



**Politecnico
di Torino**

Politecnico di Torino

Laurea Magistrale Ingegneria del Cinema e dei Media Digitali

A.a. 2025/2026

Sessione di laurea Marzo 2026

Creare contenuti audiovisivi con l'IA: La sfida dei Modelli Generativi tra Innovazione e Copyright

Relatori:

Andrea Bottino

Candidati:

Francesco Carmelo Freni

Sommario

L'avvento dell'intelligenza artificiale generativa ha innescato una trasformazione radicale nel settore audiovisivo, stravolgendone i flussi di lavoro tradizionali. Il presente lavoro di tesi esplora le potenzialità e le criticità di tale rivoluzione, attraverso la sperimentazione diretta dei modelli generativi allo stato dell'arte. L'obiettivo centrale della ricerca è stato la creazione di un'opera audiovisiva sintetica, mantenendo però un rigoroso controllo registico sulla composizione e sul linguaggio cinematografico, superando la natura spesso stocastica di tali strumenti. Parallelamente alla fase realizzativa, la tesi affronta il complesso nodo del diritto d'autore e delle implicazioni etico-giuridiche legate al copyright dei contenuti generati. L'analisi si conclude con una valutazione critica dei limiti tecnologici attuali e delle prospettive future, delineando come l'IA possa trasformarsi da elemento di rottura a strumento di co-creazione consapevole per i professionisti del settore.

Ringraziamenti

Desidero esprimere la mia più sincera gratitudine al mio relatore, per la costante disponibilità nel rispondere a ogni mio dubbio e perplessità e per avermi guidato durante questo percorso di stesura. Un ringraziamento particolare va a EDI Effetti Digitali Italiani, e in special modo al team di Generative, per la preziosa esperienza di tirocinio, che ha arricchito profondamente il mio percorso formativo e professionale. Un pensiero va ai miei amici e compagni di università, con cui ho condiviso sia momenti di gioia che di grande difficoltà. Per concludere, un ringraziamento speciale va alla mia famiglia, che mi ha sempre sostenuto e senza la quale questa laurea non sarebbe stata possibile.

Indice

Elenco delle figure	VI
1 Introduzione	1
1.1 La pipeline tradizionale di produzione	1
1.2 L'importanza degli effetti visivi	2
1.3 La pipeline tradizionale degli effetti visivi	3
1.4 La nascita di una pipeline ibrida	5
2 Stato dell'arte	8
2.1 Leggi sull'IA	8
2.2 C2PA, CAI e Content Credentials	11
2.3 Adobe	13
2.4 Google	16
2.5 RunwayML	19
2.6 Luma AI	22
2.7 OpenAI	24
2.8 KlingAI	27
2.9 Getty Images vs Stability AI	30
2.10 Artisti contro Midjourney e StabilityAI	31
2.11 Altri modelli	32
2.11.1 ByteDance	32
2.11.2 Flux	33
2.11.3 Qwen	33
2.11.4 Wan	34
2.11.5 Grok	34
2.12 ComfyUI	34
2.13 The Last Image	35
3 Pre-produzione e concept	38
3.1 Sviluppo dell'idea	39
3.2 Storia	39

3.2.1	Struttura delle scene	39
3.2.2	Elementi narrativi	40
3.2.3	Divisione in scene	40
3.3	Pitch	43
3.4	Storyboard e Animatic	44
4	Produzione e IA Generativa	45
4.1	Personaggi	45
4.1.1	Generazione dei personaggi della prima sequenza	45
4.1.2	Generazione dei personaggi della seconda sequenza	48
4.1.3	Generazione dei personaggi della terza sequenza	52
4.1.4	Generazione di altri elementi della scena	54
4.2	Creazione delle scene	56
4.2.1	Generazione della prima sequenza narrativa	57
4.2.2	Generazione della seconda sequenza narrativa	71
4.2.3	Generazione della terza sequenza narrativa	80
4.2.4	Generazione della sequenza finale	85
5	Post-produzione e montaggio	88
5.1	Montaggio	88
5.2	Altre lavorazioni	88
6	Conclusioni	90
6.1	Risultati, difficoltà e prospettivi futuri	90
	Bibliografia	92

Elenco delle figure

1.1	Fotogramma tratto da <i>Avatar</i> (2025)	3
1.2	Fotogramma tratto da <i>Rigging Intro - Blender 2.80 Fundamentals</i>	4
1.3	VFX Pipeline	5
1.4	Ripresa aerea generata di Torino	6
2.1	I livelli di rischio definiti dall' AI Act	9
2.2	Content Credentials	11
2.3	Durable Content Credentials	12
2.4	Schermata Home - Adobe Firefly	13
2.5	Adobe Content Authenticity	16
2.6	Generazione di immagini tramite Nano Banana su Gemini	17
2.7	Generazione di immagini e video tramite Flow	18
2.8	Runway - Sezione 'Video Apps'	21
2.9	LumaAI Video-To-Video	24
2.10	Storyboard - Sora	26
2.11	Kling Omni	29
2.12	Fotogramma estratto da 'The Last Image'	36
2.13	Integrazione di elementi realizzati in CG	36
2.14	Lavorazioni di compositing e finishing	37
3.1	Alcune tavole di Storyboard tratte dalla prima sequenza narrativa	44
4.1	Turn-around del personaggio del bambino	47
4.2	Turn-around del personaggio della madre	48
4.3	Prima generazione del personaggio del pugile	49
4.4	Turn-around del personaggio del pugile	50
4.5	Step di generazione del personaggio del robot minaccioso	51
4.6	Step di generazione del personaggio del robot	52
4.7	Step di generazione del personaggio della terza sequenza narrativa	53
4.8	Step di generazione dell'asset dell'orsacchiotto	54
4.9	Generazione dell'ambiente della seconda sequenza narrativa	56

4.10	Da storyboard a generazione	58
4.11	Primo e ultimo frame Scena 1 Shot 02	60
4.12	Step di generazione per la prima Still della Scena 1 Shot 05	63
4.13	Primo e ultimo frame Scena 1 Shot 05	65
4.14	Alcuni step di generazione falliti - Scena 1 Shot 07	67
4.15	Lavorazioni in Adobe Photoshop - Scena 1 Shot 07	67
4.16	Step di generazione - Scena 1 Shot 11	70
4.17	Versioning su Flow - Scena 1 Shot 11	70
4.18	Step di generazione - Scena 05 Shot 04	73
4.19	Face swap e cambio dell'espressione facciale - Scena 05 Shot 04 . . .	74
4.20	Step di generazione preliminari - Scena 05 Shot 06	75
4.21	Step di generazione iterativi - Scena 05 Shot 06	76
4.22	Creazione del First Frame e del Last Frame - Scena 05 Shot 06 . . .	77
4.23	Sequenza narrativa - Scena 05 Shot 06/ Shot 07	77
4.24	First e last frame - Scena 08 Shot 08	78
4.25	Video-To-Video - Scena 08 Shot 08	80
4.26	Still - Scena 06 Shot 05	81
4.27	Sequenza narrativa Shot 06 - Shot 01/02/03	83
4.28	Generazione video in Multi-Shot Shot 03 - Shot 01/02	84
4.29	Da fotorealistico a fumetto	86
5.1	Intervento sulla pupilla del bambino	89

Capitolo 1

Introduzione

Negli ultimi anni l'Intelligenza Artificiale Generativa ha registrato un rapido avanzamento, segnato da un'accelerazione esponenziale nei mesi più recenti. L'impatto sulla filiera audiovisiva è stato dirompente: integrandosi all'interno della pipeline di produzione, l'Intelligenza Artificiale (IA) ha ridefinito i tradizionali paradigmi operativi sia nella realizzazione dell'opera, dall'idea al prodotto finale, sia nelle specifiche lavorazioni legate agli effetti visivi (VFX).

In questo scenario di profonda trasformazione tecnologica si colloca il presente elaborato, sviluppato a seguito dell'attività di tirocinio svolta presso EDI Effetti Digitali Italiani, leader nel settore dei VFX in Italia, e la realizzazione di un prodotto audiovisivo generato, con controllo registico, supervisionato, già nelle prime fasi della pre-produzione, dal Team di Generative dell'azienda.

L'obiettivo primario del progetto è stato la realizzazione integrale di uno spot, dalla definizione del concept e del relativo storyboard, fino alla generazione delle effettive sequenze video. L'intero processo è stato guidato dalla necessità di mantenere elevati standard qualitativi, gestendo al contempo le criticità legate al diritto d'autore e ai filtri di sicurezza dei modelli generativi.

1.1 La pipeline tradizionale di produzione

La produzione di un'opera audiovisiva segue un modello lineare e sequenziale, comunemente diviso in tre macro fasi.

La pre-produzione include tutto ciò che precede le riprese sul set, o nel caso di un prodotto di animazione (2D, 3D o in Stop Motion) tutto ciò che precede l'inizio delle effettive lavorazioni. L'idea iniziale viene elaborata, scrivendo prima una logline, poi un soggetto e un trattamento. Per supportare visivamente l'idea è pratica comune associare una moodboard (raccolta di reference volta ad esprimere l'atmosfera e l'emozione che si vuole ricreare), realizzare dei disegni di concept

art, di character design e impacchettare il tutto dentro un pitch da sottoporre a eventuali produzioni. Approvato il progetto viene scritta la sceneggiatura e realizzato lo storyboard. Una pratica comune è anche quella di montare, sulla timeline di un qualsiasi programma di editing, le tavole di storyboard, così da creare un animatic. Inizia adesso una vera e propria fase di pianificazione: le varie maestranze eseguono uno spoglio della sceneggiatura, utile a stilare il budget; si dà il via a una fase di scout location, casting per gli attori, realizzazione di scenografie e costumi (nel caso di animazioni la creazione dei personaggi e degli asset necessari) fino ad arrivare alla creazione di documenti come l'Ordine del Giorno e la Shotlist, fondamentali durante le riprese.

La produzione è il momento in cui il progetto "prende vita". Sono momenti frenetici in cui ogni secondo rappresenta un costo. Tante maestranze sono coinvolte, tutto deve essere perfettamente organizzato affinché il lavoro proceda in maniera efficace.

A seguire la fase di post-produzione. Le riprese realizzate durante la produzione vengono editate dal montatore designato. Raggiunto il 'blocco del montaggio' (Picture Lock), il progetto viene inoltrato ai vari reparti, così che il lavoro del sound designer e il lavoro dei VFX, sulle scene in cui sono richiesti, possa iniziare in parallelo, dove possibile. Gli output uscenti dai diversi software verranno conformati e inviati al colorist che si occuperà della color correction e del color grading. Infine, si accede alla fase di Mastering e Delivery, in cui vengono effettuati gli interventi di finishing e i diversi export con formati e codec specifici, in funzione di dove verrà distribuito il prodotto.

1.2 L'importanza degli effetti visivi

Con il termine effetti visivi o VFX si fa riferimento a una qualsiasi immagine creata, alterata o migliorata, per un film o un qualsiasi altro media, che non possa essere ottenuta sul set durante le riprese. Questa è la differenza sostanziale con il termine più comunemente usato 'effetti speciali' (SFX), che nel gergo tecnico si riferisce più propriamente agli effetti che vengono eseguiti sul set, come ad esempio la macchina del fumo o della pioggia. Esistono tre motivi principali per cui ricorrere ai VFX:

- quando non esiste nessun modo pratico sul set di arrivare al risultato voluto;
- quando esiste un modo pratico ma potrebbe compromettere la sicurezza di qualcuno;
- quando ricorrere agli effetti visivi è meno costoso.

La diffusione degli effetti visivi all'interno dei prodotti multimediali è incrementata costantemente nel tempo. Oggigiorno sono presenti in quasi la totalità

dei prodotti, spesso in maniera invisibile: dalle piccole cancellazioni di oggetti (clean-up) fino ad arrivare a lavorazione più complesse, in cui i VFX non rappresentano solamente un elemento complementare ma il cuore pulsante dell'intera idea progettuale [1].

Basti pensare a serie TV iconiche come *Stranger Things* (Netflix), *Fallout* (Prime Video) o film del calibro di *Dune*, per rendersi conto di come gli effettivi visivi siano fondamentali nella creazione del World-Building, dando vita all'universo in cui avviene la narrazione. Un caso esemplare di questo concetto è rappresentato dalla saga di *Avatar*, in cui James Cameron, avvalendosi delle più avanzate tecnologie, ha la possibilità di muovere fisicamente una camera virtuale (Virtual Camera), riprodotta su un set virtuale previsualizzato (il mondo di Pandora). All'interno di questo spazio le movenze degli attori vengono catturate tramite tute di motion capture e trasferite ai modelli 3D dei Na'vi, personaggi protagonisti della storia.



Figura 1.1: Fotogramma tratto da *Avatar* (2009)

Fonte: [Link al fotogramma](#)

1.3 La pipeline tradizionale degli effetti visivi

Nonostante sia frequente associare gli effetti visivi esclusivamente alla post-produzione, il loro intervento all'interno del ciclo di vita di un prodotto audiovisivo ha inizio già in fase di pre-produzione. Sin dallo spoglio della sceneggiatura e dall'analisi dello storyboard, è possibile individuare le tipologie di lavorazione necessarie e stabilire un dialogo costante con la regia. Questa sinergia permette di pianificare le riprese in funzione dei VFX, garantendo l'efficacia tecnica senza compromettere la direzione artistica. Tale pianificazione preventiva consente, inoltre, di allocare

risorse e tempo per effettuare dei test tecnici, fondamentali qualora si debbano affrontare effetti complessi o workflow sperimentali. Successivamente, sul set in fase di produzione, il regista verrà affiancato da un VFX Supervisor, il quale si accerterà che le riprese vengano effettuate in una maniera corretta, tenendo a mente le lavorazioni successive.

Talvolta, prima di iniziare le riprese può anche essere chiesta una previs, ovvero una rappresentazione in 3D di quello che è lo storyboard, utilizzando modelli 3D grezzi e camere virtuali così da testare la sequenzialità e il ritmo della narrazione. Successiva alle riprese è invece la postvis, ovvero un'integrazione grezza tra girato ed elementi 3d, così da avere scene già più complete in fase di montaggio e capire se tutto funzioni correttamente.

Come si può desumere, la fase di modellazione 3D ha inizio già dopo lo spoglio della sceneggiatura, una volta individuati gli asset necessari, procedendo spesso in parallelo alle riprese. A seguire, i modelli verranno poi riggati, animati e passeranno da una fase di texturing, look development (look-dev) e lighting, fino a giungere al rendering finale.

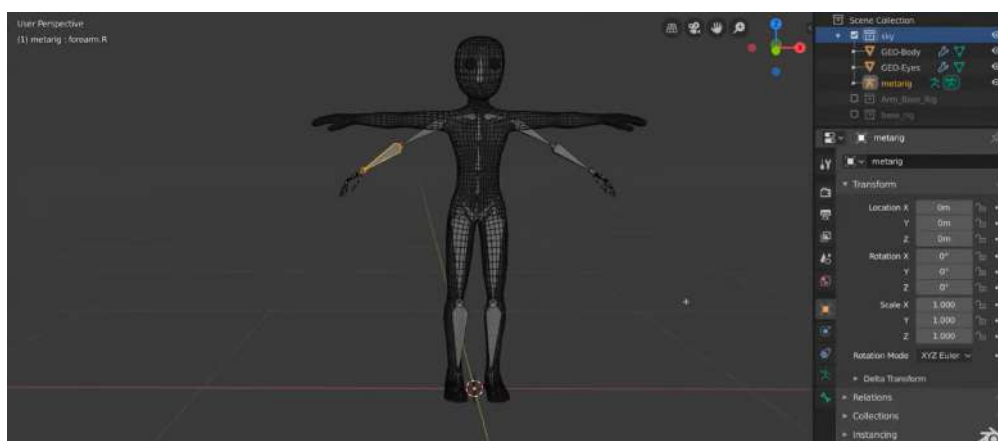


Figura 1.2: Fotogramma tratto da *Rigging Intro - Blender 2.80 Fundamentals*
Fonte: [Link al video](#)

Nel frattempo il girato reale viene pulito dagli elementi indesiderati(clean-up), ed eventualmente esteso, qualora necessario, tramite tecniche di Digital Matte Painting (DMP), spesso realizzate in 2.5D. Vengono estratte le camere così da avere un tracking della scena da riprodurre nell'ambiente virtuale, permettendo al team di layout di preparare gli ingombri che verranno poi sostituiti dagli elementi CG (Computer Generated).

Nel caso in cui siano richiesti fenomeni naturali o fisici, come, ad esempio la presenza di fuoco e fumo, esplosioni o masse d'acqua da realizzare in CG, entrerà

in gioco il team di Effects (FX) incaricato di eseguire le simulazioni dinamiche necessarie.

Il tutto confluirà, infine, verso il team di Compositing che si occuperà di integrare perfettamente i vari elementi sintetici con il girato reale, mirando a conferire al footage un senso di uniformità indispensabile per preservare la sospensione dell'incrudelità dello spettatore.

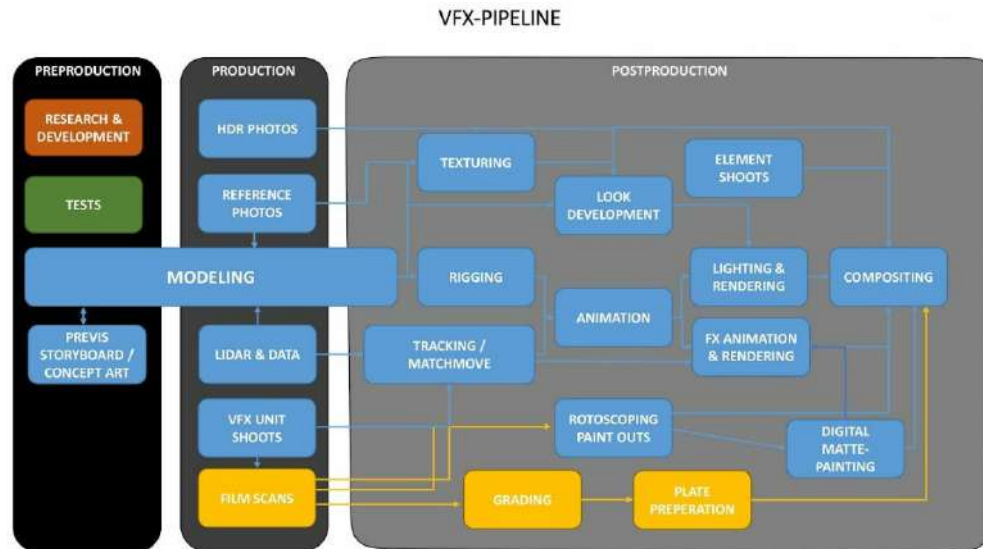


Figura 1.3: VFX Pipeline

Fonte: [Link all'articolo](#)

1.4 La nascita di una pipeline ibrida

Analogamente a quanto riscontrato in altri settori, l'Intelligenza Artificiale ha fatto irruzione nel panorama della creazione dei contenuti audiovisivi, scardinando i flussi di lavoro consolidati a favore di metodologie più rapide e snelle. La nascita di modelli operativi ibridi, in cui l'apporto creativo umano si fonde con l'automazione neurale, riflette la necessità impellente di adattamento tecnologico. In questo contesto, l'evoluzione delle competenze non rappresenta solo una scelta strategica ma un requisito imprescindibile per mantenere la propria competitività professionale e rispondere efficacemente alle nuove domande del mercato.

