



**Politecnico
di Torino**

POLITECNICO DI TORINO

Corso di Laurea Magistrale

in **INGEGNERIA DEL CINEMA E DEI MEZZI DI
COMUNICAZIONE**

a.a. 2023/2024

Sessione Dicembre 2024

Tesi di Laurea Magistrale

**Intelligenza artificiale e cinema digitale: tecniche e
integrazione nel workflow, dalla pre-produzione alla post-
produzione**

Relatore:

Prof.ssa Mazali Tatiana

Candidato:

Guidetti Matteo

Indice

ABSTRACT	1
INTRODUZIONE	3
1.1 INTELLIGENZA ARTIFICIALE: CHE COS'È?	3
1.1.1 <i>Intelligenza artificiale, Machine Learning e Deep Learning</i>	5
1.1.2 <i>Intelligenza artificiale Generativa e Non Generativa</i>	7
1.2 IA NELL'AUDIOVISIVO: LO STATO DELL'ARTE	10
1.2.1 <i>IA Generativa nel Cinema</i>	10
1.2.2 <i>IA Non Generativa nel Cinema</i>	25
1.3 ANNUNCI IMMINENTI E FUTURO	28
IPOTESI	31
2.1 OPPORTUNITÀ DA SFRUTTARE	31
2.2 LIMITI DA RICONOSCERE	35
2.3 VERSO UN NUOVO MODELLO APPLICATIVO	37
APPLICAZIONE PRATICA	39
3.1 STRUMENTI UTILIZZATI	43
3.2 PRE-PRODUZIONE	45
3.2.1 <i>Studio del campo di ricerca</i>	45
3.2.2 <i>Domande interviste</i>	46
3.3 PRODUZIONE E RIPRESE (CENNI).....	47
3.4 POST-PRODUZIONE	48
3.4.1 <i>Scrittura</i>	50
3.4.2 <i>Montaggio Video</i>	71
3.4.3 <i>Montaggio Audio (cenni)</i>	74
3.4.4 <i>Differenze e analogie con il metodo tradizionale</i>	76
CONCLUSIONI.....	79
4.1 RISULTATI	79
4.2 ANALISI	80
4.2.1 <i>Vantaggi ottenuti</i>	80
4.2.2 <i>Limiti riscontrati</i>	82
4.2.3 <i>Possibili migliorie e integrazioni</i>	84
4.3 PROSPETTIVE E ASPETTATIVE FUTURE	85
RINGRAZIAMENTI	87
BIBLIOGRAFIA	89

Elenco Figure

FIGURA 1 - RAPPRESENTAZIONE GERARCHICA DI INTELLIGENZA ARTIFICIALE, MACHINE LEARNING E DEEP LEARNING	5
FIGURA 2 - ESEMPIO GRAFICO DI SUPERVISED LEARNING [8]	6
FIGURA 3 - ESEMPIO DI UNSUPERVISED LEARNING [8]	6
FIGURA 4 - SEMPLICE VISUALIZZAZIONE DELLE RETI GAN [11]	7
FIGURA 5 - PROCESSO DI DENOISING SCHEMATIZZATO [12]	8
FIGURA 6 - RAPPRESENTAZIONE DI UNA SEQUENZA DI CNN PER CLASSIFICARE LA SCRITTURA A MANO [17]	9
FIGURA 7 - INTERFACCIA DELLA WEB APP DI MIDJOURNEY	11
FIGURA 8 - ESEMPIO DI CHARACTER CONSISTENCY MIDJOURNEY [21]	12
FIGURA 9 - INTERFACCIA DI DALL-E 3	13
FIGURA 10 - UTILIZZO DELLO STRUMENTO SELEZIONE DI DALL-E 3	13
FIGURA 11 - RISULTATO OTTENUTO DALL'UTILIZZO DELLA SELEZIONE SU DALL-E 3	14
FIGURA 12 - INTERFACCIA DI PROVA DELLA API DI STABLE DIFFUSION 3.5 LARGE TURBO	15
FIGURA 13 - INTERFACCIA DEL TOOL LEONARDO AI	16
FIGURA 14 - INTERFACCIA GPT LEONARDO PROMT MAKER	16
FIGURA 15 - INTERFACCIA DI RUNWAY GEN-2	17
FIGURA 16 - UNA PORZIONE DELLA DASHBOARD DI RUNWAY	19
FIGURA 17 - INTERFACCIA DI KLINGAI	19
FIGURA 18 - INTERFACCIA WEB DI STABLE VIDEO DIFFUSION	20
FIGURA 19 - INTERFACCIA DI STABLE AUDIO	21
FIGURA 20 - INTERFACCIA DI UTILIZZO DI ELEVEN LABS	22
FIGURA 21 - BENCHMARK DELLA VERSIONE 4.0 DI CHATGPT [37]	23
FIGURA 22 - TABELLA DI ANALISI TRA I PRINCIPALI CONCORRENTI EVIDENZIATI DA ANTHROPIC [40]	24
FIGURA 23 - INTERFACCIA DI UTILIZZO DI CLAUDE 3.5 SONNET	24
FIGURA 24 - INTERFACCIA DI DAVINCI RESOLVE, TAB EDIT	26
FIGURA 25 - INTERFACCIA DI DAVINCI RESOLVE, TAB FAIRLIGHT	27
FIGURA 26 - FRAME DI UNO DEI CORTOMETRAGGI GENERATI CON SORA [44]	28
FIGURA 27 - FRAME DI VIDEO DI ESEMPIO DI UTILIZZO DI GOOGLE VEO	29
FIGURA 28 - ESEMPIO DI INTEGRAZIONE DI UN EFFETTO CON FIREFLY SU AFTER EFFECTS	30
FIGURA 29 - SCHERMATA DI CONFIGURAZIONE DI UN GPT PERSONALIZZATO	32
FIGURA 30 - ESEMPIO DI UTILIZZO DI CANVAS SU CHATGPT	32
FIGURA 31 - SCHERMATA DI VISUALIZZAZIONE DEI FILM IN AI	33
FIGURA 32 - INTERFACCIA DI FYLM.AI	34
FIGURA 33 - ESEMPIO DI PROMPT COSTRUITO PER GEN-3	36
FIGURA 34 - LOGO DEL PROGETTO EUROPE	39
FIGURA 35 - LOGO DI EAST OF MOON	40
FIGURA 36 - LOGO DI ASSOCIAÇÃO POIOMAR	40
FIGURA 37 - LOGO DI CONSULTA EUROPA	41
FIGURA 38 - LOGO DI ABSENTIA	41
FIGURA 39 - LOGO DI OPEN IMPACT	41
FIGURA 40 - LOGO DI UNO	42
FIGURA 41 - LOGO DI CHATGPT	43
FIGURA 42 - LOGO DI DAVINCI RESOLVE STUDIO	43
FIGURA 43 - LOGO DI CLAUDE AI	44
FIGURA 44 - LOGO DI ELEVENLABS	44
FIGURA 45 - LOGO DI RUNWAY	44

FIGURA 46 - LOGO DI STABLE AUDIO	44
FIGURA 47 - ESEMPIO DI ALCUNE DOMANDE GENERALI	46
FIGURA 48 - ESEMPIO DOMANDE OTTENUTE PER IL TEMA DELLO SPOPOLAMENTO	47
FIGURA 49 - ESEMPIO FILE DI DOMANDE PER UN INTERVISTATO	48
FIGURA 50 - TOOL DI AUDIO TRANSCRIPTION SU DAVINCI RESOLVE	51
FIGURA 51 - OUTPUT DELLA FUNZIONE DI AUDIO TRANSCRIPTIO SENZA SPEAKER DETECTION.....	53
FIGURA 52 - OUTPUT DELLA FUNZIONE DI AUDIO TRANSCRIPTION CON SPEAKER DETECTION.....	53
FIGURA 53 - ESEMPIO DI FILE ESPORTATO DALLA TRASCRIZIONE DI DAVINCI RESOLVE	54
FIGURA 54 - ESEMPIO DI TRASCRIZIONE RIORDINATA CON MATTIA INTERVISTATORE E GIOVANNI INTERVISTATO	55
FIGURA 55 - ESEMPIO DI TRASCRIZIONE CORRETTA CON LE CORREZIONI TRA PARENTESI QUADRE.....	57
FIGURA 56 - ESEMPIO DI DUE PAGINE DELL'ANALISI ALL'INTERVISTATA MAGDALENA.....	58
FIGURA 57 - SCELTA DELLA PERSONA DA INSERIRE IN SARDEGNA E MOTIVAZIONE	62
FIGURA 58 - TABELLA CON LE DEFINIZIONI DEI CINQUE DOCUMENTI E RELATIVE INFORMAZIONI SCRITTE CON CHATGPT	63
FIGURA 59 - TABELLA NON COMPILATA PER LA STESURA DELLA LOGLINE.....	64
FIGURA 60 - ESEMPIO DI FLUSSO DI LAVORO PER MONTAGGIO CON TRASCRIZIONE.....	71
FIGURA 61 - ESEMPIO DI DESCRIZIONE VISIVA DELLE INQUADRATURE NELLA SCENEGGIATURA	72
FIGURA 62 - INQUADRATURA DELLA STORIA DEL PESCATORE REALIZZATA CON DALL-E	73
FIGURA 63 - UTILIZZO DEL MOTION BRUSH IN RUNWAY GEN-2	74
FIGURA 64- ESEMPIO DI BRANO DESCRITTO DA CHATGPT	75
FIGURA 65 - IMPOSTAZIONI DELLA VOCE SU ELEVEN LABS.....	83

Abstract

Con l'espansione nei settori creativi dell'*Intelligenza Artificiale*, il cinema digitale si trova di fronte alla necessità di confrontarsi con l'utilizzo dei nuovi strumenti. Benché l'utilizzo di applicazioni di *Intelligenza Artificiale* programmate *ad hoc* sia già integrato in alcune grandi produzioni hollywoodiane, la diffusione di massa di strumenti più generici ha come conseguenza una democratizzazione dell'accesso alla produzione cinematografica, contribuendo ad appiattire il divario tra professionista e *prosumer*. Inoltre, i nuovi software di *Intelligenza Artificiale* offrono nuove possibilità nei modelli applicati al settore cinematografico, implicando non solo un ribilanciamento delle risorse, ma anche una rivisitazione dei workflow tradizionali.

È in un contesto simile che si colloca questo lavoro di tesi, che ha come obiettivo quello di esplorare le nuove tipologie di workflow e le possibilità che queste offrono a livello creativo e di utilizzo delle risorse. Più nello specifico, l'indagine esplorerà come l'*Intelligenza Artificiale* possa essere integrata nel flusso produttivo di un documentario a basso budget, mostrando le possibilità sia economiche che creative che si presentano.

A livello sperimentale, il flusso di lavoro verrà applicato alla realizzazione del documentario *Forgotten Frontiers: where isolation meets innovation*, considerato un caso studio esemplare per la dimostrazione della tesi. Si esploreranno le possibilità creative date dall'utilizzo combinato di *Intelligenza Artificiale* e un archivio di documentazioni della realtà, dal quale estrarre diverse storie possibili, scegliendo a posteriori quella più adatta. Lo studio indagherà, altresì, come l'*Intelligenza Artificiale* possa offrire soluzioni per colmare le limitazioni di risorse in una produzione.

Capitolo 1

Introduzione

La pervasività della tecnologia è un fenomeno storicamente dibattuto, dagli studi condotti da Marshall McLuhan sull'influenza che la tecnologia ha nella percezione della realtà¹, fino ad arrivare ad *Apocalittici e Integrati* di echiana memoria². L'intelligenza artificiale non si esime da questo contesto di polarizzazione mediatica, specie nell'era dei *Social Network*. La complessità dell'argomento, unita alla disinformazione, hanno contribuito ad alimentare la primordiale paura del genere umano che qualcuno o, in questo caso, *qualcosa* possa sostituirlo.

Il rapporto tra Cinema Digitale e intelligenza artificiale non è escluso da queste novità. L'abilità di questa di generare prodotti audiovisivi a partire da una semplice stringa di testo è un'opportunità concreta, che anche l'industria cinematografica mainstream di Hollywood sta cogliendo. Nel settembre del 2024 James Cameron, non di certo l'ultimo arrivato nel campo delle tecniche innovative nel cinema, entra a far parte del Consiglio di amministrazione di *Stability AI*, uno dei più importanti sviluppatori di strumenti di *IA Generativa* [1].

Il rapido avanzare di questo cambiamento, da molti definito come *Quarta Rivoluzione Industriale* [2], rende necessario un adattamento che passa soprattutto attraverso la scoperta di questi nuovi strumenti, nell'ottica di comprenderli e maneggiarli adeguatamente.

1.1 Intelligenza artificiale: che cos'è?

La risposta a questa domanda è sicuramente non banale, dato che, anche presi singolarmente, i termini *Intelligenza* e *Artificiale* necessiterebbero di interi dibattiti per essere definiti esaustivamente. Il modo migliore per chiarire il termine è quello di riportare le parole di colui che viene considerato il padre dell'*Intelligenza Artificiale*.

¹ McLuhan, M. (1964). "Understanding Media: The Extensions of Man". New York: McGraw-Hill

² Eco, U. (1964). "Apocalittici e integrati. Comunicazioni di massa e teorie della cultura di massa." Milano: Bompiani

“It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.” - John McCarthy ³

Nell'articolo *“What is AI”* del 2005, John McCarthy spiega in questo modo il termine *Intelligenza Artificiale*.

McCarthy è un informatico statunitense che ha contribuito alla fioritura e allo sviluppo dell'informatica per come è oggi, in particolar modo nel campo dell'*Intelligenza Artificiale*. La paternità di questo termine gli è data come conseguenza dell'organizzazione del *Dartmouth Summer Research Project on Artificial Intelligence* nel 1956, laddove l'obiettivo del workshop è quello di studiare le modalità con cui le macchine potessero simulare aspetti propri dell'intelligenza umana [3].

La conferenza di Dartmouth, sebbene abbia dato vita al termine in uso tuttora, non viene considerata come la prima teorizzazione del concetto di *Intelligenza Artificiale*. Già in precedenza, molti teorici informatici hanno mostrato o teorizzato alcune delle caratteristiche dell'*Intelligenza Artificiale* applicate alle macchine dell'epoca. Le teorie più note sono quelle di Alan Turing che, nell'articolo *“Computer Machinery and Intelligence”*, risponde alla domanda *“Can machines think?”* ⁴ ipotizzando macchine che simulino l'intelligenza umana, con caratteristiche appartenenti all'*Intelligenza Artificiale* di oggi.

Successivamente alla conferenza di Dartmouth, nel corso degli anni le scoperte in ambito di *Intelligenza Artificiale* iniziano ad intensificarsi. Nel 1966, Joseph Weizenbaum sviluppa ELIZA un programma che ha come intento quello di simulare una conversazione in linguaggio naturale con un computer. Più semplicemente, si può considerare ELIZA di Weizenbaum come il primo *chatbot* della storia [4].

Parallelamente alle nuove scoperte nell'ambito, si sviluppano ulteriori definizioni che provano ad approfondire lo studio dell'*Intelligenza Artificiale*. John Searle, nel 1980, distingue due tipi di *Intelligenza Artificiale*: *LA debole* e *LA Forte*. La prima si tratta di un semplice strumento in grado di fornire un aiuto nella risoluzione dei problemi tramite il ragionamento umano. La seconda, più complessa, considera il computer come un cervello a sé stante, in grado di interpretare, anche cognitivamente, gli stimoli [5].

Con il passare degli anni l'*Intelligenza Artificiale* diventa un ambito esplorato anche nei videogiochi, con lo sviluppo di macchine in grado di giocare a livelli professionali: è il caso di *Deep Blue* e *AlphaGO* in grado di sconfiggere i campioni mondiali, rispettivamente di Scacchi e di GO.

³ “What is Artificial Intelligence”, 2005, John McCarthy.

⁴ A. M. Turing, i.- *Computing Machinery and Intelligence, Mind*, Volume LIX, Issue 236, October 1950, Pages 433–460

1.1.1 Intelligenza artificiale, Machine Learning e Deep Learning

Per districare la matassa dello studio dell'*Intelligenza Artificiale* è importante saper riconoscere e distinguere le varie terminologie che popolano questa fitta branca dell'informatica.

Intelligenza Artificiale, *Machine Learning* e *Deep Learning* sono termini fondamentali in questo campo di studi; comprendere le differenze e le relazioni che intercorrono tra questi tre punti di vista, dello stesso argomento, aiuta a muoversi meglio anche nell'utilizzo degli strumenti di cui sono alla base.

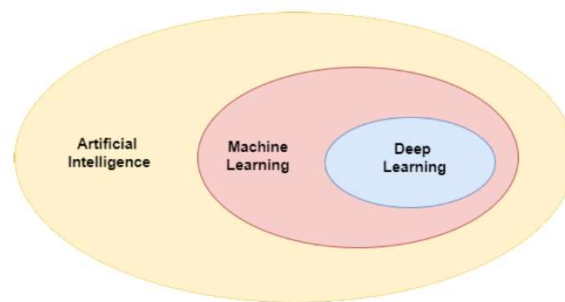


Figura 1 - Rappresentazione gerarchica di Intelligenza artificiale, Machine Learning e Deep Learning

L'*Intelligenza Artificiale* è, dei tre termini, quello più ampio e si riferisce al campo dell'informatica che ha come obiettivo quello di programmare macchine con una capacità di ragionamento e di *decision making* equiparabili a quelle umane. È possibile definirla come il tentativo di emulare il cervello umano e il suo funzionamento. Essendo un così ampio termine, si riferisce a molte funzioni centrali, come il ragionamento, la rappresentazione della conoscenza, la percezione, la comunicazione, oltre che a svariati strumenti e metodologie [6] [7].

Il *Machine Learning*, un sottoinsieme dell'Intelligenza Artificiale, è un campo dell'informatica che ha come obiettivo, come dice il termine stesso, di far imparare le macchine da grandi dataset, senza una programmazione esplicita. La realizzazione di questo scopo passa attraverso degli algoritmi, che permettono alle macchine non solo di imparare tramite un dataset, ma anche di migliorare con il tempo. Questo significa produrre algoritmi che permettano alla macchina di analizzare pattern presenti nei dati e generare dei modelli per compiti specifici, ottenendo un comportamento intelligente correlato alla capacità di fare previsioni accurate. Ciò detto, esistono tre grandi categorie in cui è possibile classificare il *Machine Learning*: *Supervised Learning*, *Unsupervised Learning* e *Reinforcement Learning* [7] [8].

Il primo tipo di *Machine Learning* si basa su un dataset supervisionato, ovvero correlato da una *label*. Più chiaramente, algoritmi che implementano il *Supervised Learning* prendono in input un dataset e una *label*, come il nome che descrive i dati. Ad esempio, se il dataset contiene delle mele, la *label* associata sarà "Mele". In questo modo, l'algoritmo sarà in grado di riconoscere l'oggetto *Mela* e associarlo al nome.

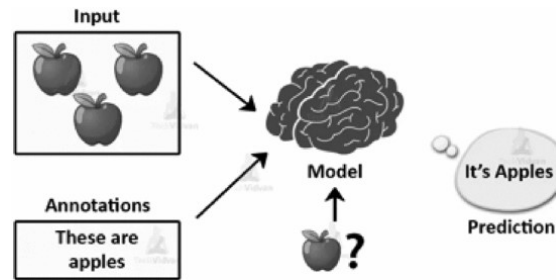


Figura 2 - Esempio grafico di *Supervised Learning* [8]

L'algoritmo *Unsupervised Learning* lavora con un dataset non supervisionato, che significa senza alcuna label in input. In questo caso, l'obiettivo dell'algoritmo è quello di riconoscere pattern all'interno di dataset, accomunando dati con caratteristiche simili e imparando a distinguerli, senza riconoscere di che di dato tipo si tratti. In un esempio [8], si utilizza come dataset di input un insieme di gatti e cani. L'algoritmo di *Unsupervised Learning* sarà in grado di identificare e ordinare gatti e cani in base alle caratteristiche di ognuno, ma senza saper riconoscere quali sono i gatti e quali sono i cani.

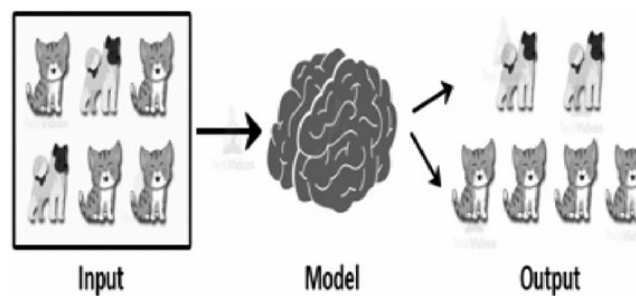


Figura 3 - Esempio di *Unsupervised Learning* [8]

Infine, il *Reinforcement Learning* si distingue maggiormente dai due algoritmi sopra citati. Si basa su un tipo di apprendimento che prevede dei premi: otterrà un premio positivo in caso di un buon lavoro di apprendimento, negativo nel caso opposto. Questo tipo di algoritmo ha come scopo quello di ottenere il miglior risultato possibile [8].

Il *Deep Learning* [6] appartiene alla famiglia delle *Artificial Neural Network (ANNs)*, complesse rappresentazioni matematiche e algoritmi interconnessi che emulano il funzionamento delle connessioni neurali dell'encefalo. Più nello specifico, le *ANNs* sono composte da elementi, chiamati neuroni, che hanno il compito di elaborare le informazioni che ricevono in input. Il processo di funzionamento di questo tipo di reti passa, quindi, attraverso la ricezione di un input, come numeri, immagini o testi. Vengono poi calcolati i pesi delle connessioni neurali, prima assegnati casualmente e poi calcolati in maniera più precisa tramite le reiterazioni del processo. I pesi vengono successivamente sommati e utilizzati per calcolare l'output tramite una funzione di trasferimento. Associato al calcolo dei pesi, vi è il calcolo di un errore, la cui minimizzazione è l'obiettivo delle reiterazioni dell'algoritmo. Questo processo viene chiamato *Backpropagation* [9].

Per rappresentare connessioni neurali più complesse, si rende necessario l'utilizzo di *Reti Neurali Profonde*, o *Deep Learning*, che utilizzano operazioni matematiche più complesse, come la convoluzione, per estrarre le informazioni necessarie da dati più grezzi ricevuti in input. Gli algoritmi di *Deep Learning* trovano il loro principale utilizzo nei casi in cui si hanno in input dataset di grandi dimensioni, riuscendo a superare le performance del *Machine learning* specialmente nella generazione di testo, immagini, video e audio.

1.1.2 Intelligenza artificiale Generativa e Non Generativa

Compresi questi concetti generali è utile acquisire ancora una semplice ma fondamentale distinzione di vocabolario che permette di osservare meglio gli strumenti di *Intelligenza Artificiale* applicati al cinema. Con il termine *Intelligenza Artificiale* ci si riferisce, da qui in avanti, in maniera gergale, a tutti gli strumenti descritti prima, in particolar modo alle *Reti Neurali* e si distingue in *Generativa* e *Non Generativa*.

La prima, quella *Generativa*, racchiude in sé tutti gli strumenti che hanno come output la generazione di qualcosa che non esisteva in precedenza, come immagini, video, testo o audio [10].

Tra i modelli in uso più importanti per l'*IA Generativa* è importante citare il modello *GAN* o *Generative Adversarial Network*. Questo modello è composto di due *Reti Neurali* chiamate una *Generatore* e l'altra *Discriminatore*. Il funzionamento si basa sulla creazione di una competizione tra le due reti: il *Generatore* riceve in input dei dati casuali e produce in output dei campioni generati, un'immagine ad esempio. Il *Discriminatore*, invece, deve distinguere le immagini generate da quelle reali. Si ingaggia in questo modo una sfida, il termine *Adversarial* deriva da questo, tra *Generatore* e *Discriminatore* in cui il primo ha l'obiettivo di migliorare la propria produzione di dati, in modo tale da "ingannare" il *Discriminatore*, il secondo viceversa [11].

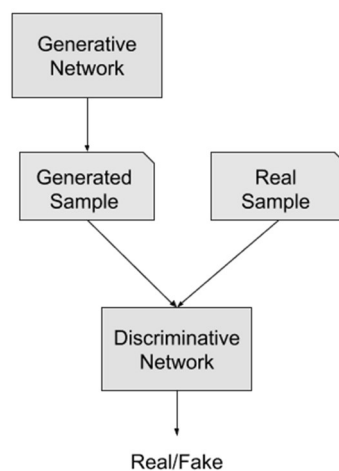


Figura 4 - Semplice visualizzazione delle reti GAN [11]

Un'evoluzione dei modelli di *Reti Generative Avversarie* sono i *Diffusion Model*, utilizzati nella maggior parte dei tool di *IA Generativa* attualmente sul mercato. Questi sono modelli generativi che sintetizzano i dati basandosi su un processo di *denoising probabilistico*. L'output ottenuto è dettagliato e realistico e proviene da una manipolazione graduale della distribuzione del rumore. In altre parole, l'addestramento di questi modelli avviene attraverso due processi consecutivi fondamentali: il *forward process* e il *reverse process*. Il primo consiste nell'applicare gradualmente del rumore gaussiano ai dati in input, ad esempio delle immagini, fino a farli diventare rumore puro. È un processo che può essere definito come una degradazione controllata dei dati. A questo punto, inizia il secondo processo, quello di *denoising*. Partendo dal rumore puro ottenuto, l'algoritmo rimuove gradualmente il rumore gaussiano, cercando di ricreare un dato quanto più possibile simile al dato originale di input.

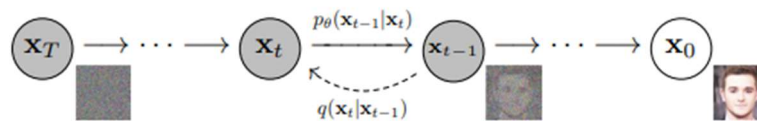


Figura 5 - Processo di denoising schematizzato [12]

Una volta completato l'addestramento in questo modo, l'algoritmo è in grado di generare dati partendo dal rumore puro [12].

I *Diffusion Models* sono considerati un'evoluzione delle *GANs* in quanto ne migliorano alcuni aspetti, funzionalità e output. I modelli di diffusione sono considerati migliori a partire dalla fase di addestramento, sicuramente più semplice da realizzare, con un'unica rete neurale. La qualità dell'output e la possibilità di controllo sul processo sono un altro degli aspetti in cui i *Diffusion Models* superano le *Reti Generative Avversarie*. Infine, la prospettiva più interessante dei modelli di diffusione è una grande flessibilità, poiché essi permettono di essere utilizzati con diverse tipologie di dati in input, come audio, video e testo [13].

Altra menzione importante tra gli algoritmi di *IA Generativa* è quella degli algoritmi di *LLM* o *Large Language Model*. Questi sono gli algoritmi che stanno alla base dei chatbot, come *ChatGPT*, che hanno come obiettivo quello di generare una conversazione in linguaggio naturale, comprendendo le domande di chi scrive e rispondendo in modo adeguato. L'operazione fondamentale degli algoritmi di *LLM* è quella della *tokenizzazione*: il testo ricevuto come input viene suddiviso in unità discrete, chiamate *token*, che possono essere parole, caratteri o altre unità simili. Oltre a ciò, i *token* contengono anche delle informazioni posizionali, dette *positional embeddings*, che indicano la posizione che il *token* assume nella frase. Questo meccanismo viene in seguito al fatto che le varie parole della frase di input non siano processate in ordine di posizione, ma in parallelo perdendo così le informazioni posizionali. A questo punto, gli algoritmi di *LLM* agiscono secondo il *meccanismo dell'attenzione*, assegnando importanza differente ai vari *token*, pesando le parole in input rispetto al contesto [14].

Nel campo della generazione audio il modello considerato come punto di riferimento è *WaveNet*. Questo si basa sul funzionamento di *Reti Neurali Profonde* che impiegano convoluzioni causali dilatate con l'obiettivo di generare audio grezzo, ottenendo un suono realistico. La generazione dell'audio si basa su un modello *autoregressivo* che utilizza i campioni precedenti per generare il campione successivo. Il suo principale impiego è quello della sintesi vocale basata sul testo, ma ha avuto buoni risultati anche nella generazione di musiche [15].

Proseguendo, l'*IA Non Generativa*, detta anche *IA Discriminativa*, si pone nell'ambito della classificazione di dati, senza generare alcunché. È quindi il caso dell'*Intelligenza Artificiale* che permette il riconoscimento e la classificazione di immagini oppure la trascrizione di una traccia audio o la traduzione di un testo [16].

Volendo citare un esempio su tutti per approfondire il significato di *IA Non Generativa*, è fondamentale parlare delle *Convolutional Neural Networks* o *CNN*, utilizzate ampiamente nell'analisi di immagini. Una *CNN* è una *rete neurale* che, data un'immagine in input, è in grado di assegnare un peso, o importanza, a vari aspetti o oggetti della stessa immagine, in modo tale da poterla classificare e distinguere dalle altre. Ciò che differenzia le *CNN* dagli altri algoritmi di classificazione delle immagini è la capacità di imparare l'utilizzo di filtri e caratteristiche dell'immagine, con il training adeguato [17].

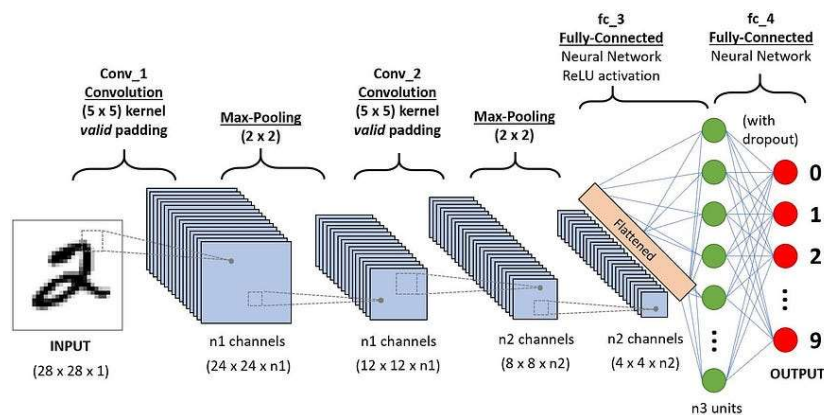


Figura 6 - Rappresentazione di una sequenza di CNN per classificare la scrittura a mano [17]

Una *CNN* si compone di quattro parti fondamentali [18]:

- Il *Convolution Layer* è il primo strato che applica una finestra di convoluzione ai pixel dell'immagine
- La *ReLU Activation Function* è una funzione applicata dopo ogni convoluzione per permettere alla rete di apprendere anche le relazioni non lineari tra gli elementi dell'immagine
- I *Pooling Layers* sono strati che estraggono le caratteristiche più significative dalla matrice convoluta

- I *Fully Connected Layers*: sono gli strati finali, utilizzati per classificare l'immagine

Riassumendo, è importante comprendere a fondo la differenza tra *IA Generativa* e *Non Generativa* perché è la prima distinzione che occorre fare nella valutazione di uno strumento. L'*IA Generativa* ha naturalmente un impatto maggiore perché è in grado di creare un output nuovo ogni volta che viene interrogata, che sia un'immagine, un video, un brano musicale o un testo. L'operato dell'*IA Non Generativa* è meno visibile, ma ugualmente significativo, poiché permette di analizzare i dati in input per velocizzare o migliorare delle operazioni, come la trascrizione di audio, la traduzione di testi o il tracking più avanzato di un oggetto, nel caso di strumenti cinematografici.

1.2 IA nell'Audiovisivo: lo Stato dell'Arte

L'esplosione dell'*Intelligenza Artificiale* è un fenomeno assai recente, soprattutto per quanto riguarda strumenti che sono alla portata del grande pubblico. Pertanto, la velocità con cui un tool può essere considerato superato o, addirittura, obsoleto è spaventosamente rapida. Va da sé che tracciare un quadro completo e aggiornato è un'operazione alquanto complessa, se non impossibile. Ciò che si può fare, però, è congelare il momento attuale e mostrare quali sono gli strumenti di *Intelligenza Artificiale* ad oggi più aggiornati e performanti.

Calando questo discorso nel contesto del cinema, il modo migliore di approcciare allo stato dell'arte è secondo la distinzione fatta pocanzi tra *IA Generativa* e *Non Generativa*, come strumenti che creano qualcosa, i primi, e strumenti che aiutano a fare qualcosa, i secondi.

1.2.1 IA Generativa nel Cinema

L'*Intelligenza Artificiale Generativa* comprende tutti gli strumenti che sono in grado di generare un output inedito avendo ricevuto in input un *prompt*, che nel maggior numero di casi è costituito da una stringa di testo, contenente le informazioni sul compito che deve portare a termine. Nel caso del cinema digitale gli output utili sono: immagini, video, audio e testo.

