankings>
- [48] **Grötschla F. et al.** (2025, 23 giugno). *Music Generation Audio Metrics: Reference-free evaluation and subjective preference*. ICASSP 2025.
https://tik-old.ee.ethz.ch/file/e2394c9aba4e90f07e4f1bd8eedc823e/Music_Generation_Audio_Metrics_ICASSP_2025.pdf
- [49] **Recut.** (2024). *Recut: Automatic Video Editing Software*. <https://getrecut.com/>
- [50] **Mentimeter.** *Mentimeter*. <https://www.mentimeter.com/>
- [51] **OpenAI.** *ChatGPT*. <https://chatgpt.com/>
- [52] **Anthropic.** *Claude*. <https://claude.ai/>
- [53] **Google.** *Gemini*. <https://gemini.google.com/app>
- [54] **Wikipedia contributors.** *Christopher Vogler*. https://it.wikipedia.org/wiki/Christopher_Vogler
- [55] **Wikipedia contributors.** *Struttura restaurativa in tre atti*.
https://it.wikipedia.org/wiki/Struttura_restaurativa_in_tre_atti
- [56] **Wikipedia contributors.** *Hayao Miyazaki*. https://it.wikipedia.org/wiki/Hayao_Miyazaki
- [57] **Wikipedia contributors.** *Studio Ghibli*. https://it.wikipedia.org/wiki/Studio_Ghibli
- [58] **LAIKA Studios.** *LAIKA Official Website*. <https://www.laika.com/>
- [59] **Wikipedia contributors.** *Coraline e la porta magica*. (2009). Directed by: Henry Selick. Produced by: Laika Entertainment.
https://it.wikipedia.org/wiki/Coraline_e_la_porta_magica
- [60] **Notion Labs Inc.** *Notion*. <https://www.notion.com/>
- [61] **Wikipedia contributors.** *Luis Sepúlveda*.
https://it.wikipedia.org/wiki/Luis_Sep%C3%BAlveda
- [62] **Wikipedia contributors.** *Beatrix Potter*. https://it.wikipedia.org/wiki/Beatrix_Potter
- [63] **Wikipedia contributors.** *La storia di Peter Coniglio*. (1902). Written by: Beatrix Potter.
https://it.wikipedia.org/wiki/La_storia_di_Peter_Coniglio
- [64] **Wikipedia contributors.** *Alice nel Paese delle Meraviglie (film 1951)*. Directed by: Clyde Geronimi, Hamilton Luske, Wilfred Jackson. Produced by: Walt Disney Productions.
[https://it.wikipedia.org/wiki/Alice_nel_Paese_delle_Meraviglie_\(film_1951\)](https://it.wikipedia.org/wiki/Alice_nel_Paese_delle_Meraviglie_(film_1951))
- [65] **Letterboxd.** *The Ones Who Never See The Moonlight*. (2024). Directed by Solène Marché et al.
<https://letterboxd.com/film/the-ones-who-never-see-the-moonlight/>
- [66] **Blender Foundation.** *Blender Demo Files*. <https://www.blender.org/download/demo-files/>
- [67] **Blender Studio.** *Dogwalk Project*. <https://studio.blender.org/projects/dogwalk/>

- [68] **Wikipedia contributors.** *Flow – Un mondo da salvare.* (2024). Directed by: Gints Zilbalodis. [https://it.wikipedia.org/wiki/Flow - Un mondo da salvare](https://it.wikipedia.org/wiki/Flow_-_Un_mondo_da_salvare)
- [69] **Wikipedia contributors.** *Vita di Pi (film).* (2012). Directed by: Ang Lee. [https://it.wikipedia.org/wiki/Vita di Pi \(film\)](https://it.wikipedia.org/wiki/Vita_di_Pi_(film))
- [70] **OpenAI.** *Sora.* <https://sora.chatgpt.com/explore>
- [71] **Google Labs.** *Flow.* <https://labs.google/flow/about>
- [72] **Midjourney, Inc.** *Midjourney.* <https://www.midjourney.com/home>
- [73] **Alexandra Sierra Rativa, Marie Postma, Menno van Zaanen** (2022, 16 marzo). *The uncanny valley of a virtual animal.* Wiley the online library. Volume 33, Issue 2. <https://onlinelibrary.wiley.com/doi/full/10.1002/cav.2043>
- [74] **V. Schwind. K. Leicht. S. Jäger. K. Wolf c. N. Henze** (2018, marzo). *Is there an uncanny valley of virtual animals? A quantitative and qualitative investigation* International Journal of Human-Computer Studies. Volume 111. Pages49-61 <https://www.sciencedirect.com/science/article/abs/pii/S1071581917301593>
- [75] **Josh Terry.** (2020, 21 febbraio). *The biggest problem with ‘Call of the Wild’ is the dog.* Desert News. <https://www.deseret.com/entertainment/2020/2/21/21146354/call-of-the-wild-cgi-dog-buck-harrison-ford/>
- [76] **Joshua Rivera.** (2019, 11 luglio). *The Lion King Review: Out of the Uncanny Valley.* GQ Culture. <https://www.gq.com/story/the-lion-king-review>
- [77] **The GIMP Team.** *GIMP – GNU Image Manipulation Program.* <https://www.gimp.org/>
- [78] **Runway.** *RunwayML.* <https://runwayml.com/>
- [79] **Kling AI.** *Kling AI Global.* <https://klingai.com/global/>
- [80] **Pika Labs.** *Pika.* <https://pika.art/>
- [81] **Google.** *Veo – Video Generation.* <https://gemini.google/it/overview/video-generation/?hl=it>
- [82] **Suno.** *Suno AI.* <https://suno.com/home>
- [83] **Loudly.** *Loudly AI Music Generator.* <https://www.loudly.com/>
- [84] **ElevenLabs.** *ElevenLabs.* <https://elevenlabs.io/>
- [85] **Wikipedia.** *Luca Ward.* [https://it.wikipedia.org/wiki/Luca Ward](https://it.wikipedia.org/wiki/Luca_Ward)
- [86] **Blackmagic Design.** *DaVinci Resolve.* <https://www.blackmagicdesign.com/it/products/davinciresolve>
- [87] **Blackmagic Design.** (2025, 4 aprile). *DaVinci AI Neural Engine – What’s New.* <https://www.blackmagicdesign.com/it/media/release/20250404-02>