Nella fase di pre-produzione, l'impiego dei Large Language Models (LLM) offre agli sceneggiatori un supporto fondamentale, permettendo di esplorare variazioni semantiche, raffinare lo stile di esposizione o ottimizzare la stesura di logline e soggetti. Parallelamente, nel campo della Concept Art, l'IA Generativa abilita una visualizzazione immediata delle idee attraverso il prompting testuale: questo

processo di iterazione rapida permette di generare molteplici varianti di un concept, offrendo spesso spunti creativi inattesi. Anche la realizzazione degli storyboard beneficia di questa accelerazione, schizzi preliminari possono essere trasformati in tavole rifinite con estrema rapidità. Tuttavia, il valore dell'esperienza artistica rimane centrale, dato che una solida base tecnica nel disegno manuale garantisce un controllo superiore sugli output generati, facilitando la successiva transizione verso l'animazione. Infine, la creazione di moodboard evolve da una semplice attività di ricerca di reference a una vera e propria generazione di scenari visivi inediti, personalizzati sulle specifiche esigenze estetiche del progetto.

Gli strumenti di IA Generativa possono essere impiegati per elaborare previs qualitativamente più avanzate, ottimizzando la pianificazione delle riprese. Di particolare interesse sono le prime sperimentazioni volte a integrare il girato con l'intelligenza artificiale, come coprire shots mancanti con la generazione di dettagli riempitivi (fegatelli) o realizzare particolari scene a minor costo, di cui un esempio ne sono la generazione di aerial shots al posto dell'uso dei droni o le riprese sott'acqua senza dover costruire complessi set da allagare. Inoltre, a seconda della natura dell'opera, l'intelligenza artificiale generativa può emanciparsi dal ruolo di mero supporto tecnico per configurarsi come un medium autonomo, in grado di generare contenuti sintetici integralmente sostitutivi delle riprese cinematografiche tradizionali.



Figura 1.4: Ripresa aerea generata di Torino

Fonte: Google Gemini

In fase di post-produzione, gli strumenti di Intelligenza Artificiale Generativa diventano utili per ogni reparto, integrandosi nelle lavorazioni esistenti e velocizzandole: i DMP artits e i Compositors possono generare direttamente dei 'contributi' in funzione delle loro necessità, da integrare poi in software come Adobe Photoshop o Nuke; gli FX artits possono focalizzare le proprie conoscenze, meno dal punto vista

tecnico e più dal punto di vista artistico, delegando alla generazione la creazione di effetti che in alternativa avrebbero richiesto lunghe e complesse simulazioni fisiche. Quest'ultime non sono assenti dal ciclo di produzione, ma rimangono essenziali, dato che gli strumenti generativi non consentono ancora un controllo puntuale. Software come Houdini sono lo stato dell'arte e in un prossimo futuro un supporto valido per creare asset sui cui addestrare i propri modelli personalizzati; infine, per quanto riguarda la modellazione 3D, la creazione di asset partendo da disegni 2D o prompt testuali non è propriamente ottimizzata, ma ben presto sarà uno strumento che consentirà agli artisti di avere dei prodotti quasi completi, da migliorare in topologia e look development, in funzione delle proprie esigenze.

Sebbene l'Intelligenza Artificiale possieda un potenziale effettivamente sostitutivo in diverse fasi tecniche, la sfida per il professionista risiede nella capacità di integrarla strategicamente nel proprio workflow. Il valore aggiunto dell'apporto umano diventa dunque il discrimine fondamentale: saper guidare lo strumento attraverso la propria sensibilità artistica per non restare esclusi da un mercato in rapidissima evoluzione.

Capitolo 2

Stato dell'arte

2.1 Leggi sull'IA

Viene utilizzato il termine AI Act [2] per fare riferimento al regolamento europeo, entrato in vigore il 1° agosto 2024, sullo sviluppo e la diffusione responsabile dell'Intelligenza Artificiale. Vengono affrontati i potenziali rischi per la salute, la sicurezza e i diritti fondamentali dei cittadini, chiedendo, tra i vari punti, trasparenza ai fornitori dei modelli di IA. La maggior parte delle norme definite, come stabilito dall'articolo 113 (UE) 2024/1689, entreranno dal 2 agosto 2026, così da creare un periodo transitorio in cui raffinare le modalità di attuazione e permettere agli sviluppatori di IA di rispettare gli obblighi stabiliti prima delle scadenze legali (AI Pact) [3]. Sarà compito del comitato europeo per l'intelligenza artificiale garantirne l'esecuzione uniforme in tutti gli Stati membri dell'UE [4]. La legge sull'IA definisce 4 livelli di rischio per i sistemi [5]:

- Rischio inaccettabile: tutti i sistemi di IA considerati una chiara minaccia per la sicurezza. In questo caso l'UE pone dei veri e propri divieti;
- Rischio elevato: casi d'uso dell' IA che possono comportare gravi rischi per la salute, la sicurezza o i diritti fondamentali. Sarà necessario soddisfare degli obblighi rigorosi prima di poter immettere questi sistemi sul mercato;
- Rischio limitato: si tratta dei rischi associati alla necessità di trasparenza in merito all'uso dell'IA, così da garantire che gli esseri umani siano informati quando necessario. Si impone, inoltre, che i contenuti generati dall' IA siano identificabili ed etichettati in modo chiaro e visibile;
- Rischio minimo o nullo: non vengono presi provvedimenti in merito. Un esempio ne sono i videogiochi abilitati all'intelligenza artificiale o i filtri antispam.

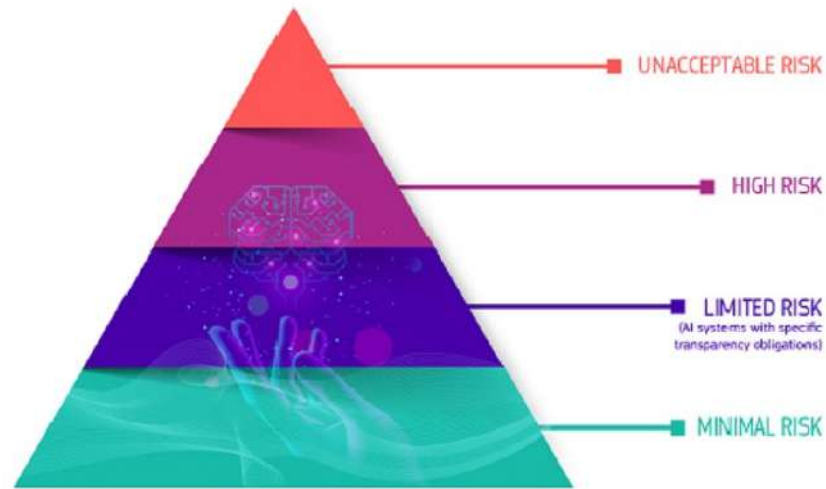


Figura 2.1: I livelli di rischio definiti dall' AI Act

Fonte: [Link all'articolo](#)

In conformità al Regolamento UE 2024/1689, i fornitori di modelli di IA per finalità generali sono tenuti a redigere e aggiornare una documentazione tecnica dettagliata che illustri i processi di addestramento e i risultati delle valutazioni, nel rispetto dei segreti commerciali. Sul fronte della trasparenza verso l'esterno, essi devono pubblicare una sintesi dei contenuti utilizzati per il training e implementare policy rigorose per il rispetto del diritto d'autore. In particolare, deve essere garantito ai titolari dei diritti il potere di opt-out, ovvero di poter scegliere se omettere i propri contenuti dai dataset di training dei modelli. Infine, per contrastare la disinformazione, i contenuti generati o manipolati (testi, audio, immagini o video) devono essere marcati con metadati o filigrane digitali che ne rendano esplicita l'origine artificiale [6].

Tali norme sono notevolmente più stringenti qualora si parlasse di modelli a rischio sistemico, ovvero i modelli considerati più grandi e impattanti, in cui la presenza di errori o falle di sicurezza potrebbero comportare pericoli su tutta l'Unione Europea. L'Unione Europea ha fissato la soglia di 10^{25} FLOPs (Floating Point Operations) per valutare lo sforzo computazionale utilizzato per addestrare i modelli e classificarli come a rischio sistemico. Quando si parla di modelli di così grandi dimensioni è imprescindibile non parlare del mercato cinese e statunitense.

Modelli cinesi, come Qwen di Alibaba rilasciato sotto licenza Apache 2.0, o i modelli di IA Generativa di ByteDance, azienda proprietaria di TikTok, sono modelli ormai ampiamente diffusi che includono un elevato numero di parametri. Inoltre, queste aziende hanno diverse infrastrutture e data center tra Europa e USA. Un esempio ne sono i già citati Alibaba Cloud con data centers tra Francoforte, Londra

e Parigi e ByteDance con data centers in Irlanda e Norvegia. Molti modelli pur avendo un numero inferiore di parametri vengono comunque considerati impattanti e soggetti a maggiori controlli, oltre al fatto che nuove startup, considerando la crescita computazionale nel tempo e la sempre maggiore complessità, potrebbero presto superare la soglia fissata di 10^{25} FLOPs. Nel momento in cui l'AI Act entrerà in vigore a tutti gli effetti sarà significativo osservare come i modelli cinesi si adatteranno, soprattutto per quanto riguarda il tema della trasparenza dei dataset di addestramento e del rispetto del copyright.

Per quanto riguarda il mercato statunitense in mancanza di una legislazione federale sull'IA, la California e lo Stato di New York hanno emanato le proprie regolamentazioni di sicurezza. Il California's SB 53 e il New York RAISE Act sono molto simili all' AI Act Europeo, ma si differenziano nel non imporre agli sviluppatori la mitigazione dei rischi, limitandosi a richiedere il rispetto di rigidi vincoli di trasparenza e reporting. Inoltre, la soglia fissata per i 'frontier models' è di 10^{26} FLOPs, soglia non raggiunta attualmente da nessun modello cinese e che concede maggiore respiro [7].

L'Italia si uniforma all'AI Act con la legge 132/2025, entrata in vigore dal 10 ottobre 2025. Così come per il regolamento Europeo, alcuni suoi punti prenderanno completa validità nel corso del tempo. Alcune introduzioni significative sono:

- viene riconosciuto il diritto d'autore per i prodotti prevalentemente fatti da un'essere umano che utilizzano l'Intelligenza Artificiale come strumento. L'IA deve essere vista come un tool integrato nel workflow e non come principale artefice del prodotto;
- è possibile omettere i contenuti sui cui si ha il diritto d'autore dai dataset di training dando esplicita negazione (opt-out).

Sebbene sia eticamente corretta e una forma di protezione necessaria per gli artisti, vengono sollevate profonde incertezze interpretative. Il nodo centrale risiede nella quantificazione dell'apporto creativo umano. Definire il confine oltre il quale un'opera cessa di essere 'generata' e diventa 'creata' resta complesso. Un progetto come quello analizzato a seguire in questa tesi dovrebbe ricadere sotto il diritto d'autore, essendo stato realizzato a seguito di un raffinamento continuo dei prompt testuali, una manipolazione manuale delle immagini tramite software come Photoshop e una generazione di animazioni video secondo una precisa direzione artistica, poi montate e post-prodotte. Per opere meno stratificate provare il proprio contributo potrebbe risultare difficoltoso. Parallelamente, l'efficacia dell'opt-out rimane dubbia. Nonostante siano previsti veti e sanzioni è difficile applicarli ai paesi esteri, rischiando così di limitare lo sviluppo dei modelli generativi nostrani (l'addestramento dei modelli richiede dataset di grandi dimensioni, escludendo contenuti si rischia di limitarli), senza riuscire effettivamente a proteggere i contenuti degli artisti dal web scraping [8, 9, 10].

2.2 C2PA, CAI e Content Credentials

La Coalition for Content Provenance and Authenticity (C2PA) è un progetto non-profit ospitato dalla Joint Development Foundation (facente parte della Linux Foundation), organizzazione che fornisce l'infrastruttura legale e amministrativa affinché diverse aziende concorrenti possano contribuire alla definizione di standard comuni volti ad affrontare la diffusione di informazioni fuorvianti online e certificare l'origine e la provenienza dei contenuti multimediali. La coalizione include un numero crescente di membri, con uno Steering Committee costituito da Adobe, BBC, Google, Meta, Microsoft, OpenAI, Publicis Groupe, Sony e Truepic [11].

La Content Authenticity Initiative (CAI) è un'iniziativa fondata da Adobe nel 2019 che include oltre 5000 membri, tra cui alcuni di spicco come BBC, Canon, Getty Images, Leica, Microsoft, Nikon, Nvidia, Truepic e Panasonic. Si occupa di sviluppare tools open source per verificare la provenienza di ogni media digitale, includendo contenuti generati in IA. Creati dalla CAI nel rispetto degli standard C2PA, i Content Credentials sono un modo di etichettare i contenuti con dei metadati che includono dettagli sul creatore del contenuto (come nome o account social media), data di creazione e come quel contenuto sia stato realizzato, se catturato da una videocamera, generato con l'IA, o editato usando un determinato software. Tali metadati sono visibili sulle piattaforme che lo permettono e nei tools di ispezione, e permettono di fare optout, richiedendo che i modelli di IA generativa non includano quei contenuti nei loro dataset di addestramento.

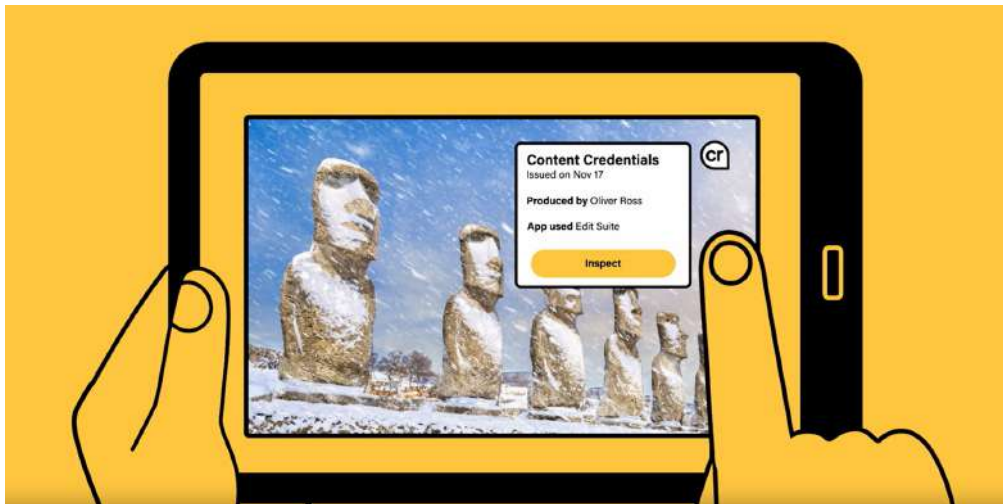


Figura 2.2: Content Crendtials

Fonte: [Link all'articolo](#)

Il content credentials offre un modo agli artisti di mostrare come sia stato realizzato un prodotto e quali lavorazioni abbiano seguito la generazione, così da

mettere in evidenza il contributo umano e dare la possibilità di ottenere i dovuti crediti per il proprio lavoro. Se qualcuno dovesse fare una modifica successiva al contenuto in maniera impropria sarà possibile determinare quale alterazioni siano avvenute [12, 13, 14].

Per rendere il Content Credentials più robusto è stata creata una combinazione di:

- **Metadati:** informazioni che vengono incorporate all'interno del contenuto stesso in un modo che non possa essere alterato senza lasciare traccia.
- **Watermarking:** filigrana nascosta e non rilevabile agli esseri umani che incorpora nel contenuto informazioni decodificabili. I watermark più attuali sono resistenti ad alterazioni come il ritaglio, la rotazione o l'aggiunta di rumore e possono sopravvivere a tentativi di ritrasmissione come gli screenshot che al contrario rimuovono i metadati;
- **Fingerprint:** è un modo per creare un codice unico basato su pixel, fotogrammi o forme d'onda audio e viene memorizzato separatamente dal contenuto come parte del Content Credential. In fase di verifica dell'autenticità, l'identificativo viene ricalcolato e confrontato con i dati presenti nel registro di riferimento. Tale metodologia ha il vantaggio di essere resiliente alla rimozione dolosa, poiché non si basa su informazioni sovrapposte, ma sulle caratteristiche intrinseche del file stesso.

		Durable against metadata stripping	Durable against removal attacks	Strong cryptographic binding	Detectable on devices and apps	No network required
Metadata						
Fingerprint						
Watermark						
Combined						

Figura 2.3: Durable Content Credentials

Fonte: Link all'articolo

Queste tre tecniche, se isolate, presentano delle debolezze. I metadati possono essere rimossi deliberatamente o accidentalmente, un watermark può essere rimosso o contraffatto, ed è comunque limitato alla quantità di dati codificabili senza degradare qualitativamente il contenuto, e il fingerprint non garantisce una perfetta corrispondenza (fuzzy matching).

Per prima cosa viene applicato il watermark, ottimizzato in base al tipo di media (audio, immagini o video), e ne viene sfruttata la limitata banda andando a inserire indicazioni sul dove si possa recuperare il manifest C2PA o un qualsiasi altro tipo di metadata. Può trovarsi nel cloud del Content Credentials come in una blockchain. Viene poi processato il fingerprint, ovvero il descrittore che permette di abbinare il contenuto al suo Content Credential. Un consumatore potrà controllare un contenuto guardando solo il manifest nel caso in cui i dati siano intatti. Qualora non fosse così, verrà decodificato il watermark recuperando l'identificatore che permette di risalire al Content Credential memorizzato nel cloud, e calcolato il fingerprint per accertarsi che ci sia una corrispondenza tra il contenuto e il Content Credential, in modo da evitare problemi di watermarks contraffatti.

Combinando le tre soluzioni si ottiene quello che viene definito come 'Durable Content Credentials', soluzione più robusta e permanente per proteggere i propri lavori [15].