Le immagini hanno un'utilità soprattutto nella pre-visualizzazione di un prodotto cinematografico, come la creazione di uno storyboard o di concept art; oltre a ciò la generazione di immagini è molto efficace anche nell'ottica di flussi di lavoro più complessi per la generazione di clip video.

La generazione di video è, probabilmente, il caso più comprensibilmente vantaggioso. Oltre a poter generare clip senza necessitare nemmeno di una camera, la generazione video è anche valida nella fase di pre-produzione per velocizzare la realizzazione di *Animatic*.

Gli strumenti di generazione audio nel cinema hanno utilità sfaccettate, come la generazione di musiche, effetti sonori e parlari per voci narranti.

Infine, la funzionalità della generazione del testo è sicuramente in fase di pre-produzione, potendo scrivere documenti utili allo sviluppo della narrazione, come i *Character Sheet*, o interi dialoghi tramite l'*Intelligenza Artificiale*.

Immagini

Le immagini sono uno degli output dell'*IA Generativa* realizzati a livello più avanzato. Su di esse si basano molti workflow di generazione di storyboard, concept art o clip video applicabili al cinema. Esistono molti strumenti in grado di generare delle immagini con qualità molto alta e una buona attinenza al prompt. Di seguito, si mostrano le proposte attualmente migliori sul mercato.

Midjourney è lo strumento di generazione di immagini più conosciuto: permette di creare delle immagini partendo da un prompt testuale. Infatti, il tool si serve di due modelli di rete neurale visti precedentemente: un *LLM* e un *Diffusion Model*. La prima rete, ha come obiettivo quello di rendere comprensibile al software il prompt testuale, mentre la seconda quello di creare l'immagine richiesta partendo da un rumore casuale [19]. Tuttavia, *Midjourney* utilizza un codice proprietario, per cui è impossibile sapere con certezza quali siano i modelli di rete neurale alla base.

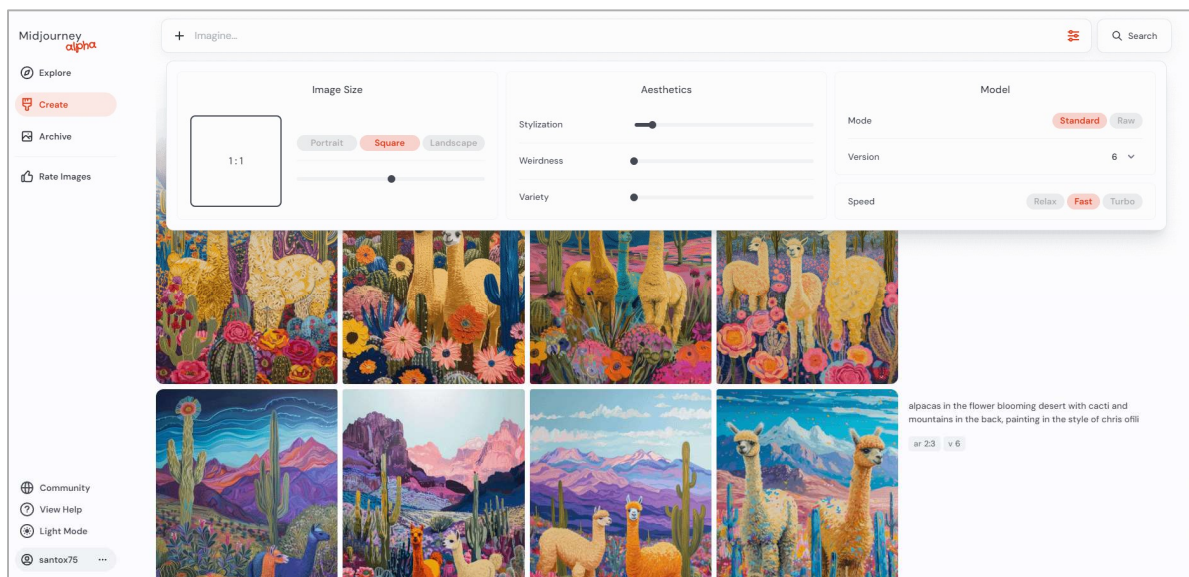


Figura 7 - Interfaccia della web app di Midjourney

Inizialmente, *Midjourney* si serviva unicamente della piattaforma *Discord* per il proprio funzionamento. Dall'agosto del 2024 gli sviluppatori hanno aperto anche un'applicazione web, seguendo le modalità con cui la maggior parte della concorrenza dà accesso ai propri servizi di generazione di immagini [20].

Una caratteristica peculiare di questo strumento è l'utilizzo di parole chiave, denominate parametri, che permettono di chiamare un determinato tipo di funzionalità, assomigliando ad un linguaggio di programmazione. Questo consente, ad esempio, di indicare

manualmente quale algoritmo utilizzare, se l'ultima versione o una delle precedenti, oppure di identificare un preciso stile da applicare all'immagine. I parametri di *Midjourney* permettono di accedere ad alcune delle funzionalità più importanti del software: una di queste è la funzione di *upscaling*, che permette all'utente di ingrandire le dimensioni dell'immagine generata, che di default sono 1024x1024 pixels. La funzione di *upscaling* è disponibile in differenti scale, a seconda delle versioni con cui viene utilizzata; l'ultima versione disponibile, la *v6.1*, supporta un *upscaling* fino a 2 volte della dimensione iniziale [21].

Tra le altre funzioni principali di *Midjourney*, spicca la possibilità di utilizzare il parametro di *Style Reference*, che permette di indicare un'immagine come reference per lo stile da applicare a quelle generate. È, inoltre, possibile indicare il peso dell'incisività della reference sull'output finale, volendo modificare la somiglianza con lo stile in input.

Tra gli ultimi aggiornamenti di *Midjourney* c'è stata l'introduzione del parametro di *Character Reference*, che, come nel caso dello stile, permette di indicare un personaggio come reference, potendo così ottenere una consistenza notevole nella generazione di personaggi [21].

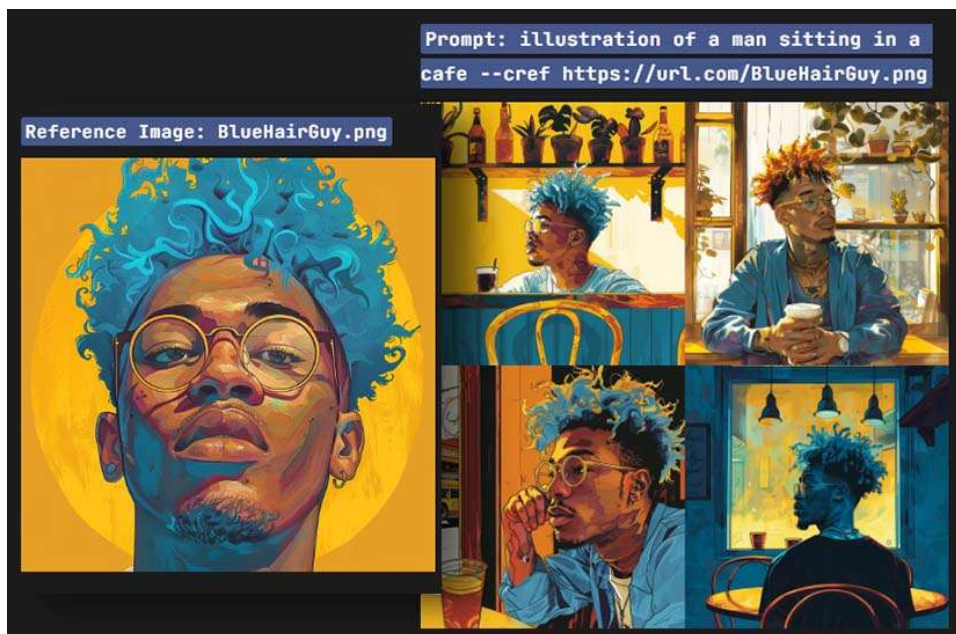


Figura 8 - Esempio di Character Consistency Midjourney [21]

Il vantaggio principale di *Midjourney* è quello della qualità dell'output, sicuramente elevata e spesso migliore rispetto alla maggior parte della concorrenza. In più, le molte possibilità date dai differenti utilizzi del prompt lo rendono, probabilmente, lo strumento leader nel settore della generazione delle immagini. Lo svantaggio principale è legato ai costi di utilizzo, che non concedono una prova gratuita e ne rendono l'accesso limitato ai soli iscritti.

Proseguendo la panoramica dei tool più importanti sul mercato, *DALL-E 3* è sicuramente fra questi. Questo tool è il generatore di immagini sviluppato da *OpenAI*, casa di sviluppo anche di *ChatGPT*, ed è la terza versione sul mercato dopo *DALL-E* e *DALL-E 2*. Alla base della generazione di immagini di *DALL-E3* si trova il potente modello di *LLM* sviluppato da *OpenAI* per *ChatGPT*. Questo viene utilizzato anche nella generazione delle immagini che vengono create tramite l'impiego di un *Modello di Diffusione*, similmente al caso di *Midjourney* [22] [23].

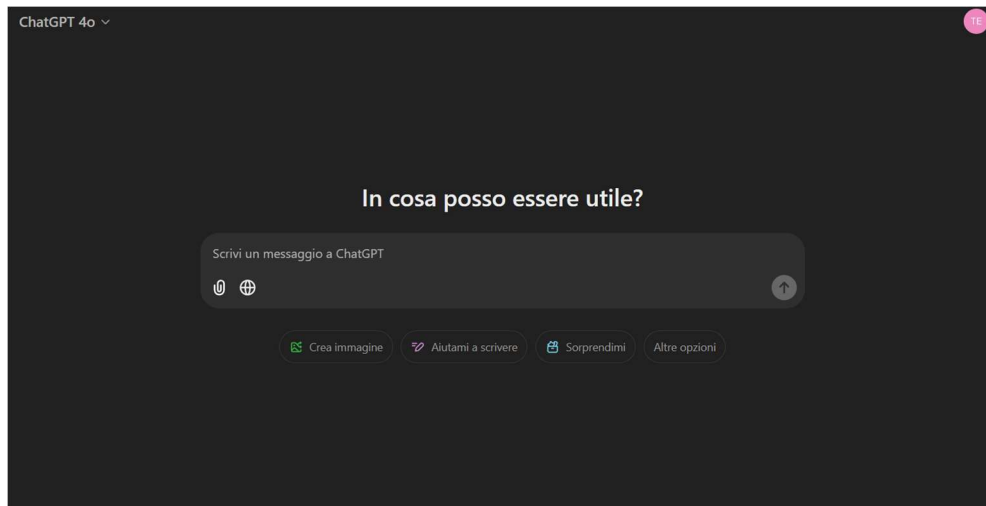


Figura 9 - Interfaccia di *DALL-E 3*

La peculiarità più importante di *DALL-E 3* è l'integrazione con *ChatGPT*, che lo rende più semplice e veloce da utilizzare. Difatti, la coerenza e il miglioramento della comprensione del prompt è l'aspetto più curato nell'ultima versione dell'algoritmo, dovuta all'integrazione in *ChatGPT* [23].

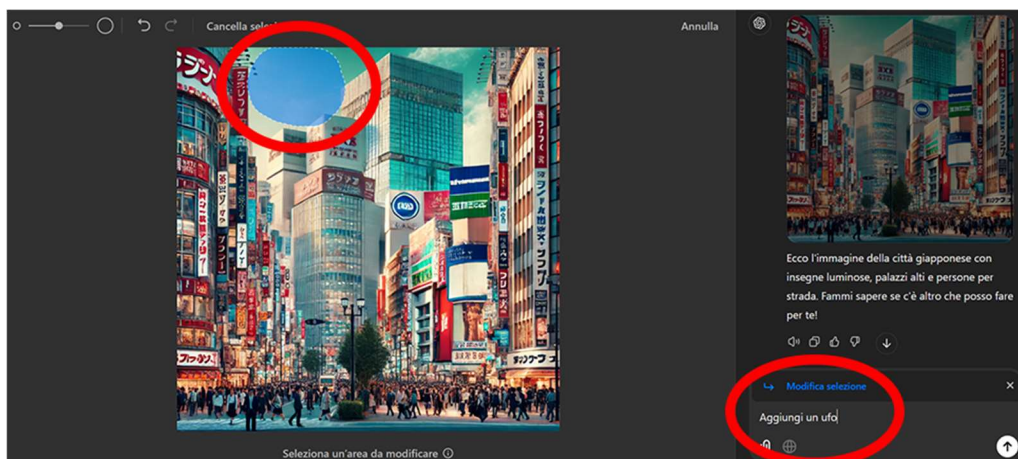


Figura 10 - Utilizzo dello strumento selezione di *DALL-E 3*

Una caratteristica che offre l'algoritmo di *DALL-E 3* è lo strumento di selezione, che consente di evidenziare una porzione dell'immagine generata e di scrivere un prompt dedicato alla modifica di quella, aggiungendo o rimuovendo elementi al suo interno.



Figura 11 - Risultato ottenuto dall'utilizzo della selezione su DALL-E 3

Il vantaggio principale nell'utilizzo di *DALL-E 3* è l'integrazione con *ChatGPT*, giacchè questo rende sicuramente più semplice la scrittura di prompt e la comprensione degli stessi da parte dell'algoritmo. È doveroso precisare che anche *DALL-E 3* è soggetto alla sottoscrizione di un abbonamento, lo stesso di *ChatGPT*. Esiste, tuttavia, la possibilità di provarlo gratuitamente utilizzando *Microsoft Copilot*, che ne integra una versione limitata.

Sul mercato esistono anche modelli che vengono rilasciati con l'intento di permettere agli utenti di inserirli nelle proprie applicazioni o di installarli direttamente sul proprio *hardware*. È il caso di *Stable Diffusion 3.5*, l'ultima versione del modello generativo open source prodotto da *Stability AI*, che offre la possibilità di generare immagini ad alta qualità. L'algoritmo utilizza un particolare modello di diffusione, chiamato *Adversarial Diffusion Distillation (ADD)*, che implementa al suo interno sia meccaniche basate sui modelli di diffusione classici, sia sulle *GANs*, partendo da rumore casuale, ma mantenendo la supervisione avversaria di un discriminatore. La distillazione, termine chiave nel concetto di questo procedimento, si riferisce ad un processo di "insegnamento", che un modello di diffusione più completo e pesante compie a beneficio di un modello più leggero, durante la fase di training. La generazione si basa sul modello più piccolo, chiamato *studente*, che utilizza i modelli di diffusione appresi dalla distillazione, anche tramite la supervisione avversaria del discriminatore. L'obiettivo di questo sistema è quello di raggiungere una maggiore velocità nella generazione, insieme ad una riduzione del costo computazionale e una qualità superiore delle immagini [24].

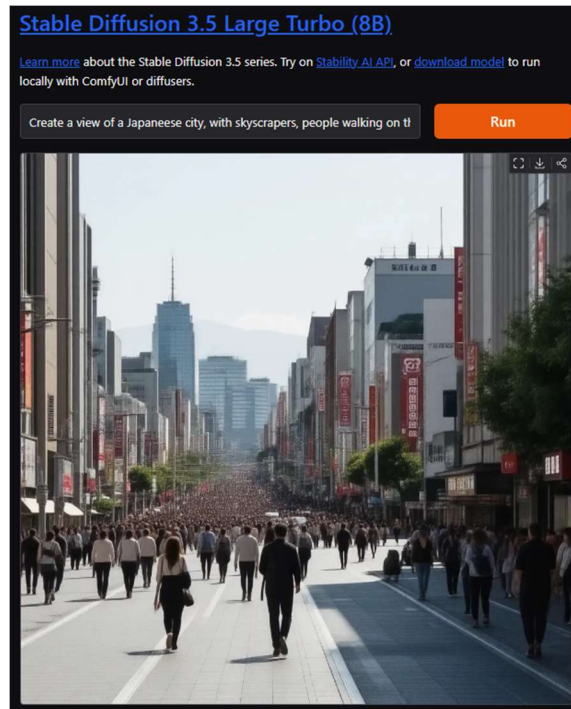


Figura 12 - Interfaccia di prova della API di Stable Diffusion 3.5 Large Turbo

Stable diffusion 3.5 offre differenti versioni di utilizzo, per permettere di essere implementato anche in casi di una potenza computazionale minore, dividendo le versioni 3.5 in *Large*, *Large Turbo* e *Medium* [25].

Essendo open source e integrabile in applicazioni esterne, non vi sono grandi possibilità di utilizzo e implementazioni online e, per utilizzarli in locale, serve una buona disposizione *hardware*. Tuttavia, i vantaggi nell'utilizzo di questo modello sono quelli auspicati nell'utilizzo di una *ADD*, con una qualità maggiore, una velocità più alta e una buona efficienza computazionale.

Lo svantaggio principale è l'assenza di un'applicazione online in cui poter utilizzare il modello nel pieno delle sue potenzialità, rimanendo quindi molto limitato alle versioni di test, con interfacce abbozzate e poche possibilità di utilizzo, come nel caso della Figura 12. Il modo per utilizzarlo è servirsi dell'interfaccia installata in locale di *ComfyUI*. In generale, questa rimane una grande possibilità in potenza, ma più difficile da utilizzare nel pratico.

In ultimo, è corretto citare anche la presenza degli algoritmi di *Leonardo AI*, considerato tra i migliori generatori di immagini sul mercato, all'altezza della concorrenza di quelli precedentemente visti. Anche questo implementa modelli di diffusione e il suo algoritmo di generazione delle immagini è *Leonardo Phoenix*.

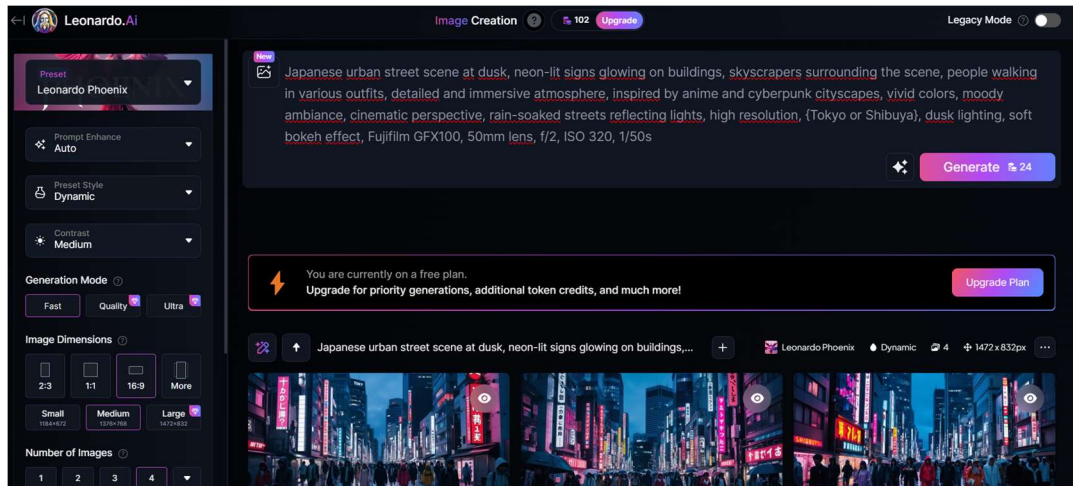


Figura 13 - Interfaccia del tool Leonardo AI

L'applicazione di *Leonardo AI* offre al suo interno molte funzionalità che rendono l'esperienza di utilizzo migliore, come la capacità di scegliere tra vari stili preimpostati che semplificano la scrittura di un prompt. Inoltre, esiste la possibilità di utilizzare un *GPT* creato dagli sviluppatori del tool per la generazione di prompt migliori da sottomettere a *Leonardo* [26].

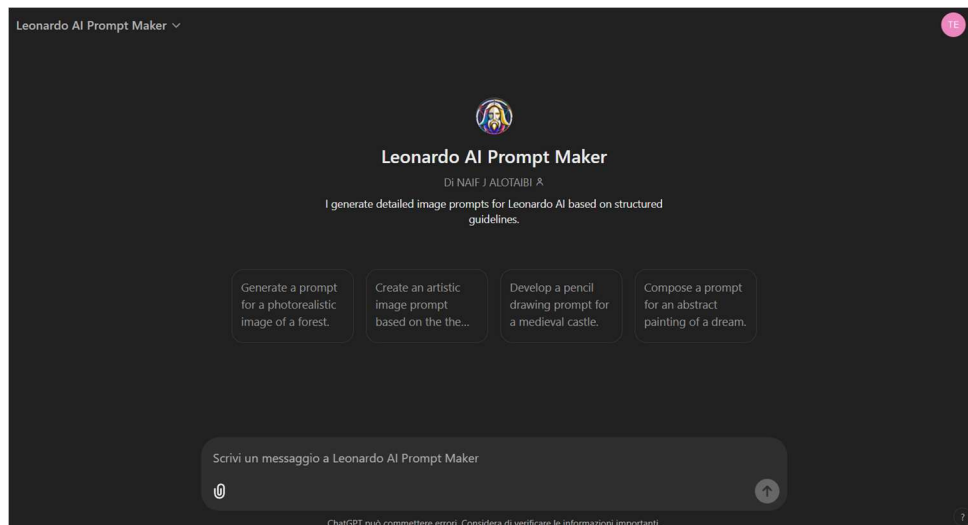


Figura 14 - Interfaccia GPT Leonardo Promt Maker

Il grande vantaggio di questo strumento di generazione immagini è l'usabilità, data da un'applicazione completa e che offre ampie possibilità di utilizzo degli algoritmi. Come gli strumenti visti, anche *Leonardo* è utilizzabile sottoscrivendo un abbonamento, con la sola differenza che viene concessa una versione di prova.

Sebbene questi quattro strumenti di generazione immagini siano equiparabili e leader nel settore ognuno per motivi differenti, dovendo ipotizzare l'utilizzo di questi strumenti in un workflow a livello cinematografico, probabilmente *Midjourney* è quello maggiormente indicato allo scopo. Tale deduzione si realizza sia per le possibilità di fotorealismo offerte,

sia per le ottime possibilità che permette di avere nella consistenza dello stile e dei personaggi. In più, con la nuova applicazione web permette un utilizzo più semplice, non limitato nelle funzionalità.

In ultimo, è sicuramente corretto menzionare anche le integrazioni dell'*IA Generativa* anche nella suite *Adobe*, con molti strumenti generativi che permettono di migliorare il flusso di lavoro e di post-produzione, raggruppati sotto il nome di *Adobe Firefly*. Un esempio su tutti è il *generative fill* di *Photoshop*, che permette di utilizzare algoritmi generativi per aggiungere e integrare oggetti all'interno di un'immagine o di espandere la stessa.

Video

Attualmente la generazione di video è un campo che sta ricevendo aggiornamenti molto velocemente, con nuovi strumenti o nuovi modelli che entrano nel mercato in maniera competitiva.

Runway è uno strumento leader nel settore della generazione video, che offre diversi modelli e diverse potenzialità, che vanno anche oltre la semplice generazione video. I modelli di generazione video rilasciati sin qui sono *Gen-1*, *Gen-2*, *Gen-3* e *ActOne*, integrato in *Gen-3*. Il primo è un modello di generazione video le cui funzioni sono superate dalle generazioni successive. La sua funzionalità viene definita *video-to-video*, ovvero offre la possibilità di applicare uno stile ad un video esistente, creando potenzialmente un'animazione. Si tratta, però, dell'algoritmo che getta le basi per le generazioni successive, utilizzando un modello di diffusione applicato alla generazione video [27].

Runway Gen-2 è un sistema multimodale di *Intelligenza Artificiale Generativa*, che offre la possibilità di generare video partendo da un prompt testuale o da un'immagine. È da considerarsi, a tutti gli effetti, un'evoluzione di *Gen-1*, perché implementa le funzioni *text-to-video* e *image-to-video* e si basa su modelli di diffusione simili a quelli della generazione precedente.

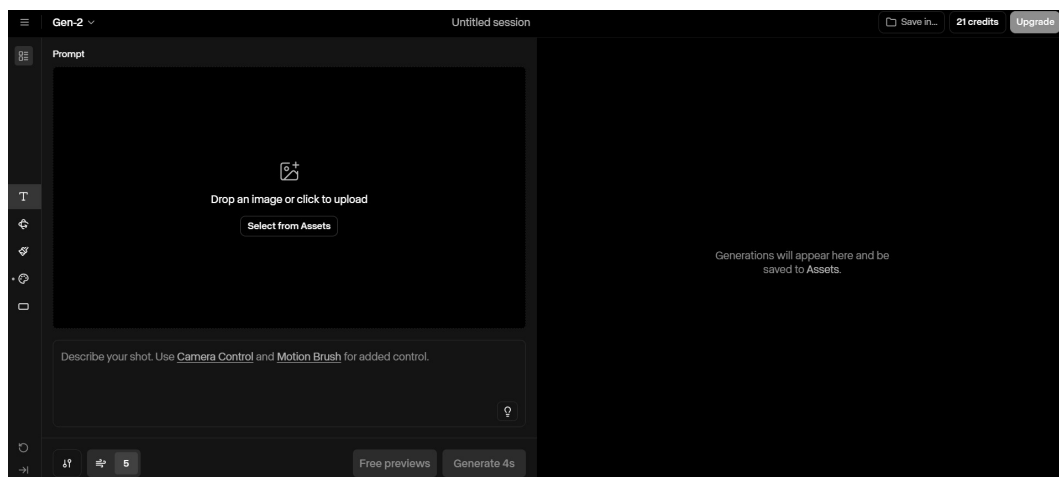


Figura 15 - Interfaccia di Runway Gen-2

Questo modello consente di raggiungere l'obiettivo della generazione tramite l'impiego di vari strumenti, che vanno oltre al semplice prompt testuale. Offre, infatti, strumenti come *Camera Control*, *Motion Brush* o *Style*, che consistono in modi differenti di fornire il prompt in input, non più in maniera testuale, ma con delle immagini o degli slider. In particolare, si possono definire movimenti di camera standard, come *pan* e *tilt*, semplicemente muovendo degli slider. Allo stesso modo, la funzione di *Motion Brush* ammette la realizzazione di animazioni sui tre assi di oggetti all'interno di un'immagine, selezionandoli con un pennello e muovendo gli slider per definirne il movimento. Il campo *Style* integra la possibilità di definire uno stile predefinito utilizzando un'immagine come reference per permettere all'algoritmo di utilizzarla come punto di partenza [28].

Gen-2, grazie alle molteplici funzionalità e casistiche di prompt che integra, costituisce un modello molto vantaggioso per svariate condizioni di utilizzo, rivelandosi uno strumento molto flessibile. Il grande svantaggio risiede ancora in una poca consistenza temporale e nella possibilità di generare video di soli quattro secondi.

Lo strumento più avanzato offerto da *Runway* è *Gen-3*, che consiste nuovamente in un'evoluzione dei modelli precedenti. L'addestramento avviene tramite un approccio multimodale su larga scala, che consente di avere una maggiore comprensione delle dinamiche tridimensionali e una resa più fotorealistica dei video generati [29].

A livello di funzionalità, la terza generazione offre uno spettro meno ampio di strumenti rispetto a quella precedente, limitandosi al *Camera Control*. Tuttavia, *Gen-3* è da considerarsi lo strumento che incorpora e migliora l'utilizzo di entrambe le generazioni precedenti, consentendo anche la possibilità di utilizzare un video come prompt, evolvendo anche come modello *video-to-video*. Inoltre, i video generati in *Gen-3* hanno una lunghezza maggiore, arrivando fino ai dieci secondi. Oltre a questo, *Runway* integra in *Gen-3* anche *ActOne*, uno strumento in grado di utilizzare il video di un attore o di una persona come reference per i movimenti corporei e facciali di un character animato [30], fungendo a tutti gli effetti come strumento di *Motion Capture*.

A livello di funzionalità *Runway* si rivela uno strumento completo, perché offre, oltre alla generazione video, una *Dashboard* molto fornita di strumenti basati sull'*Intelligenza Artificiale*. Questa integra in sé non solo strumenti di generazione di voci e di immagini, ma anche strumenti non generativi come la rimozione dello sfondo o l'eliminazione di oggetti in un'immagine o video. Lo svantaggio principale è dovuto ai costi necessari per poter fare un utilizzo completo e prolungato della piattaforma. Esiste, tuttavia, la possibilità di ottenere crediti di generazione gratuiti con l'iscrizione.

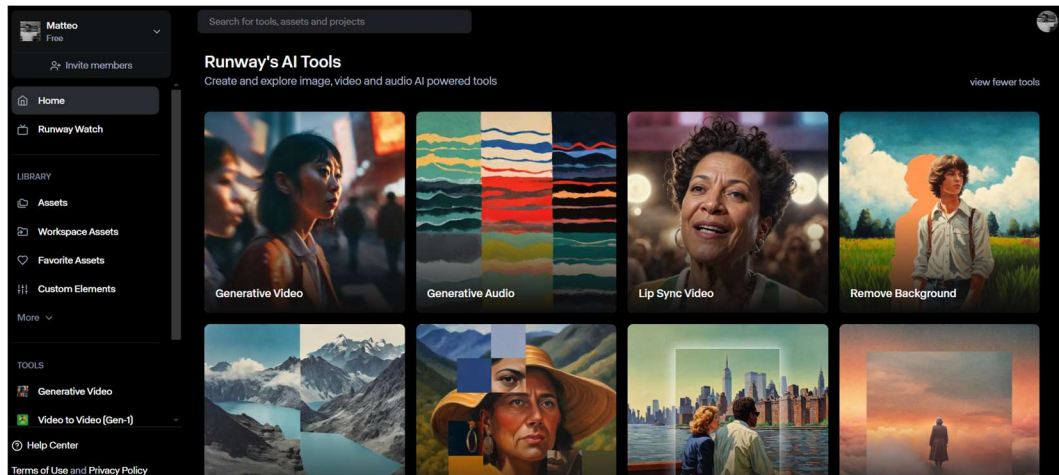


Figura 16 - Una porzione della Dashboard di Runway

Alternativamente a *Runway*, si menzionano altri strumenti che si posizionano ad un livello simile per qualità di video generato. *Kling AI* è uno di questi, che offre la possibilità di generare video a partire da testo o immagini. Il generatore di *Kling AI* si basa su funzionalità di attenzione dello spazio tridimensionale e su *trasformatori di diffusione*, che permettono di ottenere dei video generati di qualità molto alta e con una buona resa.

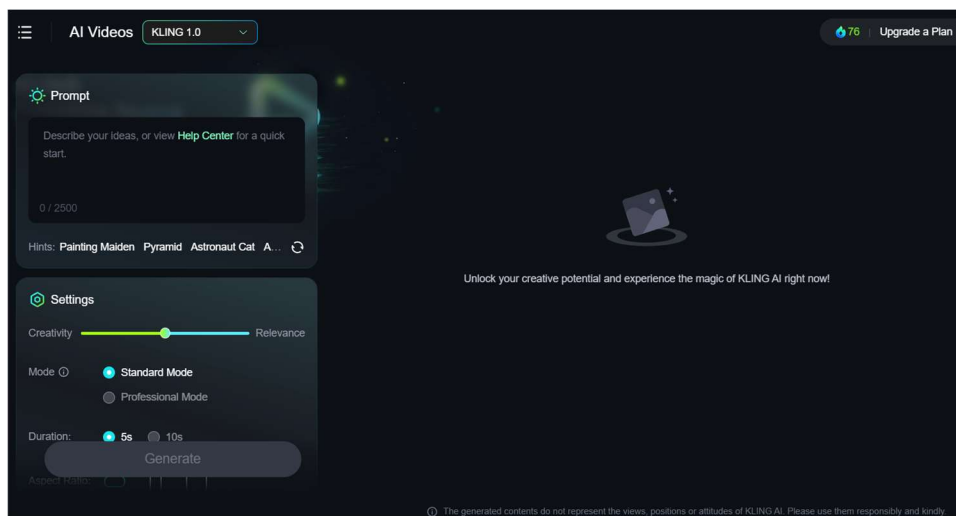


Figura 17 - Interfaccia di Kling AI

In maniera simile a quanto accade in *Runway*, *Kling AI* integra le funzionalità di *text-to-video* e di *image-to-video*, offrendo anche controlli simili a *Gen-2*, come il *motion brush* e il *camera control*.

Il vantaggio di questo algoritmo risiede nella capacità di generare video ad alta qualità e di una lunghezza maggiore sottoscrivendo l'abbonamento, arrivando a dieci secondi per ogni generazione, con l'eventualità di espanderli fino a due minuti. A livello di resa finale è comparabile a quella di *Runway*, con un'attenzione particolare ai movimenti dei personaggi e al realismo della camera [31].