Ulteriori contributi

Lev Manovich e Emanuele Arielli. (2021/2024). *Artificial Aesthetics: Generative AI, Art and Visual Media*. <https://manovich.net/index.php/projects/artificial-aesthetics>

Gazzetta Ufficiale della Repubblica Italiana. (2025, 25 settembre). *Legge italiana sul diritto d'autore e intelligenza artificiale (25G00143)*.

<https://www.gazzettaufficiale.it/eli/id/2025/09/25/25G00143/sg>

Guidetti Matteo. (2024, dicembre). *Intelligenza artificiale e cinema digitale: tecniche e integrazione nel workflow, dalla pre-produzione alla postproduzione*. Politecnico di Torino, WebThesis.

<https://webthesis.biblio.polito.it/secure/33934/1/tesi.pdf>

Azzarelli A. et al (2024). *Intelligent Cinematography: A review of AI research for cinematographic production*. arXiv. <https://arxiv.org/pdf/2405.05039>

Vittorio A. Dublino. (2025). *Quando le macchine imparano a generare il bello. Intelligenza artificiale e immaginario estetico*.

https://vittoriodublinoblog.org/wp-content/uploads/2025/12/intelligenza-artificiale-e-immaginario-estetico_fin.pdf

Ketan Totlani. (2023). *The Evolution of Generative AI: Implications for the Media and Film Industry* <https://www.ijfmr.com/papers/2023/5/7537.pdf>

Mattia Morgut. (2025, luglio). *IA e Arte: narrare al pubblico l'impatto emotivo delle performance musicali attraverso video generati con strumenti di intelligenza artificiale generativa*. Politecnico di Torino, WebThesis. <https://webthesis.biblio.polito.it/36335/1/tesi.pdf>

Simone Botosso. (2025, luglio). *IA e Arte Narrare al pubblico l'impatto emotivo delle performance musicali attraverso video generati con strumenti di intelligenza artificiale generativa*. Politecnico di Torino, WebThesis. <https://webthesis.biblio.polito.it/36331/1/tesi.pdf>

Natalia Mesa. (2023, settembre 29). *The uncanny valley, explained: Why you might find AI creepy*. National Geographic. <https://www.nationalgeographic.com/science/article/ai-uncanny-valley>

Robotics24. *Uncanny Valley*. <https://robotics24.net/blog/glossary/uncanny-valley/>

De Masi, V., Li, S., & Pessoa Serafim, F. (2025, 8 settembre). *Revolutionizing Cinema. How AI is Democratizing Production and Deepening Creative Content*.

<https://riviste.unimi.it/index.php/conessioniremote/article/view/28718>

Commissione europea. (2024). *Legge sull'IA*.

<https://digital-strategy.ec.europa.eu/it/policies/regulatory-framework-ai>

StudioBinder. *The world's leading production software*. <https://www.studiobinder.com/>

RINGRAZIAMENTI

Quest'ultima pagina di questo percorso è dedicata a ringraziare tutte le persone che hanno reso possibile, e più leggero, ogni passo di questa lunga strada.

Il primo grazie, il più grande, va ai miei genitori. Per aver creduto nella mia scelta anche quando io per prima ne dubitavo e per avermi supportata anche quando avevo perso le speranze.

Un grazie va anche a zii e cugini, che hanno sempre seguito il mio percorso... anche quando spiegare cosa stessi studiando richiedeva più tempo della tesi stessa!

Vorrei ringraziare anche la Professoressa Mazali ed il Professor Bottino per l'opportunità del progetto e per tutto quello che ho imparato nelle loro aule, e Claudia e Roberto di Narrazioni Parallele, così come tutto il team a supporto, per avermi accompagnata durante questi mesi e arricchito il mio percorso con nuove scoperte.

Un pensiero speciale va ai compagni del Politecnico, con cui ho condiviso ansie pre-esame, soddisfazioni inaspettate, ripetuti tentativi per passare “quegli” esami impossibili e le notti svegli sui progetti con le consegne all'una di notte quando il termine erano le famose 23:59.

E non posso non citare gli amici di A.I. Caramba (o I.A I.A OH o come vogliamo chiamarci), che mi hanno dimostrato come buttarsi in progetti strani sia *sempre* una buona idea.

Agli amici di lunga data, Giulia e Mattia: grazie per avermi supportata (e sopportata) durante i periodi di esami, per le videochiamate con Minecraft nei momenti giusti e le nostre tranquille serate fuori.

Un ringraziamento anche ai Torsanlorenzesi, che ogni estate, da sempre, mi fanno staccare un po' anche quando ho Chimica e Metodi Matematici da preparare.

Infine, un posto speciale è riservato ai miei fratelli e sorelle a quattro zampe, che negli anni mi hanno insegnato e dato tanto: Micia, Dixi e Pixi, Trilli, Oliver e Angel, Vie.