2.3 Adobe

Adobe Firefly è l'applicazione di intelligenza artificiale generativa facente parte dell'Adobe Creative Cloud.

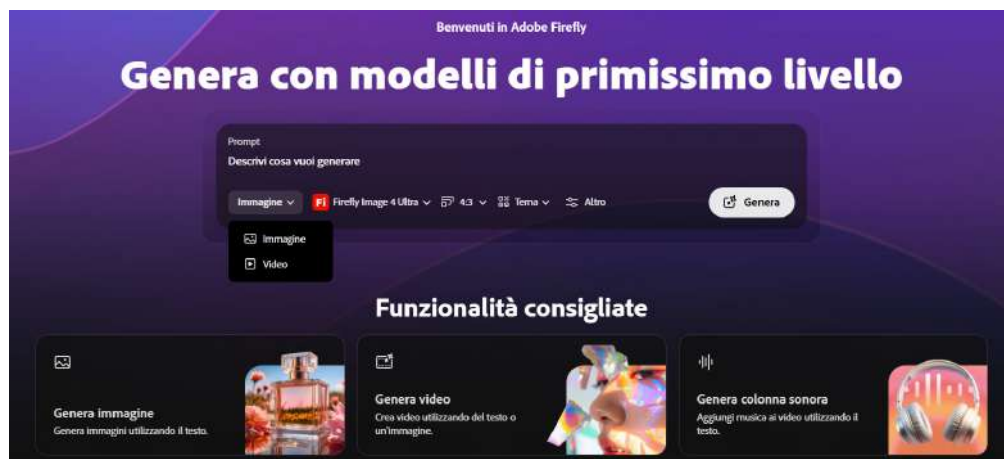


Figura 2.4: Schermata Home - Adobe Firefly

Fonte: [Link al sito](#)

Al suo interno offre diverse funzionalità tra cui, per quanto concerne la generazione di immagini, la possibilità di spaziare tra il text to image (da prompt testuale a immagine), all' image to image (da prompt testuale e immagine di reference a nuova immagine). Nello specifico, il modello proprietario allo stato dell'arte, Firefly Image 4 Ultra introduce un sistema di controllo basato su reference esterne per la composizione e lo stile, permettendo all'utente di modulare l'influenza di tali input tramite slider dedicati. Sebbene l'integrazione di preset stilistici, illuminotecnici e prospettici ampli il ventaglio creativo, la possibilità di agire solo su composizione e stile rende complessa la gestione della coerenza del personaggio e degli ambienti da una immagine all'altra. Il successore, Firefly Image 5 (attualmente in fase di anteprima), tenta di semplificare l'approccio attraverso un'unica immagine di riferimento, garantendo una maggiore coerenza ma fornendo ancora poca flessibilità nel manipolare i singoli elementi costruttivi della scena. Per superare tale criticità, Adobe ha introdotto i Custom Models, in anteprima e tramite accesso regolato da una lista d'attesa. Questa procedura permette di effettuare un fine-tuning del modello proprietario attraverso la somministrazione di un dataset ristretto (composto da 10 a 30 immagini correlate da descrizioni), consentendo l'estrapolazione e la replica sistematica di uno stile specifico o di un personaggio ricorrente [16].

Per quanto riguarda l'editing di immagini, Adobe si distingue tra i leader del settore, non solo per le funzionalità integrate in Firefly, come la rimozione dello sfondo o il ridimensionamento dell'immagine tramite tecniche di 'outpainting' o 'inpainting', ma in particolar modo per il 'Generative Fill' (Riempimento Generativo) implementato anche in Photoshop (ed esente dal consumo di crediti generativi). Attraverso la selezione di una porzione dell'immagine, anche minima, il modello di IA generativa permette di rimuovere elementi esistenti o di produrre diverse alternative per la loro sostituzione. Tale strumento garantisce la massima versatilità nella manipolazione dei contenuti, risultando particolarmente efficace se integrato nei workflow di Digital Matte Painting (DMP).

Firefly offre anche il modello Firefly Video, capace di generare output con un frame rate di 24 fps e una risoluzione fino a 1080p. Il sistema offre diversi gradi di controllo sul processo generativo: oltre al text to video basato su prompt testuali, è possibile vincolare la generazione attraverso l'inserimento di due immagini fungenti da keyframe iniziale e finale (first and last frame), garantendo una coerenza visiva precisa nello sviluppo temporale della clip. Ulteriori evoluzioni dello strumento hanno introdotto l'uso di video di riferimento per guidare il modello secondo due modalità principali:

- Reference di composizione: video di riferimento che guida l'organizzazione spaziale degli elementi, escludendo però la possibilità di definire il primo e l'ultimo frame tramite immagini statiche;
- Reference di fotocamera: video di riferimento che funge da guida per il

movimento della cinepresa, permettendo di mantenere il frame iniziale ma precludendo l'impiego di un'immagine per quello finale.

I modelli proprietari non sono l'unica possibilità, vengono inclusi, sia per le immagini che per i video dei modelli partner esterni come Nano Banana (Gemini 2.5, 3 e 3.1), GPT Image 1.5, Flux (1.1, 2 e Kontext) e Runway Gen-4 Image per le immagini, mentre Veo 3.1 di Google, Ray3.14 di Luma, Sora2 di OpenAI, Pika 2.2 e Runway Gen-4.5 per i video. Particolarmente significativa la possibilità di fare Video to Video con i modelli di Runway Gen-4 Aleph e Ray3.14 Modify usando un video di riferimento e il primo ed ultimo frame per generare un nuovo video con dei modelli che sono attualmente lo stato dell'arte [17, 18].

A differenza dell'ecosistema Adobe, che pone la compliance etica e la tutela del copyright al centro del proprio sviluppo, questi modelli operano spesso su dataset più vasti e meno filtrati. Se da un lato ciò comporta minori garanzie sotto il profilo della proprietà intellettuale, dall'altro offre prestazioni superiori in termini di versatilità e fotorealismo, superando i vincoli relativi a modelli addestrati su un quantitativo di dati limitati. E' bene sottolineare come Firefly sia stato addestrato prevalentemente sui contenuti di Adobe Stock, ovvero centinaia di milioni di foto e illustrazioni su cui Adobe detiene i diritti (in alternativa sono stati usati contenuti i cui diritti fossero scaduti o assenti) garantendo un modello che non viola copyright o altri diritti di proprietà intellettuale. E' stato tuttavia appurato, a seguito di un report fatto da Bloomberg [19] che almeno il 5% del dataset fosse costituito da dati sintetici provenienti da Midjourney, a causa di una 'contaminazione' nelle immagini caricate su Adobe Stock. L'azienda si dimostra comunque sicura del proprio modello, garantendo che ogni immagine usata nel dataset di addestramento di Firefly sia stata accuratamente moderata, assicurandosi che non ci sia alcuna violazione del diritto d'autore. Per rafforzare questo messaggio, Adobe si offre di dare copertura, anche economica, qualora dovessero insorgere problemi legali dovuti ad errori del proprio modello, se non usato in maniera impropria (da questa copertura sono esenti le generazioni ottenute dai modelli partner presenti) [20]. A comprova dell'impegno etico sostenuto da Adobe bisogna ricordare come l'azienda sia fondatrice del CAI, e applichi automaticamente il Content Credentials ai suoi output. Viene, inoltre, conferito il diritto di fare opt-out e messo a disposizione l'Adobe Content Authenticity, ovvero un tool di ispezione per visionare i Content Credentials, usabile in maniera diretta o come estensione del browser. Attualmente in beta, riconosce correttamente un'immagine, ma fatica a distinguere le reference utilizzate per generarla, non attribuendo loro a priori il content credentials, anche quando è effettivamente presente.

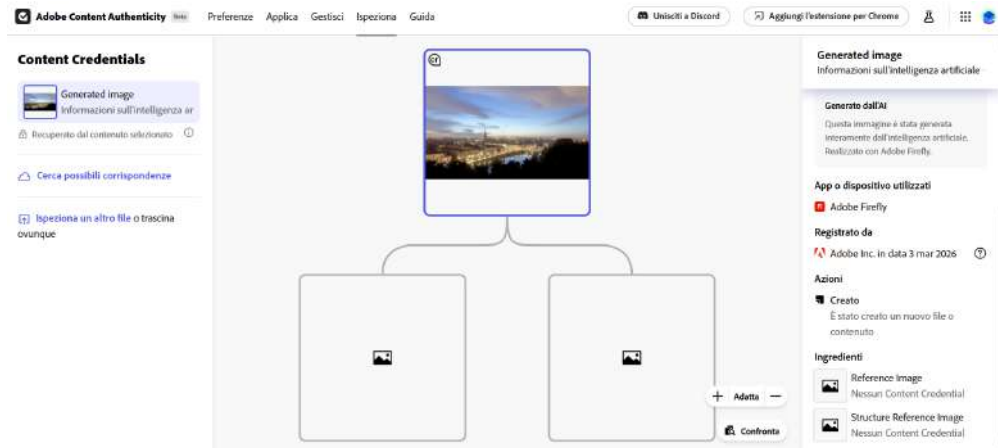


Figura 2.5: Adobe Content Authenticity

Fonte: Link al sito

Adobe mette a disposizione anche la possibilità di generare delle colonne sonore dando in ingresso un video, o di ottenere da prompt testuale degli effetti sonori. Queste funzionalità, insieme alla nuova modalità di modifica video (attualmente in beta), che mette a disposizione una timeline per l'editing video, mostrano una voglia crescente di Adobe di creare una suite omnicomprensiva, anche se al momento carente di diverse funzionalità e con basso grado di controllo.

L'accesso alle tecnologie di Adobe Firefly è regolato da un sistema di crediti generativi, il cui consumo varia in base alla complessità del modello utilizzato e ai parametri computazionali richiesti. Tali crediti vengono erogati mensilmente attraverso i vari piani di abbonamento.

2.4 Google

I modelli offerti da Google si posizionano tra i più avanzati nel panorama attuale per il rapporto tra performance e sicurezza. Il lancio di Nano Banana (ufficialmente denominato Gemini Flash Image) nell'agosto del 2025 ha rivoluzionato le metodologie di generazione ed editing delle immagini. I successivi aggiornamenti, Nano Banana Pro e il più recente Nano Banana 2, che unisce l'efficienza di Nano Banana Pro con i rapidi tempi di elaborazione del modello base, hanno spinto la frontiera della generazione di immagini verso nuovi standard qualitativi. Nano Banana è attualmente integrato all'interno di Gemini (attualmente alla versione 3.1 Pro), accessibile tramite il proprio account Google, trasformando il mero fornire istruzioni ad una macchina, tramite prompt testuali, ad una vera e propria conversazione discorsiva con un Large Language Model (LLM).

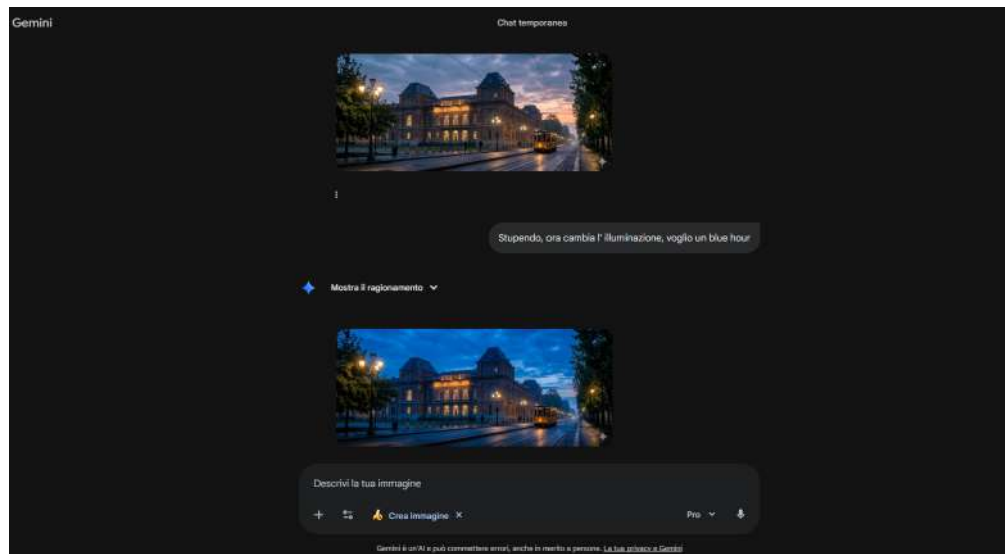


Figura 2.6: Generazione di immagini tramite Nano Bana su Gemini

Fonte: Link al sito

Il modello, inoltre, offre un'elevata comprensione degli input, consentendo di condizionare l'output con più immagini di reference, anche se, per evitare possibili allucinazioni, è buona pratica non usarne più di 2 o 3 (da specificare come il modello stia migliorando in tal senso versione dopo versione). L'assenza di costi marginali in termini di crediti generativi favorisce un processo di raffinamento continuo del contenuto visivo; tuttavia, è necessario considerare i limiti intrinseci dei processi di editing sequenziale. Un numero eccessivo di iterazioni sulla medesima immagine può infatti indurre una degradazione qualitativa dell'output.

A seguito dell'ultimo aggiornamento, anche in Italia, è possibile utilizzare Gemini 3 per accedere a Veo 3.1 Fast, modello per la generazione video di Google, così da generare un numero giornaliero limitato di video, numero che varia a seconda del proprio piano di abbonamento, in una risoluzione standard di 1280x720 px, un frame rate di 24 fps e una durata di 8s.

In alternativa è possibile accedere a tali modelli tramite la piattaforma Flow, sviluppata da Google Labs. Per quanto sia leggermente meno fluida rispetto a una chat aperta con Gemini, consente la scelta opzionale di più parametri.



Figura 2.7: Generazione di immagini e video tramite Flow

Fonte: Google Gemini

Oltre a poter organizzare i propri contenuti in progetti, è possibile scegliere se generarli in 16:9 o 9:16 e generare fino a 4 output in parallelo, funzionalità molto comoda per fare 'versioning'. Le immagini generate possono essere scaricate in risoluzione standard di 1k (1376x768 px) o 'upscale' in 2k e 4k, mentre i video, oltre a fornire la possibilità di condizionare l'animazione con un primo e ultimo frame, vengono normalmente generati con un frame rate di 24 fps in 1280x720 px, con la possibilità di upscale a 1920x1080 px. Viene, inoltre, conferita la possibilità di scegliere il modello da utilizzare, cosa che non impatta particolarmente la produzione di immagini, rimanendo il costo di Nano Banana fisso a 0 crediti generativi, ma significativo per la produzione di video, potendo variare tra Veo 3.1 Fast Low Priority (attualmente solo per gli abbonamenti Ultra, con un costo di 0 crediti a generazione), Veo 3.1 Fast e Veo 3.1 Quality, con un consumo crescente dei crediti [21].

Una significativa qualità offerta può contare su dataset di addestramento molto ampi. Per quanto il dataset effettivo, per ovvi motivi, sia ignoto, Google si impegna in una politica di trasparenza, sia nel fornire dettagli di come il dataset sia costruito, sia nello sforzo intrapreso nella mitigazione dei rischi. Degna di nota è la messa in atto di strategie pre-training e post-training, come l'applicazione di filtri atti ad associare ad ogni contenuto multimediale delle descrizioni testuali e rimuovere i contenuti potenzialmente pericolosi dal dataset o la rimozione di output che violano i termini e le condizioni d'uso. Lo scopo è quello di ovviare a due categorie di rischi:

- Usi controversi e intenzionali del modello: creazione di contenuti che incentivano la disinformazione, facilitano frodi, promuovono odio, o tentano di ottenere immagini sessualizzate ritraenti adulti non consensienti o minori;

- Usi non intenzionali del modello: generazione del modello di output sgradevoli o dannosi davanti prompt benevoli, come ad esempio può esserlo un'immagine che ritrae eccessiva violenza a seguito di una richiesta di ricreare, a scopo istruttivo, una certa battaglia storica.

Sono quindi target della politica di sicurezza di Google, aree quali lo sfruttamento e l'abuso sessuale di minori, l'odio, le molestie, la disinformazione, i deep fakes, contenuti sessualmente espliciti, la violenza e il gore. Affinchè tali elementi vengano resi inaccessibili e la politica sulla sicurezza intrapresa venga rispettata, Google ha adottato misure di valutazione stringenti durante lo sviluppo dei propri modelli, coinvolgendo team di specialisti interni ed esterni pronti a scoprire eventuali vulnerabilità (Red Teaming) e valutandone i rischi prima del rilascio [22, 23, 24, 25].

Nonostante i rischi siano minimizzati, rimane la possibilità che il dataset non sia immacolato. In passato, nonostante Google avesse messo a disposizione Google Extended, un user agent che consentiva i siti web di proteggere i propri contenuti in maniera granulare dai web crawler, modificando il file robots.txt, è stato poi ammesso in sede processuale che quell'opt-out non venisse rispettato, almeno in contesti di indicizzazione e analisi per il motore di ricerca. Tale discrepanza ha creato un precedente e sollevato interrogativi sull'effettiva efficacia degli strumenti di tutela dei creatori di contenuti e sulla coerenza tra le politiche di responsabilità aziendale e le pratiche di addestramento dei modelli [26, 27]. Ulteriori controversie sono nate a seguito di un report della CNBC secondo cui Google abbia usato un sottoinsieme del vasto catalogo di YouTube (oltre 20 miliardi di video) per addestrare Veo3, senza chiedere un consenso esplicito agli utenti né fornendo loro un adeguato compenso [28, 29].

Pur considerando le criticità evidenziate, l'impegno da parte di Google verso la sicurezza e la trasparenza rimane concreto. L'azienda aderisce allo standard C2PA e dai più recenti aggiornamenti integra i Content Credentials al SynthID, ovvero il sistema di watermarking, sviluppato da Google DeepMind. Quest'ultimo viene direttamente incluso all'interno dei file multimediali e progettato in modo da essere resistente alle modifiche [30, 31, 32]. Sul piano della governance dei dati, Google dichiara di escludere i contenuti generati dagli utenti dai propri dataset di addestramento [33] e offre una clausola di tutela legale (indemnification) ai propri clienti corporate, tutelandoli in caso di controversie giudiziarie legate alla proprietà intellettuale degli output [34, 35].