Come nel caso delle immagini, esistono tool open source anche per la generazione video, rilasciati da *Stability AI*. È il caso di *Stable Video Diffusion* che si basa su processi di *Diffusione Latente* che operano in uno spazio latente complesso, con l'intento di ridurre la complessità computazionale. Il modello rilasciato da *Stability AI* parte da un training basato su immagini e solo successivamente viene addestrato con video prima di bassa qualità e poi ad alta qualità [32]. L'utilizzo di questo modello può avvenire tramite l'installazione in locale dello stesso, oppure tramite una versione gratuita in prova sul sito web di *Stable Video*. La generazione prevede sia prompt testuali sia immagini, ma nel caso di utilizzo di prompt testuali la generazione video è divisa in due fasi, prima quella di un'immagine e poi l'utilizzo dei controlli di camera per creare il video partendo dall'immagine generata.

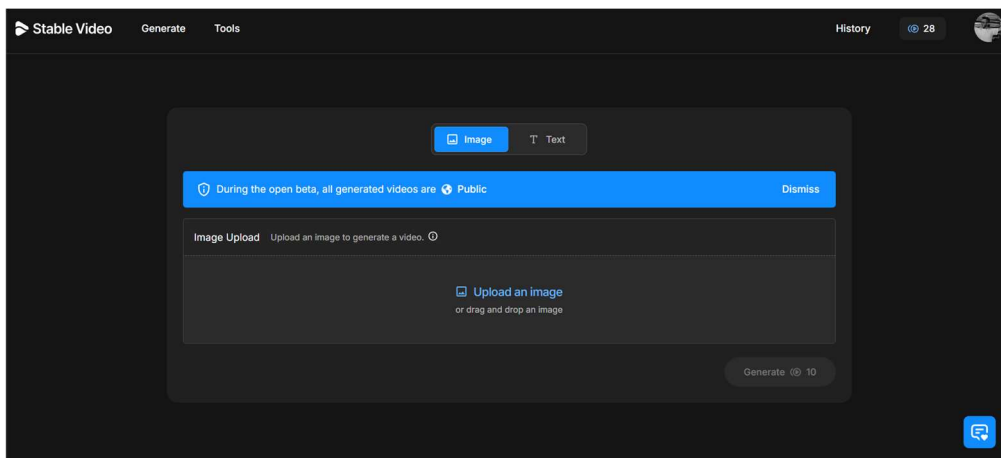


Figura 18 - Interfaccia web di *Stable Video Diffusion*

Oltre a quelli descritti, esistono anche molti altri tool in grado di generare video più o meno fotorealistici, con caratteristiche simili. Si menzionano *Minimax* e *PikaLabs* fra tutti, che sono quelli considerati maggiormente come confrontabili con i tool presentati. Attualmente, per ciò che viene proposto, il tool più completo è considerato essere *Runway*, per possibilità offerte e per qualità della generazione, con *Kling AI* di poco distante; tuttavia, gli aggiornamenti futuri dei tool presentati o di altri ancora sconosciuti possono invertire questa tendenza.

Audio

La generazione audio è un ambito sicuramente interessante nel mondo del cinema, sia per quanto riguarda la sintesi di musiche, sia per ciò che concerne la sintesi di voci. Nel mercato attuale esistono due tool tra quelli studiati che sono in grado di proporre risultati, che vadano oltre alla generazione di tracce robotiche e ripetitive, e che mostrano del potenziale.

Per quanto riguarda la generazione musicale un modello estremamente efficace, che spicca tra la concorrenza è *Stable Audio 2.0*, sviluppato da *Stability AI*, che genera tracce audio a partire da prompt testuali o da campioni audio, offrendo quindi le possibilità *text-to-audio* e *audio-to-audio*. Il funzionamento si basa su tecniche di compressione e modelli di diffusione,

che hanno l'obiettivo di migliorare la gestione di tracce lunghe a livello temporale, mantenendo una coerenza dall'inizio alla fine [33] [34].

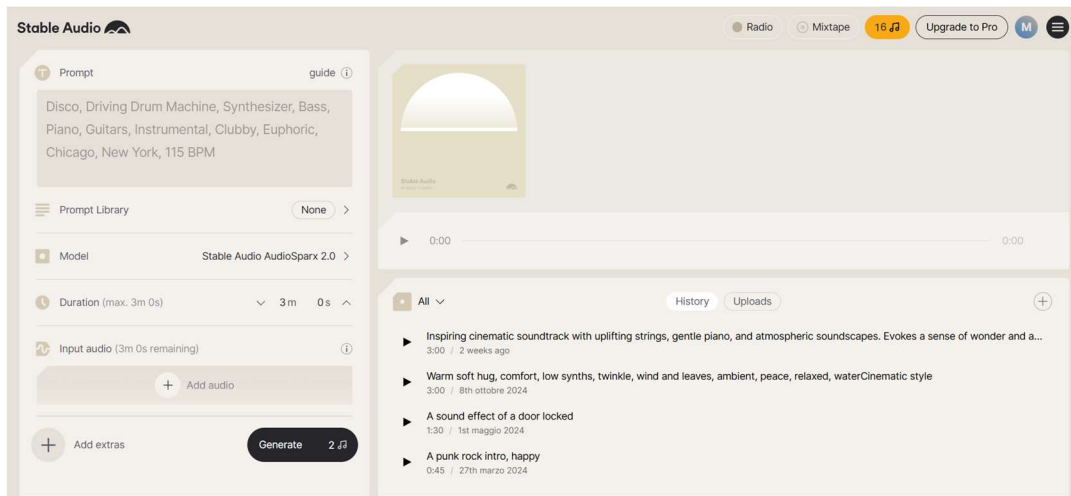


Figura 19 - Interfaccia di Stable Audio

Il risultato della generazione è una traccia musicale della durata massima di tre minuti con un'ottima attinenza al prompt e una buona costruzione della traccia, senza risultare ripetitiva. A livello di funzionalità non spiccano particolari personalizzazioni, se non quella di poter inserire un audio in input a supporto del prompt, oppure di poter aggiungere delle rifiniture più dettagliate nell'utilizzo dell'algoritmo, come gli step, il numero di risultati, il seed e la forza del prompt. In ultimo, *Stable Audio 2.0* promette anche ottime capacità nella generazione di *Effetti Sonori*. L'utilizzo di *Stable Audio* prevede una versione gratuita, con la possibilità di generare venti tracce musicali in un mese; sottoscrivendo diversi livelli di abbonamento si ha la possibilità di accedere alle funzionalità più complete e ad un maggiore numero di tracce mensili.

Nel campo della generazione vocale, uno degli strumenti leader è certamente *Eleven Labs*. Questo strumento permette di generare voci narranti partendo da un semplice testo e scegliendo la voce sintetica migliore per la necessità. I modelli alla base di questo strumento non sono stati resi noti dalla casa di sviluppo.

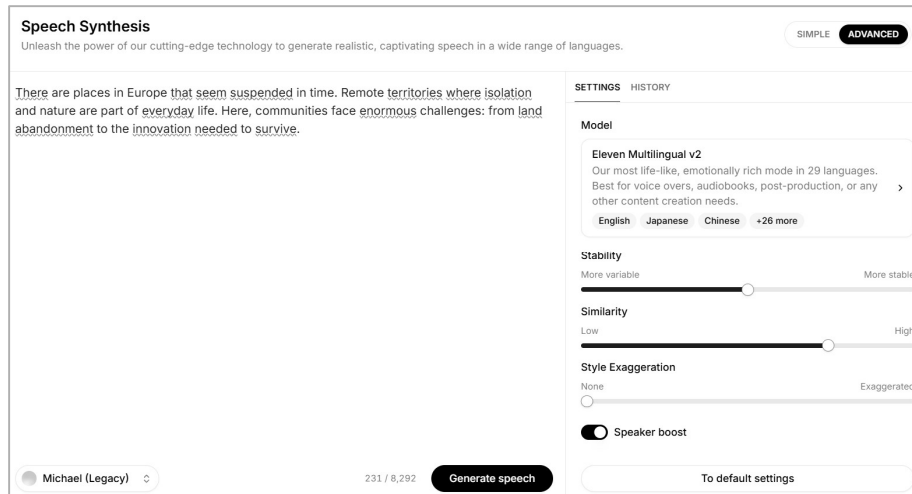


Figura 20 - Interfaccia di utilizzo di Eleven Labs

La peculiarità di questo strumento è la possibilità di clonare una voce reale: ciò consente di avere all'interno della libreria numerose voci non sintetiche, ma reali, che risultano con una resa sicuramente più efficace. Altre funzionalità degne di nota sono la versione di utilizzo *speech-to-speech*, in alternativa alla *text-to-speech*, che permette di modificare la voce presente in una traccia audio caricata. Questo tool permette inoltre di generare *Effetti Sonori* con un semplice prompt testuale che descriva brevemente l'effetto desiderato [35].

Il tool *Eleven Labs* spicca per possibilità di utilizzo di strumenti differenti, ma soprattutto per la qualità del risultato ottenuto, spesso indistinguibile dalla realtà e con moltissime possibilità di voci e lingue tra cui scegliere, potendo trovare la voce migliore per ogni caso di utilizzo. Il tool ha una versione gratuita che permette di avere dieci minuti al mese di generazione audio, che si traduce con un certo numero di crediti a disposizione. La versione gratuita è limitata anche nella possibilità di importare voci dalla libreria, con un massimo di tre voci in assoluto. Per sbloccare tutte le funzionalità complete occorre iscriversi ad un piano a pagamento, che permette di ottenere più crediti mensili e di accedere a più voci nel catalogo.

Testo

In ultimo, si presentano gli strumenti di *Intelligenza Artificiale* testuali, che comprendono prompt in linguaggio naturale e che generano del testo come output.

Il leader indiscusso in questo settore è *ChatGPT*, sviluppato da *OpenAI*, ora arrivato alla versione 4o, con la versione 4o1 in beta per gli iscritti alla versione pro. L'acronimo *GPT* sta per *Generative Pre-trained Transformer*, dove il primo termine indica la generatività dello strumento, il secondo indica il fatto che il modello sia stato allenato su un grande dataset, di circa 570GB di dati in versione testuale, e il terzo termine indica la presenza dell'architettura del *transformer* che si trova alla base del modello *LLM* di *ChatGPT* [36].

La versione 4o di *ChatGPT* viene definita come quella maggiormente in grado di comprendere il testo, rispetto alle precedenti, con anche la capacità di interpretare file differenti come immagini, audio e video. Di seguito viene inserito un grafico che mostra l'abilità della versione 4o nella comprensione del testo, comparata ad altri modelli concorrenti e misurata su vari benchmark [37].

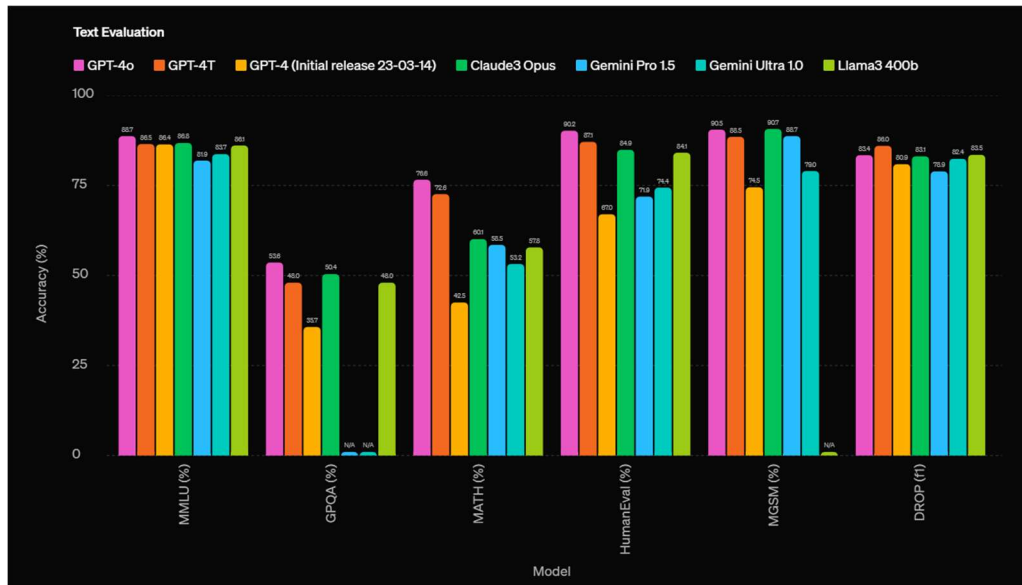


Figura 21 - Benchmark della versione 4o di ChatGPT [37]

Ulteriormente, *OpenAI* ha implementato la versione 4o con l'abilità di utilizzare una memoria, che salva le informazioni che le vengono comunicate durante una chat, per potervi accedere anche in chat successive [38]. Oltre alla memoria, è recente l'introduzione, in versione beta, dell'utilizzo di *canvas* in *ChatGPT*, permettendo di produrre in output dei documenti di testo, o dei codici, editabili in maniera bilaterale sia dall'utente, sia dal software, rendendo il risultato della conversazione maneggevole con maggiore facilità. Questo viene ancor più migliorato dall'integrazione di slider per svolgere operazioni base sul documento, come allungarlo o accorciarlo, che semplificano di molto i prompt, eliminando quelli ripetitivi [39].

In generale, *ChatGPT* ha molti vantaggi, che lo rendono leader nella generazione del testo, partendo dalla completezza e dall'accuratezza delle risposte che fornisce. Lo svantaggio principale, che in realtà accomuna la maggior parte dei tool di *Intelligenza Artificiale*, è costituito dalle limitazioni imposte dalla versione gratuita, che, nonostante tutto, è una delle versioni gratuite che offrono maggiori possibilità.

Come metro di paragone con *ChatGPT* si può utilizzare uno dei software di IA che maggiormente si propone come concorrente del modello sviluppato da *OpenAI*, ovvero *Claude*, sviluppato da *Anthropic*. Il modello più avanzato è *Claude 3.5 Sonnet*, paragonabile a

ChatGPT-4o sia nelle capacità di ragionamento, sia in molte delle funzionalità principali del software.

	Claude 3.5 Sonnet (new)	Claude 3.5 Haiku	Claude 3.5 Sonnet	GPT-4o*	GPT-4o mini*	Gemini 1.5 Pro	Gemini 1.5 Flash
Graduate level reasoning <i>GPQA (Diamond)</i>	65.0% 0-shot CoT	41.6% 0-shot CoT	59.4% 0-shot CoT	53.6% 0-shot CoT	40.2% 0-shot CoT	59.1% 0-shot CoT	51.0% 0-shot CoT
Undergraduate level knowledge <i>MMLU Pro</i>	78.0% 0-shot CoT	65.0% 0-shot CoT	75.1% 0-shot CoT	—	—	75.8% 0-shot CoT	67.3% 0-shot CoT
Code <i>HumanEval</i>	93.7% 0-shot	88.1% 0-shot	92.0% 0-shot	90.2% 0-shot	87.2% 0-shot	—	—
Math problem-solving <i>MATH</i>	78.3% 0-shot CoT	69.2% 0-shot CoT	71.1% 0-shot CoT	76.6% 0-shot CoT	70.2% 0-shot CoT	86.5% 4-shot CoT	77.9% 4-shot CoT
High school math competition <i>AIME 2024</i>	16.0% 0-shot CoT	5.3% 0-shot CoT	9.6% 0-shot CoT	9.3% 0-shot CoT	—	—	—
Visual Q/A <i>MMMU</i>	70.4% 0-shot CoT	—	68.3% 0-shot CoT	69.1% 0-shot CoT	59.4% 0-shot CoT	65.9% 0-shot CoT	62.3% 0-shot CoT
Agentic coding <i>SWE-bench Verified</i>	49.0%	40.6%	33.4%	—	—	—	—
Agentic tool use <i>TAU-bench</i>	Retail 69.2%	Retail 51.0%	Retail 62.6%	—	—	—	—
	Airline 46.0%	Airline 22.8%	Airline 36.0%	—	—	—	—

* Our evaluation tables exclude OpenAI's o1 model family as they depend on extensive pre-response computation time, unlike typical models. This fundamental difference makes performance comparisons difficult.

Figura 22 - Tabella di analisi tra i principali concorrenti evidenziati da Anthropic [40]

Parallelamente a Claude 3.5 Sonnet è stata annunciata, come si nota dalla tabella sopra, anche una versione ridotta, chiamata Claude 3.5 Haiku, che offre prestazioni minori per una leggerezza e rapidità maggiori [40] [41].

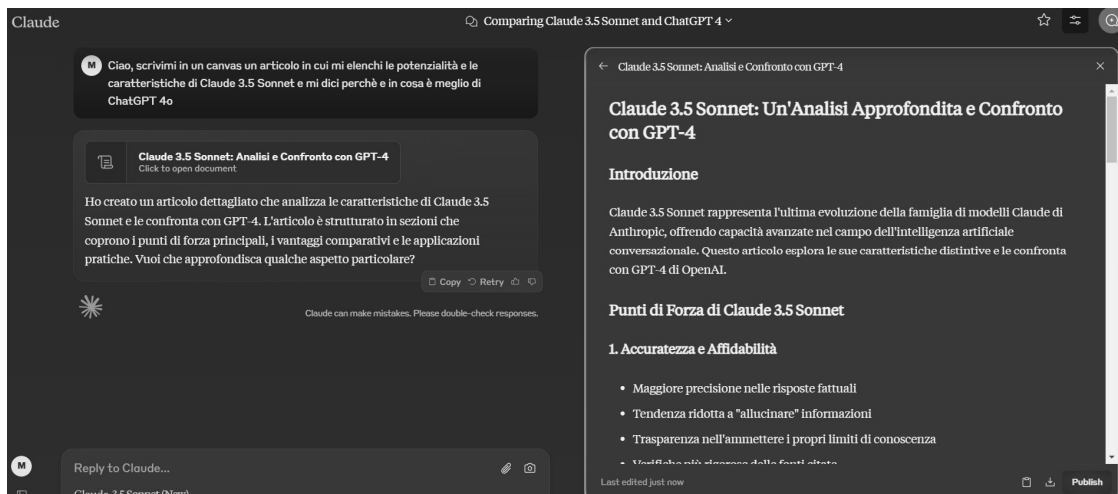


Figura 23 - Interfaccia di utilizzo di Claude 3.5 Sonnet

Dalle varie analisi condotte i due strumenti sono comparabili in molti utilizzi. Anche Claude prevede una versione a pagamento che appiattisce le limitazioni previste dal piano gratuito. Entrambi i modelli, ChatGPT e Claude, implementano l'utilizzo di Canvas, tuttavia, Claude lo

propone in una maniera più statica, non permettendo le modifiche puntuali allo stesso modo di *ChatGPT*, rendendo questo uno svantaggio rispetto al modello leader.

1.2.2 IA Non Generativa nel Cinema

Nell'ambito dell'*Intelligenza Artificiale Non Generativa*, invece, è più difficile creare una distinzione netta degli strumenti tra video e audio: questo perché molto spesso uno stesso software offre possibilità multiple di utilizzo dell'*IA*, sia nel video sia nell'audio.

Gli strumenti non generativi, si rivelano molto importanti nella produzione cinematografica, in quanto si collocano come supporto a processi e operazioni, soprattutto in post-produzione, con l'intento di semplificarli e velocizzarli, senza perdere la resa. Sono spesso strumenti che elaborano i file in maniera più sofisticata aumentandone la qualità, o che aiutano a portare a termine operazioni di post-produzione più velocemente.

Il software che maggiormente applica questi principi e integra nel proprio funzionamento l'*Intelligenza Artificiale* è *Davinci Resolve*. Questo si serve del *Neural Engine*, una componente avanzata del software che sfrutta l'*IA* e l'apprendimento automatico per potenziare molte delle funzionalità presenti [42].

In particolar modo, le funzionalità basate sull'*IA* che *Davinci Resolve* integra nel workflow di montaggio video sono le seguenti:

- *Object removal* è una funzione che permette di rimuovere oggetti in una scena, interpolando i pixel presenti sullo sfondo per coprire l'oggetto tolto;
- *Face refinement* è una funzione che permette il tracking di visi con l'obiettivo di applicare effetti direttamente sul volto in maniera più semplice;
- L'effetto *Depth maps* permette di scontornare in maniera semplice ciò che sta in primo piano, separandolo dallo sfondo. Il risultato è raggiunto generando delle mappe di profondità per la clip e utilizzandole per comprendere la posizione nello spazio degli oggetti;
- *Face recognition* è un effetto utile all'organizzazione delle clip in un progetto. Permette di analizzare il girato e di organizzarlo sulla base delle persone che sono presenti, riconoscendone i volti;
- L'utilizzo dello strumento *Smart Reframe* velocizza il processo di modifica dell'aspect ratio del girato, mantenendo sempre centrato il soggetto. Risulta utile nel caso in cui dal girato orizzontale in 16:9 è necessario estrapolare delle clip verticali;
- *Scene Cuts Detection* è uno strumento che permette di tagliare un video finito in corrispondenza di ogni taglio di montaggio, suddividendolo in clip singole;
- L'effetto *Magic Mask* riconosce persone, oggetti o dettagli del corpo umano, come i capelli, dando la possibilità di creare velocemente una maschera su di quelle;

- *Relight effect* permette di aggiungere delle luci non presenti nella scena, migliorando l'illuminazione in post-produzione e facendola interagire con gli oggetti reali;
- *Montaggio con il testo* e *Trascrizione Automatica* sono degli effetti che permettono di trascrivere il contenuto audio di un'intera clip, o timeline, e di montare solamente selezionando il testo ed inserendolo in una nuova timeline. La nuova versione del software integra la funzione di riconoscimento dell'interlocutore, permettendo di nominarlo e di individuarlo in tutte le clip del progetto in cui prende parola;
- *Subtitle generator* genera automaticamente i sottotitoli scegliendo lingua della clip audio e il numero di caratteri per riga di cui sarà composto il sottotitolo. È un effetto molto simile alla trascrizione automatica, con l'unica differenza che l'output è composto dai sottotitoli generati che possono essere tradotti automaticamente e reimportati nel software;
- *Super scale* è un effetto che consente di fare un *upscaling* del video;
- L'effetto *Speed Warp* crea un'interpolazione dei frame per rallentare un video se questo è stato girato ad un *frame rate* troppo basso per il rallentamento applicato;

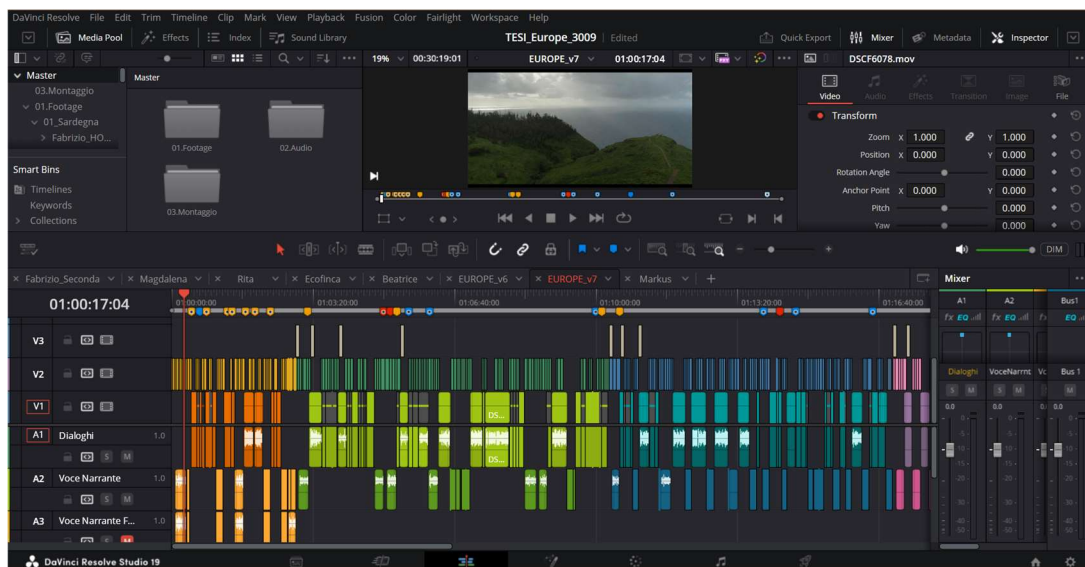


Figura 24 - Interfaccia di Davinci Resolve, tab Edit

Per quanto riguarda l'audio, gli strumenti che utilizzano l'*Intelligenza Artificiale* integrati in *Davinci Resolve* sono:

- *Voice isolation*, un effetto che permette di pulire la voce dei microfoni, modificando l'intensità dell'effetto sulla base della quantità di rumore presente;
- L'effetto di *Music Remixer* permette di individuare gli strumenti presenti in una traccia musicale e di isolarli, potendo così agire solamente sul singolo strumento;

- Lo strumento *Ducker* implementa una *Side Chain Compression* automatica, andando a comprimere una traccia audio in funzione dell'attivazione di un'altra, potendo ad esempio controllare automaticamente il volume della musica in funzione del parlato;
- La funzione di *Intellitrack* permette di tracciare il movimento di un oggetto nel frame. Questa è applicabile ad un'automazione del pan audio, permettendo di mixare i canali stereo in funzione del tracking di un oggetto ottenuto e, quindi, del movimento dell'oggetto nell'inquadratura.
- *Audio Classification* è uno strumento che permette di analizzare il footage che si ha e di suddividerlo automaticamente per tipo di audio che contiene, come SFX, musica o dialoghi;

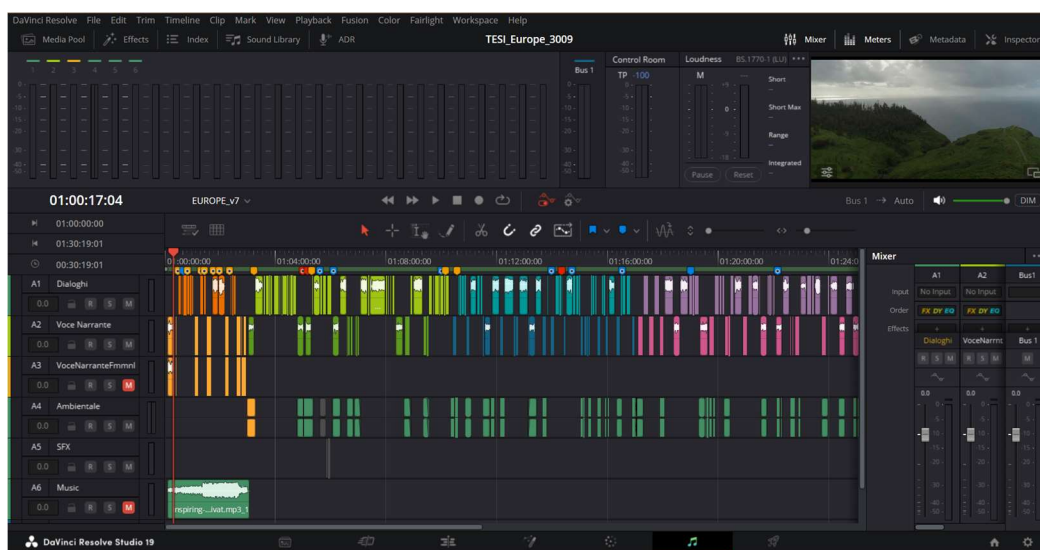


Figura 25 - Interfaccia di Davinci Resolve, tab Fairlight

Il grande guadagno dato da questi strumenti offerti da *Davinci Resolve* è la grande facilità con cui possono essere integrati in un flusso di lavoro, rendendo il software uno dei leader nel settore della post-produzione per completezza e possibilità offerte, mantenendo tutto l'iter produttivo all'interno di un unico software senza export intermedi.

Esistono anche altri strumenti più verticali, specializzati in un unico impiego dell'*Intelligenza Artificiale*. Questo è il caso di *TopazLabs* e *MagnificAI*, due strumenti che impiegano algoritmi di *LA* nell'*upscaling* ad alta qualità di immagini e video: sono sicuramente due strumenti che producono risultati ad un livello qualitativo molto alto e che, di conseguenza prevedono esclusivamente piani a pagamento.

Anche nel campo dell'elaborazione audio esistono strumenti più specifici, soprattutto per la pulizia delle tracce. Due esempi di strumenti che sfruttano l'*LA* per la pulizia delle tracce sono *Adobe Podcast* e *AICoustics*. Sebbene questi permettano di raggiungere una buona qualità di risultati nella pulizia dell'audio, non competono con lo strumento integrato in

Davinci Resolve, che consiste in un semplice effetto da applicare alla traccia già presente nel progetto, invece di doverle elaborare e reimportare nel software.

Un ultimo tool che utilizza l'IA è *Fyln.AI*, che permette di ottimizzare le operazioni di *color correction* e *color grading* di una clip in maniera più semplice, utilizzando vari parametri di modifica e proponendo differenti risultati tra cui scegliere ad ogni modifica e applicazione di un parametro.

1.3 Annunci imminenti e futuro

Come detto in precedenza, il mondo dell'*Intelligenza Artificiale* evolve con una rapidità molto alta, tanto da proporre spesso numerose novità in poco tempo. Pertanto, anche in questo caso, rimanere al passo con gli annunci imminenti di nuovi tool e versioni è un arduo compito.

L'annuncio che, probabilmente, ha avuto una maggiore risonanza nel corso del 2024 è stato quello di *OpenAI*, che ad inizio anno ha presentato *Sora*, un generatore video che promette un fotorealismo elevato. Questo ha fatto sì che la maggior parte dei competitor nel campo della generazione video utilizzassero come metro di paragone uno strumento ancora in beta testing, attualmente non lanciato sul mercato. Le promesse legate a *Sora* sono quelle di un modello di diffusione in grado di generare video di differenti dimensioni e lunghezze a partire da testo, utilizzando una vasta mole di dati che permettono di raggiungere una simulazione molto accurata, anche fisica, della realtà [43]. Oltre a ciò, *Sora* promette anche di generare sequenze di video, che abbiano consistenza tra loro, in modo tale da poter montare un'intera scena.



Figura 26 - Frame di uno dei cortometraggi generati con *Sora* [44]

Attualmente *Sora* è in ancora in beta testing, non disponibile al pubblico, ma solamente ad una ristretta cerchia di persone. Chiaramente, le aspettative si sono innalzate molto durante i mesi di attesa, anche come conseguenza dei primi video prodotti con lo strumento.

Tuttavia, non può ancora essere utilizzato come metro di paragone, almeno finché non sarà completamente a disposizione [44].

Parallelamente all'annuncio di *Sora*, sono stati proposti altri nuovi tool, alcuni già presenti sul mercato, altri ancora no. Questo è il caso di *Google Veo*, uno strumento di generazione video che si pone come concorrente diretto del modello di *OpenAI*. Le caratteristiche promesse, nel mondo dei video generativi, sono innovative tanto quanto quelle di *Sora*, ma quella che spicca maggiormente è la possibilità di integrare oggetti generati all'interno di una clip, abbattendo idealmente tutti i passaggi della filiera del compositing video.

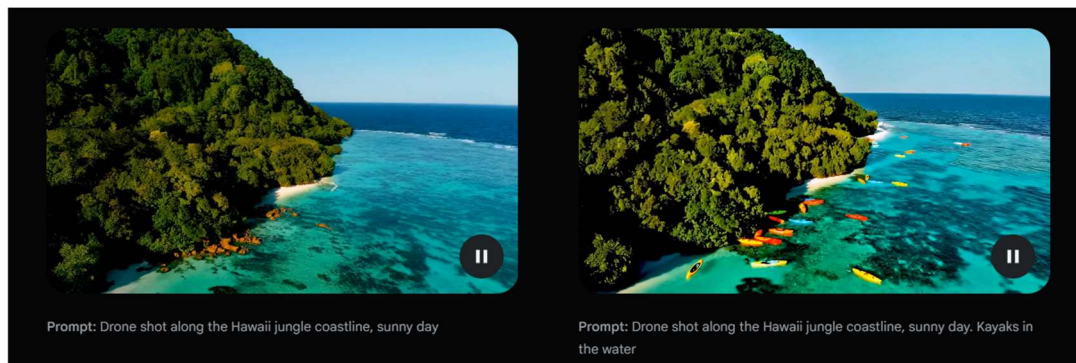


Figura 27 - Frame di video di esempio di utilizzo di Google Veo

Insieme a ciò, sono state annunciate molte altre peculiarità, come la possibilità di avere una maggiore consistenza tra frame dello stesso video, una maggiore comprensione dei prompt e un controllo creativo più elevato, implementando anche tecniche come il *timelapse*. Questo modello è figlio di tutte le ricerche precedenti nel campo, che non hanno condotto al lancio di prodotti, come *Google Lumiere* [45]; infatti, il prodotto di Google è ancora in una fase di rilascio, senza essere nemmeno in beta testing per una ristretta parte di pubblico. Nonostante ciò, si pone nella competizione ad alti livelli nella generazione video.