2.5 RunwayML

RunwayML è stata fondata nel 2018 da tre ricercatori della Tisch School of the Arts presso la New York University. La startup ha registrato una rapida espansione

nel mercato dell'IA generativa, consolidando la propria posizione anche grazie a ingenti finanziamenti da parte di leaders nel settore come Google e NVIDIA [36, 37]. Prima del rilascio del già citato Nana Bana, Runway veniva considerato lo stato dell'arte della generazione di immagini, grazie al suo piano di abbonamento di fascia unlimited e la possibilità di definire delle reference da richiamare testualmente nei prompt. Per quanto riguarda la generazione video, il modello più recente, Runway Gen-4.5, lanciato nel dicembre del 2025, ha definito un sostanziale miglioramento rispetto ai precedenti sul piano della performance e della qualità visiva [38]. Uno dei grandi vantaggi è quello di avere un'ampia selezione di rapporti d'aspetto (aspect ratio) tra cui scegliere, fornendo anche un formato tipicamente cinematografico come il 21:9, e la possibilità di scegliere in maniera mirata la durata delle animazioni da un minimo di 2s fino a un massimo di 10s, conferendo così una maggiore libertà di controllo. Al contrario, un punto a sfavore è quello di avere a disposizione soltanto il first frame per condizionare l'animazione. Sarà necessario accedere alla funzione 'Animate Frames' nella sezione 'App' per poter usufruire anche del last frame, con, però, una scelta dei parametri a disposizione più ridotta. La sezione 'App' è un'altra caratteristica peculiare di Runway, aggiungendo funzionalità che possono essere integrate in diversi workflows. Per quanto riguarda le immagini, il Character Render che da un look 3d a un'immagine 2D flat, il Cinematic Brainstorm, il quale partendo da un'immagine genera 9 inquadrature cinematicamente correlate, e lo Story Panels, che partendo da un'immagine sequenzializza la generazione dello shot successivo, sono alcune tra le funzionalità più utili. Tuttavia, è sulla gestione dei video che Runway si differenzia, includendo:

- Act-Two rappresenta una soluzione avanzata per l'animazione di avatar digitali, permettendo di infondere voce, espressioni facciali e dinamismo motorio ai soggetti virtuali. Operando in modo analogo a un sistema di motion capture (MoCap) basato su video, lo strumento è in grado di estrapolare da una clip sorgente (della durata massima di 30 secondi) i tratti biomeccanici e mimici dell'attore, traslandoli sul personaggio digitale di destinazione;
- Backdrop: è una funzionalità avanzata di rotoscoping basata su intelligenza artificiale che permette la segmentazione e l'isolamento del soggetto in movimento. Lo strumento consente di sostituire integralmente il background ambientale preservando la fedeltà visiva del soggetto originale;
- Character Swap: funzione speculare del backdrop che permette di sostituire qualsiasi personaggio all'interno di una sequenza, mantenendone inalterati i movimenti, l'illuminazione e la perfetta integrazione in scena;
- Lighting: sfrutta il calcolo delle normali per relightare la scena 2D in modo dinamico. Permette di modificare direzione e intensità delle luci mantenendo l'integrità originale degli elementi e della composizione;

- Motion Sketch: Trasforma immagini statiche in animazioni controllate attraverso semplici schizzi direzionali. Definendo traiettorie e flussi di movimento direttamente sull'immagine è possibile guidare l'IA nella generazione delle animazioni video.
- Time of Day: modifica dinamicamente l'illuminazione globale di una sequenza video per simulare diverse fasce orarie. Uno strumento indispensabile per gestire transizioni day-to-night o alterare l'atmosfera luminosa mantenendo la coerenza della scena.
- Weather: permette di trasformare radicalmente le condizioni atmosferiche di una clip. Grazie all'intelligenza artificiale generativa, integra pioggia, neve, nebbia o tempeste adattando dinamicamente i riflessi, la visibilità e le interazioni della luce sulle superfici.

Queste sono alcune delle più importanti funzionalità presenti. Da specificare che, a differenza dalla pura generazione, l'editing dei video si basa sul modello Runway Gen-4 Aleph [39].

Sebbene i modelli proprietari di Runway siano all'avanguardia, il panorama dell'IA corre velocemente. Per garantire sempre il massimo della qualità, Runway ha adottato una strategia aperta, integrando i migliori modelli partner del mercato. Gli utenti possono così spaziare dalla precisione di Nano Banana, Flux, GPT Image e Seedream per le immagini, fino alla potenza cinematografica di Kling, Veo, Sora e Wan per le generazioni video.

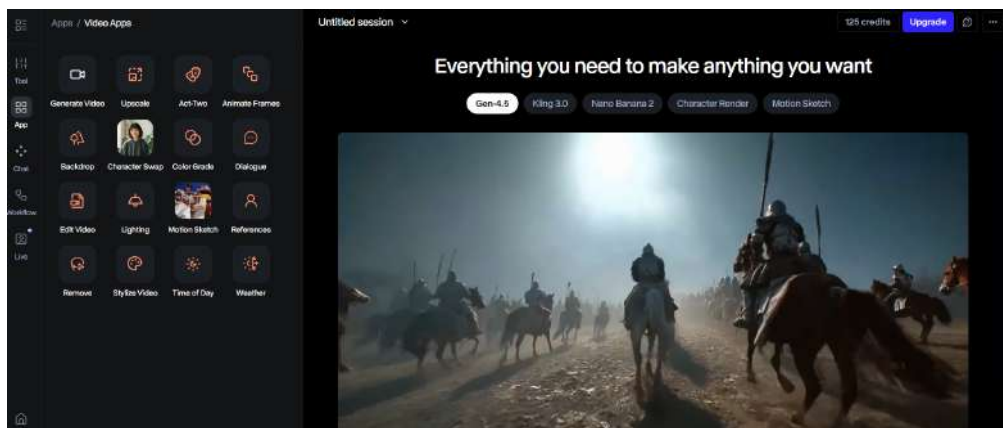


Figura 2.8: Runway - Sezione 'Video Apps'

Fonte: [Link al sito](#)

La forte volontà di ritagliarsi uno spazio nel mercato viene messa in evidenza anche dalla recente collaborazione con Lionsgate, società focalizzata sulla creazione e distribuzione di contenuti. Per dare contesto, tale azienda si è occupata della

distribuzione di titoli come 'The Hunger Games', 'John Wick', e la saga di 'Twilight' e possiede un catalogo di oltre 20.000 titoli. Lo scopo della collaborazione, instaurata nel 2024, era quello di sviluppare un nuovo modello di IA generativa personalizzato su tale catalogo, così da creare nuove opportunità per i filmmakers, i registi o altri professioni del settore. Dopo poco più di un anno di collaborazione, il progetto si è arenato, dimostrando come il numero di titoli posseduti da Lionsgate fosse insufficiente per addestrare in maniera appropriata un modello [40, 41, 42].

Runway garantisce ai propri clienti, in particolare in ambito enterprise, elevati standard di sicurezza. L'azienda ha infatti ottenuto l'attestazione SOC 2 Type II, che certifica l'efficacia e la continuità temporale dei controlli per la protezione dei dati. Questo standard viene sviluppato dall'American Institute of Certified Public Accountants (AICPA), la principale organizzazione professionale indipendente statunitense, che definisce standard e criteri per valutare come le aziende gestiscano e proteggano la riservatezza dei dati affidati loro. Inoltre, Runway applica policy interne molto stringenti sul proprio personale per assicurare che le informazioni dei clienti non vengano né condivise né vendute in maniera impropria. Dal punto di vista dell'integrità, Runway aderisce agli standard C2PA e aggiunge un watermark invisibile su ogni contenuto generato [43, 44, 45, 46].

Nonostante il rigoroso impegno sul fronte della sicurezza informatica, Runway, analogamente ad altre aziende leader nello sviluppo di intelligenza artificiale generativa, non è stata esente da controversie di natura etica e legale legate al diritto d'autore. A seguito di un leak interno, un documento ottenuto dalla testata giornalistica 404 Media ha rivelato che l'azienda avrebbe addestrato il proprio modello Gen-3 attraverso il web scraping non autorizzato di migliaia di video tratti da YouTube. Tra le fonti utilizzate figurerebbero contenuti protetti da copyright appartenenti a major dell'intrattenimento come Netflix, Disney, Nintendo e Rockstar Games, oltre a materiale proveniente da piattaforme di streaming pirata come KissCartoon [47, 48, 49]. Runway è stata inoltre coinvolta, seppur in misura marginale, nell'azione legale collettiva intrapresa da un gruppo di artisti contro Stability AI, Midjourney e DeviantArt. Questo coinvolgimento deriva dai legami storici dell'azienda con Stability AI, avendo Runway partecipato come co-sviluppatore alla creazione della versione originaria del modello text-to-image Stable Diffusion [50].

2.6 Luma AI

Fondata nel 2021 da un collettivo di esperti provenienti dai vertici dell'industria tecnologica, Luma AI ha inizialmente focalizzato la propria ricerca sulla generazione di asset tridimensionali a partire da input testuali [51]. A seguito della collaborazione con HUMAIN, colosso dell'intelligenza artificiale sostenuto dal fondo per gli investimenti pubblici dell'Arabia Saudita [52] e con l'introduzione dell'architettura

Ray, nelle sue versioni 2, 3 e la più recente 3.14, l'azienda si è consolidata come punto di riferimento nella sintesi e manipolazione video. Allo stato attuale, le soluzioni di Video-to-Video proposte da Luma AI rappresentano l'avanguardia del settore, distinguendosi per una coerenza spaziale e temporale superiore e per la capacità di supportare workflow professionali grazie ad output in alta gamma dinamica (HDR).[53]

Oltre all'implementazione di parametri predefiniti per il controllo dei movimenti di camera (quali pan, tilt, push-in e molti) [54] e per il controllo della composizione (come POV, aerial shot, over-the-shoulder e molti altri) [55], Luma AI introduce una modalità di interazione multimodale avanzata. Il sistema permette infatti di guidare l'animazione della scena tramite annotazioni grafiche dirette sull'immagine di partenza. Questo approccio consente di sostituire la descrizione testuale con un'interfaccia visiva intuitiva, conferendo all'utente la facoltà di tracciare traiettorie e movimenti a mano libera, e garantendo così una precisione spaziale difficilmente raggiungibile tramite il solo linguaggio naturale [53].

Tradizionalmente, la modifica sostanziale di elementi all'interno di un'inquadratura richiede pesanti lavorazioni in compositing, per quanto possibile, se non addirittura la ripetizione delle riprese fisiche o la rigenerazione integrale dei rendering 3D. Analogamente a soluzioni come Runway Aleph, Luma AI propone un cambio di paradigma: grazie a una superiore aderenza semantica al prompt e al mantenimento delle caratteristiche strutturali del filmato originale, il modello permette di disaccoppiare il movimento dalla forma. È infatti possibile estrarre dinamiche corporee, micro-espressioni facciali e sincronizzazione labiale (lip-sync) per trasporle su nuovi soggetti o ambientazioni. Questa tecnologia riduce la necessità di infrastrutture costose come i green screen, i LED wall o le tute di motion capture, consentendo di trasformare riprese effettuate in contesti non controllati in inquadrature dal valore estetico cinematografico [56, 57].

Una delle principali criticità del Video-to-Video risiede nell'incoerenza temporale, che genera artefatti visivi noti come wobbling (oscillazioni distorsive) sulle geometrie in movimento. Al fine di mitigare tali fenomeni e favorire una corretta segmentazione degli elementi da parte del modello, risulta fondamentale l'ottimizzazione del girato sorgente. Le linee guida prevedono l'utilizzo di camere stabilizzate, evitando camere a mano tremolanti, la scelta di sfondi neutri o poco complessi e la limitazione di scene con soggetti multipli o movimenti eccessivamente rapidi. Risulta altrettanto determinante l'impiego di file ad alta risoluzione, privi di sfocature (motion blur) che potrebbero compromettere la corretta interpretazione della scena [58, 59].

**Figura 2.9:** LumaAI Video-To-Video

Fonte: [Link al sito](#)

Dal punto di vista etico non sono presenti particolari sbavature se non una breve polemica nata per un trailer, rilasciato in passato da Luma, intitolato *Monster Camp* che ritraeva sullo sfondo di una scena un personaggio molto simile a quello di 'Mike Wazowski' del film Pixar "Monsters & Co". A seguito di questo evento, non sono state aperte cause da parte di Disney/Pixar, ma sono nati dubbi da parte dell'opinione pubblica della bontà del dataset, attualmente segreto, utilizzato da Luma per addestrare i propri modelli [60, 61].

Luma AI chiede ai propri utenti il rispetto delle normative sul diritto d'autore, conformandosi esplicitamente alle disposizioni del Digital Millennium Copyright Act (DMCA). A tal fine, la società ha istituito la figura di un agente legale designato preposto alla gestione delle notifiche di violazione. In presenza di una segnalazione formale, Luma si riserva la facoltà discrezionale di rimuovere il contenuto contestato o di disabilitarne l'accesso sulla piattaforma, fatta salva l'eventuale risoluzione della controversia dinanzi alle autorità giudiziarie competenti [62].

2.7 OpenAI

OpenAI, per la generazione di immagini, si affida al suo modello DALL-E 3, costruito nativamente su ChatGPT, cosa che consente di sfruttare il Large Language Model (LLM) come partner per il brainstorming e l'affinamento dei prompt [63]. In alternativa è anche possibile utilizzare Sora, modello sviluppato per la generazione video, ma che consente la sintesi di immagini trattandole come video costituiti da un solo frame. Per quanto riguarda la pura generazione video presenta alcune funzionalità peculiari:

- Re-cut: utile in presenza di inquadrature non conformi alle aspettative all'interno di una sequenza. Invece di procedere alla rigenerazione integrale della sequenza, il sistema permette un intervento di trimming selettivo, ovvero l'utente può eliminare la porzione indesiderata e richiedere al modello la generazione mirata del solo segmento mancante, garantendo così coerenza visiva e risparmio di risorse.
- Remix: introduce la possibilità di effettuare un editing semantico del contenuto video preesistente. Attraverso l'integrazione di nuovi prompt testuali, il modello è in grado di ricalibrare e modificare elementi specifici all'interno della sequenza, quali stili, oggetti o condizioni ambientali, mantenendo però la coerenza strutturale del video originale;
- Blend: consente la fusione di due sequenze video distinte. Attraverso questo processo di interpolazione, il modello analizza le caratteristiche strutturali di entrambi i filmati sorgente per generare un terzo output inedito, capace di sintetizzare in modo armonico gli stili, i soggetti e le dinamiche di movimento di entrambi gli input;
- Loop: il modello è in grado di ricalibrare qualsiasi segmento del video per trasformarlo in un'iterazione fluida e perpetua. Tale strumento risulta particolarmente efficace per la creazione di contenuti destinati a piattaforme che richiedono una fruizione continua e priva di interruzioni visive;
- Storyboard: permette di stabilire una continuità sequenziale tra generazioni video distinte. Attraverso la disposizione dei prompt testuali di ogni generazione all'interno di una timeline non lineare, l'utente può orchestrare la narrazione visiva. Una maggiore spaziatura temporale tra i prompt consente al modello di elaborare transizioni più fluide e naturali, minimizzando le discrepanze visive ed evitando stacchi netti [64].

La nuova versione del modello Sora2 ha ottenuto notevoli miglioramenti nell'obbedire alle leggi della fisica e nel seguire istruzioni complesse. Eccelle in stili realistici, cinematografici e anime, includendo effetti sonori, dialoghi e ambienti [65].

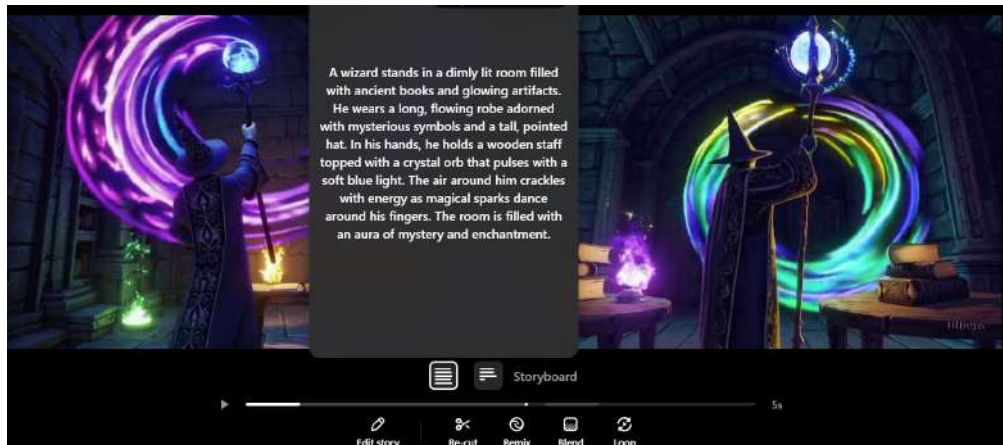


Figura 2.10: Storyboard - Sora

Fonte: [Link al sito](#)

Sora ha intrapreso un percorso di distribuzione multi-piattaforma, affiancando alla versione web l'applicazione mobile, al momento lanciata sugli store dei dispositivi iOS negli Stati Uniti e in Canada, con l'ambizione di rilasciare presto anche su Android ed estendere la copertura geografica ad Europa, Regno Unito e Australia. L'obiettivo strategico sembra convergere verso la creazione di un vero e proprio social network in cui catturare la propria immagine, generare ed animare un proprio avatar digitale (Digital Human), da inserire in scenari sintetici complessi e dare sfogo alla propria fantasia con gli strumenti generativi a disposizione [66, 65]. Essendo un'applicazione destinata anche ad adolescenti, il framework di sviluppo di Sora ha posto particolare enfasi sulla sicurezza, includendo sin dalle fasi di addestramento e di fine-tuning meccanismi di filtraggio e tecniche di mitigazione dei rischi, garantendo un ambiente digitale protetto e conforme agli standard etici di tutela dei minori. In ogni caso i profili di adolescenti non sono visibili agli adulti ed è integrato un efficiente programma di controllo parentale. Ogni immagine o video contenente i dati dell'utente, quali aspetto o voce, richiede il consenso per poter essere utilizzata da altri, permesso che può essere revocato dall'utente in qualsiasi momento. Vengono inoltre bloccati contenuti che ritraggono personaggi famosi o opere esistenti, o che contengono contenuti sessuali espliciti, violenza, propaganda politica, incitamento dell'odio, promozione di droghe o altre sostanze nocive. OpenAI, inoltre, aggiunge una filigrana invisibile ai propri contenuti e dei metadati conformi agli standard C2PA [67, 68].

Nonostante le premesse innovative, il debutto di Sora è stato accompagnato da accese controversie riguardanti la gestione della proprietà intellettuale. Come riportato da testate quali CNBC, il modello ha dimostrato la capacità di generare contenuti visivi riconducibili a marchi e personaggi protetti da copyright; tra gli esempi citati figurano asset iconici come il logo di 'Ronald McDonald' e personaggi

appartenenti a franchise d'animazione quali 'SpongeBob SquarePants', 'Rick and Morty' e 'South Park'. Tali evidenze sollevano interrogativi cruciali sulla natura dei dataset di addestramento e sulla potenziale violazione dei diritti d'autore negli output sintetici. Anche per Sora, al suo lancio, non sono mancate le polemiche. Per ovviare alle criticità legate alla proprietà intellettuale, OpenAI ha implementato sistemi di segnalazione e rimozione dei contenuti protetti, favorendo al contempo un approccio basato su accordi di licenza diretti. Un esempio emblematico è la recente partnership triennale siglata con The Walt Disney Company, che legittima l'integrazione di oltre 200 personaggi appartenenti ai franchise Disney, Marvel, Pixar e Star Wars all'interno del modello. In parallelo, Disney utilizzerà le API di OpenAI per sviluppare nuovi prodotti proprietari. OpenAI, inoltre, si impegna, in onore dell'accordo, nel continuare e rafforzare le pratiche di responsabilità e sicurezza precedentemente elencate [69, 70].

2.8 KlingAI

Kling è il modello di generazione video all'avanguardia sviluppato da Kuaishou Technology, colosso tecnologico cinese e principale competitor di TikTok. L'azienda, già nota a livello globale per l'ecosistema social Kwai, ha consolidato la sua posizione nel mercato dell'IA nel 2024, rilasciando Kling come risposta diretta ai modelli di OpenAI e Google [71]. Sebbene offra solide capacità di generazione d'immagini, Kling si è affermato come uno dei leader globali nella generazione video. Spesso messo a confronto con Veo 3 di Google per qualità e performance, il modello di Kuaishou si distingue per un ciclo di sviluppo estremamente rapido e impattante. Con i recenti rilasci di Kling O1 e Kling Video 3.0, l'azienda ha compiuto un salto generazionale, non solo perfezionando l'aderenza ai prompt e la coerenza fisica dei movimenti, ma introducendo funzionalità di editing video che stanno ridefinendo gli standard del settore, trasformando la generazione video da un semplice output tecnico a un flusso di lavoro artistico e controllato.

Superando i limiti del masking tradizionale, il modello elimina la necessità di ritagliare o selezionare manualmente le porzioni di immagine da modificare. Grazie alla sua natura nativamente multimodale, Kling interpreta direttamente il contesto visivo in funzione delle istruzioni ricevute. Inoltre, la sua architettura permette la gestione simultanea di task multipli: è possibile, ad esempio, inserire un nuovo personaggio e trasformare lo sfondo in un'unica operazione, garantendo una coerenza stilistica e fisica senza precedenti. A seguire alcune delle più importanti funzionalità introdotte:

- **Image-Element Reference:** Il sistema permette l'integrazione di un massimo di sette riferimenti esterni, definiti "Elements", che fungono da vincoli visivi per la generazione. Questi asset possono essere composti da singole immagini o da

set di quattro scatti che ritraggono lo stesso soggetto da diverse angolazioni, garantendo una ricostruzione tridimensionale coerente. La natura multimodale di Kling estende il concetto di "Element" anche al comparto audio e video: è infatti possibile associare un'impronta vocale specifica a un personaggio o utilizzare clip video come riferimento dinamico. Attraverso il prompt testuale, l'utente coordina le interazioni tra questi diversi elementi, orchestrando scene complesse in cui identità visiva e vocale rimangono costanti;

- **Transformation:** consente la manipolazione di un video sorgente, permettendo l'inserimento, la modifica o la rimozione di elementi specifici (soggetti e scenari) attraverso l'integrazione di prompt testuali e immagini di riferimento (reference images). Il sistema non si limita alla variazione degli attributi estetici, come il color grading, l'illuminazione o le condizioni atmosferiche, ma interviene sulla composizione stessa della scena. Tra le applicazioni più avanzate figurano il background replacement (estensione del concetto di green screen digitale) e il trasferimento di stile (style transfer), mantenendo la coerenza fisica e temporale del video originale;
- **Video Reference:** permette l'impiego di una clip sorgente come vincolo strutturale per la generazione di nuovi contenuti. Il modello è in grado di eseguire un'estrapolazione dei vettori di movimento, replicando fedelmente sia le dinamiche dei soggetti in scena sia i movimenti di camera (come pan, tilt o dolly shots). Quasi come una tuta di Motion Capture Markless combinata a un Camera Tracking, è possibile decodificare i movimenti umani e le dinamiche della cinepresa da un video per trasferirli su asset statici o nuove generazioni.
- **Frames:** similmente ad altri modelli, viene introdotto un sistema di controllo spaziale e temporale basato sull'inserimento di Start e End Frames. L'utente definisce lo stato iniziale e finale della sequenza caricando due immagini distinte. Il modello agisce come un motore di interpolazione intelligente, generando una transizione dinamica coerente tra le due. Attraverso il prompt testuale, è possibile descrivere l'azione specifica o la trasformazione che deve intercorrere, permettendo al sistema di calcolare traiettorie fisiche complesse e trasformazioni morfologiche che mantengono la fedeltà visiva di entrambi i fotogrammi chiave. Un grande vantaggio risiede nella gestione granulare della linea temporale, offrendo all'utente la possibilità di definire con precisione la durata di ogni clip in un intervallo compreso tra i 3 e i 15 secondi, conferendo così un controllo dell'azione che non si limita alla dimensione spaziale ma viene esteso anche a quella temporale;
- **Text-To-Video:** la generazione video viene basata esclusivamente sul prompt testuale. Più la descrizione della scena sarà dettagliata, maggiore sarà la capacità del modello di mappare variabili complesse, restituendo un output

caratterizzato da un'elevata fedeltà visiva e una narrazione scenica stratificata. Per ottimizzare l'elaborazione, si adotta una struttura sintattica specifica che comprende: definizione del soggetto (attributi fisici e identitari), dinamica dell'azione (movimenti specifici), contestualizzazione ambientale (scenografia e dettagli di sfondo), regia Virtuale (movimenti di camera come dolly o tracking shot), oltre alla definizione dell'illuminazione e del tono atmosferico;

- **Multi-Shot:** abilita la produzione di una sequenza cinematografica articolata all'interno di un unico processo generativo. Sfruttando l'estensione temporale modulabile (da 3 a 15 secondi), l'utente può frazionare la timeline e assegnare a ogni segmento un prompt testuale differente. Questo approccio permette una regia multi-focale, in cui inquadrature e prospettive diverse vengono concatenate organicamente, garantendo una fluidità di montaggio e una coerenza visiva che riducono drasticamente la necessità di una post-produzione esterna [72, 73].

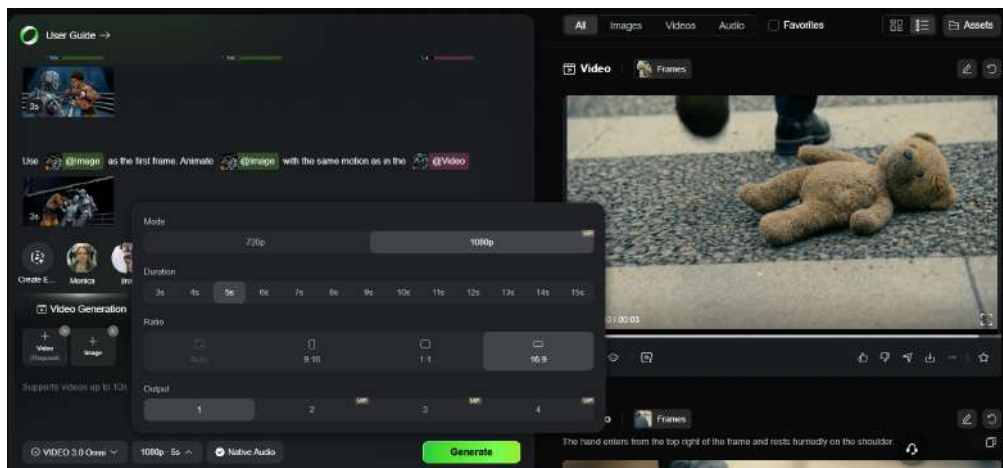


Figura 2.11: Kling Omni

Fonte: Link al sito

Nonostante Kling adotti una politica di generazione apparentemente più permissiva rispetto ai competitor occidentali (come Adobe Firefly, Google Veo o OpenAI Sora), il modello mantiene rigorosi protocolli di moderazione. Nello specifico, la piattaforma inibisce la produzione di contenuti NSFW (Not Safe For Work) e applica filtri di censura stringenti su temi di propaganda politica, in linea con le normative del Cyberspace Administration of China (CAC) [74, 75].

Sotto il profilo legale, sebbene Kuaishou riconosca formalmente agli utenti la proprietà intellettuale e i diritti commerciali su input e output, la gestione della privacy presenta zone d'ombra significative. Per gli utenti della fascia Free, i contenuti

generati perdono il carattere di riservatezza, venendo pubblicati nei feed pubblici della piattaforma e resi accessibili alla community. Inoltre, indipendentemente dal piano di abbonamento sottoscritto, le clausole contrattuali (ToS) garantiscono all'azienda il diritto di utilizzare input e output per l'ottimizzazione dei servizi. Tale disposizione crea perplessità sul re-training dei modelli, lasciando aperta la possibilità che i dati degli utenti contribuiscano attivamente all'evoluzione dell'architettura neurale di Kling, sollevando dubbi sulla reale esclusività del materiale prodotto [76].

2.9 Getty Images vs Stability AI

Nell'agosto del 2022, Stability AI ha scosso il mercato tecnologico rilasciando Stable Diffusion, un modello open source per la generazione di immagini che ha ridefinito gli standard della creatività digitale [77]. Grazie a prestazioni straordinarie e alla sua natura accessibile, la società ha saputo attirare l'attenzione di icone del cinema come James Cameron, recentemente entrato nel consiglio di amministrazione per esplorare le nuove frontiere del visual storytelling [78].

Oggi, l'influenza di Stability AI si estende ben oltre le immagini statiche, consolidandosi attraverso alleanze strategiche con i giganti dell'intrattenimento:

- Gaming: la partnership con Electronic Arts (EA) punta a integrare l'IA generativa direttamente nei flussi di sviluppo, ottimizzando la creazione di mondi virtuali e asset di gioco [79];
- Musica: La collaborazione con major del calibro di Universal Music Group e Warner Music Group è focalizzata sulla ricerca di strumenti innovativi per la composizione e la produzione sonora [80, 81].

Getty Images rappresenta una delle più vaste e autorevoli library visive al mondo, custodendo un archivio sterminato di immagini e contenuti video protetti da copyright e di altissimo valore editoriale. Giugno 2023 ha segnato l'inizio del confronto legale presso l'High Court di Londra tra Getty Images e Stability AI. Al centro della disputa, l'accusa mossa al colosso dell'IA di aver utilizzato oltre 12 milioni di immagini protette da copyright, senza alcuna autorizzazione, per l'addestramento dei propri modelli. Il processo mette in discussione tre pilastri fondamentali: la legittimità dell'uso di dati protetti in fase di training, la titolarità del diritto d'autore sugli output generati e, in ultima analisi, la stessa validità legale di Stable Diffusion. Dal canto suo, Stability AI ha risposto con una difesa decisa, sostenendo che tali apparizioni siano solo casi rari e fortuiti. La società ha inoltre cercato di spostare la responsabilità legale, affermando che gli output siano generazioni create dagli utenti e che spetterebbe a Getty l'onere di provare

l'effettiva violazione del copyright per ogni singola immagine. L'azienda sostiene, inoltre, che il modello non sia stato addestrato nel Regno Unito, fattore che lo esenterebbe dall'applicazione della legge britannica sulla proprietà intellettuale. Questa posizione ha spinto Getty Images a internazionalizzare il conflitto, aprendo un secondo fronte legale negli Stati Uniti [82, 83].

Recentemente, il procedimento ha subito una svolta significativa a favore di Stability AI. La corte ha infatti stabilito che il modello non viola direttamente il copyright per il solo fatto di essere stato addestrato su contenuti protetti, poiché Stable Diffusion non "conserva" né "riproduce" i file originali, ma si limita ad apprendere i concetti statistici che li definiscono. Questo processo è stato equiparato a una forma di apprendimento digitale piuttosto che a una duplicazione illegale. Tuttavia, a Getty Images è stata infatti concessa la possibilità di ricorrere in appello per sostenere l'idea che Stable Diffusion costituisca, di per sé, una "copia contraffatta" (infringing copy). La questione solleva un interrogativo sul diritto dei modelli di ricostruire con estrema precisione opere protette, e se essi stessi possano essere considerati delle copie illegale, pur non contenendo fisicamente un singolo pixel dell'immagini originali. L'esito di questo appello è destinato a ridefinire i confini della proprietà intellettuale, poiché una vittoria di Getty potrebbe sancire come i pesi matematici di un'intelligenza artificiale possono essere trattati legalmente alla stregua di materiale piratato [84, 85].

2.10 Artisti contro Midjourney e StabilityAI

Midjourney si conferma come uno dei modelli di intelligenza artificiale generativa per immagini più celebri e influenti a livello globale. Per lungo tempo ha dettato gli standard del settore, affermandosi come l'indiscusso stato dell'arte grazie a funzionalità pionieristiche, in particolare la possibilità di fornire in input immagini di riferimento per mantenere la coerenza dei personaggi o per replicare composizioni e stili specifici. Un suo tratto distintivo è proprio la libertà lasciata all'utente di manipolare l'influenza di queste reference, inserendo nel prompt testuale specifiche "parole in codice" e parametri numerici per bilanciarne esattamente il peso. Sebbene attualmente sia stato superato in termini di pura versatilità di editing da modelli più recenti e flessibili come Nano Banana o la famiglia Qwen, Midjourney mantiene una posizione di assoluto rilievo nel panorama creativo. La straordinaria resa estetica e la spiccata "direzione artistica" intrinseca dei suoi output lo rendono infatti una risorsa ancora validissima. Oggi si conferma uno strumento utile soprattutto per le fasi preliminari di brainstorming e per lo sviluppo rapido di concept art, fornendo ad artisti e designer una base visiva concreta e di alta qualità per l'avvio dei propri progetti.

Nel 2023, il panorama dell'intelligenza artificiale generativa ha subito un forte scossone legale quando Stability AI, DeviantArt e Midjourney sono state citate in giudizio in una massiccia causa per violazione del copyright. Questa storica class action è stata promossa da un gruppo di creativi guidato dall'illustratrice Sarah Andersen, la quale aveva scoperto che, semplicemente inserendo il proprio nome all'interno di Stable Diffusion, il modello era in grado di replicare in modo impressionante il suo inconfondibile stile grafico. L'impianto accusatorio ha dimostrato come questi sistemi, se stimolati con prompt testuali mirati, siano effettivamente in grado di restituire in output riproduzioni evidenti delle opere protette utilizzate durante il loro addestramento. La tempesta legale attorno a Midjourney si è poi ulteriormente inasprita quando colossi dell'intrattenimento del calibro di Disney, Universal e Warner Bros Discovery hanno depositato denunce formali contro l'azienda, accusandola di plagio e di sfruttamento illecito delle proprie proprietà intellettuali [86]. In parallelo, la class action degli artisti è avanzata fino alla delicata fase di discovery, un momento processuale cruciale in cui i querelanti hanno il diritto di accedere alle comunicazioni interne e ai registri di addestramento delle aziende tecnologiche. Questa complessa e imponente raccolta di prove getterà le basi per il processo vero e proprio, il cui inizio ufficiale è fissato per il settembre del 2026 [87, 88].

2.11 Altri modelli

2.11.1 ByteDance

ByteDance, il colosso tecnologico dietro il successo globale di TikTok, si è imposto come leader nel settore dell'IA generativa grazie a una suite di strumenti all'avanguardia. Tra questi spicca SeeDream, dedicato alla creazione di immagini, e soprattutto SeeDance, il rivoluzionario modello di generazione video che sta ridefinendo i confini del settore. In particolare, il recentissimo rilascio di SeeDance 2.0 ha scosso l'industria per le sue prestazioni straordinarie. Il modello offre un livello di fotorealismo senza precedenti, unito a una capacità espressiva nel "recitato" dei personaggi che ha spinto molti esperti a definire i risultati di qualità "hollywoodiana". La sua forza risiede nella gestione di scene d'azione ad alto dinamismo, dove SeeDance riesce a mantenere una coerenza visiva e una fluidità che finora sembravano irraggiungibili per un'intelligenza artificiale.

Tuttavia, questo primato tecnologico ha un rovescio della medaglia: il modello sembra operare con filtri di copyright estremamente permissivi. Questa mancanza di restrizioni ha portato ByteDance in rotta di collisione con i giganti del cinema. Major del calibro di Disney e Paramount Skydance hanno già inviato formali diffide all'azienda, accusandola di aver addestrato il modello su materiale protetto e di

permettere la generazione di contenuti che violano apertamente la loro proprietà intellettuale [89].

2.11.2 Flux

Flux è il modello di generazione d'immagini di punta sviluppato da Black Forest Labs, una startup tedesca fondata da alcuni ricercatori che avevano originariamente partecipato alla creazione di Stable Diffusion. Fin dal suo debutto, Flux è stato celebrato per una resa fotografica senza precedenti e per la sua straordinaria capacità di gestire dettagli complessi, come il rendering delle mani e la precisione del testo all'interno delle immagini. Per gran parte del 2024 e del 2025, è stato considerato il vero erede di Stable Diffusion e lo stato dell'arte del settore, venendo superato solo di recente dalle prestazioni di Nano Banana (Gemini 3 Flash Image) di Google. Nonostante l'ascesa dei modelli Big Tech, Flux rimane uno strumento di riferimento grazie alla sua architettura transformer-based che garantisce una coerenza visiva eccezionale.

Un punto di dibattito che lo riguarda è relativo alla sua complessa struttura di licenza. La versione più accessibile, denominata Schnell, è rilasciata sotto licenza Apache 2.0, il che ne favorisce l'adozione personale e lo sviluppo open source senza particolari vincoli. Al contrario, la variante Dev offre prestazioni decisamente superiori ma rimane confinata a una licenza di tipo non commerciale, rendendola uno strumento ideale esclusivamente per la ricerca accademica e la sperimentazione tecnica. Per chi intende invece sfruttare il modello commercialmente e alle massime prestazioni possibili, l'opzione prediletta è la versione Pro, che tuttavia non viene rilasciata come software scaricabile ma è accessibile tramite API a pagamento, garantendo così all'azienda il controllo sull'uso professionale della propria tecnologia.