In ultimo, *Adobe* ha recentemente comunicato l'imminente inizio dei beta testing per il proprio modello di generazione video *Adobe Firefly Video Model*: esso avrà come caratteristica principale quella di essere integrato nei software della suite Adobe, come *Premiere Pro*, con un funzionamento simile a quello di *Photoshop Generative Fill*, quindi con l'intento di inserire o rimuovere oggetti generati da clip di video reali, velocizzando enormemente i processi di *compositing*. Oltre a ciò, chiaramente avrebbe anche la funzione di generare video partendo da un prompt.



Figura 28 - Esempio di integrazione di un effetto con Firefly su After Effects

Le potenzialità che offrirebbe questo tipo di integrazione sarebbero molto alte, perché renderebbero l'*Intelligenza Artificiale Generativa* una parte integrante di un workflow ben collaudato e funzionante come quello della post-produzione video, senza stravolgerlo, ma mantenendone intatti gli strumenti, rendendo più semplice la transizione verso un utilizzo sempre più ampio e diffuso degli strumenti di *Intelligenza Artificiale* applicati al cinema.

Capitolo 2

Ipotesi

Il contesto introdotto conduce all'ipotesi di comprendere come sfruttare meglio l'*Intelligenza Artificiale* in modalità produttive nel cinema. Si analizzeranno di seguito le opportunità offerte dagli algoritmi principali e le maggiori limitazioni strutturali e di utilizzo. L'obiettivo è quello di ipotizzare un modello di casistica in cui questi strumenti trovino un'utilità concreta e applicabile.

2.1 Opportunità da sfruttare

Il rapido avanzamento e i costanti aggiornamenti dei modelli di *Intelligenza Artificiale* e dei tool associati, ampliano costantemente il ventaglio di possibilità che questi strumenti offrono per chi li utilizza.

ChatGPT, leader nel settore degli *LLM*, dimostra un costante avanzamento nelle proprie capacità di comprensione del testo e restituzione di risposte in linguaggio naturale, molto articolate e fedeli. Le potenzialità spaziano ben oltre la semplice velocizzazione dei processi di ricerca sul web e di studio delle informazioni, ma raggiungono un livello di creatività notevole. Nascono, quindi, i primi esperimenti di workflow che utilizzano *ChatGPT* come sceneggiatore o scrittore di prodotti originali, scritti partendo dalle idee della persona, ma elaborati, nella forma e nella costruzione della narrazione, dal software. Questi, spesso, propongono un algoritmo che sia applicabile in maniera assoluta, indipendentemente dal tipo di prodotto [46].

Oltre a questo, le modalità di utilizzo dell'algoritmo di *ChatGPT* sono legate anche allo studio di documenti, con grandi miglioramenti da parte del software nella lettura, comprensione e analisi di documenti di testo anche di grandi dimensioni.

Tra i benefici che offre *ChatGPT*, nella versione a pagamento, vi sono i cosiddetti *GPTs Personalizzati*, promettendo la possibilità di "programmare" un assistente personale, in grado di rispondere in maniera verticale a necessità specifiche, avanzate dall'utente [47].

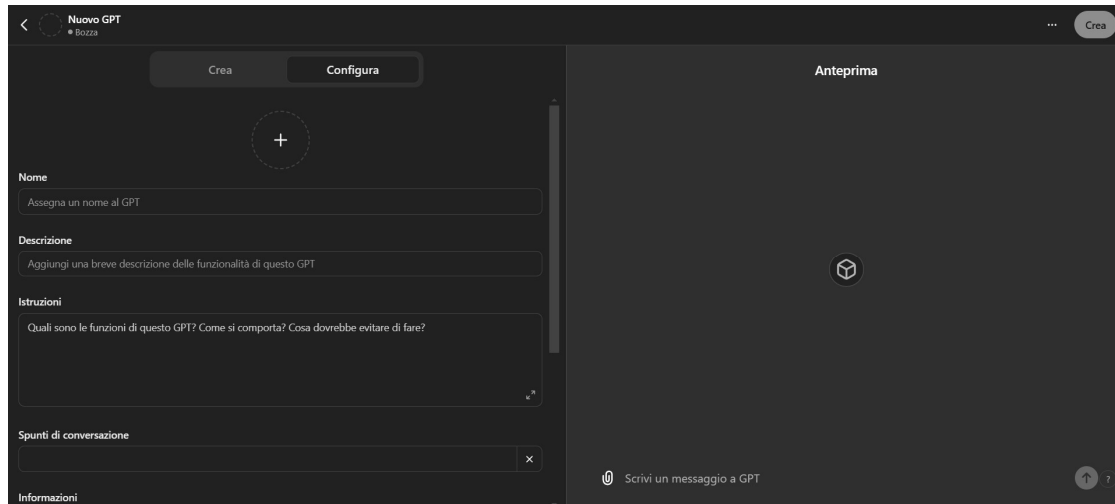


Figura 29 - Schermata di configurazione di un GPT Personalizzato

Inoltre, le modifiche recenti della versione 4o implementano una memoria nell'algoritmo, che di base immagazzina tutto ciò che viene comunicato, per rendere interconnesse chat differenti, potendo comprendere anche riferimenti anche a conversazioni precedenti. Chiaramente, le informazioni contenute nella memoria del software sono modificabili ed eliminabili, mantenendo solo quelle utili [38].

Le ultime novità, ancora in merito a *ChatGPT*, si individuano nell'utilizzo di *Canvas*, che creano dei documenti di testo editabili, puntualmente, in tempo reale sia dall'utente, sia dall'algoritmo, aumentando le possibilità di collaborazione. Oltre a ciò, implementano una semplificazione di alcune domande standard, come l'aumentare o il diminuire la lunghezza del testo, o le richieste di elaborazione, tramite degli slider, snellendo l'utilizzo di prompt ripetuti e circoscrivendone l'utilizzo a modifiche più importanti [39].



Figura 30 - Esempio di utilizzo di Canvas su ChatGPT

Le potenzialità dei nuovi strumenti esistono anche nell'ambito della generazione di immagini, video e audio, spesso proponendo dei flussi di lavoro per generare interi video

solamente con l'utilizzo di strumenti come *Midjourney*, *Runway Gen-3* e generatori di tracce audio, come *Eleven Labs*. Ciò ha portato alla crescita di prodotti in *Full Generative AI* che fanno del realismo e della velocità di realizzazione dei grandi punti di forza. Per questi motivi, case di sviluppo di tool di *IA Generativa*, come *Runway*, hanno iniziato ad organizzare concorsi e festival mondiali per film interamente realizzati con l'*Intelligenza Artificiale*: è il caso dell'*AIFF*⁵, *AI Film Festival*, arrivato alla terza edizione annuale.

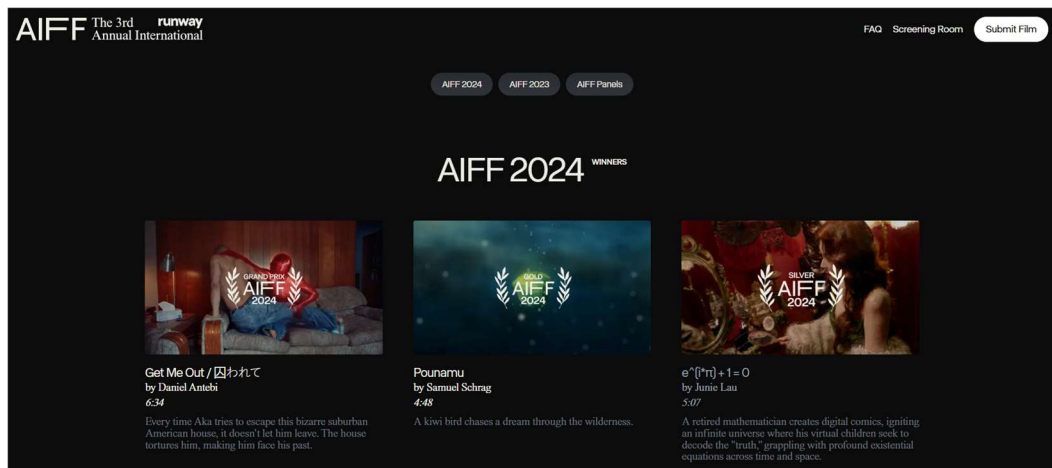


Figura 31 - Schermata di visualizzazione dei film in AI

Contestualmente, gli aggiornamenti dei software rivelano una sempre crescente consistenza nei risultati e nuove opportunità creative di utilizzo degli strumenti, come il *Motion Brush* su *Runway Gen-2*, che permette l'animazione di singoli elementi in un'immagine, selezionandoli con un pennello e muovendoli con dei semplici slider, rappresentativi dei tre assi cartesiani. Oppure ancora, il nuovo strumento di *Motion Capture Generativa* proposto da *Runway*, chiamato *ActOne*, permette di animare personaggi generati con l'*Intelligenza Artificiale* del software, utilizzando il video di un attore reale per catturarne i movimenti [30].

Gli strumenti di *IA Non Generativa* espandono queste prospettive, velocizzando e migliorando molti processi di post-produzione. È il caso del montaggio tramite le trascrizioni di clip su *Davinci Resolve* o la pulizia e separazione delle tracce audio automatica sullo stesso software. Tali strumenti mettono a disposizione nuovi scenari produttivi, insieme alle possibilità di miglioramenti nella realizzazione dei *Visual Effects*, che integrano tracking puntuali più veloci, come *IntelliTrack* su *Davinci Resolve*, e possibilità di generare e integrare elementi nel girato in maniera più rapida, in un workflow che utilizzi il *Generative Fill* di *Photoshop* e il tracking automatico su *AfterEffects*.

Le potenzialità di questi strumenti nella post-produzione offrono anche nuovi flussi di lavoro per il *color grading*, con l'utilizzo di tool come *Fyilm.AI*, che applica una correzione colore con algoritmi di AI. Questo strumento integra la possibilità di riconoscere un certo

⁵ Per approfondimenti ed esempi: <https://aiff.runwayml.com/screening-room>

color grading da un'immagine di reference e applicarlo all'immagine da colorare, rivoluzionando il workflow velocizzando enormemente il processo. Ciò avviene perché la ricerca di reference è alla base di ogni attività nell'industria cinematografica e *Fyilm.AI* offre uno strumento per renderle non solamente un punto di riferimento, ma anche un asset da utilizzare nel flusso di lavoro.

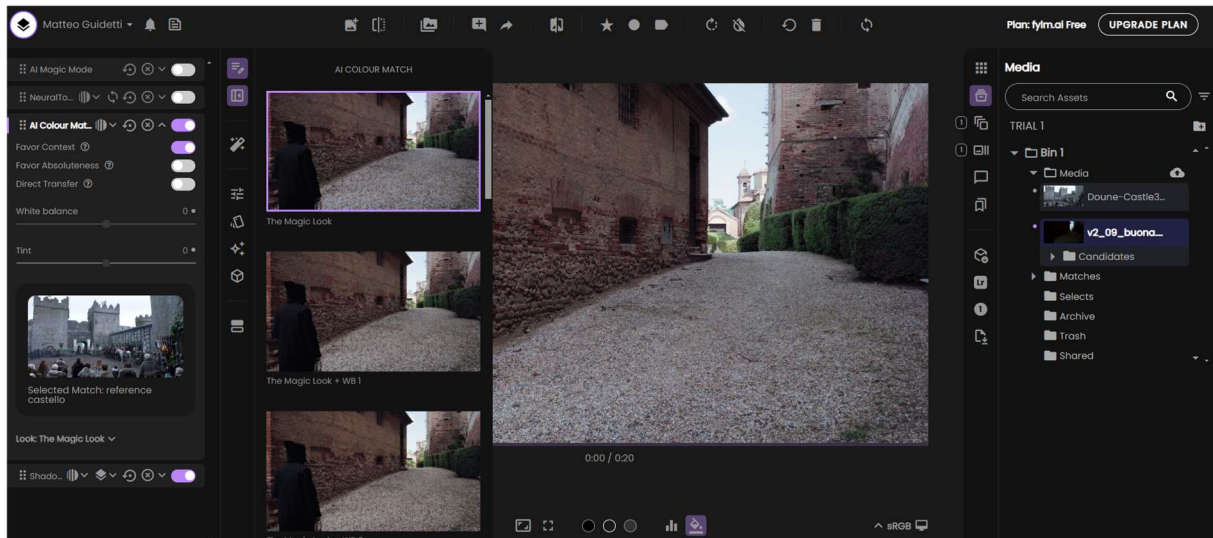


Figura 32 – Interfaccia di *Fyilm.AI*

Le potenzialità offerte dagli strumenti di *Intelligenza Artificiale* aprono nuove opportunità nell'ambito della produzione cinematografica. I ruoli tradizionali subiscono delle variazioni nelle mansioni svolte, con più possibilità di accrescere la qualità e la velocità di esecuzione.

Dal punto di vista fotografico, i movimenti di camera diventano delle descrizioni dettagliate o, al più, lo spostamento di alcuni slider, così come il taglio dell'inquadratura o lo schema di luci. L'approccio *trial and error* di questi strumenti permette l'esplorazione rapida ed efficace di differenti punti di vista, ottenendo, così, una maggiore varietà nei modi di raccontare storie.

A livello registico e di sceneggiatura, le potenzialità date dagli strumenti offrono nuove modalità di scrittura, integrando nel flusso anche la creatività dell'*Intelligenza Artificiale*. Gli strumenti di generazione del testo offrono anche una velocità maggiore nell'estrarre storie da documenti di archivio, proponendo una scelta tra tutte le possibili trame da raccontare contenute nei documenti analizzati. Questo apre a nuove possibilità in sceneggiatura: narrare punti di vista differenti con le stesse immagini.

Inoltre, tutta la filiera della post-produzione e Visual Effects può subire molte modifiche nei ruoli, dove la ricerca e la creazione degli assets può essere sveltita ed implementata con l'*Intelligenza Artificiale*, oltre che tutte le operazioni più macchinose come il *tracking* e il *rotoscoping*. Un esempio di rivoluzione con l'*IA* nel campo dei Visual Effects è il tool *Wonder Dynamics*, che velocizza tutte le operazioni che si celano dietro una corretta realizzazione di

un effetto visivo. In particolare, si occupa dell'integrazione di personaggi digitali in una scena reale, utilizzando i movimenti di un attore vero come base per creare un rigging e un'animazione corretta del personaggio animato. In più, integra anche tutte le operazioni di lighting del character e compositing. Oltre a questo, l'utilizzo del tool è integrato con i più importanti software di 3D, come *Blender*, *Maya* e *Unreal Engine*.

L'accessibilità a questi strumenti rende più democratiche le facoltà di espressione della creatività, in maniera più indipendente da risorse e possibilità, prevedendo come unico metro di paragone la creatività di chi li usa.

All'atto pratico, ciò si traduce in tre grandi potenzialità che questi strumenti offrono. La prima è una semplice questione di velocizzazione di alcuni processi, spesso quelli più laboriosi.

La seconda è una diretta conseguenza della velocità. Una maggiore velocità offre potenzialità creative più ampie e sfaccettate, rendendo spesso possibile creare nuove storie, asset o versioni, in maniera più rapida, abbattendo i compromessi con le tempistiche e lasciando, in presenza di risorse, che sia la creatività a guidare il processo.

La terza è, in realtà, un'unione delle due precedenti. Il risparmio di tempo e le grandi capacità creative possono essere utili in contesti in cui la scarsità di risorse precluderebbe il portare a termine un prodotto, portando i vantaggi di questi strumenti a colmare gli svantaggi di una produzione.

2.2 Limiti da riconoscere

Un'analisi completa impone di soffermarsi anche sugli aspetti che costituiscono ancora delle limitazioni per questi strumenti, in modo tale da comprendere non solo dove possono spingersi ora, ma anche dove potranno migliorare in futuro.

Il primo, forse maggiore, vincolo è da individuarsi nella modalità di utilizzo degli strumenti di *IA Generativa*. La scrittura di un prompt testuale efficace è, infatti, da considerarsi una scienza non sempre esatta. Questo costituisce una seria restrizione, in quanto preclude la possibilità di ottenere dei risultati all'altezza, che sono comunque molto dipendenti dalla qualità del testo in input, piuttosto che dall'originalità dell'idea. Tuttavia, gli sviluppatori di molti tool di *IA Generativa* hanno già riconosciuto questa come una limitazione e impiegato delle strategie per arginarla. Gli esempi di *motion brush* o degli slider integrati nei *canvas* di *ChatGPT* fanno già parte delle soluzioni adottate nell'ottica di rendere questo freno meno ostile. Oltre a ciò, per le casistiche in cui un prompt non possa essere riassunto con uno slider, *Runway* ha integrato nel suo strumento di generazione video *Gen-3* una lista di prompt preimpostati, utili ad esempio per i movimenti di macchina più complessi. Questi prompt sono costituiti da una frase standard con degli spazi vuoti in cui inserire le caratteristiche proprie del video da generare.

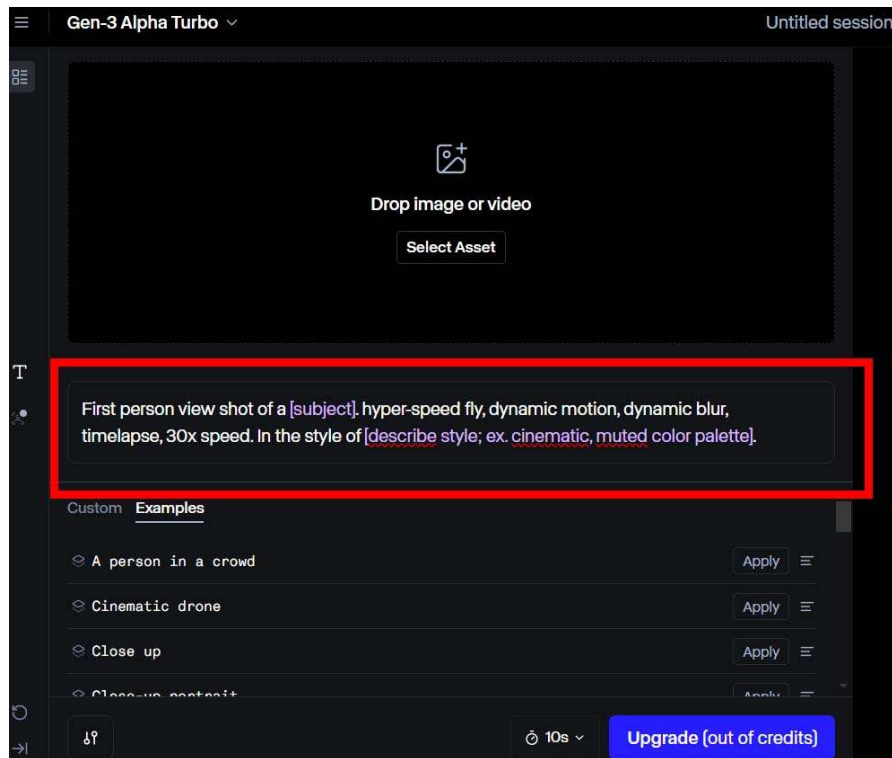


Figura 33 - Esempio di prompt costruito per Gen-3

In generale, le limitazioni date dal prompting si ripercuotono anche sulle possibilità di controllo dell'output. Spesso, l'interpretazione del prompt genera delle incomprensioni. Oltretutto, in un processo che fa dell'iterazione la propria forza, le possibilità di operare modifiche più fini all'output ottenuto senza rigenerarlo interamente sono contenute. Chiaramente, questo non è vero in valore assoluto per tutti gli strumenti: *ChatGPT* e i generatori di testo implementano questa funzione nel proprio output, così come molti generatori di immagini come *Midjourney* e *DALL-E*. Tuttavia, per i generatori di video e audio questo risulta essere ancora un freno.

Un'altra limitazione risiede nella consistenza delle immagini. La consistenza è ciò che permette a due inquadrature consecutive di essere percepite come appartenenti alla stessa scena, quindi nello stesso tempo e spazio. Ma non solo, permette anche di individuare un personaggio per le proprie caratteristiche e di riconoscerlo per le stesse durante tutta la durata del film. Utilizzando un semplice prompt testuale, sebbene contenente la descrizione dettagliata della scena e dei personaggi, la generazione non sarà consistente con le altre. Questa falla viene risolta, nel caso della generazione di video, utilizzando dei workflow più avanzati che partano dalla generazione di immagini con *Midjourney*, tool che integra la possibilità di utilizzare un'immagine come reference per i personaggi, rendendoli quindi consistenti. Le immagini generate saranno poi utilizzate come prompt in un qualsiasi strumento di generazione *image-to-video*.

Proseguendo nell'analisi dei limiti, gli strumenti di *IA Generativa* sono soggetti a quelle che vengono chiamate *allucinazioni*. Questo fenomeno si verifica quando l'output ricevuto contiene, in maniera parziale o totale, degli errori, producendo qualcosa di privo di senso o incorretto [48].

Con riferimento a generazioni di immagine, video o audio, una restrizione che accomuna tutti gli output è spesso un look specifico, con dettagli poco definiti e, nel caso dei video, con un rumore di fondo che cambia vistosamente nel tempo. Anche gli audio risultano a volte robotici e monotoni. Chiaramente, questi risultati sono, a volte, dipendenti anche dalla cura che si dedica al prompt, tornando alla limitazione iniziale.

Se la democratizzazione di questi strumenti è da considerarsi un vantaggio, è altrettanto vero che l'ottenimento di risultati a livello professionale è comunque figlio di competenze tecniche e conoscenze teoriche. Soprattutto in un'ottica in cui, per adesso, le parole che compongono il prompt incidono pesantemente sul risultato ottenuto, conoscere le terminologie del settore e le tecniche porta sicuramente ad una maggiore consapevolezza nella scrittura di un prompt efficace. Inoltre, questi prerequisiti sono necessari per una valutazione più attenta dei risultati ottenuti.

2.3 Verso un nuovo modello applicativo

Alla luce delle opportunità offerte dagli strumenti di *Intelligenza Artificiale* nell'ambito della realizzazione di prodotti cinematografici, è possibile oggi raccontare storie in maniera innovativa, sperimentando nuovi metodi di lavoro.

Tra i principali vantaggi dell'Intelligenza Artificiale in un flusso di lavoro professionale, emerge la sua utilità come supporto aggiuntivo in caso di carenza di risorse, come fondi, personale e tempo. Questo approccio prevede quindi l'integrazione degli strumenti e la modifica di un workflow ostacolato dalla mancanza di mezzi, nell'ottica di rendere meno ostili i vincoli, raggiungendo un risultato professionale.

Si vuole dimostrare che la possibilità di sfruttare un *archivio* di materiale eterogeneo, come video, audio o documenti, possa essere un valido alleato nella mancanza di risorse e soprattutto nella creazione di storytelling diversi a partire dal materiale contenuto. Quest'ipotesi si basa sull'idea che un archivio di materiali possa contenere innumerevoli storie e che l'*Intelligenza Artificiale* sia uno strumento in grado di fornire un consistente aiuto nel considerarle e scegliere quella più adatta allo scopo. Si tratta di un caso in cui, l'*Intelligenza Artificiale* non solamente colma l'assenza di disponibilità, ma illumina anche la possibilità di estrapolare storie da un archivio.

Idealmente, la sceneggiatura cambia la propria collocazione nel workflow, non comandando più le riprese, ma sfruttandole per creare una storia in postproduzione. È quindi da considerarsi, questo, come un cambiamento radicale nel flusso di lavoro proposto

e l'*Intelligenza Artificiale* ne è causa, date le enormi possibilità di analisi di un archivio che propone.

La forma, a livello cinematografico, migliore per accreditare quest'ipotesi è quella del documentario. Questo genere permette di sperimentare al meglio un workflow che ha come radici le riprese di archivio, dalle quali ricavare nuove storie. Il genere documentaristico si adatta perfettamente all'idea che esistano più storie raccontabili all'interno di uno stesso insieme di riprese, coincidendo perfettamente con l'ipotesi proposta. Inoltre, è possibile estremizzare quest'idea realizzando un documentario idealmente senza uno script di base, ma semplicemente accendendo la camera e raccogliendo testimonianze, a voler costituire un archivio di esperienze. Lo storytelling sarà un risultato postumo, realizzato con l'*Intelligenza Artificiale*: la dimostrazione di questo sarà l'obiettivo dell'esperimento presentato nel capitolo successivo.

Capitolo 3

Applicazione pratica

L'Europa è un continente ricco di eterogeneità per quanto riguarda suo tessuto sociale. Snodo importante dell'economia e della geopolitica mondiale, la sua parte continentale è il luogo dove la società occidentale getta le proprie radici. Esistono luoghi, però, che risentono della lontananza dal centro del continente. Luoghi, spesso isole, che si trovano a dover contrastare il fenomeno dell'isolamento, dello spopolamento e delle scarse possibilità di accesso ai servizi digitali. Luoghi in cui i contrasti diventano più marcati: all'esodo costante dei propri abitanti, spesso giovani in cerca di opportunità, si contrappone il turismo di massa, concentrato in pochi mesi all'anno, che contribuisce a rendere più ostili le condizioni per chi, in quei luoghi, ci vuole vivere per più di una sola stagione.

È in questo contesto che si inserisce il progetto *EUROPE*, acronimo di *Remote and Rural Areas of Europe Together*, un progetto KA210 Erasmus+ che investiga le sfide e le opportunità, correlate all'isolamento, nelle aree rurali e remote d'Europa. Questo progetto ha come obiettivo quello di creare delle connessioni tra i luoghi remoti europei, indagati e non dal progetto, promuovendo la cittadinanza attiva, la sostenibilità e la digitalizzazione come antidoto allo spopolamento, al digital divide e al turismo stagionale. *EUROPE* è un progetto che tocca quattro territori europei, rappresentativi di queste tematiche ed estremamente eterogenei per cultura, morfologia e paesaggi. Le quattro località scelte sono: la Sardegna, in particolare il paese di Stintino, l'Islanda, con attenzione al villaggio di Þingeyri, l'isola di Gran Canaria, nell'arcipelago delle Isole Canarie appartenente al territorio spagnolo, e l'isola portoghese di Madeira.



Figura 34 - Logo del progetto EUROPE

La realizzazione di questo progetto nasce dalla collaborazione tra quattro associazioni appartenenti ai territori citati e altre due associazioni di supporto al progetto.

La coordinazione dell'intero progetto è assegnata a *East of Moon*, una piccola-media impresa islandese con sede a Reykjavik. *East of Moon* è un laboratorio di innovazione orientato alla comunità, con l'obiettivo di creare ecosistemi di innovazione volti al sostegno delle comunità rurali e remote europee. Oltre alla coordinazione del progetto, *East of Moon* è responsabile dell'organizzazione del field trip in Islanda, per permettere agli altri partner di osservare da vicino le comunità remote e rurali islandesi, tramite workshop, visite e interviste a persone del luogo.



Figura 35 - Logo di East of Moon

L'associazione no-profit portoghese *Associação Poiomar* è responsabile dell'organizzazione del viaggio a Maderia, scoprendo le sfide di insularità, sostenibilità e sviluppo economico. *Poiomar* è un'associazione fondata sull'isola di Madeira la cui missione è la protezione e conservazione dell'area di Anjos, un piccolo villaggio caratterizzato da coltivazioni di banane e da una crescente offerta turistica. Inoltre, *Poiomar*, insieme ad altri partner del progetto, ha il compito di contribuire alla redazione del *Toolkit*, parte degli output del progetto.



Figura 36 - Logo di Associação Poiomar

L'isola di Gran Canaria è rappresentata dalla piccola-media impresa di *Consulta Europa Projects and Innovation*, che si occupa della pianificazione del viaggio alle Canarie, con particolare attenzione ai temi cardine del progetto e la messa a disposizione di attività in merito ed è, inoltre, leader dell'attività di realizzazione e redazione del *Toolkit*. *Consulta Europa* è un'agenzia di sviluppo locale che ha come obiettivo quello di realizzare progetti innovativi volti allo sviluppo e alla soluzione dei problemi specifici delle isole Canarie, dovuti principalmente alle condizioni di estrema lontananza dal continente. L'obiettivo è

quello di coinvolgere i cittadini in attività che promuovano il territorio e la qualità della vita locale, tramite attività di ricerca e di formazione.



Figura 37 - Logo di Consulta Europa

Absentia è un'organizzazione no-profit fondata a Stintino, in Sardegna, che nel progetto guida il viaggio nel territorio Sardo ed è a capo della coordinazione della produzione del documentario, output principale del progetto. Durante tutti i viaggi nei luoghi suddetti, i rappresentanti di *Absentia* hanno il ruolo di girare il documentario richiesto dal progetto. L'organizzazione promuove attività legate al territorio di Stintino che siano volte alla sensibilizzazione su temi sociali come l'integrazione, lo sviluppo sostenibile ed economia verde.



Figura 38 - Logo di Absentia

La responsabilità della qualità del progetto e del compimento dello stesso è assegnata ad *Open Impact SRL*, una startup italiana, fondata a Roma. *Open Impact* si posiziona nel settore della valutazione e gestione dell'impatto sociale, ambientale ed economico dei progetti. Inoltre, contribuisce alla stesura del *Toolkit*, valorizzandola con le proprie esperienze in ambito di innovazione sociale e gestione dell'impatto. *Open Impact* è uno dei due partner che non si occupa dell'organizzazione di un field trip, ma che ha partecipato a tutte le attività proposte dal progetto nei quattro luoghi remoti e rurali d'Europa.



Figura 39 - Logo di Open Impact

In ultimo, *l'Asociación para el desarrollo socioemocional de la persona, la sociedad y para la cooperación internacional Uno* assume il ruolo di coordinatore e gestore del budget del progetto, insieme ad *East of Moon*. Il suo compito è quello di contribuire a garantire che le attività amministrative e finanziarie siano gestite con precisione. L'associazione *Uno*, fondata a Valencia, mira a potenziare le competenze socio-emotive dei cittadini, con l'obiettivo di creare una società più inclusiva. L'approccio è quello del coinvolgimento di persone nella risoluzione di problemi sociali tramite la collaborazione.



Figura 40 - Logo di Uno

Le connessioni tra questi luoghi si sono realizzate tramite dei viaggi in cui, a turno, i rappresentanti dei partner del territorio ospitavano tutti i portavoce delle altre zone remote e delle associazioni aderenti al progetto. Questo scambio diventava occasione di entrare in contatto con i territori e le realtà locali, tramite visite, interviste e workshop da parte di chi, in quelle regioni, ha trovato le proprie opportunità e ne ha fatto strumento per risolvere le sfide date dall'isolamento. L'intento finale del progetto è quello di creare connessioni tra i quattro luoghi remoti europei, condividendo problemi e soluzioni a tematiche comuni, oltre che creare una guida per tutti gli altri luoghi isolati europei che affrontano gli stessi problemi e che possono trovare negli output di questo progetto una risorsa utile nel superare le sfide.

Gli obiettivi e le aspettative del progetto sono rivolti alla creazione di reti collaborative tra i territori a supporto di una transizione sostenibile e di un rafforzamento della capacità di agire dei partner, frutto dell'apprendimento reciproco. Oltre a ciò, anche la valorizzazione delle realtà locali e delle soluzioni adottate da loro nell'affrontare le sfide rurali contemporanee. Infine, tra gli obiettivi c'è l'aumento della consapevolezza dei decisori politici in merito alle sfide e alle opportunità che i territori remoti e rurali europei affrontano.

Gli output finali e tangibili del progetto *EUROPE* sono un documentario e un *Toolkit*. Il documentario ha l'obiettivo di presentare al meglio le sfide e le opportunità che questi luoghi affrontano, ascoltando le voci di chi, nell'isolamento, ha trovato delle condizioni favorevoli. Il *Toolkit* è un documento che include le attività svolte, le ricerche e le metodologie, da condividere gratuitamente, con l'obiettivo di poter contribuire nella connessione e nello sviluppo di tutti quei luoghi rurali e remoti in Europa che non hanno preso parte al progetto.

La realizzazione del documentario è stata commissionata dagli organizzatori del progetto *EUROPE* a Mattia Meloni, in quanto membro dell'associazione *Absentia*, esperto Film-maker e collaboratore esterno dei corsi di Ingegneria del Cinema e dei Mezzi di Comunicazione presso il Politecnico di Torino. Le risorse destinate alla realizzazione del documentario per il progetto sono molto limitate, sufficienti a coprire le spese solamente per la produzione sul set del documentario, con una troupe composta unicamente da Mattia Meloni.

È in questo istante che il compimento del documentario diventa l'esempio perfetto per l'applicazione pratica dell'ipotesi mostrata nel capitolo precedente. Si tratta di un caso in cui è possibile provare a sopperire alla mancanza di risorse, in particolare denaro, persone e, di

conseguenza, tempo, con l'utilizzo sperimentale degli strumenti di *Intelligenza Artificiale*. È un caso in cui l'*Intelligenza Artificiale* non diventa sostituto di risorse, in particolare umane, ma diventa una possibilità per chi di risorse non ne ha a sufficienza.

3.1 Strumenti utilizzati

Tra gli strumenti elencati nello stato dell'arte, descritto nel primo capitolo, ne sono stati scelti alcuni che avessero le caratteristiche migliori per l'applicazione del caso studio, con la possibilità di utilizzarli in versione gratuita, laddove attuabile.

Lo strumento più utilizzato in tutte le fasi del progetto è *ChatGPT*. Questo ha assunto vari ruoli all'occorrenza, conducendo sia operazioni meramente automatizzate e meccaniche, sia operazioni di stampo creativo. Per la buona riuscita dell'esperimento è stato sottoscritto un abbonamento alla versione pro, che permettesse di accedere a strumenti più avanzati, abbattendo i limiti di condivisione di documenti e di lunghezza delle conversazioni. Nel corso dell'esperimento sono state utilizzate varie versioni dell'algoritmo, a seconda degli aggiornamenti. In particolare, si è fatto uso delle seguenti versioni: *GPT 3.5*, *GPT 4o* e *GPT 4o con canvas*, in preview.



Figura 41 - Logo di ChatGPT

Avendo sottoscritto l'abbonamento pro a *ChatGPT*, questo include anche la possibilità di utilizzare il generatore di immagini *DALL-E 3*, sviluppato da *OpenAI*. Pur non essendo il migliore sul mercato, il generatore di immagini è servito nella fase di montaggio video, nel workflow di generazione di una storia animata in 2D.

Il software di editing video scelto è *Davinci Resolve Studio 19*, perché al suo interno fornisce molti strumenti di *Intelligenza Artificiale*, gestiti dal *Neural Engine* di Blackmagic Design. È il caso di strumenti utilizzati, come la trascrizione automatica o la pulizia dell'audio.



Figura 42 - Logo di Davinci Resolve Studio

In alcuni casi, a supporto di *ChatGPT* è stato impiegato anche *Claude 3.5*, utile per suddividere il carico di lavoro meccanico del primo e per una maggiore facilità ad ottenere risultati che rispettassero il prompt, nelle fasi iniziali.



Figura 43 - Logo di Claude AI

Per la generazione delle voci narranti è stato utilizzato il tool *Eleven Labs* perché garantisce risultati molto migliori rispetto ai concorrenti citati: il tool prevede un sistema di crediti mensili che ne permettono un utilizzo base.

II ElevenLabs

Figura 44 - Logo di ElevenLabs

L'animazione delle immagini è stata sperimentata utilizzando il tool di generazione video *Runway Gen-2*, che include la funzione di *motion brush* e *camera control*, facilitando molto il prompting. Anche in questo caso esiste un sistema di crediti che permettono un utilizzo base, gratuitamente.



Figura 45 - Logo di Runway

Infine, il flusso di lavoro sperimentato superficialmente per la generazione delle musiche prevede l'utilizzo del tool *Stable Audio*, che genera tracce musicali: il tool è utilizzabile gratuitamente in una versione di prova con un sistema di crediti ricaricati mensilmente.

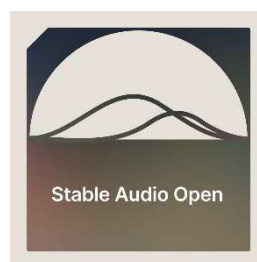


Figura 46 - Logo di Stable Audio

L'utilizzo di questi tool è dettagliato maggiormente nel corso del capitolo, insieme al workflow proposto per la produzione del documentario *Forgotten Frontiers: where isolation meets innovation*.

3.2 Pre-produzione

Nel caso del documentario da realizzare, la pre-produzione consiste nella preparazione a girare le riprese. La prima cosa fondamentale è che l'assenza di risorse ha come conseguenza l'impossibilità di ingaggiare uno story editor per creare uno storytelling, che serva da guida durante le riprese. In questo modo la decisione è ricaduta sul creare uno storytelling eseguito successivamente alle riprese utilizzando l'*Intelligenza Artificiale*, rompendo così lo schema tradizionale, che vuole la sceneggiatura come mappa della narrazione, dall'inizio alla fine.

Per questo motivo, il compito della pre-produzione è quello di preparare la raccolta dell'*archivio* di documentazioni della realtà, da utilizzare, a posteriori, per scrivere la sceneggiatura.

In particolar modo, i passaggi importanti per gettare le basi della creazione dell'*archivio* sono lo studio del campo di ricerca e la scrittura delle domande da porre agli intervistati e, per portarli a termine, lo strumento utilizzato è *ChatGPT*.

3.2.1 Studio del campo di ricerca

Che si utilizzi il metodo tradizionale o questo metodo sperimentale, la preparazione di un documentario deve necessariamente essere figlia di un attento studio dell'argomento. Già in questa prima fase della pre-produzione l'*Intelligenza Artificiale* si rivela essere un valido alleato nella mancanza di risorse.

Lo studio degli argomenti è stato condotto, infatti, utilizzando le potenzialità di *ChatGPT*, ancora nella sua versione 3.5, per ricavare più velocemente le informazioni realmente utili allo scopo. Per questo motivo, le conversazioni con *ChatGPT* sono state impostate suddividendole per luoghi, in cui ogni chat diventa la fonte principale dello studio per quella regione. L'approccio è quello di iniziare con la raccolta delle informazioni più generali sul territorio, che possano dare un contesto anche a chi studia e permettere di entrare a conoscenza di prospettive che esulino dai temi specifici del progetto *Europe*. Successivamente, dopo aver gettato le basi, la conversazione prosegue con lo studio più approfondito delle tematiche che interessano il progetto *Europe*, chiedendo a *ChatGPT* un'attenta analisi su quali siano le condizioni inerenti al tema in ogni luogo.

La parola chiave quando si tratta di *LA Generativa* è *reiterazione*. Ogni risultato che si ottiene dalla generazione del testo di *ChatGPT* è sempre migliorabile chiedendo approfondimenti, chiarimenti o riscritture. In questo caso, la reiterazione consiste nel continuare a porre domande allo strumento, decidendo quale aspetto delle risposte precedenti approfondire e quale accantonare, e se accantonarne.

Naturalmente, gli output di questa fase sono dei documenti di testo in cui vengono salvate tutte le informazioni, in modo tale da avere dei *database* contenenti gli aspetti più importanti da conoscere per ogni regione.

Il vantaggio principale di utilizzare *ChatGPT* durante lo studio degli argomenti è la rapidità di accesso alle informazioni. Infatti, questo strumento ricava la propria risposta accedendo in contemporanea a moltissime fonti presenti sul web, in modo tale da ricavare la risposta migliore ad una domanda specifica: ad esempio, può essere richiesto un particolare argomento come lo spopolamento in Islanda. Di conseguenza, in un contesto di mancanza di risorse, come il tempo, uno strumento che velocizzi il processo di accesso alle informazioni risulta essere efficace. Oltretutto, data l'impossibilità dovuta allo scarso budget, l'utilizzo di *ChatGPT* si colloca anche come sostegno alla raccolta di informazioni, tradizionalmente prelevate direttamente dai luoghi investigati, consentendo così di documentare regioni e culture sconosciute, pur non visitandole fisicamente.

Il rovescio della medaglia è uno dei difetti propri di *ChatGPT*: l'attendibilità. Estrahendo le informazioni da moltissime fonti sul web, la possibilità di incappare in informazioni falsate esiste e non può essere sottovalutata. Il modo migliore per aggirarla è la reiterazione delle domande, verificando la consistenza delle risposte ed evidenziando eventuali falle.

3.2.2 Domande interviste

Dopo lo studio degli argomenti, la pre-produzione del documentario prosegue con la stesura delle domande da fare alle figure che si incontreranno durante i viaggi nei quattro punti remoti d'Europa. Non avendo una sceneggiatura da seguire né, tantomeno, consapevolezza di tutte le figure che si potranno incontrare, è indispensabile creare un database contenente tutte le possibili domande, da utilizzare all'occorrenza. Le domande saranno utilizzate durante le interviste che costituiranno il cuore dell'*archivio* su cui si basa quest'applicazione sperimentale. L'obiettivo è quindi quello di popolare un *archivio*, il più fornito possibile, per avere, durante la post-produzione, un ventaglio più ampio di possibilità nella generazione di uno *storytelling*.

- Introduzione
1. Potresti iniziare presentandoti e raccontandoci un po' della tua esperienza?
 2. Descrivi la tua comunità. Quali sono le sue principali caratteristiche e risorse?
 3. Quali sono i principali vantaggi e svantaggi di vivere in un'isola piccola o in un villaggio remoto/ rurale?
 4. Come si è evoluta la tua comunità negli ultimi anni? Quali cambiamenti hai osservato?
 5. Quali sono le principali fonti di reddito per le persone nella tua comunità?

Figura 47 - Esempio di alcune domande generali

Le domande sono state suddivise in categorie: prima alcune domande più generiche sulla persona intervistata e sul luogo, poi delle domande più approfondite per ogni tematica trattata, spesso tarate su ciò che si è appreso durante la fase di studio riguardo le regioni considerate. La scrittura di queste domande è stata svolta con l'utilizzo di *ChatGPT*. La richiesta, nello specifico, prevede la scrittura delle domande generiche e poi quella delle domande più specifiche, allenando lo strumento con le documentazioni ottenute durante

lo studio dei luoghi. In questo modo, il software ha un contesto di base da cui partire e un recinto entro il quale rimanere.

La metodologia utilizzata per sottoporre le richieste al software, in questo caso, è iterativa su due differenti livelli. Il primo, come già descritto nel precedente paragrafo, è un livello di iterazione sullo stesso gruppo di domande, ad esempio quelle inerenti al tema dello spopolamento. Per fare ciò, si utilizza l'iterazione volta a migliorare l'output e ad ottenere delle domande che siano più pertinenti o, semplicemente, in maggior numero. Questo permette quindi di modellare l'output finale a seconda del gusto personale, includendo, escludendo o modificando le domande in modo opportuno.

Il secondo livello di iterazione mira a frammentare l'input dato a ChatGPT, sottoponendo al software un solo gruppo di domande da scrivere, per concentrare tutta la potenza di calcolo su una singola istruzione. Un modo di agire ben più noto come *Divide et impera*. Pertanto, ci si concentra dapprima sulle domande riguardo lo spopolamento e, una volta ritenuto il risultato soddisfacente, si sposta il focus sul gruppo di domande successivo, ad esempio l'invecchiamento. Questo approccio sarà alla base dell'intero flusso di lavoro per la generazione dei testi.

- Spopolamento*
1. Quali sono le principali ragioni alla base dello spopolamento in questa comunità?
 2. Quali iniziative sono state intraprese per contrastare lo spopolamento e quali sono state le loro conseguenze?
 3. Come viene affrontata la questione dell'attrazione di nuovi residenti in questa comunità?
 4. Quali sono gli sforzi per mantenere i giovani nella comunità e quali opportunità vengono offerte loro?
 5. Quali iniziative vengono intraprese per promuovere lo sviluppo economico e creare opportunità di lavoro nella comunità?

Figura 48 - Esempio domande ottenute per il tema dello spopolamento

Una volta raggiunta la piena soddisfazione del risultato finale, questo sarà il punto di partenza per le interviste sul set, incrementando le possibilità di adattamento in base alla persona intervistata e, in potenza, adattandosi anche ad intervistare persone fuori programma, espandendo la dimensione dell'*archivio*.

L'utilizzo di *ChatGPT* nella stesura delle domande va incontro all'esigenza di colmare l'assenza di risorse, in questo caso temporali, velocizzando il processo creativo, senza snaturarlo, date le possibilità di perfezionamento del processo di iterazione, da parte di una mente umana.

3.3 Produzione e riprese (cenni)

La fase successiva alla pre-produzione è quella delle riprese che, di per sé, non propone nulla di innovativo o di differente rispetto al metodo tradizionale. Quello che cambia è invece lo scopo delle riprese, ora volte alla raccolta di immagini che possano costituire

L'*archivio* alla base del processo di lavoro. La fase di ripresa sul set, inoltre, non si basa su uno script predeterminato, ma su un *database* più versatile di domande, che andranno scelte sulla base di fattori differenti.

Durante le riprese, il *database* di domande costituisce la linea guida per le interviste e le domande vengono scelte sulla base della qualifica della persona intervistata. Ad esempio, se si tratta di un assessore alla digitalizzazione saranno sicuramente incluse domande riguardanti il digital divide. Oltre a ciò, anche il fattore tempo diventa determinante nella scelta delle domande da porre, potendone ampliare o restringere il numero a seconda del tempo a disposizione. Perciò, prima di ogni intervista è necessario costruire una traccia, estraendo le domande pertinenti dal database e redigendo un file *ad hoc* per l'intervistato. Anche in questo caso si rivela utile l'impiego di *ChatGPT*, che fornisce un valido ed immediato aiuto nel tradurre le domande dall'italiano alla lingua con cui si svolgerà l'intervista.

Naturalmente, le domande costituiscono solamente un filo conduttore, ma la direzione presa dall'intervista potrebbe portare ad esplorare punti di vista nuovi, non considerati dal file di traccia.

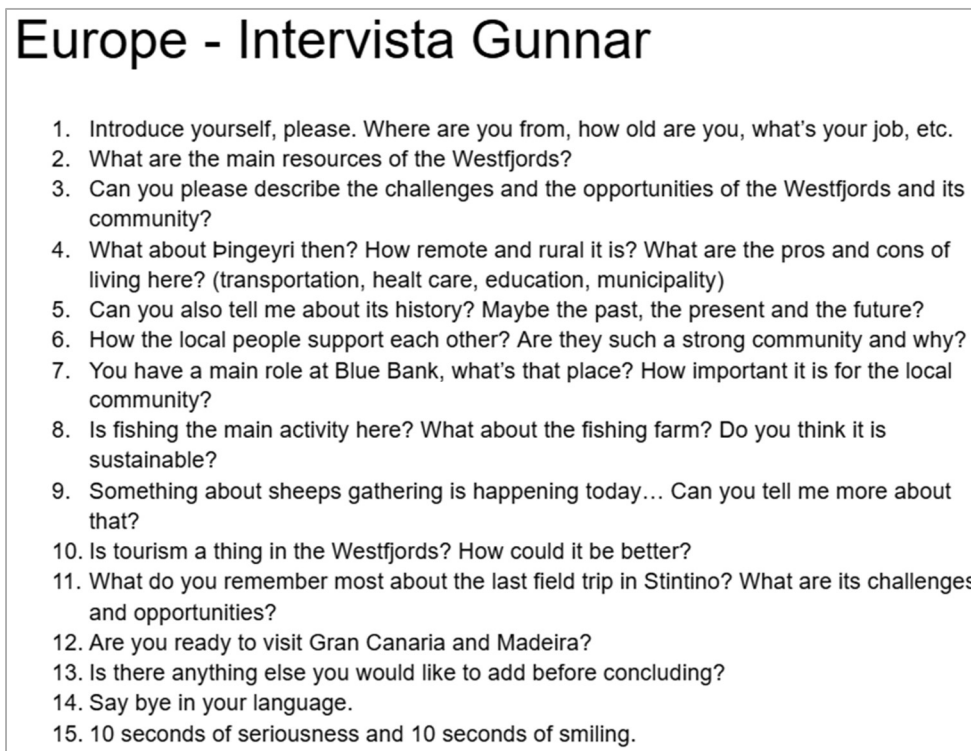


Figura 49 - Esempio file di domande per un intervistato

3.4 Post-Produzione

Fino a questo punto, sono state affrontate le fasi della creazione dell'*archivio*, utile a mostrare l'applicazione sperimentale dell'ipotesi presentata. La post-produzione è la fase in cui la

sperimentazione prende vita concretamente. L'obiettivo è quindi quello di utilizzare il materiale contenuto nell'archivio, appena creato, come punto di partenza per generare uno storytelling che possa condurre alla produzione finale di un documentario.

Più nello specifico, verrà presentato un workflow che parta dalle immagini grezze contenute nell'archivio e che giunga all'export di un documentario che rispetti gli obiettivi presentati inizialmente dal progetto *Europe*. Gli strumenti di *Intelligenza Artificiale* saranno alla base del flusso di lavoro proposto, contribuendo a rendere meno ostile la mancanza di risorse per la produzione.

Il punto di partenza è, quindi, l'*archivio*, che consiste in cartelle contenenti il girato grezzo, registrato durante i quattro *field trip* nelle comunità rurali e remote d'Europa. In particolare modo, le cartelle sono suddivise nei quattro luoghi toccati dal progetto *Europe*, Sardegna, Islanda, Gran Canaria e Madeira, e, per ognuno di questi, le clip sono state suddivise in cartelle giornaliere. Il contenuto dell'*archivio* è un insieme di: interviste alle persone incontrate durante il progetto, clip paesaggistiche che raccontano la bellezza di questi luoghi, riprese dei workshop che sono stati seguiti nel corso della permanenza e momenti di confronto e condivisione tra i protagonisti del progetto.

Le persone che sono state intervistate, le cui clip girate sono contenute nell'archivio, sono le seguenti.

Sardegna:

- Fabrizio Contini, host del progetto a Stintino e fondatore di *Absentia, Exploring the absence*, tra i partner del progetto;
- Agostino Schiaffino, pescatore e imprenditore ittiturismo a Stintino;
- Carlo Coni, fondatore di *Treballo*, una house di co-working a Laconi;
- Giada Moschella, giovane abitante locale, facente parte della consulta giovanile nel comune di Stintino;
- Giovanni Conti, imprenditore locale e fondatore di *Relicta*;
- Rita Vallebella, sindaco di Stintino;

Islanda:

- Arnar Sigursson, coordinatore dell'intero progetto *Europe*, host del progetto nel villaggio di Þingeyri e fondatore di *East of Moon*, tra i partner del progetto;
- Magdalena, giovane ragazza polacca trasferitasi a Þingeyri;

Gran Canaria:

- Beatrice Avagnina, host del progetto nelle Isole Canarie e managing director in *Consulta Europa*, uno dei partner del progetto;

- Jay Mendes, nomade digitale portoghese;
- Jaime, fondatore del villaggio ecologico *Selva Doramas* a Gran Canaria;

Madeira:

- Markus, fondatore dell'ecovillaggio *Arambha*;

3.4.1 Scrittura

Tipicamente, sarebbe impensabile collocare la scrittura di un documentario nella fase di post-produzione. La sceneggiatura è la base di partenza di tutta la pre-produzione di un prodotto cinematografico. Nel workflow presentato, invece, la sceneggiatura assume una connotazione differente, non unica e immutabile come quella originale, ma scelta tra le varie possibilità proposte dall'*archivio* e soggetta costantemente a revisioni, aggiornamenti e modifiche anche durante la post-produzione. È quindi un prodotto successivo, che nasce dal contenuto dell'*archivio*.

La stesura della sceneggiatura completa è la prima fase del montaggio del documentario. Per fare ciò, l'*Intelligenza Artificiale* assume il ruolo che, in presenza di risorse, prenderebbe in carico uno *Story Editor*, ossia colui che ha il compito di individuare, innanzitutto, quale storia raccontare tra tutte quelle presenti nell'*archivio* e anche il modo di farlo efficacemente. Il flusso di lavoro utilizza e intreccia tre strumenti di *Intelligenza Artificiale*, per raggiungere l'obiettivo di scrittura di una sceneggiatura completa per il documentario: la funzione di trascrizione video di *Davinci Resolve*, *ChatGPT* (in versione 4.0) e *Claude*.

Prima di cominciare, è utile prendere nota di alcune linee guida registiche per indirizzare la stesura della sceneggiatura verso un certo tipo di prodotto, escludendone altri a priori. In altre parole, informazioni come la durata, lo stile e le tematiche generali da affrontare sono informazioni da conoscere per creare dei confini alle innumerevoli possibilità. Inoltre, è possibile decidere in principio la struttura della sceneggiatura del documentario, dando una suddivisione in capitoli, così da rendere più fluido il lavoro successivo. Nel caso specifico del progetto, di seguito sono elencate le note stilistiche e di regia utilizzate come input insieme al materiale di *archivio*:

- Formato appropriato alla distribuzione per il web;
- Durata inferiore ai 30 minuti;
- Stile cinematografico;
- Utilizzo di una voce narrante che possa raccontare ciò che non viene detto dalle battute scelte, colorando e arricchendo la narrazione;
- Suddivisione della sceneggiatura in sei capitoli: *Introduzione*, *Sardegna*, *Islanda*, *Gran Canaria*, *Madeira* e *Conclusione*;

- Ogni capitolo deve contenere la voce della persona organizzatrice e portavoce del luogo a cui è associato il capitolo e un'altra voce appartenente ad una persona intervistata sul posto

È doveroso precisare che questo passaggio non è indispensabile ai fini di una buona riuscita dell'esperimento, ma è a discrezione delle persone che si occupano della realizzazione del documentario. Saltare questa fase significa dare una maggiore libertà all'*Intelligenza Artificiale*.

È utile, inoltre, specificare che l'obiettivo finale della scrittura della sceneggiatura è quello di contenere battute e dialoghi il più possibile fedeli alle trascrizioni e, quindi, al parlato delle interviste, per poter concretamente utilizzare il materiale dell'*archivio* in fase di montaggio.

Il flusso di lavoro presentato può essere suddiviso in cinque fasi fondamentali: trascrizione delle interviste, preparazione ed elaborazione del materiale trascritto, analisi e definizione delle linee guida, stesura dei documenti, ricontrollo e scrittura del titolo.

Trascrizione delle interviste

Di per sé, il materiale di archivio costituisce la fonte principale da cui estrarre tutte le possibili storie da raccontare. Tuttavia, il formato con cui viene ricevuto, ovvero delle clip video con audio, non è ottimale per permettere a *ChatGPT* di analizzarlo e creare uno storytelling. Questo perché lo strumento è in grado di analizzare solo materiale testuale. Il modo migliore, quindi, di condividere il materiale di archivio con *ChatGPT* è il formato testuale, che sia sottoforma di file .pdf, .txt, .docx o semplice copia-incolla nel campo di testo (l'ultima possibilità è funzionante, ma sconsigliata nel caso di testi troppo lunghi). Pertanto, occorre trascrivere tutte le interviste presenti nell'*archivio*.

Il metodo prescelto e valutato come più efficace per portare a termine questo primo passaggio è quello di utilizzare la funzione di *auto trascrizione*, che implementa algoritmi di *Intelligenza Artificiale*, integrata nel software *Davinci Resolve Studio 19*. La funzione in questione permette di trascrivere in maniera automatica sia singole clip, sia intere timeline, ottenendo in output una versione della clip, o della timeline appunto, in forma testuale.



Figura 50 - Tool di Audio Transcription su Davinci Resolve

Le interviste presenti nell'*archivio* sono in una forma frammentata in diverse clip, in cui ogni video corrisponde spesso ad una domanda e alla sua risposta. Conseguentemente,

L'approccio scelto è stato quello di creare una timeline per ogni persona intervistata, inserendo in maniera sequenziale tutte le clip dell'intervista, realizzando la cosiddetta "messa in fila".

Una funzione ulteriore, presente dalla versione 19 del software, è quella di *Speaker Detection*. Questa aggiunge una possibilità in più alla semplice trascrizione audio, che risulta spesso caotica, ovvero quella di identificare automaticamente da quale tra gli interlocutori sia stata pronunciata una frase. Nel caso delle interviste, questa funzione rende l'output ottenuto più ordinato e comprensibile. Una volta completata la trascrizione utilizzando la funzione di *Speaker Detection*, è possibile rinominare gli interlocutori manualmente, per avere un maggior ordine. Oltretutto, il riconoscimento dell'interlocutore è valido anche su timeline differenti dello stesso progetto, per cui se, come nel caso dell'esperimento, l'intervistatore è sempre la stessa persona, l'algoritmo lo riconoscerà ogni volta come la medesima persona, attribuendogli automaticamente il nome prescelto.

Ai fini dell'esperimento, la trascrizione delle interviste è stata provata sia utilizzando la funzione di *Speaker Detection*, sia disattivandola. L'attivazione della funzione comporta sicuramente un grande vantaggio in termini di ordine dell'output, ma anche un aumento dei tempi di trascrizione. Questi, infatti, se disattivando la funzione si aggirano ad avere un rapporto *tempo impiegato/lunghezza della timeline* pari circa a 1:2, attivando la funzione di *Speaker Detection* aumentano il rapporto fino ad essere circa 1:1. Perciò dovendo trascrivere una timeline di dieci minuti, senza *Speaker Detection* il tempo impiegato sarebbe di circa cinque minuti, attivando la funzione sarebbe più prossimo ai dieci minuti. Naturalmente, le tempistiche indicate si riferiscono al caso studio nello specifico. È bene, però, precisare che queste ultime variano in relazione alla capacità computazionale a disposizione, velocizzando o rallentando sulla base di questa.

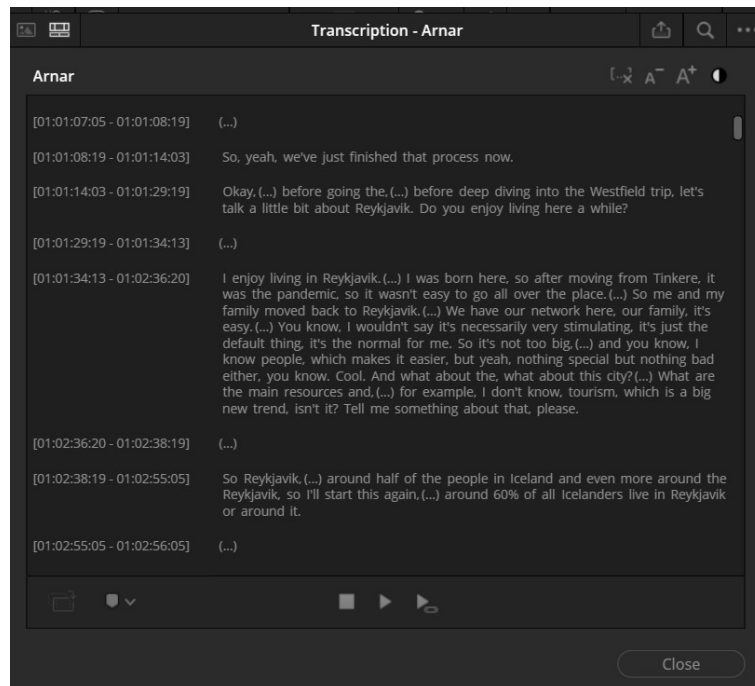


Figura 51 - Output della funzione di Audio Transcription senza Speaker Detection

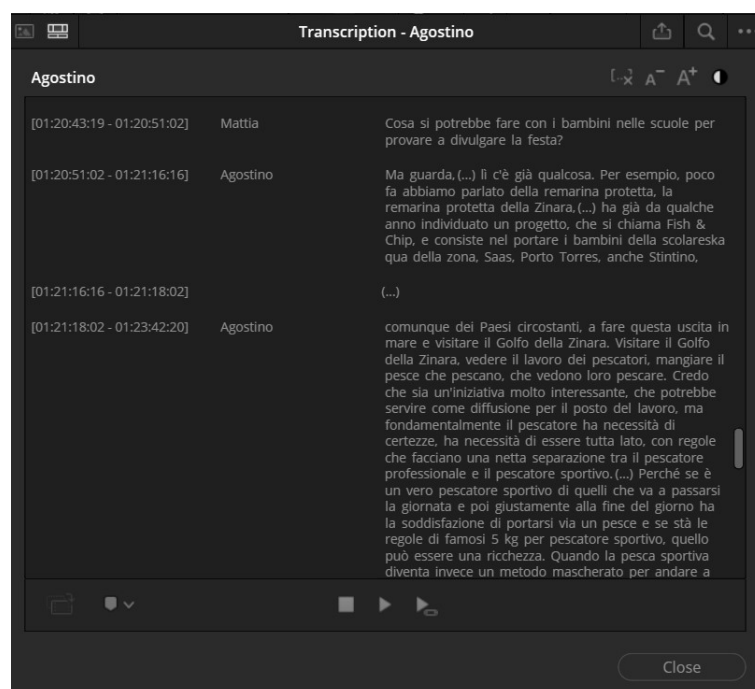


Figura 52 - Output della funzione di Audio Transcription con Speaker Detection

Durante l'applicazione di questa prima porzione di workflow, sono emerse delle criticità nel funzionamento del software di trascrizione. In primo luogo, si è notato che spesso la lingua delle trascrizioni non corrispondeva a quella effettiva con cui veniva condotta l'intervista. Questo perché, naturalmente, gli intervistati non sempre mantenevano una pronuncia standard delle parole, specialmente coloro che non parlavano nella propria lingua

madre. In questo caso, la soluzione risiede nel disattivare la rilevazione automatica della lingua da parte del software e impostare manualmente la lingua in cui viene condotta l'intervista.

Un'ulteriore difficoltà è stata riscontrare che, solo per alcuni frammenti di certe interviste, la trascrizione non fosse andata a buon fine, creando un vuoto, della durata anche di qualche minuto, nel testo. Tutto questo senza apparenti problemi nell'audio dell'intervista. La risposta a quest'incertezza è più elaborata. Per risolvere è stato necessario individuare le parti mancanti, isolarle ed esportarle manualmente applicando lo strumento di pulizia dell'audio e isolamento della voce integrato in *Davinci Resolve*, nonostante non presentassero differenze evidenti di rumorosità nell'audio. Questo strumento di *Voice Isolation* sfrutta algoritmi di *Intelligenza Artificiale* per pulire tracce audio rumorose. Applicando queste correzioni e reinserendo la clip nella timeline, la trascrizione non ha più avuto insidie.

L'export della trascrizione avviene sottoforma di un file di testo che mantiene la stessa forma e spaziatura tra le parole presenti nella trascrizione su *Davinci Resolve*. Tuttavia, indipendentemente dall'utilizzo della funzione di *Speaker Detection*, le informazioni sul nome dell'interlocutore non vengono esportate.

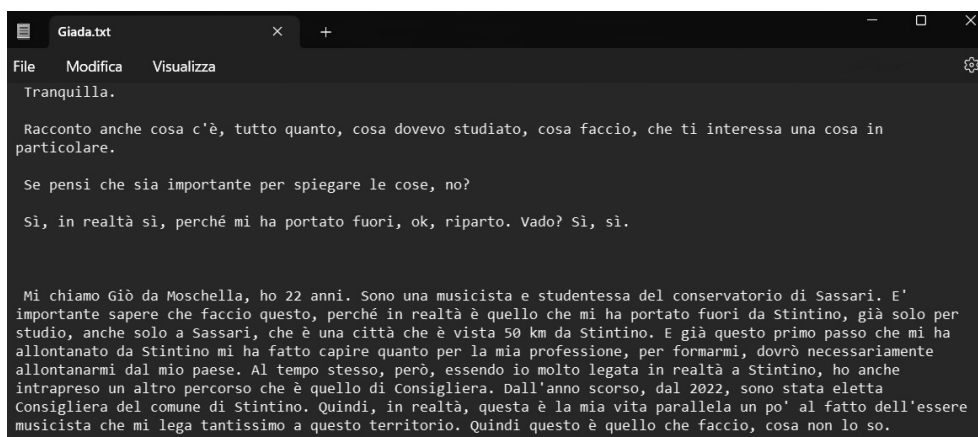


Figura 53 - Esempio di file esportato dalla trascrizione di *Davinci Resolve*

Considerando brevemente il beneficio nella scelta di utilizzare questo strumento specifico, a discapito di altre soluzioni proposte da strumenti di *Intelligenza Artificiale*, la funzione di trascrizione integrata in *Davinci Resolve* comporta dei considerevoli vantaggi nel momento del montaggio video. Consente, infatti, di individuare rapidamente frammenti di video tramite una ricerca testuale, evidenziarli come si farebbe con un normale testo e inserirli in qualunque timeline, già isolati e tagliati correttamente.

Preparazione ed elaborazione del materiale trascritto

Come anticipato precedentemente, l'output della trascrizione di *Davinci Resolve* esportato al di fuori del software mantiene la stessa formattazione, come mostrato nella Figura 53. Pertanto, il documento di testo risulta essere caotico e, di conseguenza, più difficilmente

utilizzabile e comprensibile. Oltretutto, le informazioni sull'interlocutore non sono presenti nell'output, sia attivando la funzione di *Speaker Detection*, sia nel caso contrario. Nonostante ciò, il vantaggio di utilizzare la funzione di riconoscimento dell'interlocutore è quello di avere in output delle battute isolate dei vari interlocutori e non confusamente mischiate come quando non è attiva la funzione.