2.11.3 Qwen

Qwen-Image e la sua controparte dedicata alla modifica delle immagini, Qwen-Image-Edit, sono i modelli di punta di Alibaba Cloud per la generazione e la manipolazione visiva. Queste tecnologie si sono imposte sul mercato grazie a una flessibilità operativa fuori dal comune, distinguendosi in particolare per la capacità di gestire simultaneamente molteplici immagini di riferimento in input. Questa funzione permette agli utenti di mantenere un'elevata coerenza stilistica o di soggetto, facilitando la creazione di composizioni complesse che richiedono il coordinamento di diversi elementi visivi.

Un valore aggiunto a questi modelli risiede nella precisione dei flussi di lavoro, potenziati da strumenti avanzati di inpainting che consentono di isolare e rigenerare porzioni specifiche dell'immagine senza alterarne il contesto generale. Grazie alla loro natura open source (licenziati spesso sotto Apache 2.0), i modelli Qwen sono

diventati una delle alternative più solide a soluzioni proprietarie come Adobe Photoshop o Nano Banana di Google. A differenza di questi ultimi, che operano in ecosistemi chiusi, le soluzioni di Alibaba offrono un controllo granulare sulla generazione, garantendo un livello di personalizzazione che li rende strumenti flessibili e strategici sia per la sperimentazione accademica che per le moderne pipeline di produzione professionale.

2.11.4 Wan

Wan, l'innovativo modello di generazione video sviluppato da Alibaba Cloud, rappresenta una delle soluzioni più avanzate e versatili dell'attuale panorama tecnologico. A differenza di molti competitor puramente commerciali, Wan si distingue per la sua natura open source, offrendo una suite completa che spazia dal text-to-video fino a sofisticate funzioni di video-to-video e image-to-video. Questa apertura ha permesso agli utenti non solo di operare tramite infrastrutture cloud, ma anche di implementare il modello in locale, garantendo una privacy e un controllo creativo senza precedenti. Tuttavia, l'estrema qualità cinematografica e il realismo delle generazioni è controbilanciata da una notevole pesantezza computazionale, richiedendo risorse hardware di alto profilo per esprimere il massimo potenziale del modello.

2.11.5 Grok

Grok, l'intelligenza artificiale sviluppata da xAI e integrata nella piattaforma X, si è ormai evoluta in un ecosistema multimodale capace di generare sia immagini che video. Sebbene sotto il profilo puramente tecnico possa risultare ancora un passo indietro rispetto ad alcuni competitor di fascia alta, specialmente per via della risoluzione video attualmente limitata a 720p, il modello compensa ampiamente con un'altissima velocità di esecuzione. Grok si distingue infatti per una qualità sorprendente nel fotorealismo (in particolare nella resa degli animali) e nella creazione di clip dinamiche dal forte impatto visivo, caratteristiche che lo rendono lo strumento perfetto per alimentare i contenuti virali e veloci tipici dei social media.

2.12 ComfyUI

ComfyUI rappresenta l'interfaccia grafica più avanzata e flessibile per l'utilizzo di modelli generativi open source in ambiente locale. A differenza delle soluzioni tradizionali basate su semplici caselle di testo per l'inserimento dei prompt, questo software trasforma l'esperienza utente in una vera e propria progettazione tecnica e creativa. Lo spazio di lavoro si presenta come una lavagna digitale infinita sulla

quale è possibile disporre diversi blocchi operativi, definiti nodi, ognuno dei quali è deputato a una specifica funzione, come il caricamento del modello, la decodifica dell'immagine o l'impostazione dei parametri di campionamento.

Il cuore del sistema risiede nella logica dei collegamenti: i nodi vengono connessi tra loro tramite cavi virtuali, permettendo all'utente di visualizzare in tempo reale il percorso delle informazioni e di comprendere esattamente come il dato grezzo si trasformi nel risultato finale. Questa architettura a nodi consente di combinare modelli differenti all'interno di workflow personalizzati, garantendo una libertà di intervento millimetrica su ogni fase del processo. Un ulteriore vantaggio strategico di ComfyUI risiede nella portabilità e nella condivisione dei dati, poiché l'intera struttura del flusso di lavoro viene incorporata nei metadati delle immagini generate. Ciò permette a chiunque di replicare processi incredibilmente complessi semplicemente trascinando il file dell'immagine all'interno del programma, rendendo la collaborazione e l'automazione del lavoro più immediate e potenti che mai.

2.13 The Last Image

The Last Image è il primo cortometraggio italiano totalmente realizzato con l'intelligenza artificiale generativa e VFX. Prodotto da HAI – Human & Artificial Imagination in collaborazione con EDI Effetti Digitali Italiani, e' stato presentato ad ottobre 2025 presso la Festa del Cinema di Roma e proiettato in anteprima all'Anteo di Milano. Oggi è visibile integralmente presso i social di HAI (Instagram e LinkedIn) ed è un ottimo esempio di produzione ibrida, in cui l'intelligenza artificiale, combinandosi con il valore umano e l'esperienza di diversi professionisti, ha avuto la possibilità' di esprimersi come media a se in tutte le sue opportunità'. La direzione registica di Frankie Caradonna, insieme al lavoro del team di generative e una sapiente pre-produzione, dalla realizzazione di storyboard in 2D e 3D, ad un'accurata ricerca storica dell'Inghilterra medievale, alla collaborazione con il DOP Mik Allen, hanno dato vita al soggetto realizzato dall'agenzia creativa SMALL. Il mirato lavoro di post-produzione svolto da EDI, arricchito dal montaggio del film editor Luca Angeleri, le musiche di Stabbiolo Music, il sound design di Disc To Disc Productions e la color di Danilo Vittori, ha permesso di avere un prodotto ancora attuale e di spessore. Per contestualizzare, nonostante sia uscito pochi mesi prima dal momento di scrittura di questo elaborato, nel mondo dell'intelligenza artificiale pochi mesi equivalgono a diversi anni di sviluppo tecnologico e i modelli attuali sono notevolmente più potenti rispetto a quelli utilizzati per produrre The Last Image.

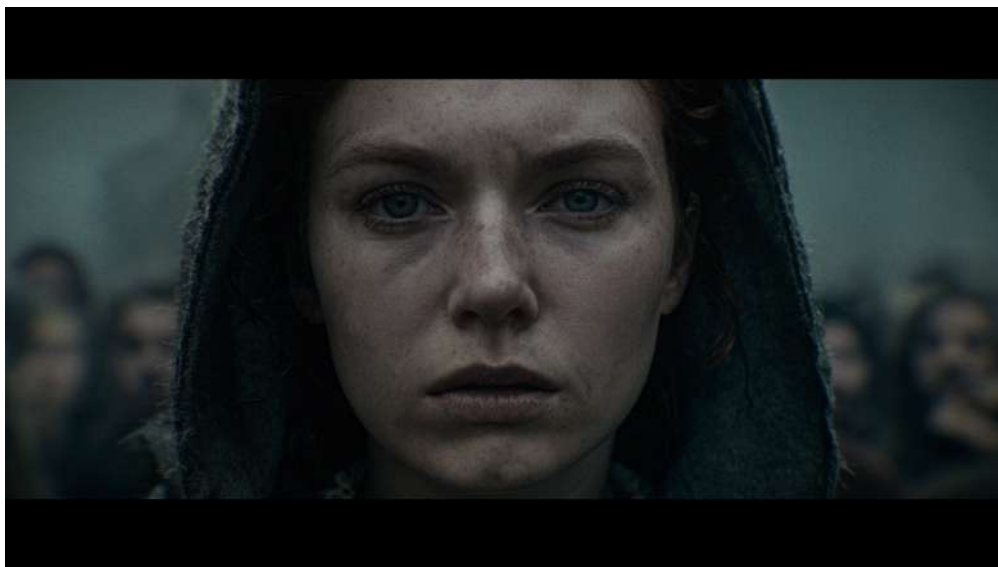


Figura 2.12: Fotogramma estratto da 'The Last Image'

Fonte: The Last Image

Per citare alcuni episodi di integrazione tra VFX e IA, con gli strumenti generativi a disposizione, era complesso mantenere il look desiderato per l'ascia del boia su ogni shot. Per soddisfare un necessario vincolo di coerenza sono entrate in gioco le lavorazioni in CG, poi finalizzate tra compositing e finishing.



Figura 2.13: Integrazione di elementi realizzati in CG

Fonte: The Last Image

Un altro episodio significativo è stato il lavoro fatto in compositing sui denti e sugli occhi del protagonista Adam, per aumentare il grado di dettaglio. Seppur l'output in uscita dalla generazione fosse di grande qualità, non era semplice ottenere l'effetto voluto su elementi così specifici.



Figura 2.14: Lavorazioni di compositing e finishing

Fonte: The Last Image

The Last Image ha rappresentato un progetto ambizioso e pionieristico, anticipando quelli che si profilano come i probabili futuri standard dell'industria cinematografica e digitale. Il lavoro ha segnato il debutto di Eva Murati, la prima attrice virtuale italiana interamente generata tramite Intelligenza Artificiale. Grazie all'esperienza di tirocinio svolta in azienda, ho avuto l'opportunità di contribuire alla fase finale della produzione, lavorando alla finalizzazione di alcuni shot e approfondendo le dinamiche di integrazione tra l'IA e la pipeline tradizionale.

Capitolo 3

Pre-produzione e concept

Indipendentemente dal medium espressivo adottato, sia esso live action, animazione tradizionale o, come nel caso presente, interamente generato tramite intelligenza artificiale, il processo creativo mantiene una radice metodologica comune: la definizione del target e lo sviluppo dell'idea. La scelta di produrre uno spot pubblicitario, piuttosto che un cortometraggio, è stata una scelta strategica scaturita da una riflessione sulla rapidità con cui i modelli allo stato dell'arte diventano obsoleti. Una produzione eccessivamente dilatata nel tempo comporterebbe, allo stato attuale della tecnologia, il rischio di una marcata disparità qualitativa tra le sequenze iniziali e quelle finali. L'incessante evoluzione degli algoritmi generativi imporrebbe, di conseguenza, un'ingente fase di rigenerazione delle clip animate per uniformare il prodotto agli standard più recenti, rendendo il flusso di lavoro inefficiente e privo di una stabilità tecnica definita. Sotto il profilo del linguaggio audiovisivo, il formato del 'commercial' si sposa molto bene, per sua natura, con le attuali capacità dei modelli generativi. La struttura tipica della pubblicità, caratterizzata da un montaggio serrato e inquadrature la cui durata oscilla mediamente tra i 2 e i 5 secondi, mitiga la durata temporale ridotta dei video generati (in genere dai 5 ai 15 secondi). Inoltre, l'efficacia comunicativa dello spot, spesso basata sul potere evocativo delle immagini e sul sound design piuttosto che su dialoghi complessi, permette di superare le criticità legate alla sincronizzazione labiale (lip-sync) e alla recitazione attoriale, valorizzando la pura estetica visiva. Inoltre, per ragioni deontologiche e di tutela del diritto d'autore, si è scelto di non realizzare uno Spec Commercial per un brand esistente, optando per un video di comunicazione sociale incentrato sul concetto stesso di Intelligenza Artificiale.

3.1 Sviluppo dell'idea

Uno dei dibattiti più accesi sull'intelligenza artificiale verte intorno una visione dicotomica che percepisce da un lato l'IA come strumento di potenziamento e dall'altro come un agente sostitutivo dell'operato umano. Questa tensione è particolarmente evidente nel settore artistico, dove la 'manualità', intesa come insieme di tecnica, esperienza e sforzo pluriennale, rischia di essere marginalizzata dalle nuove tecnologie. Se per una parte dei professionisti l'IA rappresenta una minaccia esistenziale alla propria identità lavorativa, per altri si configura come un nuovo catalizzatore creativo. In questa seconda visione, l'algoritmo non sostituisce l'artista, ma ne estende le capacità, intervenendo sin dalle fasi iniziali dell'idea e offrendo stimoli visivi inediti che fungono da base per un'ulteriore evoluzione concettuale e pratica. In un ecosistema digitale in cui l'accesso a strumenti avanzati è ormai democratizzato e universale, emerge con forza la necessità di riscoprire l'ontologia del valore umano. Risulta fondamentale, oggi più che mai, rivendicare quel contributo soggettivo che nessuna automazione può replicare: un'eredità di pensiero e intenzione che, integrata alla potenza di calcolo, consente all'essere umano di superare i propri confini creativi senza smarrire la propria identità.

In questo scenario si colloca il seguente progetto di tesi, volto a indagare la dialettica tra le potenzialità e le fragilità dell'essere umano nel confronto con le nuove tecnologie. Il progetto si traduce in un'opera visiva che esplora tre ambiti d'avanguardia in cui l'intelligenza artificiale gioca un ruolo determinante: la guida autonoma, la robotica e l'esplorazione dello spazio. Attraverso un'integrazione sinergica tra successioni di immagini e narrazione in voice-over, il lavoro intende ribadire in modo inequivocabile la centralità del fattore umano all'interno del paradigma dell'IA, evidenziando quel valore aggiunto che rimane unico e insostituibile.

3.2 Storia

3.2.1 Struttura delle scene

La narrazione si articola attraverso l'alternanza di tre micro-storie (vignette), ciascuna caratterizzata da una progressione lineare che esplora il rapporto simbiotico tra uomo e macchina:

- Introduzione dei personaggi: rapide sequenze che contestualizzano i protagonisti all'interno del proprio mondo ordinario;
- Conflitto: I personaggi si trovano ad affrontare situazioni di difficoltà o disagio derivanti dal confronto diretto con le nuove tecnologie. L'enfasi è posta sulle sensazioni di inadeguatezza e sui limiti che l'essere umano avverte nel momento in cui l'intelligenza artificiale sembra sovrapporsi al suo operato;

- Superamento e risoluzione: La situazione critica viene risolta riscoprendo il valore intrinseco dell'essere umano e ribadendo il ruolo fondamentale e insostituibile dell'umanità nel processo tecnologico.

3.2.2 Elementi narrativi

Al fine di accentuare il carico emotivo della storia e conferire maggiore profondità tematica, il progetto si avvale di tre specifici espedienti narrativi:

- Voice Over: una voce narrante femminile accompagna le sequenze visive verso la risoluzione finale, esplicitando il tema dell'intelligenza artificiale e dando voce alla natura dicotomica dell'animo umano, sospeso tra timore e aspirazione;
- Effetto pixelato: un effetto particellare che si sovrappone agli elementi della scena, agendo come metafora visiva della trasformazione introdotta dalle nuove tecnologie. Questo disturbo digitale non solo rappresenta il cambiamento, ma richiama simbolicamente il processo di iterazione dei prompt tipico degli strumenti generativi, in cui la realtà viene progressivamente manipolata e riscritta dall'algoritmo;
- Stile cartoon: in una scena verrà alternato uno stile cartoon al fotorealismo, con lo scopo di enfatizzare il valore della creatività umana e dare una forma visiva all'immaginazione pura. Questo richiamo al mondo del fumetto e dell'illustrazione funziona anche da ponte narrativo verso il colpo di scena finale.

3.2.3 Divisione in scene

Definita la struttura e gli elementi narrativi, la storia è stata sviluppata e divisa in scene:

- Scena 1: la scena si apre su un attraversamento pedonale, una madre e il suo bambino si apprestano ad attraversare la strada, mano nella mano;
- Scena 2: un pugile, da solo nello spogliatoio, si prepara al suo prossimo combattimento;
- Scena 3: due astronauti si incamminano verso lo shuttle che li porterà nello spazio;
- Scena 4: continuando la prima scena, la madre e il figlio attraversano la strada. Un urto accidentale con un passante causa la caduta di un vecchio orsetto

di pezza rammendato, che il bambino stringeva a sé. L'oggetto, con le sue cuciture a vista e la pelliccia logora, diventa l'archetipo di un passato che rischia di rimanere indietro. Il bambino, mosso dal legame emotivo, si svincola dalla presa materna e corre verso il centro della carreggiata per recuperare il giocattolo. Proprio nel momento in cui sta per afferrarlo, interviene l'effetto particellare: una sovrapposizione digitale pixelata avvolge il peluche, facendolo dissolvere nel nulla. Questa sparizione improvvisa simboleggia l'inevitabilità di un cambiamento incalzante, un'evoluzione tecnologica che non ammette soste e che obbliga l'individuo a un passo frenetico per non restare escluso. Mentre il semaforo pedonale vira bruscamente al rosso, la tensione si sposta sulla strada. Un'auto a guida autonoma, programmata su algoritmi di pura efficienza, accelera procedendo verso le strisce pedonali dove la madre, nel frattempo, ha raggiunto il figlio per fargli scudo con il proprio corpo. L'avanzata imminente del veicolo proietta nello spettatore un profondo senso d'ansia: la tecnologia, se non governata o se inseguita senza la necessaria prontezza, assume i connotati di una forza travolgente capace, metaforicamente e letteralmente, di "investire" l'essere umano. Nello specifico dell'intelligenza artificiale, essa ridefinisce il concetto stesso di competenza, rendendo obsoleti modelli di pensiero e manualità che hanno richiesto generazioni per essere perfezionati. In questo senso, essere 'investiti' dall'IA significa subire uno shock ontologico: il pericolo non è solo la perdita del primato tecnico, ma il disorientamento di fronte a uno strumento che, privo di una guida etica e intenzionale, rischia di procedere verso obiettivi di pura efficienza, ignorando le necessità e le fragilità che definiscono la nostra umanità;

- Scena 5: continuando la seconda scena, l'azione si sposta all'interno di una palestra deserta, un ambiente che amplifica il senso di isolamento e concentrazione del protagonista. Un pugile, nel pieno della sua preparazione atletica, si appresta a salire sul ring per affrontare una sfida. Ad attenderlo al centro del quadrato non ci sono avversari umani, ma tre automi dall'estetica essenziale e moderna. Tuttavia, il confronto subisce un'improvvisa distorsione: l'effetto particellare interviene violentemente, sovrapponendosi ai robot e alterandone la struttura. Le proporzioni vengono amplificate, l'aspetto diventa più minaccioso. La determinazione sul volto del pugile svanisce, lasciando spazio a un'espressione di puro terrore. Questa metamorfosi visiva rappresenta l'impossibilità dell'uomo di competere sul piano della pura performance evolutiva: sfidare la tecnologia su un terreno di forza e velocità significa, inevitabilmente, esserne sopraffatti. Mentre l'essere umano è vincolato ai propri limiti biologici e ai tempi lenti dell'apprendimento manuale, l'intelligenza artificiale itera e migliora in continuazione, diventando un avversario fuori portata per chiunque tenti di affrontarla con gli strumenti del passato;