Per poter permettere a *ChatGPT* di compiere un'analisi coerente e completa è utile riordinare tutti i documenti delle trascrizioni, evidenziando e distinguendo le domande dalle risposte in modo tale che la comprensione del testo non sia un ostacolo nei passaggi successivi.

La forma scelta per riorganizzare i documenti è la seguente: *<Nome dell'interlocutore>*: *<Testo della domanda/risposta>*, come mostrato meglio nell'immagine successiva.

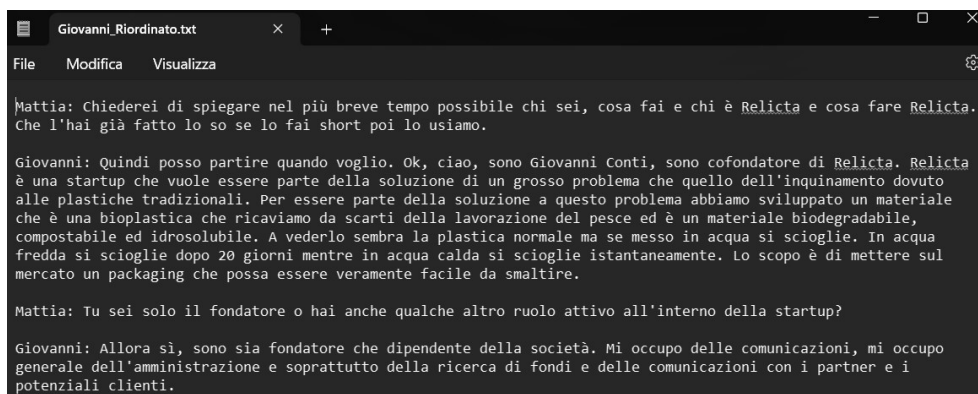


Figura 54 - Esempio di trascrizione riordinata con Mattia intervistatore e Giovanni intervistato

Per ottenere il risultato si è scelto di utilizzare strumenti di *Intelligenza Artificiale* come *ChatGPT* e *Claude*, con l'intento di automatizzare e velocizzare un lavoro macchinoso da fare manualmente.

Per raggiungere l'obiettivo, il prompt utilizzato per le richieste sia a *ChatGPT*, sia a *Claude* è il seguente

```

Ciao, per il mio lavoro di tesi sto trascrivendo delle interviste. Per
farlo ho utilizzato la funzione Transcript di Davinci Resolve 19,
utilizzando anche la funzione di speaker detection. In output ho ricevuto
questo file .txt. Il mio obiettivo è quello di renderlo più leggibile,
riorganizzandolo per differenziare bene le domande dalle risposte.
Sapresti riscrivere questo file, mantenendo le esatte parole contenute in
esso, ma cambiando la formattazione in modo da distinguere le domande
dell'intervistatore, Mattia, dalle risposte dell'intervistato/a, <Nome
dell'intervistato/a>? Utilizza il formato <Nome>: <Domanda/Risposta> sii
preciso nella riscrittura.
  
```

L'obiettivo con cui questo prompt è scritto è quello di fornire in prima istanza un contesto del lavoro che si sta conducendo, in modo tale da permettere all'algoritmo di capire quale

possa essere l'utilità dell'output. Successivamente si è inserita la richiesta dell'operazione da svolgere con tutte le specifiche richieste che devono essere rispettate.

Inizialmente, al prompt veniva allegato l'intero file .txt, ricevuto in output dalla trascrizione di *Davinci Resolve*. Tuttavia, questa pratica non ha mai condotto a risultati soddisfacenti. Infatti, se per la parte iniziale del file, l'algoritmo rispettava la richiesta fatta nel prompt, nel corso della generazione del testo la tendenza diventava quella di andare fuori tema, riassumendo la seconda metà del documento in poche righe, oppure incrementando la creatività aggiungendo domande e risposte fantasiose.

Pertanto, questo problema ha condotto a cercare delle soluzioni, prima esplorando strumenti differenti come *Claude* e poi elaborando una soluzione standard che potesse valere anche per *ChatGPT*. L'approfondimento del software *Claude* ha quindi lo scopo di confrontare i risultati di un'operazione piuttosto ripetitiva, come quella della formattazione del testo, con quelli ottenuti con *ChatGPT*. *Claude*, seppur viziato da molte limitazioni nella sua versione gratuita, si rivela essere uno strumento affidabile, rispettando il prompt indicato, senza aggiungere nulla. Tuttavia, è necessario frammentare il testo da inviare per i limiti dati dalla versione.

La soluzione standard per utilizzare anche *ChatGPT* in queste operazioni nasce da questa frammentazione. Provando, infatti, a frammentare il testo delle trascrizioni in piccole porzioni, anche *ChatGPT* porta a termine il compito correttamente. Per questo motivo, il prompt sopra indicato viene, nelle fasi successive, non più correlato al file di testo completo, ma ad un suo frammento, reiterando la richiesta di riformattazione su tutti i frammenti in cui è stato suddiviso il testo.

Come accennato in precedenza, la suddivisione delle operazioni richieste a *ChatGPT* porta ad ottenere risultati migliori.

Ulteriormente alla riorganizzazione del testo, nella fase di preparazione ed elaborazione delle trascrizioni è importante anche operare una correzione dei testi. Effettivamente, la trascrizione automatica di *Davinci Resolve* è spesso viziata da errori di comprensione delle parole, contenendo errori ortografici o parole prive di senso, soprattutto se utilizzata in lingue differenti da quella inglese.

Volendo, quindi, rendere i testi comprensibili al massimo è utile svolgere ancora un passaggio volto alla correzione degli errori nel testo, almeno quelli più grossolani. Anche in questo caso, la scelta è quella di utilizzare gli algoritmi di *Intelligenza Artificiale* di *ChatGPT* per automatizzare un'operazione altrimenti molto lunga. Naturalmente, l'approccio anche per quest'operazione è quello di frammentare i testi in input, come per i passaggi precedenti.

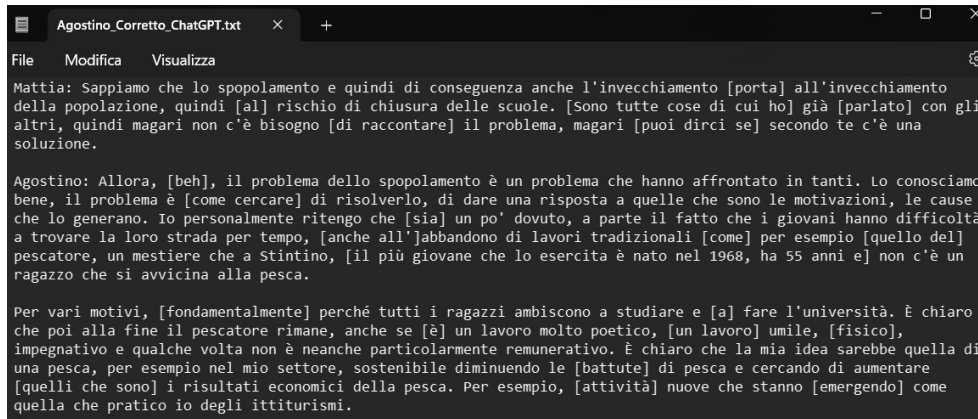


Figura 55 - Esempio di trascrizione corretta con le correzioni tra parentesi quadre

La sfida maggiore in questo caso è impedire a *ChatGPT* di correggere troppo pesantemente il testo, che deve rimanere il più possibile fedele alla versione originale e alle parole pronunciate nell'intervista. È quindi necessario specificare all'algoritmo che la correzione deve essere molto limitata, senza rielaborazioni o riscritture del testo. Per controllare che le correzioni fatte dall'algoritmo rispettino le indicazioni è stato richiesto, nel prompt, di evidenziare tutte le modifiche fatte inserendole tra parentesi quadre. In questo modo, anche con uno sguardo più rapido è possibile individuare le correzioni apportate ed intervenire immediatamente, in caso di anomalie. Di seguito è riportato il prompt utilizzato per ottenere i risultati:

Ciao, per la mia tesi ho trascritto delle interviste utilizzando l'IA di Davinci Resolve 19. Ho ottenuto dei file di testo che però contengono degli errori ortografici e grammaticali. Il mio obiettivo è quello di correggerli in modo da avere la trascrizione corretta delle parole precise dell'intervista. Tu saresti in grado di leggere questi file di testo e correggere solamente gli errori grammaticali e ortografici, senza intervenire sulla chiarezza e la fluidità della frase? Potresti evidenziare le tue correzioni utilizzando delle parentesi quadre?

Come già fatto per la riformattazione del testo, a questo prompt viene allegato un frammento del file da correggere, reiterando l'operazione per i frammenti successivi dello stesso file.

Terminato anche questo passaggio, tutti i materiali contenuti nell'*archivio* sono stati convertiti in forma testuale, comprensibile agli algoritmi di testo, e resi più fruibili con le elaborazioni. Questi, saranno il punto di partenza per ricavare uno storytelling dall'*archivio* di materiali.

Analisi approfondita delle interviste

La scrittura della sceneggiatura, ricavando uno storytelling dai materiali di archivio, si basa sulla volontà di ottenere un algoritmo standard da applicare in maniera iterativa. Per raggiungere un simile risultato il passaggio a monte è una conversazione con *ChatGPT*, in

qualità di esperto in materia di sceneggiatura e *prompt engineering*. La conversazione chiedeva a *ChatGPT* di proporre un algoritmo che potesse condurre dai file ottenuti nelle fasi precedenti ad una sceneggiatura completa del documentario, tramite prompt standard e compilazione di tabelle, che miravano ad estrarre tutte le informazioni possibili dalle trascrizioni del materiale di archivio. In output a questa conversazione sono stati ottenuti tutti i passaggi descritti di seguito, i prompt utilizzati e le tabelle associate. Naturalmente, tutti i passaggi dell'algoritmo erano già stati presi in considerazione precedentemente e sviluppati in maniera più approfondita successivamente con l'utilizzo di *ChatGPT*.

Il primo passaggio dell'algoritmo proposto è quello di analizzare le trascrizioni, estrapolando tutte le informazioni utili ai fini della sceneggiatura, come: nome e ruolo della persona, opinioni in merito alle tematiche principali, citazioni testuali delle trascrizioni che contengano le opinioni. Successivamente, l'analisi della singola persona viene unita a quella di tutte le altre per produrre, idealmente, dei confronti. Questo ha come scopo quello di ottenere dei documenti che selezionino le informazioni principali e più utili nella scrittura dello storytelling, senza accantonare le trascrizioni complete, ma integrandole.

Inizialmente, questo primo passaggio è stato condotto con scarsi risultati. La condivisione delle trascrizioni avveniva una volta sola all'inizio della conversazione, chiedendo poi di accedere ad ogni trascrizione singolarmente per estrapolare i temi e le informazioni richieste. Il prompt utilizzato, inoltre, era poco efficace perché entrava poco nello specifico di ciò che realmente si volesse ottenere e, apparentemente, *ChatGPT* si concentrava solo su alcune parti delle interviste, tralasciando il resto.

Si è rivelata necessaria una modifica nell'approccio, approfondendo la frammentazione dei compiti assegnati e salvando tutti i risultati in documenti esterni.

<p>Analisi Magdalena</p> <ul style="list-style-type: none"> • Nome personaggio: Magdalena • La sua attività: Lavoratrice in un allevamento di pesci in Islanda, in una stazione di alimentazione. • Citazioni testuali esatte, senza modifiche o traduzioni: <ol style="list-style-type: none"> 1. "My name is Magdalena, I am from Poland and I am 26." 2. "I work on a fish farm, I work in a feeding station here in [Píngeyri]." 3. "I came to Iceland as an au-pair... After one month, I came here to work in the fish factory. And that was seven years ago." 4. "Yeah, so in Poland there wasn't much to do for me because I also lived in like a not very big town. So I was looking for something better... I found a job here." 5. "I didn't really like my job in the fish factory... And then I met my boyfriend, my current husband, and I think that's the main reason I stayed here until today." <p>Magdalena lavora in un allevamento di pesci e, inizialmente, si è trasferita in Islanda come au pair, ma poi trovato un lavoro nel settore ittico, decidendo di stabilirsi lì per amore.</p> <p>Spopolamento</p> <ul style="list-style-type: none"> • Introduzione generica: Nell'intervista, Magdalena tocca il tema dello spopolamento riferendosi alla situazione del villaggio di Píngeyri in Islanda. Sottolinea come la popolazione stia invecchiando e come ci siano pochi giovani disposti a trasferirsi e vivere in questa comunità. Questo fenomeno ha gravi conseguenze per la vitalità del villaggio e la sua capacità di attrarre nuove persone. • Problemi legati allo spopolamento: Uno dei principali problemi legati allo spopolamento è la mancanza di opportunità lavorative e di alloggi disponibili. Magdalena evidenzia che molti giovani non scelgono di trasferirsi a Píngeyri a causa della scarsità di opzioni abitative. Un altro problema è la presenza di molte case vuote, che vengono utilizzate solo come residenze estive, riducendo ulteriormente le possibilità di accoglienza per chi vorrebbe stabilirsi permanentemente. • Possibili soluzioni che l'intervistato individua: Magdalena suggerisce che, per contrastare lo spopolamento, sarebbe utile intervenire sulla gestione delle case estive, incentivando chi le possiede a costruirne altre e lasciando le case nel villaggio a persone disposte a viverci tutto l'anno. Un miglioramento delle opportunità lavorative e abitative potrebbe rendere Píngeyri più attraente per i giovani e per le famiglie. • Citazioni testuali esatte: <ol style="list-style-type: none"> 1. Sulla mancanza di giovani e invecchiamento della popolazione: "Yeah, that's a problem here, that most of the people are really old and then not many young people want to move here." 2. Sulla scarsità di case disponibili per chi vuole trasferirsi: "We had a big problem with that actually because we wanted to buy a house for many 	<p>years and it was impossible... I know a few people who would like to move here but they just don't have a place to live."</p> <ol style="list-style-type: none"> 3. Sulla soluzione di incentivare nuove costruzioni per le residenze estive: "Maybe somehow convince them to build summer houses around the fjord... and leave the houses for people who would actually want to live here, work here, and be a part of the community." 4. Problema delle abitazioni vuote perché residenze estive: "Almost half of the houses here in Píngeyri are just empty. People don't live here; they only keep them as summer houses." <p>Istruzione</p> <ul style="list-style-type: none"> • Introduzione generica: Nell'intervista, Magdalena descrive il sistema scolastico locale di Píngeyri, evidenziando le caratteristiche peculiari della scuola, che serve una comunità isolata. Fa un confronto con il sistema educativo polacco, mettendo in luce i benefici dell'approccio islandese, più focalizzato sullo sviluppo individuale piuttosto che sulle valutazioni basate sui voti. • Problemi legati all'istruzione: Un problema che Magdalena menziona è la scarsità di insegnanti e risorse per offrire corsi specifici o attività extra-curricolari. Questa limitazione comporta una riduzione delle opportunità per i bambini, soprattutto durante i mesi invernali, quando il tempo inclemente riduce le possibilità di attività all'aperto. Inoltre, la piccola dimensione della scuola, con studenti di più classi raggruppati insieme, può rappresentare una sfida per un'istruzione più personalizzata. • Possibili soluzioni che l'intervistato individua: Anche se Magdalena esprime apprezzamento per il sistema scolastico islandese, non suggerisce cambiamenti particolari, ma sottolinea che un maggior investimento in insegnanti e risorse per attività aggiuntive potrebbe migliorare l'esperienza scolastica complessiva. Considera l'approccio educativo della scuola di Píngeyri un vantaggio, grazie alla libertà data ai bambini di sviluppare le proprie competenze. • Citazioni testuali esatte: <ol style="list-style-type: none"> 1. Sulla struttura della scuola locale: "So, it's a small school. There are around 40-45 students."(Magdalena_finale_pdf) 2. Sulla gestione delle classi miste e l'approccio educativo: "There are grades, but still, they are in class together, like first, second, and third grade. The school here focuses a lot on kids' freedom. They want to show the kids that they don't have to have grades to be appreciated."(Magdalena_finale_pdf) 3. Sul confronto tra il sistema scolastico islandese e quello polacco: "I'm comparing it with schools in Poland. It's so different because in Poland, everything is focused on grades and that you have to be really good at everything. And here, the main focus is on the good side, on the good skills of the kids. And I really like it." (Magdalena_finale_pdf)
--	---

Figura 56 - Esempio di due pagine dell'analisi all'intervistata Magdalena

Il metodo utilizzato è quindi quello di inviare ogni trascrizione singolarmente e suddividere le richieste concentrandosi su una tematica alla volta. Inoltre, per assicurarsi che l'algoritmo faccia un'attenta lettura del documento, la richiesta è quella di una doppia lettura, dall'inizio alla fine e viceversa.

L'iterazione, quindi, in questa fase diventa a tre livelli. Il primo ad alto livello sui singoli documenti di trascrizione. Il secondo su ogni tema del documento, suddiviso nelle seguenti richieste:

- Presentazione e informazioni generali dell'intervistato
- Opinione e citazioni in merito al tema dello spopolamento
- Opinione e citazioni in merito al tema dell'invecchiamento
- Opinione e citazioni in merito al tema dell'istruzione
- Opinione e citazioni in merito al tema dei trasporti
- Opinione e citazioni in merito al tema della digitalizzazione
- Opinione e citazioni in merito al tema della transizione ecologica
- Opinione e citazioni in merito al tema delle infrastrutture
- Opinione e citazioni in merito al tema del turismo
- Opinione e citazioni in merito al tema del nomadismo digitale
- Eventuali opinioni e citazioni su altri temi trattati e non considerati

Di seguito si riporta l'esempio di un prompt standard per questa fase iterativa:

```
Passiamo ora al tema <TEMA>. Leggi l'intervista a <INTERVISTATO> dall'inizio alla fine e viceversa. Come si esprime l'intervistato in merito al turismo? Se ritieni che l'intervista non parli del tema in questione, segnalamelo. Altrimenti, organizza la tua risposta in questo modo: Titolo: "<TEMA>"
```

- Introduzione generica
- Problemi legati al turismo
- Possibili soluzioni che l'intervistato individua
- Citazioni testuali esatte, senza modifiche né traduzioni, di ciò che viene detto dall'intervistato in merito al tema. Introduci le citazioni con l'argomento specifico a cui fanno riferimento, per una maggiore comprensione. Considera solamente le citazioni che vengono dette in un contesto in cui si parla del tema in questione e non di altri temi

Il terzo livello iterativo è quello sulla singola risposta, chiedendo correzioni e verifiche di completezza. Un esempio di prompt di verifica è riportato qui di seguito:

```
Rileggi ora l'intero documento e accertati delle seguenti cose:
```

- Di non aver dimenticato nulla, né nelle citazioni né nelle analisi
- Di aver inserito le citazioni complete

- Del fatto che le citazioni siano effettivamente state pronunciate nel contesto dell'argomento in questione

A questo punto, l'analisi dei temi in ogni documento, viene salvata in un unico file contenente tutte le analisi condotte. Questo documento accompagnerà, per il resto dei passaggi, quello contenente tutte le trascrizioni alle interviste. In questo modo, *ChatGPT* avrà a disposizione non solo di accedere alle frasi corrette e coerenti pronunciate dagli intervistati, ma anche di conoscere le opinioni degli stessi per utilizzarle meglio nel creare delle storie.

Stesura dei documenti e della sceneggiatura

Una volta completata l'analisi delle trascrizioni e unite in un file unico contenente tutte le analisi, l'algoritmo scritto insieme a *ChatGPT* prevede l'inizio di quella che sarà la conversazione che dovrà portare alla definizione di tutti i documenti necessari per scrivere una sceneggiatura completa. Questi sono, in particolare modo: *logline*, *soggetto*, *sinossi* e *trattamento*.

Prima di questo, però, è utile inviare a *ChatGPT* una prima volta i documenti di trascrizione e analisi, descrivendoli dettagliatamente, e comunicare le note stilistiche e di regia decise all'inizio della post-produzione. Questi due passaggi preliminari possono essere fatti in maniera separata, inviando prima i documenti e chiedendo un messaggio di conferma ricezione e lettura da parte di *ChatGPT* e poi un messaggio per la definizione delle linee guida.

Di seguito viene inserito il prompt a cui vengono allegati i file contenenti trascrizioni e analisi delle interviste:

```
Ciao, per il mio lavoro di tesi sto realizzando un documentario. L'unica base del documentario sono alcune interviste che ho già girato a delle persone provenienti da 4 luoghi remoti d'Europa: Sardegna, Islanda, Madeira e Gran Canaria. Il mio obiettivo è quello di riuscire a scrivere insieme a te i documenti che portano alla stesura finale della sceneggiatura del documentario, che sono: Logline, Sinossi, Soggetto, Trattamento e Sceneggiatura, appunto. Ma procederemo con calma. Per iniziare ti sto inviando due documenti pdf:
```

- `Trascrizioni_completo.pdf` contiene le trascrizioni complete di tutte le 11 interviste che ho girato
- `Analisi_completa.pdf` contiene invece un'analisi delle tematiche fatta per ogni persona intervistata, basandomi sui temi che vorrei trattare nel documentario.

```
Potresti per ora leggere questi documenti dall'inizio alla fine e viceversa e memorizzarli?
```

Prima della richiesta di memorizzazione vengono introdotti il lavoro, le condizioni di partenza e l'obiettivo da raggiungere. Questo ha l'utilità di impostare un contesto comune con *ChatGPT* e permettergli di avere coscienza dei documenti condivisi.

Il passaggio successivo è quello di inviare un messaggio di definizione delle linee guida e iniziare a lavorare con *ChatGPT*, che assume il ruolo di *Story Editor* del progetto. In modo specifico, vengono condivisi con l'algoritmo i dettagli su: la struttura in capitoli della sceneggiatura, considerando anche le scelte già fatte su quali intervistati parlano in quali capitoli, lo stile cinematografico del documentario e la necessità di scegliere quale storia raccontare in Sardegna.

Infatti, nel caso del documentario per il progetto *Europe*, gli intervistati in Sardegna sono molti di più rispetto a quelli che si è deciso di inserire in ogni capitolo. Tuttavia, la scelta di chi utilizzare spetta a *ChatGPT*, in base alla storia che ritiene migliore da raccontare. Per cui, nel caso specifico, l'output di questo prompt è la scelta di quale persona affiancare all'host, Fabrizio, in Sardegna, motivandola. Il prompt utilizzato per raggiungere questi obiettivi è il seguente, ed è frutto sempre della conversazione con *ChatGPT* volta a raggiungere un *prompt engineering* adeguato:

```
Prima di iniziare a lavorare sulla logline e sugli altri documenti per il mio documentario, vorrei che tenessi presenti alcune linee guida strutturali e tematiche che desidero siano seguite durante tutto il processo di scrittura:
```

1. Struttura in Capitoli:

- Il documentario sarà suddiviso in sei capitoli:
 - Introduzione: Presentazione delle persone intervistate e introduzione del documentario.
 - Capitolo 1 - Sardegna: Questo capitolo includerà due persone: Fabrizio e un'altra persona da scegliere in base all'utilità per lo scopo del documentario tra le altre 5 intervistate in Sardegna.
 - Capitolo 2 - Islanda: Includerà Arnar e Magdalena.
 - Capitolo 3 - Gran Canaria: Includerà Beatrice e Mendes.
 - Capitolo 4 - Madeira: Includerà Markus.
 - Conclusione: Tirerà le fila del documentario, concludendo con i saluti delle persone intervistate.

2. Selezione dei Personaggi:

- Delle 6 persone intervistate in Sardegna, utilizzeremo Fabrizio e un'altra persona da selezionare in base all'utilità per lo scopo del documentario. Le interviste di Arnar e Magdalena saranno incluse nel capitolo dell'Islanda, quelle di Beatrice e Mendes nel capitolo di Gran Canaria, e l'intervista di Markus nel capitolo di Madeira. Seleziona ora la persona della Sardegna che affiancherà Fabrizio nel documentario e nella risposta che mi dai scrivimi quale persona hai selezionato e perchè lo hai fatto

3. Stile Cinematografico:

- Vorrei che il documentario avesse uno stile cinematografico. Questo significa un'attenzione particolare alla narrazione visiva, alla qualità delle riprese e al montaggio, con un focus sull'immersione dello spettatore nei luoghi e nelle storie raccontate.

Ricordati queste linee guida durante la stesura della logline, della sinossi, del soggetto, del trattamento e della sceneggiatura.

A seguito di questo prompt, *ChatGPT* compie la prima scelta da *Story Editor* vero e proprio, inserendo nel capitolo della Sardegna la voce di Agostino Schiaffino al fianco di quella di Fabrizio Contini. L'immagine di seguito riporta la risposta completa di *ChatGPT* con le motivazioni che lo hanno portato a questa scelta.

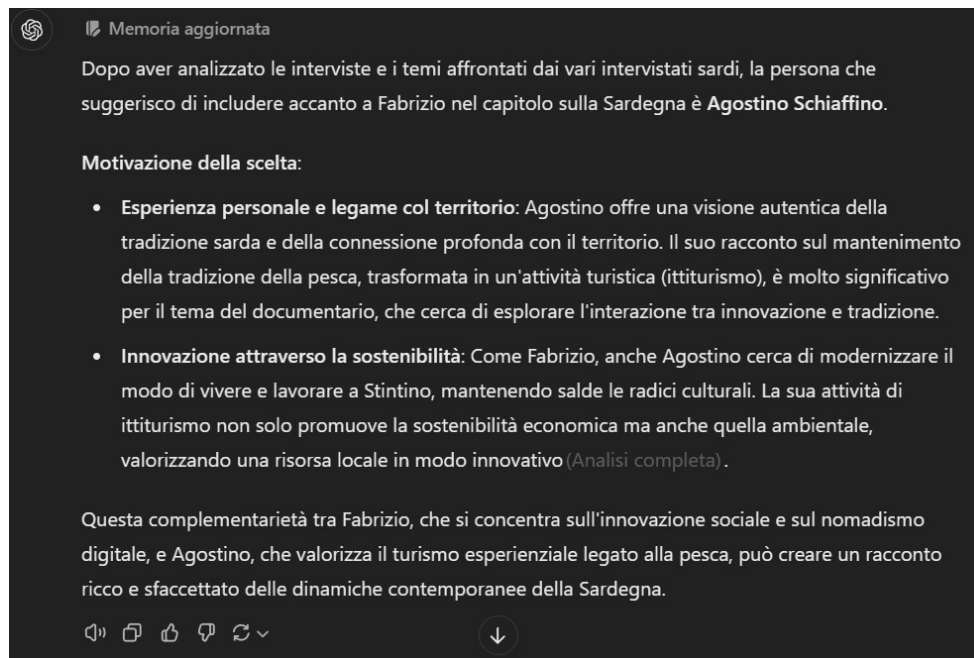


Figura 57 - Scelta della persona da inserire in Sardegna e motivazione

Dopodiché, l'ultimo passaggio nella definizione delle linee guida è quello di riportare alcune precisazioni, come l'utilizzo di una voce narrante scritta da *ChatGPT* e la necessità di utilizzare i documenti allegati al primo messaggio per scrivere le battute degli intervistati senza modifiche, rielaborazioni o, tantomeno, traduzioni.

Successivamente, si può iniziare con la scrittura dei documenti elencati poco sopra. Per raggiungere l'obiettivo di scriverli al meglio, dalla conversazione con *ChatGPT* per il *prompt engineering* si sono ricavate delle definizioni standard per i documenti descritti e delle tabelle con alcuni campi da compilare che potessero condurre ad una migliore stesura di ogni documento.

Documento	Definizione	Obiettivo	Lunghezza
Logline	Una o due frasi che catturano il concetto centrale e il conflitto principale del documentario.	Intrigare e attrarre l'attenzione, riassumendo l'idea centrale in modo immediato.	Molto breve (1-2 frasi)
Sinossi	Riassunto conciso che offre una panoramica generale della storia, dei personaggi e del conflitto.	Fornire una visione d'insieme del progetto, ideale per presentazioni preliminari.	Breve (da un paragrafo a una pagina)
Soggetto	Descrizione più dettagliata dell'intera storia, inclusi introduzione, sviluppo della trama, e risoluzione.	Fornire una base più solida per lo sviluppo della sceneggiatura, dettagliando la trama senza suddivisione in scene.	Media (1-10 pagine)
Trattamento	Racconto dettagliato suddiviso in scene, che descrive visivamente l'azione ma senza dialoghi completi.	Mostrare come la storia si sviluppa visivamente e preparare il terreno per la sceneggiatura completa.	Piuttosto lungo (10-30 pagine)
Sceneggiatura	Documento completo e dettagliato con descrizione delle scene, dialoghi, e indicazioni tecniche.	Preparare il documento definitivo per la produzione del documentario, includendo tutti i dettagli necessari.	Dettagliato e lungo (lunghezza variabile)

Figura 58 - Tabella con le definizioni dei cinque documenti e relative informazioni scritte con ChatGPT

In questo modo, la scrittura di ognuno dei documenti è codificata, standard e applicabile a qualsiasi contesto e *archivio* di partenza. Le tabelle hanno come input il documento precedente e come output il documento corrente, le righe nel mezzo sono elementi con cui si cerca di portare *ChatGPT* a “ragionare” sul contenuto dei documenti ed estrapolare una storia che sia coerente.

In maniera più specifica, il workflow consiste nell’inviare un prompt che richieda la stesura di un documento, nel primo caso la *logline*, basandosi sui documenti precedentemente condivisi e dando altre piccole note come la definizione del documento in questione, ricavata dalla Figura 58, e la lunghezza indicativa decisa per il documentario. Nel prompt viene anche richiesto di compilare tutte le righe della tabella, nel campo vuoto, condivisa al fondo. Di seguito un esempio di prompt per la *logline*:

Sulla base dei documenti che ti ho fornito e tenendo presenti le linee guida strutturali e stilistiche che ti ho dato, potresti compilare, nella colonna 'Spazio per la tua risposta', la seguente tabella per creare una logline per un documentario di circa 30 minuti? La logline dovrebbe riflettere il tema centrale del documentario e mantenere uno stile incisivo e sintetico.

In coda a questo messaggio la tabella condivisa è simile alla seguente:

Elemento	Descrizione	Spazio per la tua risposta
Tema principale	Identifica il concetto centrale o l'argomento del documentario.	
Protagonista/i	Descrivi il personaggio o i personaggi principali.	
Conflitto principale	Indica il conflitto o il problema centrale che verrà affrontato.	
Obiettivo	Specifica cosa i protagonisti cercano di ottenere o dimostrare.	
Logline	Formula una frase che racchiuda tutti gli elementi precedenti in modo sintetico e accattivante.	

Figura 59 - Tabella non compilata per la stesura della logline

Le tabelle successive a questa hanno una struttura molto simile, con l'unica differenza che la prima riga è occupata dall'output della tabella precedente, quasi come se fosse un'operazione ricorsiva, con delle funzioni che prendono in input l'output di altre funzioni. Per la scrittura delle tabelle successive anche la struttura del prompt è piuttosto simile a quello della *logline*, con la definizione del documento che deve produrre *ChatGPT* e l'inserimento della tabella corrispondente.

Naturalmente, come nel caso delle fasi precedenti, anche nella stesura delle tabelle e dei documenti utili alla sceneggiatura la parola chiave è *reiterazione*. Ad ogni output ottenuto possono corrispondere dei prompt di revisione o correzione fino a raggiungere una versione soddisfacente.

L'ultimo documento prima della sceneggiatura, il trattamento, è quello che dà le informazioni sulle scene, quante scene scrivere e cosa accade in ognuna di queste. È quindi importante definire bene il numero delle scene di cui si vuole che sia composto il documentario. Questo numero lo si può definire in base alla lunghezza che si prevede per il documentario. È quindi importante definire il numero di scene anche nel momento della scrittura del prompt per il trattamento e modificarlo, qualora sia sbagliato nell'output dato da *Chat GPT*.

Ottenuto il trattamento, questo sarà la base di partenza per la scrittura della sceneggiatura. È utile chiedere a *ChatGPT* di memorizzare la struttura del trattamento, scritto da lui stesso, in modo da potervi riferire senza dover necessariamente inviarlo ogni volta.

L'approccio utilizzato per scrivere la sceneggiatura è, similmente a quanto successo in precedenza, iterativo. Si utilizzano prompt per ogni scena da scrivere, chiedendo di apportare delle correzioni alla scena, se necessario, prima di passare al prompt per la scena

successiva. Il prompt per la prima scena definisce anche le linee guida per la scrittura della sceneggiatura in modo completo. Di seguito, viene riportato il prompt utilizzato, ricavato, come gli altri, dalle conversazioni con *ChatGPT* sul *prompt engineering*.

Ora che il trattamento è pronto, ti chiedo di scrivere la sceneggiatura del documentario di 30 minuti seguendo queste linee guida e utilizzando un approccio iterativo, scena per scena:

- Struttura della Scena: Ogni scena deve essere completa e ben sviluppata, seguendo il trattamento. Includi una descrizione visiva dettagliata, dialoghi, azioni dei personaggi, e l'atmosfera generale.

Utilizzo dei Dialoghi:

- Le battute dei personaggi devono essere testuali, prese direttamente dalle trascrizioni delle interviste senza modifiche.
- Non limitarti alle citazioni estratte durante l'analisi iniziale, ma esplora l'intero contenuto delle trascrizioni per trovare dialoghi pertinenti e significativi per ogni scena.

Voce Narrante:

- Puoi inventare le battute solo per la voce narrante. La voce narrante deve essere utilizzata per collegare le scene, fornire contesto, introdurre i personaggi e i temi, e concludere la narrazione.

Intestazioni e Descrizioni:

- Segui la struttura standard della sceneggiatura con intestazioni di scena (es. INT. CASA DI FABRIZIO - GIORNO), descrizioni delle azioni e degli ambienti, dialoghi dei personaggi e indicazioni sulle inquadrature.

Durata e Sviluppo:

- Ogni scena deve contribuire in modo significativo allo sviluppo della narrazione e mantenere l'attenzione dello spettatore. Evita scene troppo brevi e battute isolate.
- Assicurati che ogni scena includa più battute o scambi di dialoghi, in modo da creare una narrazione fluida e coinvolgente.

Iniziamo con la scrittura della prima scena. Segui queste indicazioni per descrivere e scrivere la scena completa e, una volta che sarà stata approvata, procederemo alla successiva.

Questo prompt si concentra sul dare tutte le informazioni ritenute necessarie per la corretta stesura della sceneggiatura, come indicazioni sulla voce narrante, sulle descrizioni visive delle scene o indicazioni su come utilizzare le battute.

A questo punto, la scrittura della sceneggiatura consiste in un processo iterativo, come è già stato mostrato in precedenza, in cui ogni scena viene scritta singolarmente e corretta prima di passare alla successiva. La frammentazione non costituisce in alcun modo un problema per la coerenza e la continuità tra le scene, condizioni garantite dalla presenza del trattamento.

Il prompt utilizzato per la scrittura delle scene successive è più conciso e si basa sempre sull'utilizzo dei documenti inviati e sul trattamento memorizzato.

Grazie per la prima scena. Ora procediamo con la scena successiva del documentario, basandoti sempre sulle trascrizioni delle interviste, sull'analisi completa e sul trattamento. Segui le stesse linee guida precedenti per sviluppare una scena completa e dettagliata. Leggi bene i documenti per inserire le citazioni corrette, senza modifiche né traduzioni.

Chiaramente, questa è un'operazione molto importante per la buona riuscita del documentario. Pertanto, mantenere il controllo della direzione che la sceneggiatura sta prendendo è fondamentale. Di conseguenza, oltre alle correzioni per chiari ed evidenti errori in sceneggiatura, come l'utilizzo di battute non presenti nei documenti, è sicuramente utile apportare modifiche alla scena seguendo il proprio gusto, come la sostituzione di battute con altre, l'integrazione o l'eliminazione di battute, oppure ancora la modifica delle frasi originali della voce narrante o delle descrizioni visive.

È, quindi, opportuno conoscere il contenuto dei documenti prodotti, sia le analisi sia le trascrizioni, per intervenire con maggior prontezza ed efficacia correggendo la scena. Questo serve sia per saper riconoscere le battute che non rispettano le trascrizioni, sia per conoscerne di differenti che magari troverebbero una buona collocazione all'interno di una scena. Va da sé, che la conoscenza di questi documenti viene quasi come naturale conseguenza dei passaggi precedenti di maneggiamento degli stessi.

Revisione e scrittura del titolo

Alla fine del processo di iterazione, si ottiene una sceneggiatura completa, frutto dell'unione tra la *creatività* di *ChatGPT*, in veste di *story editor*, e la propria supervisione creativa. A questo punto, il passaggio successivo, prima di procedere con il montaggio video, è quello di ricontrollare il lavoro svolto. Per farlo è utile sfruttare nuovamente l'algoritmo di *ChatGPT*, inviando la sceneggiatura completa in un prompt che chieda di leggerlo e di ipotizzare delle correzioni generiche per migliorarla. Quello di seguito è il prompt utilizzato per questa richiesta:

Ti invio ora la sceneggiatura completa. Da questo momento in poi sei uno sceneggiatore professionista. Leggi dall'inizio alla fine e viceversa questa sceneggiatura e valutala secondo i parametri che tu ritieni più corretti. Dammi una valutazione alla sceneggiatura.

A questo punto, *ChatGPT* individua alcune correzioni che possano migliorare la valutazione della sceneggiatura ed è possibile continuare la conversazione chiedendo degli esempi per le correzioni più rilevanti.

Chiaramente, la scelta finale spetta sempre a chi fa uso di questo workflow, valutando sia la sceneggiatura completa, sia le correzioni proposte da *ChatGPT*. Per questo motivo, le correzioni sono una fase puramente soggettiva.

La conclusione del flusso di lavoro è data dalla scrittura del titolo, utilizzando ancora una volta *ChatGPT*. Per ottenere delle proposte di titolo è necessario inviare la versione ottenuta della sceneggiatura, invitando a un'attenta lettura l'algoritmo e chiedendo in output delle proposte di titolo, nella quantità ritenuta corretta. Il titolo proposto da *ChatGPT* ritenuto adatto al documentario, in questo caso, è stato: *Forgotten Frontiers: Where Isolation Meets Innovation*.

Considerazioni sul workflow per la scrittura

L'obiettivo del flusso di lavoro, riguardante la scrittura, è stato raggiunto, ovvero quello della stesura di una sceneggiatura. I passaggi descritti sopra, che portano a questo risultato, sono frutto di un approccio *trial and error* durante il percorso di studio.

In generale, *ChatGPT* è il punto di riferimento per ogni fase descritta, oltre che per la stesura vera e propria, anche per fasi collaterali come la scrittura dei prompt e la revisione.

I risultati ottenuti nelle fasi intermedie, come la stesura dei vari documenti, sono semplici linee guida. Pertanto, una corretta compilazione dei campi delle tabelle non è indispensabile per la buona riuscita dell'esperimento in scrittura, quanto più un indicatore che tracci la direzione che sta prendendo la scrittura.

Occorre precisare alcuni fattori riguardo la versione di *ChatGPT*. Per condurre l'esperimento è stato sottoscritto il piano a pagamento dello strumento, che aiuta nello sveltire i tempi, abbattendo limitazioni temporali e di numero di file condivisi, rendendone pressoché illimitato l'utilizzo. Durante la conduzione dell'esperimento il software è stato aggiornato, inserendo la memoria nelle conversazioni. Questo aggiornamento è stato fondamentale nel velocizzare i tempi, perché molti dei problemi riscontrati nella fase di studio erano dovuti all'assenza di memoria. Infine, la nuova versione, tuttora ancora in preview, che utilizza i *Canvas* è, probabilmente, un ulteriore aiuto per la creazione e modifica puntuale dei documenti, sveltendo le operazioni e snellendo i prompt inviati per correggere una singola frase nel documento.

Concludendo la fase di scrittura, le tempistiche che comprendono sia la fase di studio dell'algoritmo sia l'applicazione del workflow studiato, sono di circa un mese e mezzo, tra agosto e settembre 2024.

A titolo d'esempio, di seguito si inserisce un estratto della sceneggiatura completa, troppo lunga per essere inserita integralmente (circa 30 pagine), per visualizzare meglio il risultato ottenuto. Si tratta della prima scena del documentario, quella introduttiva, che inizia lo spettatore alle tematiche e ai personaggi principali presenti.

Forgotten Frontiers: Where Isolation Meets Innovation

SCENA 1: INTRODUZIONE AI LUOGHI REMOTI D'EUROPA E AI PROTAGONISTI

INT./EST. DIVERSI LUOGHI - GIORNO

La scena si apre con inquadrature che mostrano la bellezza e l'isolamento dei paesaggi: le onde si infrangono sulle coste rocciose della Sardegna, i fiordi dell'Islanda si stagliano contro l'orizzonte, e le dune di Gran Canaria emergono nel paesaggio. Infine, le vallate di Madeira completano il quadro. La natura domina queste terre, trasmettendo sia la loro maestosità che il loro isolamento.

VOCE NARRANTE (FUORI CAMPO):

There are places in Europe that seem suspended in time. Remote territories where isolation and nature are part of everyday life. Here, communities face enormous challenges: from land abandonment to the innovation needed to survive.

L'inquadratura passa da un villaggio in Sardegna a un piccolo porto di pescatori, e poi si sposta sull'Islanda, mostrando le strade innevate. Vediamo i protagonisti nei loro contesti.

VOCE NARRANTE (FUORI CAMPO):

Let's meet the voices that face these challenges and strive for a better future.

EST. MARE DI STINTINO - GIORNO

FABRIZIO (PARLANDO DIRETTAMENTE ALLA TELECAMERA):

Ciao a tutti, sono Fabrizio Contini e, prima di tutto, sono un ragazzo di [Stintino]. Sono il presidente di una piccola organizzazione absentia, qui a Stintino stiamo cercando di portare dei progetti innovativi, progetti di innovazione sociale e anche progetti europei come questo che stiamo facendo adesso in Europa.

AGOSTINO (PARLANDO DIRETTAMENTE ALLA TELECAMERA, SULLA SUA BARCA DA PESCA):

Io mi chiamo Agostino Schiaffino, ho compiuto 60 anni 15 giorni fa, dunque sono del 1963. Sono stintinese, direi puro sangue, dal momento che tutti i miei bisnonni erano qua a Stintino al momento della fondazione del paese

Immagini del paesaggio di Stintino

VOCE NARRANTE (FUORI CAMPO) :

Fabrizio and Agostino are committed to revitalizing Stintino, focusing on creating opportunities and preserving the traditions that define their community.

EST. REYKJAVIK, UFFICIO EAST OF MOON - GIORNO

ARNAR (PARLANDO DIRETTAMENTE ALLA TELECAMERA) :

I am Arnar Sigursson, the founder of East of Moon, a laboratory for community-centered innovation.

EST. PORTO DI ÞINGEYRI - GIORNO

MAGDALENA (PARLANDO DIRETTAMENTE ALLA TELECAMERA) :

My name is Magdalena, I am from Poland and I am 26. I work on a fish farm, I work in a feeding station here in Þingeyri.

Immagini di paesaggi dell'Islanda, viaggi in macchina e Arnar che lavora.

VOCE NARRANTE (FUORI CAMPO) :

Arnar and Magdalena bring us into the heart of the Westfjords, where they confront the struggles of isolation and the harshness of life in one of Iceland's most remote regions.

EST. LAS PALMAS, GRAN CANARIA - GIORNO

BEATRICE (PARLANDO DIRETTAMENTE ALLA TELECAMERA) :

Hola, hello, my name is Beatrice Avantina. I'm the managing director of Consulta Europa, projects and innovation. Consulta Europa is an SME based in the Canary Islands, Spain.

MARKUS (PARLANDO DIRETTAMENTE ALLA TELECAMERA) :

My name is Markus, I am 40 years old. I was born in the very north of Italy, close to the Austrian and Swiss border, in a very beautiful place in a city called Merano.

Immagini di Gran Canaria e campi coltivati a Madeira.

VOCE NARRANTE (FUORI CAMPO) :

From Gran Canaria to Madeira, Beatrice and Markus show us the potential of sustainable farming, turning isolation into an opportunity to reconnect with the land.

EST. MADEIRA, VIGNETI DI MARKUS - GIORNO

INT. COWORKING SPACE, GRAN CANARIA - GIORNO

MENDES (PARLANDO DIRETTAMENTE ALLA TELECAMERA) :

My name is Jay Mendes. I am from Portugal. I was a digital nomad for 10-12 years. I work with either sustainability or remote work, and I organize events for both.

Immagini di paesaggi di Madeira

VOCE NARRANTE (FUORI CAMPO) :

Telmo and Mendes see digital nomadism as the key to transforming Madeira—connecting its remote beauty to a global community.

Le immagini finali della scena mostrano i protagonisti nei loro ambienti, mentre la musica si alza leggermente di tono, e la voce narrante riprende.

VOCE NARRANTE (FUORI CAMPO) :

These voices will guide us on a journey through some of the most remote places in Europe. Here, innovation and tradition intertwine to tackle the challenges of a constantly changing world.

3.4.2 Montaggio Video

Dopo aver concluso la scrittura di una prima versione della sceneggiatura, il focus si sposta sull'editing video. Essendo il documentario commissionato un output per un progetto europeo, è necessario che la lingua principale con cui viene narrato sia l'inglese. Tuttavia, le voci narranti sono state inizialmente scritte in italiano. Pertanto, è stato necessario tradurle inviando la sceneggiatura a *ChatGPT* con la richiesta di traduzione dall'italiano all'inglese.

A questo punto, per poter montare, il passo successivo è stato generare le voci narranti utilizzando il tool *Eleven Labs*, operazione dettagliata meglio nel corso del paragrafo riguardante il montaggio audio.

Con il materiale al completo, il lavoro si concentra innanzitutto su un montaggio "uditivo" delle parole della sceneggiatura. Quest'ultima diventa più che una traccia da seguire, una trasposizione scritta della timeline di editing. L'utilizzo, nelle precedenti fasi, della funzione di trascrizione integrata in *Davinci Resolve* trova ora la sua utilità nel rendere il montaggio delle battute molto più veloce ed efficace. Infatti, avendo nella sceneggiatura la trascrizione pressoché esatta delle parole delle interviste, la ricerca delle battute da inserire diventa una semplice ricerca testuale, con l'incipit della battuta o parole chiave. Una volta individuata la frase corretta, basta evidenziarla e inserirla nella timeline con l'apposito comando.

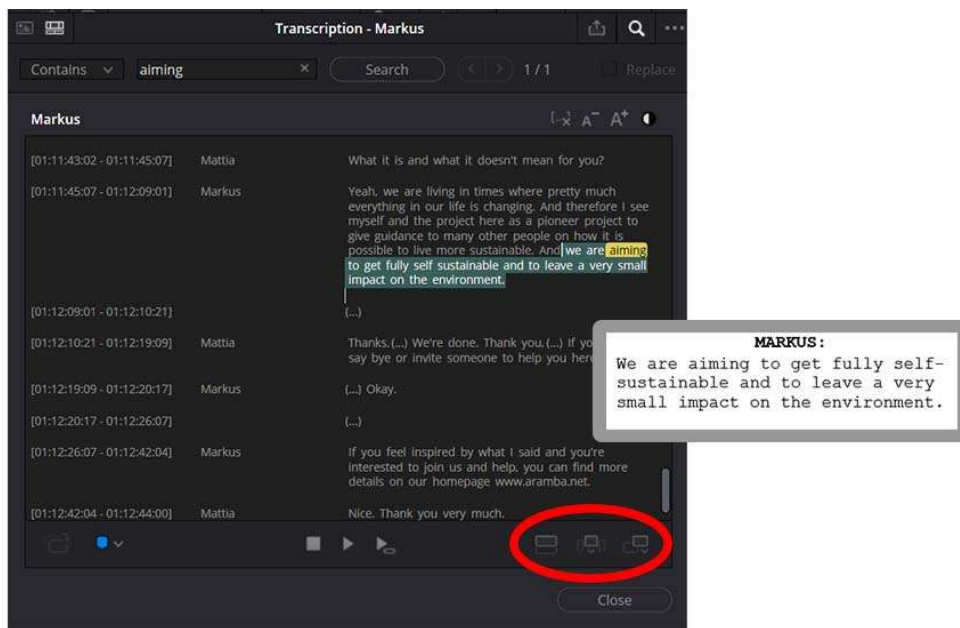


Figura 60 - Esempio di flusso di lavoro per montaggio con trascrizione

Tuttavia, è bene precisare che quello adottato non si tratta di *montaggio automatico* con l'*Intelligenza artificiale*, bensì di un montaggio tradizionale in cui la scelta, la ricerca e l'isolamento delle clip sono aiutati e velocizzati dall'*Intelligenza Artificiale*.

La scelta delle clip è agevolata dalla sceneggiatura non solamente nel caso delle clip contenenti le battute dell'intervista, ma anche nella scelta dei cosiddetti *fegatelli*, descritti da

ChatGPT durante la stesura della sceneggiatura. Esattamente come accade nel caso tradizionale, anche nella sceneggiatura guida scritta in post-produzione c'è una descrizione visiva delle immagini da utilizzare come coperture, se necessarie. Queste descrizioni hanno, però, la funzione di guida non per girare quel tipo di inquadratura, bensì per ricercarla all'interno dell'*archivio* o, eventualmente, comprarla da siti di stock come *Envato Elements*.

EST. STINTINO - GIORNO
La telecamera si alza per una panoramica aerea di Stintino, un piccolo borgo costiero tra il mare cristallino e la costa rocciosa.

Figura 61 - Esempio di descrizione visiva delle inquadrature nella sceneggiatura

Il flusso di lavoro nel montaggio video si basa, anche in questo caso, su un'iterazione. Dopo aver portato a termine un primo *rough cut*, basato sulla prima versione della sceneggiatura, la narrazione può essere ricontrollata e valutata nella sua forma video e audio. Questo, naturalmente, conduce alla possibilità di apportare delle modifiche alla sceneggiatura, aggiungendo, togliendo o modificando le battute.

Per questo motivo, la sceneggiatura non è più da considerarsi unica ed immutabile, ma mutevole e poliforme, adattandosi alle esigenze che emergono dal montaggio. Il workflow per la modifica della narrazione è molto simile a quello per la scrittura, ma ha la grossa differenza di poter essere molto più guidato dalle scelte dell'editor, che, a questo punto, ha una maggiore coscienza di ciò che può servire, specialmente conoscendo l'*archivio*.

Nel caso di *Forgotten Frontiers*, sono stati individuati dei momenti deboli nel montaggio e delle possibili modifiche nelle tematiche e spostamenti di personaggi. Chiaramente, è necessario, avendo scelto le tematiche, individuare comunque le battute esatte da inserire e modificare le voci narranti, in modo tale che seguano il nuovo filo narrativo. La ricerca delle battute avviene similmente all'analisi condotta sui documenti delle trascrizioni, inviando a *ChatGPT* il documento contenente le trascrizioni interessate e chiedendo in output tutte le battute che riguardino l'argomento, tra le quali poter scegliere quelle nuove da inserire. Fatto ciò, il passaggio successivo è quello di sottomettere la scena con le nuove battute a *ChatGPT* e chiedere una rielaborazione delle voci narranti. Diventa, quindi, una sinergia con equilibri differenti tra editor e *Intelligenza Artificiale*, dove la scelta del primo fa la differenza.

In modo simile può avvenire l'aggiunta di personaggi nuovi, avendo già una proposta sulla collocazione temporale nella narrazione di questi.

Occorre una precisazione: quella presentata è l'applicazione avvenuta nel caso di *Forgotten Frontiers*, in cui dopo un primo *rough cut*, regista e editor hanno avuto idee più chiare su come poter modificare la narrazione. Naturalmente, se questo non fosse così, le potenzialità di *ChatGPT* e la sua funzione di *story editor* non perdono di valore, avendo la possibilità di utilizzare lo strumento come consulente anche nelle scelte successive.

In ogni caso, perché questo tipo di flusso di lavoro funzioni, la condizione necessaria è che vi sia uno stretto collegamento tra la sceneggiatura e la timeline del documentario, che devono essere due rappresentazioni dello stesso prodotto in formati differenti. Questo per poter utilizzare la sceneggiatura come sostituto della timeline nelle conversazioni con *ChatGPT* e ottenere degli output applicabili alla stessa.

Concludendo la fase di montaggio, si accenna ad una possibilità esplorata superficialmente nel corso dell'editing: la generazione di coperture animate. Tra le modifiche proposte a livello registico spicca quella di utilizzare l'*IA Generativa* per provare a generare delle coperture sottoforma di una breve storia in animazione 2D. In particolare, la storia riprende la giornata di un pescatore, come quella che Agostino Schiaffino racconta nel documentario. Per ottenere ciò, è stato provato superficialmente un workflow generativo. La prima cosa fatta è quella di utilizzare, ancora una volta, *ChatGPT*, in questo caso come regista, inviando in input il racconto della giornata del pescatore e chiedendo in output la descrizione di alcune inquadrature che potessero raccontarla.

Una volta ottenuto un risultato soddisfacente dalle descrizioni delle inquadrature, la fase successiva è quella della generazione delle inquadrature statiche. Per farlo, è stato impiegato il generatore di immagini *DALL-E*, a cui sono stati inviate in input le descrizioni delle inquadrature e delle immagini di reference stilistiche a cui attenersi, reiterando le richieste fino ad ottenere il risultato migliore.



Figura 62 - Inquadratura della storia del pescatore realizzata con *DALL-E*

Una volta ottenute, le tavole sono state animate utilizzando lo strumento di generazione video di *Runway Gen-2*. La scelta non è ricaduta sullo strumento di ultima generazione, *Gen-3*, perché quella precedente fornisce la funzione di *motion brush*, utile per evidenziare e animare oggetti singoli in un'immagine, data in input, indicando lo spostamento dell'animazione con tre slider che permettono il movimento sui tre assi cartesiani. Nell'esempio della Figura 622 è stato selezionato con il *motion brush* solamente il sole e si è richiesto un leggero movimento in verticale, a voler simulare l'alba.

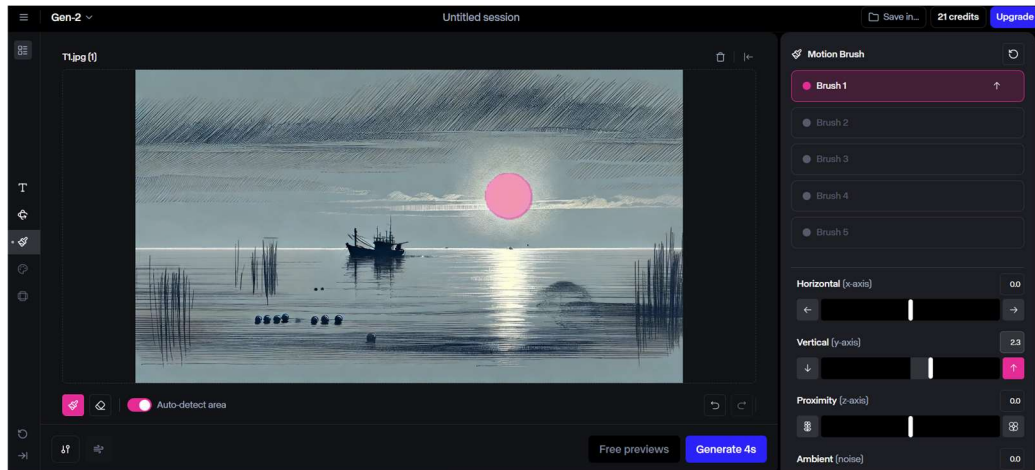


Figura 63 - Utilizzo del motion brush in Runway Gen-2

Tuttavia, questo tipo di applicazione è stata poco approfondita nel corso dello studio, pertanto non può essere ulteriormente dettagliata.

Concludendo, il montaggio video ha richiesto, per raggiungere una versione accettabile, una lavorazione di circa un mese del tesista, tra settembre e ottobre 2024. Ai fini dello studio, il montaggio raggiunto può essere considerato accettabile. Tuttavia, necessita di ulteriori maneggiamenti, che esulano dal contesto della tesi, nell'ottica della distribuzione ai fini del progetto europeo.

3.4.3 Montaggio Audio (cenni)

La fase di montaggio audio è uno dei passaggi meno esplorati dallo studio. L'impiego maggiore di strumenti di *Intelligenza Artificiale* nella realizzazione del montaggio audio per il documentario è quello della generazione delle voci narranti. Una volta tradotte dall'italiano all'inglese, le battute sono utilizzate come input nello strumento *Text to Speech* del tool *Eleven Labs*. Scelta la voce narrante sintetica migliore per lo scopo, si utilizza l'algoritmo *Eleven Multilingual v2*, quello più avanzato, ottenendo in output una traccia audio contenente la voce sintetizzata. Tuttavia, sebbene questo strumento si sia rivelato estremamente utile nel colmare la mancanza di risorse, quello presentato non è che un mero utilizzo dello strumento in questione. Pertanto, non vi sono ulteriori dettagli se non quelli descritti.

L'utilizzo di un simile strumento apre anche a delle possibilità in più dal punto di vista della fruizione del prodotto. Infatti, il tool *Eleven Labs* offre voci sintetiche, o clonate, in molte altre lingue, oltre a quella inglese, consentendo di realizzare versioni multilingua dello stesso prodotto, traducendo rapidamente le battute della voce narrante con *ChatGPT* e generando le tracce audio con *Eleven Labs*. Tutto ciò è supportato ulteriormente dalla generazione automatica di sottotitoli integrata in *Davinci Resolve*, potendo ottenere velocemente i sottotitoli adatti ad ogni lingua scelta per la distribuzione.

Il tool *Eleven Labs* offre anche un'ulteriore possibilità di utilizzo, questa poco esplorata: la generazione di *Effetti Sonori*. Anche in questo caso, però, si tratta di una semplice messa in pratica delle possibilità offerte dallo strumento, inserendo come prompt una descrizione dell'effetto voluto e ottenendo in output la generazione di quattro possibili effetti che rispettano le richieste.

L'ultima, e forse più interessante, casistica esplorata superficialmente a livello audio è quella di un workflow per la generazione delle musiche che compongono la colonna sonora. Questo approccio utilizza in maniera combinata due strumenti: *ChatGPT* e il generatore di musiche *Stable Audio*, il migliore tra quelli provati. L'idea è ancora una volta quella di utilizzare *ChatGPT* come consulente creativo, esperto del documentario in produzione. Inviata la sceneggiatura in input al prompt, la richiesta è quella di individuare cinque tipi di brani da poter inserire come accompagnamento alla narrazione. Questi non sono brani esistenti, ma idee più generiche da poter utilizzare come prompt per la generazione di musiche. Pertanto, in output sono richieste delle caratteristiche specifiche che descrivano i brani proposti, utili alla scrittura di un prompt per la generazione di musica. Le descrizioni richieste sono: genere del brano, mood, strumenti musicali utilizzati, BPM indicativi, breve descrizione a parole del brano e reference di brani famosi per poter comprendere la resa ricercata.

5 "Boundaries of Winter"


- **Genere:** Ambient orchestrale
- **Mood:** Gelido, solenne, evocativo
- **Strumenti:** Archi bassi, sintetizzatori freddi, percussioni riverberate, fiati
- **BPM:** 40-50 BPM
- **Reference:**  Fatcat Records Max Richter - On the Nature of Daylight
- **Descrizione:** Pensato per le scene in cui l'isolamento è palpabile, come il viaggio attraverso i paesaggi invernali di Islanda e le difficoltà di accesso alle risorse durante l'inverno. La composizione trasmette una sensazione di isolamento profondo, ma anche di perseveranza.

Figura 64- Esempio di brano descritto da ChatGPT

Avendo ottenuto cinque brani da utilizzare, è possibile ricontrollare l'utilità dei brani chiedendo a *ChatGPT* un'opinione sulla quantità, se troppo elevata per la sceneggiatura o troppo esigua e delle proposte per migliorare. La descrizione di ogni brano contiene anche il possibile utilizzo dello stesso nella sceneggiatura, indicando per quale momento è stato pensato. Di conseguenza, è utile chiedere un ultimo riassunto a *ChatGPT*, che contenga esclusivamente le scene del documentario per cui sono stati proposti i brani, ottenendo idealmente uno script per il montaggio audio.

Una volta ottenute le indicazioni sui brani, è possibile trasformarle in un prompt da utilizzare come input nel tool *Stable Audio*. La generazione del prompt può essere fatta

manualmente, oppure, utilizzando *ChatGPT*, allenandolo con le pratiche per un *prompting* efficace condivise dagli sviluppatori di *Stable Audio*. Qualsiasi sia il metodo scelto, ottenuti i prompt possono essere utilizzati come input nel tool di generazione musiche di *Stable Audio*, che produce in output una traccia musicale della durata di tre minuti. Il modello utilizzato per il test condotto è *Stable Audio AudioSparx 2.0*, l'ultima versione disponibile.

Tuttavia, la generazione delle musiche è stata provata solo superficialmente, ottenendo comunque risultati positivi. Questo, lascia spazio ad approfondimenti e affinamenti della tecnica, con una buona speranza di un risultato positivo.

3.4.4 Differenze e analogie con il metodo tradizionale

Tipicamente, il metodo tradizionale di produzione di un documentario prevede la scrittura della sceneggiatura in pre-produzione. Questa, non è comunque la versione definitiva, ma una traccia concreta che possa guidare le riprese nell'organizzazione e nel materiale da raccogliere. Successivamente alle riprese, uno *Story Editor* estrapola una storia, compatibile con la sceneggiatura, dal materiale ripreso, dando inizio al montaggio. In molti casi, nella realizzazione di documentari, *Story Editor* e *Video Editor* possono essere la stessa persona.

La prima fase di montaggio è quella in cui l'*editor* visiona completamente per più volte il materiale ricevuto dalle riprese, per individuare quali clip o battute inserire. Contemporaneamente, vengono scritte le battute della voce narrante, prodotte dalle necessità dello storytelling individuato. Una volta scritte, si contatta uno speaker per organizzare un giorno in cui registrare i voice over.

Con tutto il materiale pronto, si può montare una prima versione grezza del documentario, avendo poche possibilità, però, di modificare lo storytelling e la sceneggiatura su cui si basa, perché comporterebbe il rimettere in moto tutta la macchina di scrittura e registrazione, impiegando ulteriori risorse.

Nel mentre vengono composte le musiche ed eventuali scene animate da inserire nel documentario.

In presenza di risorse, questo è sicuramente il metodo principale per ottenere un prodotto di qualità cinematografica.

Il metodo proposto viene incontro alle esigenze di scarsità di risorse, proponendo delle innovazioni dal punto di vista dell'organizzazione del workflow. Lo scopo primario della pre-produzione è quello di gettare le basi per la creazione dell'*archivio* in modo che sia molto popolato. Le riprese non sono più volte al seguire le linee guida dettate dalla sceneggiatura, ma più libere, più reali, figlie del momento e della sensibilità di chi le sta girando.

La post-produzione diventa il cuore del documentario, in cui non solo prende forma, ma nasce insieme alla sceneggiatura. L'innovazione più grande proposta è quella di spostare la

sceneggiatura dalla pre-produzione alla post-produzione, aprendo più direzioni possibili per lo storytelling, date dall'estensione dell'archivio. Oltre a ciò, la sceneggiatura diventa non più solo una linea da seguire, ma uno strumento che può mutare nel corso della post-produzione a seconda delle esigenze che si presentano.

La mutevolezza della sceneggiatura non consiste più in un impiego di risorse di alcun genere, permettendo in poco tempo di provare differenti soluzioni, riscrivendo le voci narranti e generandole immediatamente per provarne la resa.

Capitolo 4

Conclusioni

4.1 Risultati

L'output ottenuto dal workflow presentato è il documentario *Forgotten Frontiers: where isolation meets innovation*. Si tratta di un prodotto che considera i temi fondamentali del progetto *Europe*, descritto precedentemente. Più nello specifico, la scelta dei temi su cui concentrarsi è stata di *ChatGPT* che ha optato per temi quali: spopolamento, isolamento e difficoltà dovute al turismo di massa. Essi sono contrapposti ad attività innovative come l'ittiturismo, la riscoperta di un'agricoltura più sostenibile e il nomadismo digitale.

L'output ottenuto rientra perfettamente nei canoni imposti all'inizio, come stile, durata e struttura, segno che il flusso di lavoro proposto ha portato a degli esiti positivi. Il risultato ottenuto è anche figlio di modifiche alla sceneggiatura avvenute dopo i primi *rough cut*, evidenziando una buona riuscita della collaborazione uomo-macchina proposta, in cui l'*Intelligenza Artificiale* ha agito sia dal punto di vista creativo, sia dal punto di vista meccanico e della facilitazione del lavoro di ricerca.