- Scena 6: continuando la terza scena, lo shuttle viene lanciato nello spazio. Uno dei due astronauti si allontana dal proprio compito per osservare la Terra da un piccolo oblò. Il suo sguardo è intriso di una nostalgia profonda e silenziosa: un senso di distacco che non è solo fisico, ma ideologico. Questa sequenza, nella sua essenzialità, cristallizza il sentimento di alienazione che molti provano nei confronti dell'intelligenza artificiale. Lo shuttle diventa una sorta di esilio concettuale, mentre l'astronauta rappresenta quella parte di umanità che, spaventata o sopraffatta dall'onnipresenza algoritmica, coltiva il desiderio di fuggire verso un "altrove" incontaminato. Lo sguardo rivolto al pianeta lontano simboleggia l'anelito a un mondo in cui l'esistenza conservava la propria misura a ridosso dell'umano, e in cui l'individuo poteva sentirsi al riparo da un'innovazione percepita come estranea o minacciosa.e estranea o minacciosa;
- Scena 7: questa scena rappresenta la risoluzione emotiva del primo segmento. In continuità con la quarta scena, l'imminente impatto viene scongiurato da un intervento improvviso: l'effetto particellare, descritto in precedenza, avvolge l'autovettura prima del contatto, facendola dissolvere nel nulla. Al centro della strada, immobili e incolumi, rimangono la madre e il bambino, stretti in un tenero abbraccio che squarcia la tensione del momento. In questa fase, la pixelizzazione della macchina cessa di essere un elemento di disturbo per trasformarsi nel simbolo di un cambiamento non necessariamente negativo, ma protettivo e adattivo. L'attenzione si sposta così sulla potenza dell'abbraccio, ovvero una manifestazione di umanità pura ed empatica che nessuna macchina può replicare o sottrarre all'individuo. Quest'immagine pone l'accento su ciò che definisce realmente l'essenza umana, elementi come le emozioni, l'istinto di protezione e il legame affettivo. La scena invita a riscoprire quel valore che ci distingue dall'automazione, offrendo una riflessione necessaria da cui ripartire per ridefinire il nostro ruolo in un mondo tecnologicamente sempre più avanzato;
- Scena 8: questa scena rappresenta la risoluzione emotiva del secondo segmento. In continuità con la quinta scena, la sfida sul ring subisce una metamorfosi stilistica e narrativa. Il pugile abbandona i rigidi schemi della boxe tradizionale per adottare uno stile di lotta imprevedibile e "sregolato": le sue movenze trascendono le leggi della fisica, trasformandolo in una sorta di eroe da fumetto capace di sferrare colpi spettacolari e ai limiti del possibile che abbattano i robot nemici. Visivamente, questa sequenza è caratterizzata da un'alternanza ritmica tra il fotorealismo e l'estetica cartoon, che trasfigura il combattimento in una dimensione puramente fantasiosa. Questa transizione stilistica mira a porre in primo piano la creatività elevandola a risorsa fondamentale per l'essere umano. Se la macchina eccelle nella logica, e nel calcolo dei dati,

l'uomo risponde con il pensiero laterale, l'astrazione e la capacità di sognare l'impossibile. L'irruzione dello stile cartoon simboleggia proprio questo "vantaggio competitivo": l'intelligenza artificiale può replicare la realtà, ma è l'intuito umano a poterla reinventare;

- **Scena 9:** questa scena rappresenta la conclusione del terzo segmento. Continuando la sesta scena, lo shuttle prosegue la sua navigazione nello spazio. Attraverso una serie di match cut ritmici, il veicolo transita fluidamente tra diversi scenari galattici, pittoreschi e remoti, suggerendo visivamente un viaggio che trascende i confini dell'esplorazione umana contemporanea. La tecnologia dello shuttle e l'intenzionalità degli astronauti si fondono in un unico movimento armonico verso l'ignoto. Questa sequenza rappresenta il superamento del conflitto tra uomo e intelligenza artificiale. L'obiettivo è dimostrare come dalla sinergia tra intuito umano e potenza tecnologica sia possibile valicare i limiti biologici e cognitivi attuali. Lo spazio profondo diventa così la metafora delle nuove frontiere che l'umanità può raggiungere non più fuggendo dalla macchina, ma collaborando con essa. La tecnologia cessa di essere una minaccia per trasformarsi nel veicolo che abilita l'uomo a esplorare dimensioni altrimenti inaccessibili, definendo un nuovo paradigma di progresso condiviso.
- **Scena 10:** Attraverso un match cut finale, la vastità dello spazio profondo si restringe fino a coincidere con la vignetta di un fumetto. L'albo è tra le mani del bambino della prima scena: l'intera odissea vissuta finora non è stata che il frutto della sua fantasia. Questa chiusura circolare riconnette l'epilogo al prologo, svelando come le sequenze descritte siano nate dalla mente del giovane protagonista. Questa scena chiude il cerchio della storia e chiarisce il messaggio del progetto: la fantasia non è una fuga dalla realtà, ma lo strumento pratico per gestire l'innovazione. Ritrovare la capacità di sognare tipica dell'infanzia significa rimettere l'uomo al centro del processo creativo. In questo modo, l'intelligenza artificiale smette di essere una minaccia che ci "investe" o ci sostituisce e diventa un mezzo nelle nostre mani. Il bambino che legge il fumetto rappresenta l'essere umano che, grazie al proprio intuito, trasforma il mezzo tecnologico in uno strumento per esprimere e realizzare la propria fantasia.

3.3 Pitch

Successivamente, è stato elaborato un pitch con l'obiettivo di sintetizzare l'idea progettuale, l'approccio narrativo e lo sviluppo della storia. Il ruolo del pitch, tuttavia, non si esaurisce nella mera condensazione degli elementi, ma mira a trasmettere

l'impatto emotivo e l'atmosfera che caratterizzeranno la messa in scena. A tal fine, è stata condotta una ricerca iconografica su piattaforme specializzate quali Shotdeck (<https://shotdeck.com/>) e Filmgrab (<https://film-grab.com/>), celebri per la loro vaste raccolte di fotogrammi tratti da diverse produzioni cinematografiche globali. L'analisi di queste references professionali è stata fondamentale per trarre ispirazione e definire con maggiore precisione e coerenza stilistica aspetti quali la composizione dell'inquadratura, l'illuminazione, il color grading e il mood complessivo dell'opera.

3.4 Storyboard e Animatic

Seguendo un workflow tradizionale, è stato realizzato uno storyboard volto a definire le inquadrature, la disposizione degli elementi in scena e la composizione visiva. Per la fase di disegno è stata utilizzata l'applicazione Procreate su iPad, sebbene il processo si sarebbe potuto avvalere di software analoghi come Photoshop o Clip Studio Paint. In questa fase, lo storyboard non è stato concepito come un elaborato artistico fine a se stesso, quanto piuttosto come uno strumento funzionale e propedeutico per guidare la successiva generazione delle immagini. Anche disegni in bozza, purché chiari nella struttura, si sono rivelati sufficienti per trasmettere l'idea progettuale. Successivamente, le tavole sono state montate su una timeline di DaVinci Resolve per la creazione di un animatic, fondamentale per testare il ritmo delle azioni e la sequenzialità narrativa delle inquadrature.

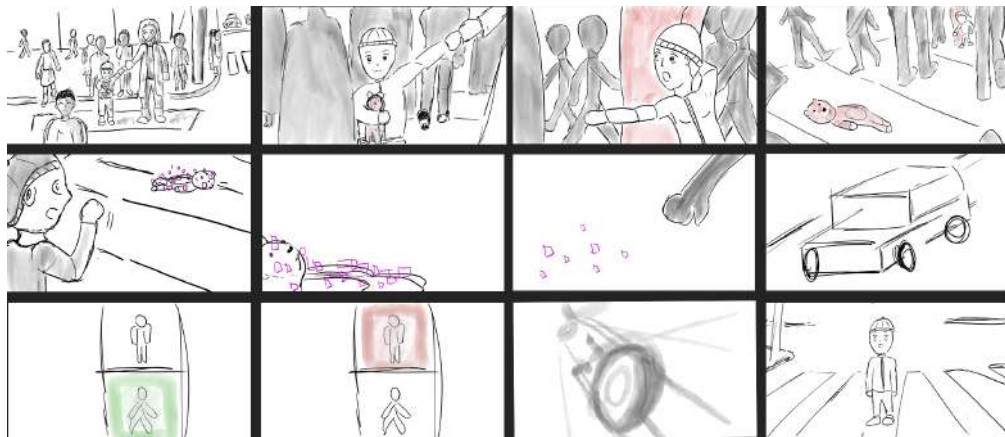


Figura 3.1: Alcune tavole di Storyboard tratte dalla prima sequenza narrativa

Capitolo 4

Produzione e IA Generativa

4.1 Personaggi

La fase embrionale del progetto si è focalizzata sulla caratterizzazione dei personaggi. Sebbene in sede di pitch fossero state individuate delle reference visive, si è scelto di non impiegarle direttamente nel processo di generazione per prevenire potenziali criticità legate alla violazione del copyright. Gli asset sono stati pertanto generati ex novo attraverso un paradigma text-to-image, demandando l'originalità degli output ai protocolli di addestramento e ai filtri etici del modello.

Per l'elaborazione è stata adottata la piattaforma Gemini, sfruttando la sinergia tra le capacità di comprensione semantica del Large Language Model (LLM) e la potenza computazionale del motore grafico Nano Banana. Per quanto concerne l'interazione con il modello, si è optato per un approccio linguistico ibrido. Nonostante i recenti aggiornamenti della suite Google (incluso l'ambiente Flow) abbiano esteso il supporto alla lingua italiana attraverso sistemi integrati di traduzione e contestualizzazione, l'evidenza empirica ha confermato una maggiore precisione del modello nell'elaborazione di prompt redatti in lingua inglese. Di conseguenza, la stesura dei comandi è avvenuta alternando le due lingue in base alla complessità semantica richiesta.

4.1.1 Generazione dei personaggi della prima sequenza

La configurazione del bambino è stata definita mediante la seguente sequenza di prompt:

Iterazione 1:

Prompt: "Generate a 6-year-old-boy, 110 centimeters tall, brown eyes, naive expression, straight brown hair with some strands escaping from the wool cap he is wearing. The slender build is covered by a winter jacket and a pair of

jeans. He is wearing sneakers on his feet".

Analisi: il primo prompt è stato strutturato per definire le componenti volumetriche e stilistiche fondamentali, quali la fisionomia del personaggio, la tipologia di vestiario e i tratti somatici macroscopici. In questa fase il colore dei vestiti non era di particolare rilevanza, essendo facilmente cambiabile in un'iterazione successiva. Sebbene, quindi, la palette cromatica fosse destinata a un'ottimizzazione di secondo livello, l'output ottenuto ha restituito immediatamente un risultato gradevole e con una predominanza di sfumature di blu in linea con l'idea originale, non richiedendo quindi particolari passaggi successivi in tale direzione.

Iterazione 2:

Prompt: "Remove the strand of hair poking out from behind the head. Put the child on a white background".

Input: Immagine generata all'iterazione 1.

Analisi: la prima iterazione presentava il difetto di avere alcune ciocche di capelli anomale, uscenti da dietro la testa. Si è quindi proceduto a una correzione mirata del prompt. Il personaggio è stato inoltre isolato su uno sfondo neutro per evitare contaminazioni semantiche o cromatiche; l'assenza di dettagli ambientali garantisce una maggiore pulizia visiva, rendendo l'immagine una reference ideale da usare in generazioni successive.

Iterazione 3:

Prompt: "Do a turn around of this child".

Input: Immagine generata all'iterazione 2.

Analisi: Una volta consolidata la resa estetica del soggetto, si è proceduto alla generazione di un character turn-around comprendente le quattro viste principali. Questa operazione risulta fondamentale già in questa fase iniziale, così da stabilire una visual reference esaustiva, garantendo la coerenza del personaggio nelle successive fasi di produzione.

Iterazione 4:

Prompt: "Do a turn-around of the child in a close up".

Input: Immagine generata all'iterazione 3.

Analisi: risulta inoltre essenziale disporre di un'inquadratura in primo piano (close-up), finalizzata a stabilire un riferimento dettagliato per la fisionomia del volto. Tale accorgimento garantisce la coerenza del soggetto nelle inquadrature a focale ridotta, dove i tratti somatici costituiscono il fulcro visivo. Parallelamente, la strutturazione di un turn-around nelle quattro viste canoniche fornisce al modello una base geometrica completa, facilitando l'interpolazione corretta delle pose intermedie, come le inquadrature di tre quarti.



Figura 4.1: Turn-around del personaggio del bambino

Una volta definito il design del bambino, il passaggio successivo è stato rappresentare una figura materna che risultasse visivamente coerente e credibile:

Iterazione 1:

Prompt: "Generate this child's mother. 165 centimeters tall, brown hair tied back in a ponytail with some strands left loose at the front, falling sideways on her face(right and left). The woman has some slight wrinkles on her face. She is wearing a coat, under which she has a sweater, trousers and a pair of ankle boots".

Input: Turn around del bambino in full shot.

Analisi: al fine di avere una somiglianza genetica tra madre e figlio, è stata utilizzata in input una reference del bambino. Sono stati definiti nel prompt testuale indicazioni sui tratti somatici della madre e il suo vestiario.

Iterazione 2:

Prompt: "The trousers and boots must be the same style but in a darker color, tending towards dark blue. Put the woman on a white background, still full-body".

Input: immagine generata all'iterazione 1.

Analisi: in questo caso è stato cambiato il colore di alcuni vestiti, per farli avvicinare all'immaginario desiderato. Anche in questo caso il personaggio è stato messo su sfondo bianco, mostrando anche come all'interno di uno stesso prompt si possano richiedere al modello più modifiche in contemporanea.

Iterazione 3:

Prompt: "Change the color of the sweater, it must be the same color as the child's wool cap".

Input: Turn around del bambino in close up ed immagine generata all'iterazione 2.

Analisi: per cercare di creare un match cromatico tra bambino e madre, si è cercato di associare il berretto del bambino al maglione della madre.

Iterazione 4:

Prompt: "Do a turn around of this woman".

Input: Immagine generata all'iterazione 3.

Analisi: così come fatto per il bambino, è stato utile creare un turn-around della donna, in vista delle generazioni future.

Iterazione 5:

Prompt: "Do a turn around of this woman's face (close up). Mantain consistency with her face and with her clothes as seen in the full-body figure as much as possible".

Input: Immagine generata all'iterazione 4.

Analisi: anche in questo caso caso è stato creato un close up della donna.



Figura 4.2: Turn-around del personaggio della madre

Durante questa fase sono state realizzate anche alcune espressioni facciali per il bambino e la madre, come prima prova di acting e da utilizzare come reference in caso di necessità.

4.1.2 Generazione dei personaggi della seconda sequenza

Il secondo blocco narrativo vede come protagonista il pugile, il personaggio che affronterà i robot nel corso del combattimento:

Iterazione 1:

Prompt: "Generate a boxer of color, around 20 years old, 180 centimeters tall, with curly, dark hair and a squared jawline. He is wearing dark gloves with a white wrist strap. Shorts with white flames on a dark background. Dark boxing boots with white laces".

Analisi: oltre a indicare i tratti somatici del pugile, sono stati dati dettagli anche sui vestiti, così da ottenerne una prima versione.



Figura 4.3: Prima generazione del personaggio del pugile

Iterazione 2:

Prompt: "Take the first boxer in the image. Put him on a white background. Change the color of the flames on the shorts; It must be the same color as the child's wool cap".

Input: immagine generata all'iterazione 1 e e turn-around del bambino in close up.

Analisi: dall'immagine iniziale era necessario estrapolare solo una vista, è stata scelta la prima essendo quella più vicina all'immaginario previsto. Volendo poi dare un'elemento di continuità con la scena precedente è stato utilizzato il colore del cappello del bambino per i dettagli sui pantaloncini del pugile.

Iterazione 3:

Prompt: "Do a turn-around of the boxer in full-body: front view, profile, back view, other profile".

Input: immagine generata all'iterazione 2.

Analisi: la generazione precedente, influenzata anche dal turnaround dell'immagine del bambino, ha restituito più di una vista del pugile. Il turn-around è però incorretto presentando due volte il frontale ripetuto e poi una vista da dietro del pugile.

Iterazione 4:

Prompt: "In the last image, the boxer's feet are turned the wrong way compared to the rest of the body. Fix them".

Input: immagine generata all'iterazione 3.

Analisi: il turn-around è stato generato correttamente, a parte un'anomalia presente in una delle viste del profilo, in cui la parte superiore del corpo è ruotata correttamente, mentre la parte posteriore è ruotata al contrario. Il prompt di quest'iterazione dimostra come, dando gli opportuni riferimenti, sia possibile influenzare il modello ad agire in maniera mirata, modificando solo il problema di interesse e lasciando il resto inalterato.

Iterazione 5:

Prompt: "Do the turn-around of this boxer's face(close-up). Maintain consistency with his face as see in the full-body figure as much as possible".

Input: immagine generata all'iterazione 4.

Analisi: anche in questo caso è stato generato il close-up così da assicurarsi la consistenza del volto nelle future generazioni. Generandolo a partire dall'immagine in figura intera, il volto appare leggermente diverso rispetto alla prima generazione, ma più in linea con l'aspetto immaginato per il personaggio.

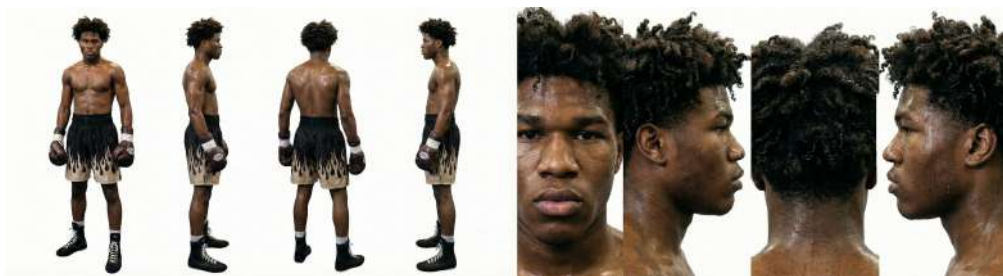


Figura 4.4: Turn-around del personaggio del pugile

Per la creazione del robot è stato adottato un workflow differente, basato sulla reference specifica di K-2SO (*Rogue One: A Star Wars Story*, 2016). Invece di sottoporre direttamente l'immagine al modello, è stato applicato un processo di reverse-engineering testuale: è stato chiesto a Gemini di descrivere il personaggio per analizzare come l'IA interpretasse i suoi tratti distintivi. La descrizione ottenuta non è stata copiata pedissequamente, ma utilizzata come base analitica per redigere un prompt originale, selezionando e reinterpretando solo le caratteristiche fondamentali necessarie al design finale.