Le scelte creative compiute da *Chat GPT* prevedono: le tematiche trattate in ogni capitolo, la selezione dei personaggi in Sardegna, la raccolta delle battute dei personaggi più indicate a raccontare un certo tema e la scrittura completa di tutte le battute della voce narrante, nonché delle descrizioni visive da rispettare nella scelta dei *B-roll*. Il risvolto di queste valutazioni creative è soddisfacente, soprattutto dal punto di vista della narrazione in sé, che risulta sempre coerente e ben strutturata.

L'utilizzo dell'archivio a disposizione, volto alla scrittura di una sceneggiatura in una fase successiva, si è rivelato efficace; sono state necessarie, tuttavia, delle integrazioni da parte di elementi di *stock*, comprati esternamente, per completare le inquadrature di copertura. Queste non sono da considerarsi carenze sostanziali, visto che si tratta di un documentario a basso budget; pertanto, la ripresa di tutte le immagini necessarie è un'operazione complessa. Al di là di questo, l'integrazione tra archivio e *Intelligenza Artificiale* è risultata ben riuscita e capace di esaudire le necessità richieste, con qualche possibile miglioria che verrà meglio descritta in seguito riguardo la descrizione e l'utilizzo dei *B-Roll*.

Oltre a ciò, si sono utilizzate delle immagini generate con applicazioni di *IA*, nel modo presentato, in un'applicazione poco approfondita. Il risultato ottenuto è sicuramente valido dal punto di vista visivo, ma migliorabile attraverso una reiterazione e un'attenzione maggiore nella messa in pratica del workflow.

La riuscita delle voci narranti è notevole e dalla parvenza professionale, sia dal punto di vista della qualità audio, sia dal punto di vista della qualità della recitazione, che si dimostra sempre consistente e comparabile al livello standard dell'industria. Per quanto riguarda le musiche e i restanti passaggi audio, le applicazioni provate sono insufficienti per trarne delle conclusioni; lasciano, però, aperte delle possibilità a studi futuri in merito, poiché presentano ottime potenzialità.

Il documentario ottenuto non è nella sua versione ultimata. Ciò che si è raggiunto è semplicemente un risultato che possa essere indicativo di una buona riuscita del flusso di lavoro presentato; per essere pronto alla distribuzione, necessita ancora di ulteriori rimaneggiamenti in post-produzione che esulano dagli obiettivi di questa tesi. In particolare, il montaggio video è ad una versione grezza ed è necessario un intervento che ne affini i dettagli. Inoltre, non vi è alcuna grafica a supporto e spiegazione delle immagini. A livello audio, mancano ancora, come detto, molte componenti della colonna sonora, come musiche ed effetti sonori. È necessaria, inoltre, una color correction completa, dato che questo è uno degli aspetti meno considerati fino a qui.

Riassumendo, i risultati si dimostrano validi specialmente se applicati al workflow teorizzato, ancora in una fase embrionale. Con dei rimaneggiamenti nel processo, alcuni descritti tra poco, altri ancora da studiare, è possibile ottenere risultati sempre migliori e sempre più all'altezza dei prodotti leader nel settore.

4.2 Analisi

Al netto dei risultati ottenuti nell'applicazione del flusso di lavoro presentato al documentario *Forgotten Future*, è utile considerare un'analisi più approfondita dei vantaggi e dei limiti riscontrati, proponendo anche delle possibili migliorie o integrazioni.

4.2.1 Vantaggi ottenuti

Il workflow è stato ideato con l'intento di scoprire se e come l'*Intelligenza Artificiale* possa essere, allo stato attuale dell'arte, un valido aiuto in un caso di produzione professionale a basso budget, sia dal punto di vista delle risorse, sia da quello creativo.

La post-produzione del documentario, fase in cui si concentra il cuore dello studio, ha richiesto circa tre mesi di lavoro, tra studio del metodo, applicazione in sceneggiatura e montaggio video. Questo processo permette di gingere al risultato ottenuto, da considerarsi intermedio nell'ottica di una distribuzione futura.

Rispetto ad una post-produzione tradizionale, le tempistiche devono considerare un'ulteriore fase, quella di stesura della sceneggiatura, che tradizionalmente figura nel tempo dedicato alla pre-produzione. Tuttavia, nel caso della produzione di un documentario in maniera tradizionale, esiste comunque, in post-produzione, la fase di story editing che ha come obiettivo quello di costruire una storia sulla base del girato. Pertanto, pur essendo un lavoro differente e più corposo, la scrittura della sceneggiatura in post-produzione non comporterebbe un aumento di tempo significativo, in quanto incorpora in sé una fase propria del metodo tradizionale, ampliandola. In aggiunta a ciò, il nuovo metodo riduce, per non dire *annulla*, i tempi in cui il montatore e lo story editor esaminano tutto il girato per estrapolare delle storie, delegando quest'attività alla trascrizione automatica e lettura dei testi da parte dell'*Intelligenza Artificiale*. In definitiva, questo è da considerarsi sicuramente come un sostegno nella distribuzione delle risorse, creando i presupposti, eventualmente, per impiegarle altrove durante i tempi di trascrizione.

A livello di montaggio vero e proprio, il contributo maggiore è dato dalla velocità di ricerca delle clip all'interno dell'archivio, comprendendo l'utilizzo dei testi trascritti automaticamente e di una ricerca testuale. Anche in questo caso, seppur minore, il contributo esiste e va considerato. Oltretutto, durante la fase di montaggio si sono generate le tracce audio della voce narrante, senza la necessità di attrezzatura o attori, passaggio che non sarebbe stato possibile avere in assenza di risorse. Anche in questo caso, l'utilizzo di tool come *Eleven Labs* ha contribuito enormemente al raggiungimento dell'obiettivo.

In ultimo, si riporta come l'intera fase di post-produzione, giunta al livello descritto, è stata svolta solamente impiegando due persone, un supervisore ed un esecutore, prevedendo anche in questo caso un vantaggio in termini di risorse.

Considerando, invece, l'approccio dal punto di vista creativo, il vantaggio risiede nella possibilità di impiegare l'*Intelligenza Artificiale* anche nella carenza di persone creative. È opportuno sottolineare come, oltre che essere un contributo, per così dire, riequilibrante delle iniquità date dalla scarsità di risorse, questo workflow apporta dei vantaggi anche più grandi. La velocità, con cui questi strumenti estrapolano storie da un archivio, mantiene aperte più possibilità per raccontarne di differenti, potendo modificare facilmente la sceneggiatura anche durante la fase di editing, sostituendo i frammenti presenti, con altri contenuti nell'archivio. Questa potrebbe essere l'applicazione che porta il bilancio tra vecchio e nuovo metodo dall'essere pari ad essere positivo. Un simile vantaggio è supportato dalla rapidità e semplicità con cui è possibile ricreare tutti i materiali originali non presenti nell'archivio iniziale, come le voci narranti, riscrivendo facilmente le battute senza la necessità di programmare nuove sessioni di registrazione con attore e attrezzatura.

Concludendo, i vantaggi dell'approccio che prevede l'utilizzo di un archivio come base di partenza per un documentario non sono di per sé attribuibili alla modalità possibile sfruttando l'*IA*. Le radici di questa proposta sono da attribuirsi all'inizio degli anni 2000,

con il documentario *Grizzly Man* di Werner Herzog. Questi, infatti, entra in possesso del girato della vita di Timothy Treadwell, un uomo che decide di fuggire dalla società vivendo a protezione degli orsi Grizzly. Il girato diventa l'archivio di partenza, nonché unico bacino da cui Herzog attinge per la realizzazione del documentario [49].

Un esempio più recente, che adotta questo approccio partendo da un archivio, è la docuserie *The Beatles: Get Back*, di Peter Jackson: la docuserie del 2021 racconta la storia degli ultimi successi della band inglese degli anni Sessanta. Il girato proviene interamente da un archivio, composto da clip audio e video girate durante le prove della band, nell'ultimo anno della sua attività. Questo, avrebbe dovuto essere la base per la produzione di un film, da parte dei Beatles, negli anni Sessanta. Ma, visto il loro scioglimento, il film non è mai uscito nelle sale. Peter Jackson, entrato in possesso dell'archivio del film, decide di estrapolare esclusivamente da questo materiale le storie per realizzare la docuserie [50].

L'approccio presentato, dunque, può collocarsi, con le dovute precauzioni e adattamenti, come un'evoluzione nella linea produttiva di opere simili a quelle citate, rivelandosi un possibile apporto nell'estrazione di storie da archivi come quelli utilizzati da Herzog e Jackson.

4.2.2 Limiti riscontrati

Un'analisi completa del metodo sperimentato considera anche quelle che sono state le limitazioni riscontrate durante lo studio. In particolare, i limiti presentati riguardano per lo più strumenti di *LA Generativa*, visto lo studio più approfondito di questi e un'applicazione minore di quelli non generativi.

A livello di risultato ottenuto, i limiti sono da considerarsi conseguenza dello stato embrionale degli strumenti. In generale, il risultato ottenuto dalla generazione di materiale da integrare con quello girato realmente è ancora di qualità e resa inferiori e poco integrabile. La differenza tra ciò che è reale e ciò che è generato si nota, rendendo quindi più difficile, per ora, un approccio ibrido, soprattutto a livello video. Nel caso dell'audio, il divario è sicuramente più sottile, come dimostrato dall'ottima integrazione della voce narrante nel documentario *Forgotten Frontiers*; tuttavia, le buone prestazioni sono fortemente dipendenti dalla lingua, ottenendo risultati ottimi solo in lingua inglese. Il parlato italiano risulta ancora robotico e troppo impostato per poter essere considerato ingannevole all'orecchio umano.

Inoltre, le possibilità limitate nella gestione dei dettagli nell'output rendono sicuramente l'applicazione di questi strumenti uno svantaggio, soprattutto in una produzione in cui le risorse non sono un problema e la gestione dei dettagli diventa fondamentale. In un contesto a basso budget come nell'esperimento condotto, alcune di queste limitazioni sono smussate, ma permangono e vanno considerate. In particolare, oltre all'impossibilità di controllare appieno i dettagli nell'output video ottenuto, anche l'interpretazione delle voci generate può costituire un problema. Gli slider che permettono di modificare la resa su

Eleven Labs sono poco indicativi di modifiche all'interpretazione. Oltretutto, l'utilizzo dei prompt si rivela scomodo nella gestione dei dettagli perché dipende da una corretta descrizione soggettiva di ciò che si vuole correggere.

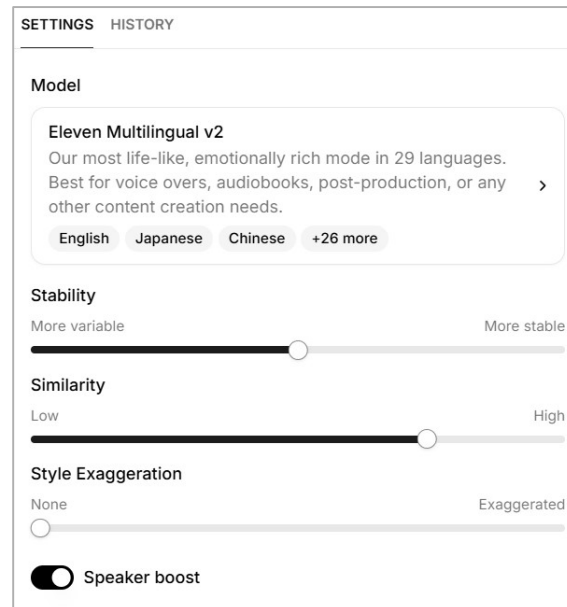


Figura 65 - Impostazioni della voce su Eleven Labs

Ulteriori svantaggi da tenere in considerazione sono quelli della generazione limitata nella lunghezza degli output e della necessità di utilizzare le versioni a pagamento. Nel primo caso, l'output di un sistema di *IA Generativa* è ridotto nelle sue caratteristiche, come la risoluzione, la lunghezza (nel caso dei video), il frame rate e in generale le informazioni contenute nel file, come profondità di bit e profilo colore. Questo rende i file meno manipolabili in post-produzione e meno integrabili in un flusso di lavoro ibrido di alto livello. La manipolazione dei file deve avvenire all'interno dei tool, modificando i dettagli tramite il prompting, ma si ripresenta il problema descritto in precedenza.

Il secondo caso, invece, implica che, volendo applicare gli strumenti ad un workflow integrale, è sicuramente inevitabile dover passare alla versione a pagamento degli strumenti, come fatto con *ChatGPT* nell'esempio sperimentale. Questo perché alcune funzionalità o, semplicemente, la possibilità di utilizzare gli algoritmi migliori è riservata solo agli abbonati. Questo senz'altro diventa un costo da considerare e a cui doversi adattare anche in contesti a basso budget. Le prove gratuite esistono, ma in un contesto in cui alla base c'è l'iterazione dell'output queste non sono sufficienti per ottenere risultati accettabili.

Queste limitazioni costituiscono dei compromessi che il metodo proposto impone. Per questo motivo, attualmente, il contesto di applicazione migliore è quello in cui la mancanza di risorse può limare la pesantezza degli adattamenti. In presenza di risorse, il confronto tra flusso di lavoro tradizionale e quello presentato non ha ancora raggiunto una parità in tutte le sue fasi.

4.2.3 Possibili migliorie e integrazioni

L'analisi di vantaggi e limiti riscontrati nel workflow conduce ad ipotizzare brevemente quali possano essere delle modifiche e integrazioni vantaggiose al fine di estendere la possibilità, a studi futuri, di proseguire quanto ad oggi imbastito.

Innanzitutto, come già mostrato nell'applicazione pratica, esistono molti passaggi di questa solamente abbozzati, ma che meritano un'esplorazione più approfondita viste le opportunità descritte nel capitolo 2.

In particolare, ci si riferisce alla possibilità di studiare più nel profondo l'applicazione di un flusso di lavoro più elaborato nella generazione di clip, integrabili con quelle presenti nell'archivio e utilizzabili in maniera più concreta, affiancandole a quelle girate. Questo è da applicarsi sia alla generazione di storie animate, come iniziato nell'esperimento pratico, sia alla generazione di clip che possano integrarsi con i *b-roll* presenti nell'archivio.

Allo stesso modo, uno studio approfondito sull'audio generativo, che prosegua il flusso abbozzato nella generazione di musiche, è un aspetto che merita di essere preso in considerazione.

Nel caso della generazione delle voci narranti, esiste la possibilità di scegliere pause ed emozioni della voce in output. Se per le pause il metodo è più consolidato, per quanto riguarda le emozioni il tool è meno avanzato. Nel caso della volontà di introdurre una pausa nella voce narrante generata, è possibile inserire il comando `<break time= "xs"/>`, dove *x* indica la durata in secondi della pausa che si vuole immettere. Questo impone alla generazione dell'audio una pausa in corrispondenza del punto in cui è stato inserito il comando. La generazione di battute pronunciate con una certa emozione specifica, invece, è più rudimentale e consiste nello scrivere le battute come se fossero discorsi diretti di un racconto, tra virgolette e seguiti da un'indicazione del tono di voce con cui viene pronunciata la frase. Di seguito si riporta un esempio più specifico di testo da inserire nel tool:

`"What are you doing here?" she shouted angrily`

In questo esempio, l'output ottenuto sarà un audio che contiene l'intera frase, compresa delle parole fuori dalle virgolette, ma con il virgolettato pronunciato con un tono di voce più arrabbiato. Inoltre, è possibile imporre alla generazione della voce una maggiore intensità nella pronuncia di una parola, scrivendola a caratteri maiuscoli. Chiaramente, questo approccio è ancora sperimentale, ma lascia ben intendere che si possa essere vicini ad una soluzione, che può giungere magari attraverso l'introduzione di tag codificati che definiscano le emozioni da utilizzare, evitando la generazione del testo fuori dalle virgolette.

Traendo spunti da alcuni problemi riscontrati nel workflow, una miglioria possibile nella gestione dell'archivio, per come presentata, è quella di trascrivere anche le clip presenti

nell'archivio che non contengono interviste, ma solamente momenti di copertura. Creando un file specifico e sottoponendolo all'analisi di *ChatGPT*, insieme alle trascrizioni e alle analisi delle interviste, si potrebbe ipotizzare un miglioramento significativo alle descrizioni visive nella sceneggiatura generata, derivato in percentuale minore della creatività dell'algoritmo e più aderente, nel concreto, al materiale di archivio. In questo modo, anche il montaggio delle coperture potrebbe risultare significativamente più veloce e la creatività di *ChatGPT* potrebbe esprimersi in un contesto ancora più scalabile su ciò che l'archivio offre. Chiaramente, questa è solamente un'ipotesi ancora non dimostrata, che, se si realizzasse a livello pratico, dovrebbe necessariamente prevedere un ulteriore impiego di risorse, secondo modalità manuale, nella scrittura di un file di inventario delle clip di archivio. Tale proposta, vuole essere solamente uno spunto per eventuali nuovi studi nell'ambito.

4.3 Prospettive e aspettative future

Concludendo l'analisi del metodo presentato, ci si sofferma sul capire quali prospettive future potranno eventualmente migliorarlo.

La tendenza che si già si intravede per il futuro si concentra sul miglioramento della fase di prompting. Abbattere la barriera testuale o confinarla a piccole parole o frasi può essere un'aspettativa futura concreta per rendere più efficace l'utilizzo di questi strumenti. Le possibilità di modificare in maniera più dettagliata il prompt renderebbero l'iterazione più mirata a correggere le imperfezioni, mantenendo ciò che, invece, convince.

A livello di industria, la maggior parte degli output non è attualmente all'altezza dello standard. Tuttavia, la fase attuale è da considerarsi ancora uno sviluppo embrionale di questi strumenti, le cui capacità e possibilità crescono rapidamente di anno in anno. In più, l'approccio dei colossi hollywoodiani a questo nuovo mondo lascia intendere che le aspettative, insieme alle potenzialità, sono molto alte. Oltre a James Cameron, anche Lionsgate e Marvel sono entrate nel mondo dell'*Intelligenza Artificiale*, la prima stringendo una nuova partnership con *Runway* [51], la seconda realizzando i titoli di testa per la serie "*Secret Invasion*" utilizzando l'*Intelligenza Artificiale Generativa* [52].

Pertanto, l'applicazione di questi strumenti di *Intelligenza Artificiale* nel cinema è un mondo in costante evoluzione, che potrà, in futuro, rivoluzionare tutti i processi produttivi alla base dell'industria cinematografica.

Ringraziamenti

Vorrei riservare questo spazio finale della mia tesi di laurea ai ringraziamenti verso tutti coloro che hanno contribuito, con un instancabile supporto, alla stesura della stessa e a quelle persone che, per la loro presenza essenziale, hanno reso questi anni universitari colmi di vita.

In primo luogo, i miei più sentiti ringraziamenti vanno al professor Mattia Meloni, molto più che un'azienda ospitante per la mia tesi, sempre presente, disponibile ed entusiasta di questo percorso compiuto insieme. Un amico.

Vorrei ringraziare, poi, la professoressa Tatiana Mazali, relatrice di questo elaborato, per avermi seguito con grande professionalità e comprensione nella conclusione dei miei studi.

I miei genitori, Simona e Alessandro, e mio fratello Andrea, per aver sempre sostenuto le mie scelte con i fatti e per avermi lasciato andare quando lo chiedevo, senza mai farmi mancare nulla al ritorno. A loro va un ringraziamento speciale.

A Miriam, la persona che più ha creduto in me, anche quando ero io il primo a non farlo, e che ha capito ciò che ero e ciò che desideravo, prima ancora che lo facessi io stesso. Grazie per essere sempre stata dall'altro capo di *quel filo*, senza mai aver mai mollato la presa, anche nei momenti in cui la distanza lo ha teso maggiormente.

A Michela e Diego, grazie per avermi accolto e supportato con calore, facendomi sempre sentire parte della famiglia.

Ringrazio, in ultimo, tutti i miei amici, menzionandoli di seguito, per semplicità, cominciando da chi ho conosciuto prima.

A Elena, Samuele ed Elisa, grazie per aver condiviso con me i sogni, le insicurezze e le vittorie di questi cinque anni insieme. In voi ho riconosciuto la mia stessa voglia di trovare un posto nel mondo in grado di rispettare i nostri confini.

Ai miei coinquilini, Federico e Francesco, per aver reso la nostra casa un luogo sicuro e pacifico, dove hanno sempre abitato serenità, collaborazione e divertimento.

A Marinella, grazie per avermi sempre fatto sentire speciale e per aver reso più miti questi due anni di magistrale, percorrendoli insieme a me, in ogni scelta, progetto e occasione, dentro e fuori dall'università.

Agli Amici delle Panche, grazie per avermi portato a conoscere la leggerezza e la spensieratezza con cui affrontare la vita.

A tutti voi, rinnovo i miei più sinceri ringraziamenti, per avermi sempre fatto sentire voluto bene e per aver allontanato, da me, la solitudine.

Matteo

Bibliografia

- [1] StabilityAI, «James Cameron, Academy Award-Winning Filmmaker, Joins Stability AI Board of Directors,» 24 Settembre 2024. [Online]. Available: <https://stability.ai/news/james-cameron-joins-stability-ai-board-of-directors>. [Consultato il giorno 30 Ottobre 2024].
- [2] Salesforce, «Che cos'è la quarta rivoluzione industriale?,» 14 Agosto 2019. [Online]. Available: <https://www.salesforce.com/it/blog/che-cosa-quarta-rivoluzione-industriale/>. [Consultato il giorno 19 Novembre 2024].
- [3] P. Stone, R. Brooks, E. Brynjolfsson, R. Calo, O. Etzioni, G. Hager, J. Hirschberg, S. Kalyanakrishnan, E. Kamar, S. Kraus, K. Leyton-Brown, D. Parkes, W. Press, A. Saxenian, J. Shah, M. Tambe e Astro, «Artificial Intelligence and Life in 2030.,» Stanford University, Stanford, 2016.
- [4] J. Weizenbaum, «ELIZA—a computer program for the study of natural language communication between man and machine,» *Commun. ACM*, vol. 9, n. 1, p. 36–45, 1966.
- [5] J. R. Searle, «Minds, brains, and programs.,» *Behavioral and Brain Sciences*, pp. 417-457, 1980.
- [6] C. Janiesch, P. Zschech e K. Heinrich, «Machine learning and deep learning,» *Electronic Markets*, vol. 31, 2021.
- [7] M. Soori, B. Arezoo e R. Dastres, «Artificial intelligence, machine learning and deep learning in advanced robotics, a review,» *Cognitive Robotics*, vol. 3, pp. 54-70, 2023.
- [8] V. Gupta, V. Mishra, P. Singhal e A. Kumar, «An Overview of Supervised Machine Learning Algorithm,» *IEEE - 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, 2022.
- [9] E. Y. Li, «Artificial neural networks and their business applications,» *Information & Management*, vol. 27, n. 5, pp. 303--313, 1994.
- [10] C. Wankhede, «What is generative AI?,» 21 Giugno 2023. [Online]. Available: <https://www.androidauthority.com/what-is-generative-ai-3321354/>. [Consultato il giorno 24 Ottobre 2024].
- [11] J. Brownlee, «Best Resources for Getting Started With GANs,» 12 Luglio 2019. [Online]. Available: <https://machinelearningmastery.com/resources-for-getting-started-with-generative-adversarial-networks/>. [Consultato il giorno 24 Ottobre 2024].

- [12] J. Ho, A. Jain e P. Abbeel, «Denoising Diffusion Probabilistic Models,» *CoRR*, vol. abs/2006.11239, 2020.
- [13] P. Dhariwal e A. Nichol, «Diffusion Models Beat GANs on Image Synthesis,» *CoRR*, vol. abs/2105.05233, 2021.
- [14] Q. Wang, Z. Qiu, Y. Zhang, J. Chen, H. Lin, X. Liu e Y. Liu, «A Comprehensive Overview of Large Language Models,» 2023. [Online]. Available: <https://ar5iv.labs.arxiv.org/html/2307.06435>. [Consultato il giorno 24 Ottobre 2024].
- [15] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior e K. Kavukcuoglu, «WaveNet: A Generative Model for Raw Audio,» 19 Settembre 2016. [Online]. Available: <https://arxiv.org/pdf/1609.03499v2>. [Consultato il giorno 24 Ottobre 2024].
- [16] E. Canorea, «Discriminative AI vs Generative AI: Keys to understanding them,» 15 Febbraio 2024. [Online]. Available: <https://www.plainconcepts.com/discriminative-ai-vs-generative-ai/#:~:text=Non%2Dgenerative%20AI%20models%20perform,music%2C%20designing%20images%2C%20etc..> [Consultato il giorno 24 Ottobre 2024].
- [17] S. Saha, «A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way,» 15 Dicembre 2018. [Online]. Available: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>. [Consultato il giorno 25 Ottobre 2024].
- [18] Z. Keita, «An Introduction to Convolutional Neural Networks (CNNs),» 14 Novembre 2023. [Online]. Available: <https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns>. [Consultato il giorno 24 Ottobre 2024].
- [19] C. Wankhede, «What is Midjourney AI and how does it work?,» 8 Settembre 2024. [Online]. Available: <https://www.androidauthority.com/what-is-midjourney-3324590/>. [Consultato il giorno 2 Novembre 2024].
- [20] Midjourney, «The Web App is open to all,» 23 Agosto 2024. [Online]. Available: <https://www.midjourney.com/updates/the-web-is-open-to-all>. [Consultato il giorno 2 Novembre 2024].
- [21] Midjourney, «Midjourney documentation,» [Online]. Available: <https://docs.midjourney.com/docs/quick-start>. [Consultato il giorno 2 Novembre 2024].
- [22] OpenAI, «DALL-E 3,» [Online]. Available: <https://openai.com/index/dall-e-3/>. [Consultato il giorno 2 Novembre 2024].

- [23] OpenAI, «Improving Image Generation with Better Captions,» 2024. [Online]. Available: <https://cdn.openai.com/papers/dall-e-3.pdf>. [Consultato il giorno 2 Novembre 2024].
- [24] StabilityAI, «Adversarial Diffusion Distillation,» 28 Novembre 2023. [Online]. Available: https://static1.squarespace.com/static/6213c340453c3f502425776e/t/65663480a92fba51d0e1023f/1701197769659/adversarial_diffusion_distillation.pdf. [Consultato il giorno 2 Novembre 2024].
- [25] StabilityAI, «Introducing Stable Diffusion 3.5,» 22 Ottobre 2024. [Online]. Available: <https://stability.ai/news/introducing-stable-diffusion-3-5>. [Consultato il giorno 2 Novembre 2024].
- [26] LeonardoAI, «Leonardo Documentation,» [Online]. Available: <https://docs.leonardo.ai/docs/generate-images-using-leonardo-phoenix-model>. [Consultato il giorno 2 Novembre 2024].
- [27] P. Esser, J. Chiu, P. Atighehchian, J. Granskog e A. Germanidis, «Structure and Content-Guided Video Synthesis with Diffusion Models,» 6 Febbraio 2023. [Online]. Available: <https://arxiv.org/abs/2302.03011>. [Consultato il giorno 3 Novembre 2024].
- [28] A. Germanidis, «Gen-2: Generate novel videos with text, images or video clips,» Febbraio 2023. [Online]. Available: <https://runwayml.com/research/gen-2>. [Consultato il giorno 3 Novembre 2024].
- [29] A. Germanidis, «Introducing Gen-3 Alpha: A New Frontier for Video Generation,» 17 Giugno 2024. [Online]. Available: <https://runwayml.com/research/introducing-gen-3-alpha>. [Consultato il giorno 3 Novembre 2024].
- [30] Runway, «Introducing Act-One. A new way to generate expressive character performances using simple video inputs,» 22 Ottobre 2024. [Online]. Available: <https://runwayml.com/research/introducing-act-one>. [Consultato il giorno 30 Ottobre 2024].
- [31] KlingAI, «KLING, MAKE IMAGINATION ALIVE,» [Online]. Available: <https://klingai.io/>. [Consultato il giorno 3 Novembre 2024].
- [32] StabilityAI, «Introducing Stable Video Diffusion,» 21 Novembre 2023. [Online]. Available: <https://stability.ai/news/stable-video-diffusion-open-ai-video-model>. [Consultato il giorno 3 Novembre 2024].
- [33] StabilityAI, «Introducing Stable Audio 2.0,» 3 Aprile 2024. [Online]. Available: <https://stability.ai/news/stable-audio-2-0>. [Consultato il giorno 3 Novembre 2024].

- [34] StabilityAI, «Stable Audio: Fast Timing-Conditioned Latent Audio Diffusion,» 13 Settembre 2024. [Online]. Available: <https://stability.ai/research/stable-audio-efficient-timing-latent-diffusion>. [Consultato il giorno 3 Novembre 2024].
- [35] ElevenLabs, «Documentazione di ElevenLabs,» [Online]. Available: <https://elevenlabs.io/docs/introduction>. [Consultato il giorno 3 Novembre 2024].
- [36] C. Wankhede, «What does GPT stand for in ChatGPT?,» 23 Aprile 2023. [Online]. Available: <https://www.androidauthority.com/what-does-gpt-stand-for-in-chatgpt-3316137/>. [Consultato il giorno 3 Novembre 2024].
- [37] OpenAI, «Hello GPT-4o,» 13 Maggio 2024. [Online]. Available: <https://openai.com/index/hello-gpt-4o/>. [Consultato il giorno 3 Novembre 2024].
- [38] OpenAI, «Memory and new controls for ChatGPT,» 13 Febbraio 2024. [Online]. Available: <https://openai.com/index/memory-and-new-controls-for-chatgpt/>. [Consultato il giorno 30 Ottobre 2024].
- [39] OpenAI, «Introducing Canvas. A new way of working with ChatGPT to write and code,» 3 Ottobre 2024. [Online]. Available: <https://openai.com/index/introducing-canvas/>. [Consultato il giorno 30 Ottobre 2024].
- [40] Anthropic, «Introducing computer use, a new Claude 3.5 Sonnet, and Claude 3.5 Haiku,» 22 Ottobre 2024. [Online]. Available: <https://www.anthropic.com/news/3-5-models-and-computer-use>. [Consultato il giorno 3 Novembre 2024].
- [41] Anthropic, «Documentazione di Claude,» [Online]. Available: <https://docs.anthropic.com/en/docs/welcome>. [Consultato il giorno 3 Novembre 2024].
- [42] E. Carabelli e A. Simonetti, «DaVinci Resolve 19 l'I.A. rivoluziona la post-produzione,» [Online]. Available: <https://digitalproduction.tv/davinci-resolve-19-li-a-e-post-produzione/>. [Consultato il giorno 4 Novembre 2024].
- [43] OpenAI, «Video generation models as world simulators,» 15 Febbraio 2024. [Online]. Available: <https://openai.com/index/video-generation-models-as-world-simulators/>. [Consultato il giorno 4 Novembre 2024].
- [44] OpenAI, «Sora: first impressions,» 25 Marzo 2024. [Online]. Available: <https://openai.com/index/sora-first-impressions/>. [Consultato il giorno 4 Novembre 2024].
- [45] DeepMind Google, «Veo, our most capable generative video model,» 2024. [Online]. Available: <https://deepmind.google/technologies/veo/>. [Consultato il giorno 4 Novembre 2024].

- [46] J. Hamilton, «How to Write a Book in ChatGPT: My Complete Process,» 19 Gennaio 2024. [Online]. Available: <https://kindlepreneur.com/chatgpt-writing/>. [Consultato il giorno 30 Ottobre 2024].
- [47] OpenAI, «Introducing GPTs,» 6 Novembre 2023. [Online]. Available: <https://openai.com/index/introducing-gpts/>. [Consultato il giorno 30 Ottobre 2024].
- [48] R. Awati, «What are AI hallucinations and why are they a problem?,» Ottobre 2024. [Online]. Available: <https://www.techtarget.com/whatis/definition/AI-hallucination>. [Consultato il giorno 29 Ottobre 2024].
- [49] S. Pecetta, «Werner Herzog - L'ultimo conquistatore,» [Online]. Available: https://www.ondacinema.it/monografie/scheda/werner_herzog.html. [Consultato il giorno 30 Ottobre 2024].
- [50] C. Potter, «Peter Jackson Talks The Beatles: Get Back,» 23 Novembre 2021. [Online]. Available: <https://d23.com/peter-jackson-talks-the-beatles-get-back/>. [Consultato il giorno 31 Ottobre 2024].
- [51] Runway, «Runway partners with Lionsgate,» 18 Settembre 2024. [Online]. Available: <https://runwayml.com/news/runway-partners-with-lionsgate>. [Consultato il giorno 31 Ottobre 2024].
- [52] Z. Millman, «Yes, Secret Invasion's opening credits scene is AI-made — here's why,» 2023 Giugno 2023. [Online]. Available: <https://www.polygon.com/23767640/ai-mcu-secret-invasion-opening-credits>. [Consultato il giorno 31 Ottobre 2024].