Iterazione 1:

Prompt: "Generate a fearsome robot 200 centimeters tall, very slender particularly in the limbs. Its torso is big, almost like a armor. The primary color is a dark charcoal gray or black used across the majority of its body, giving it a stealthy and somewhat intimidating look. Its small head has two sensor as a eyes".

Analisi: il prompt definisce i tratti distintivi del personaggio, lasciando al modello un margine di autonomia interpretativa per gli elementi secondari.

Iterazione 2:

Prompt: "Remove the muscles, make it more skinny".

Input: immagine generata all'iterazione 1.

Analisi: il personaggio ottenuto al precedente passaggio mostrava una fisionomia molto robusta e muscolosa, segno anche di una contaminazione quasi nulla da parte della reference usata come ispirazione, che al contrario ritrae un robot molto longilineo.

Iterazione 3:

Prompt: "Add just the torso part. Like its an armor".

Input: immagine generata all'iterazione 2.

Analisi: nel tentativo di rendere il personaggio più slanciato, il modello ha omissso la corazza superiore definita nella prima iterazione. Questo evidenzia come il processo di generazione sia un raffinamento continuo, in cui la correzione di un singolo attributo può inaspettatamente alterarne altri già consolidati.

Iterazione 4:

Prompt: "Put the robot on a white background and remove the red eyes".

Input: immagine generata all'iterazione 3.

Analisi: raffinamento conclusivo e posizionamento del personaggio all'interno di uno sfondo neutro prima di procedere al solito turn-around.



Figura 4.5: Step di generazione del personaggio del robot minaccioso

Per la versione meno minacciosa del robot, si è optato per un design caratterizzato da linee morbide e un'estetica essenziale, ispirata a modelli reali come l'Agibot A2. Anche in questo caso è stata adottata la medesima metodologia in cui Gemini è stato utilizzato per analizzare la reference e sintetizzarne gli elementi chiave in una descrizione testuale. Tale output non è stato impiegato direttamente come prompt, ma è servito da base analitica per redigere un testo originale, garantendo coerenza tra l'ispirazione reale e la generazione text-to-image.

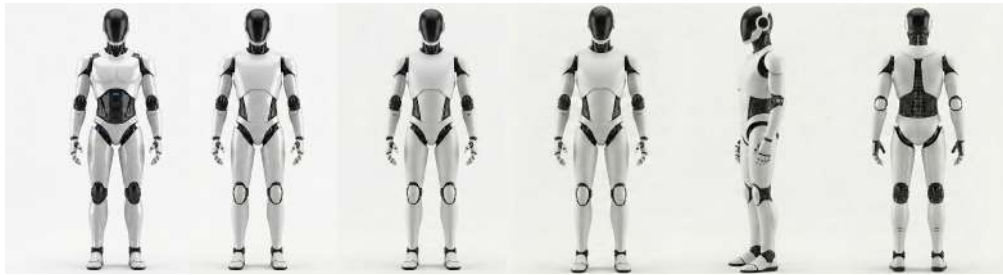


Figura 4.6: Step di generazione del personaggio del robot

4.1.3 Generazione dei personaggi della terza sequenza

Sebbene la storia mostri due astronauti, l'inquadratura che li ritraeva entrambi in volto è stata rimossa in fase di montaggio. Questa scelta sposta l'intero peso emotivo sul singolo astronauta che, colto in un momento di profonda nostalgia, osserva la Terra dall'oblò dello shuttle.

Iterazione 1:

Prompt: "Close-up of a middle-aged man with asian almond-shaped eyes, slightly cropped hair, and a hint of a beard. Put him on a clean white background".

Analisi: essendo la scena in cui guarda fuori dall'oblò significata, l'attenzione, anche in fase di generazione ricade sul primo piano della persona.

Iterazione 2:

Prompt: "Put this man in full figure".

Input: immagine generata all'iterazione 1.

Analisi: al fine di poter vestire il personaggio da astronauta, è più comodo metterlo adesso in figura intera. Allargando il campo alcuni dettagli del viso prima definiti si perderanno, ma in linea di massima i tratti somatici dovrebbero rimanere invariati.

Iterazione 3:

Prompt: "Put this man inside an astronaut suit. His expression is serious, he is not smiling".

Input: immagine generata all'iterazione 2.

Analisi: dopo aver definito l'abbigliamento da astronauta, è stato necessario vincolare l'espressione a un tono serio. Questo accorgimento è stato fondamentale per ovviare a precedenti tentativi in cui il personaggio appariva sorridente, garantendo così la coerenza emotiva necessaria alla successiva scena di nostalgia in cui verrà rappresentato.

Iterazione 4:

Prompt: "Remove any logo from the suit."

Input: immagine generata all'iterazione 3.

Analisi: per prevenire potenziali violazioni di copyright, è stato rimosso ogni segno distintivo dalla tuta. Durante le iterazioni, il modello ha mostrato una marcata tendenza a riprodurre il logo NASA, evidenziando un bias specifico nei dati di addestramento che è stato necessario correggere per ottenere un design neutro.

Iterazione 5:

Prompt: "Close up of the man".

Input: immagine generata all'iterazione 4.

Analisi: si è proceduto con una nuova generazione focalizzata sul volto dell'uomo: un close-up dettagliato che lo ritrae finalmente nel suo outfit da astronauta, consolidando il look del personaggio.

Iterazione 6:

Prompt: "Make a turnaround of this astronaut in close-up".

Input: immagine generata all'iterazione 5.

Analisi: così come per gli altri personaggi, si è preferito preparare un turnaround anche per l'astronauta.



Figura 4.7: Step di generazione del personaggio della terza sequenza narrativa

4.1.4 Generazione di altri elementi della scena

Un elemento cardine della prima sequenza è l'orsacchiotto. Il suo aspetto logoro costituisce un punto di riferimento visivo fondamentale, la cui coerenza deve essere preservata in ogni inquadratura:

Iterazione 1:

Prompt: "A 3D character reference sheet showing a vintage, taupe-brown plush teddy bear in front, back, and two profile views, in a row. The bear has a well-loved, matted texture, bead eyes, and no mouth".

Analisi: Inizialmente, l'asset dell'orsacchiotto è stato generato senza i distintivi rammendi e le cicatrici asimmetriche che ne caratterizzano l'aspetto vissuto.

Iterazione 2:

Prompt: "add stitching and patches to the teddy bear to indicate signs of aging".

Input: immagine generata all'iterazione 1.

Analisi: sono stati aggiunti all'orsacchiotto i diversi rattoppi e cuciture.

Il workflow ha previsto la generazione di diverse varianti dell'asset da utilizzare come 'livelli' in Adobe Photoshop. Sovrapponendoli e operando con delle maschere, è stato eseguito un intervento mirato per scegliere, quali rammendi lasciare visibili e quali rimuovere, garantendo il massimo controllo.



Figura 4.8: Step di generazione dell'asset dell'orsacchiotto

Spostandoci alla seconda sequenza narrativa, essendo all'interno di uno spazio circoscritto, a priori, insieme ai personaggi è stata definita l'ambientazione:

Iterazione 1:

Prompt: "Establishing shot: A long, elevated white walkway leads to a boxing ring, which is even more elevated. The only things illuminated in the scene are the ring and the walkway; the attention must fall on them. The rest is in darkness; nothing is visible, but it is implied that the space is large. The lighting comes from spotlights on the ceiling that emit cold light. The ring is made of ropes of different colors, white and black. The rubber floor of the ring is colored in a faded dark blue".

Analisi: il prompt si concentra sulla gestione dello spazio, sulla descrizione precisa degli oggetti in scena e sulla creazione di un'atmosfera specifica. L'uso di un'illuminazione selettiva permette di definire un ambiente vasto pur lasciandolo quasi interamente nel buio, concentrando il focus narrativo sul percorso e sul ring.

Iterazione 2:

Prompt: "The walkway has to be lower; do not make it elevated. Put white steps at the end of the walkway to climb onto the ring. Remove the railings from the walkway. Make the environment even darker; not much should be visible beyond the ring and the walkway. Do not show the building walls; keep them in shadow".

Input: immagine generata all'iterazione 1.

Analisi: l'inquadratura precedentemente generata risulta sbilanciata. La passerella sopraelevata che conduce al ring attira troppo l'attenzione, sottraendo centralità al ring stesso. Inoltre, la profondità di campo è eccessiva, rendendo visibili le pareti di fondo e rompendo l'atmosfera ricercata. È necessario abbassare il punto di vista della pedana e oscurare o sfumare lo sfondo per mantenere il focus esclusivamente sul ring.

Iterazione 3:

Prompt: "Enlarge the bleachers along the side walls of the venue".

Input: immagine generata all'iterazione 2.

Analisi: gli spalti, inizialmente sottodimensionati, sono state ampliati per dare maggiore respiro alla struttura e successivamente scuriti in fase di post-produzione per enfatizzare l'atmosfera della scena.

Oltre ai prompt principali sono state realizzate più varianti, poi assemblate su Adobe Photoshop attraverso tecniche di digital matte painting. Questo approccio ha permesso di rifinire minuziosamente ogni singolo elemento della composizione, garantendo il pieno controllo creativo sul risultato finale.

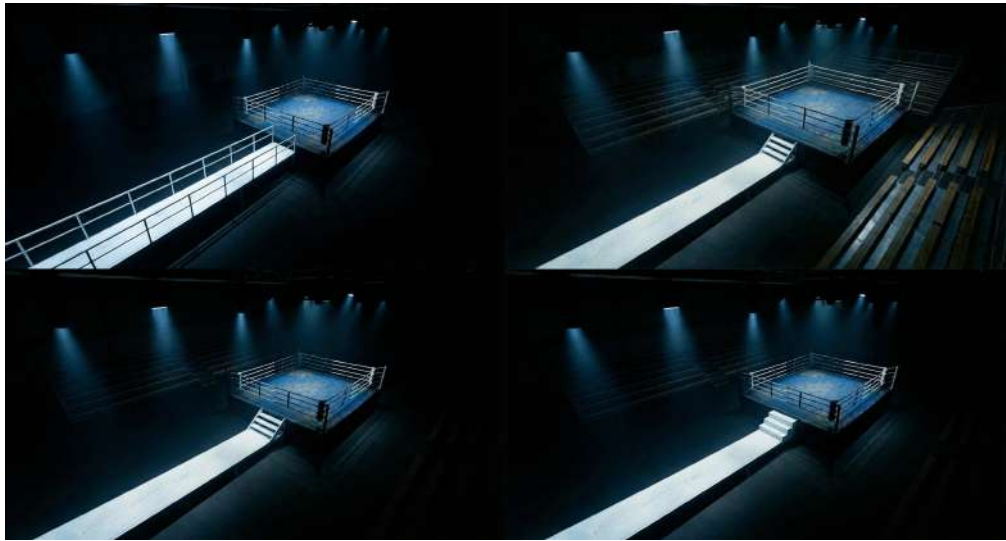


Figura 4.9: Generazione dell'ambiente della seconda sequenza narrativa

4.2 Creazione delle scene

È importante sottolineare che l'ordine di generazione del materiale non riflette la cronologia della timeline finale. In questo processo, il montaggio diventa una fase integrata alla produzione piuttosto che un semplice passaggio di post-produzione: l'accostamento continuo delle clip generate permette una revisione registica costante. Questo approccio iterativo consente di valutare immediatamente l'efficacia della narrazione e di individuare eventuali lacune, guidando la generazione dei contenuti successivi in base alle reali necessità del montaggio.

Il workflow non si è basato su una semplice generazione text-to-video, ma su un approccio più controllato volto a massimizzare la precisione dell'inquadratura e della regia. Sono state preventivamente prodotte delle immagini chiave (stills) da utilizzare come input per le animazioni. Per le sequenze lineari è stato sufficiente l'impiego di un First Frame, mentre per le clip più dinamiche è stato necessario guidare le azioni utilizzando sia il primo che l'ultimo fotogramma (First e Last Frame). La creazione delle immagini ha visto l'impiego di Nano Banana, gestito tramite Gemini e Flow per ottimizzare l'output visivo. Sul fronte video, sono stati alternati Veo 3.1 e KlingAI (2.6, 3.0 e l'architettura Omni), scegliendo di volta in volta il modello capace di interpretare al meglio il movimento e il realismo richiesti dalla scena. Per evitare eccessive ridondanze verranno riportati a seguire i processi di lavorazione degli shot più significati.

4.2.1 Generazione della prima sequenza narrativa

La prima sequenza narrativa ha presentato una sfida complessa a causa della presenza del personaggio del bambino. I rigidi protocolli di sicurezza di Flow hanno reso impossibile l'upload e l'elaborazione delle immagini, con blocchi sistematici applicati ai soggetti minori; di conseguenza, anche l'utilizzo di Veo 3.1 per l'animazione è risultato precluso. Si è reso quindi necessario, per la generazione delle immagini, un workflow alternativo tramite l'interfaccia di Gemini. Sebbene il modello di base fosse sempre Nano Banana, Gemini ha dimostrato una migliore contestualizzazione semantica e filtri di safety meno restrittivi. Tuttavia, il processo è rimasto oneroso, richiedendo un accurato prompt tuning e frequenti reset delle sessioni per superare i fenomeni di over-blocking e portare a termine la generazione. Le generazioni video sono invece state affidate a KlingAI, che non ha riscontrato criticità con i filtri di sicurezza legati alla presenza di minori. Questa flessibilità ha permesso di animare le scene con il bambino senza i blocchi sistematici incontrati in precedenza, garantendo la continuità produttiva necessaria per completare la prima sequenza narrativa.

A differenza dei personaggi, sviluppati inizialmente tramite un approccio text-to-image, la generazione delle stills è partita da basi visive preesistenti. Nello specifico, sono stati forniti al modello sia i character design precedentemente consolidati, sia lo storyboard di riferimento; quest'ultimo è stato fondamentale per orientare la composizione della scena e garantire il rispetto delle inquadrature pianificate.

SCENA 1 - SHOT 03

FIRST FRAME

Iterazione 1:

Prompt: "Re-render this image in a photorealistic way. The child is crossing the road in the middle of the crowd, holding his mother's hand".

Input: tavola di storyboard corrispondente, turn-around in figura intera del bambino, asset dell'orsacchiotto.

Analisi: l'obiettivo consiste nel rendere fotorealistica la tavola di storyboard, aggiungendo gli asset previamente generati e mantenendone la composizione.

Iterazione 2:

Prompt: "The mother should be this one here".

Input: immagine ottenuta all'iterazione precedente e asset della madre in figura intera.

Analisi: Poiché la madre appare solo di scorcio e non si desidera mostrarne il viso, si è deciso di non inserire il suo asset come input nel prompt

precedente, così da non generare ambiguità nel modello. L'attenzione deve concentrarsi esclusivamente sulla corretta resa dei suoi vestiti.

Iterazione 3:

Prompt: "The child is looking ahead, past the camera.".

Input: immagine ottenuta all'iterazione 2.

Analisi: il layout originale dello storyboard ha influenzato la posa nella prima iterazione, portando il bambino a rivolgere lo sguardo direttamente in camera. Per correggere questa staticità e conferire maggiore naturalezza allo shot, si è deciso di reindirizzare il suo sguardo altrove.

Le generazioni ottenute non costituiscono il risultato finale, ma rappresentano asset idonei per essere importati come livelli in Adobe Photoshop e sottoposti a ulteriore lavorazione. Spesso, infatti, l'output generativo non dev'essere definitivo, ma prodotto nell'ottica di una successiva post-produzione. Nello specifico, si è reso necessario intervenire sulla folla alle spalle dei personaggi chiave: se nella prima iterazione la disposizione dei passanti rispettava fedelmente lo storyboard, nelle successive la folla risultava involontariamente diradata. È stato quindi generato un clean plate della folla corretta (previa rimozione del bambino e della madre) e, successivamente, questo plate è stato integrato su Photoshop per ripristinare il corretto background alle spalle dei personaggi.



Figura 4.10: Da storyboard a generazione

Utilizzando questa still come primo frame (first frame), è stata effettuata una prima generazione video su Kling 2.6 basata sul prompt: 'The child walks

through the crowd while holding his mother's hand'. L'elemento critico mancante era l'impatto, o almeno la sua preparazione, con il passante che avrebbe poi causato la caduta dell'orsacchiotto dalle mani del bambino. Per risolvere questa carenza, è stato preparato un ultimo frame (last frame) manipolando un fotogramma estratto dalla prima generazione video, garantendo così il movimento dei passanti sullo sfondo dei personaggi principali. Se la folla fosse rimasta statica tra il primo e l'ultimo frame, la generazione video avrebbe potuto introdurre artefatti o loop indesiderati. Per estrarre un frame da un video si possono adottare molteplici workflow, dall'uso di tool come FFmpeg, all'implementazione di librerie come OpenCV su Python, fino a soluzioni più immediate come, in questo caso, l'utilizzo di DaVinci Resolve.

LAST FRAME

Iterazione 1:

Prompt: "Aggiungi nel wide shot un'altra donna con il cappotto nero, vista di spalle, che entra in scena dalla sinistra".

Input: frame estratto dalla precedente generazione video.

Analisi: per rapidità di esecuzione, il prompt testuale è stato formulato in italiano. Come già esplicitato in precedenza, l'utilizzo di una lingua diversa dall'inglese nell'interazione con il Large Language Model non comporta differenze significative. L'obiettivo era l'inserimento di una nuova figura femminile in primo piano che entrasse in scena durante la clip.

L'immagine ottenuta è poi stata raffinata, con piccoli interventi in Adobe Photoshop. Tra il primo e l'ultimo frame sussistono, inevitabilmente, differenze di risoluzione. Se la prima immagine è una generazione nativa, la seconda è una manipolazione derivata da un fotogramma video in uscita da Kling, con risoluzione 1920x1080 px. Tale discrepanza non inficerà la successiva generazione video. Inoltre, l'eventuale presenza di bande nere non costituisce un problema, poiché la timeline definitiva prevista è in formato 2:35:1, rendendo questa criticità del tutto trascurabile.



Figura 4.11: Primo e ultimo frame Scena 1 Shot 02

VIDEO

- *Modello:* KlingAI 2.6
- *Prompt:* "The child walks through the crowd while holding his mother's hand"
- *Input:* First frame e last frame.
- *Analisi:* nonostante i test effettuati con prompt differenti, la riproposizione del prompt originale della prima generazione video ha garantito i risultati migliori. Sebbene l'impatto fisico tra la donna e il bambino non venga esplicitamente raffigurato, la donna si avvicina camminando fino a occluderlo parzialmente. Questa interazione, pur senza contatto, consente di veicolare efficacemente la dinamica dell'azione attraverso un montaggio mirato, prevedendo un cut opportuno verso l'inquadratura successiva.

SCENA 1 - SHOT 05

FIRST FRAME

Iterazione 1:

Prompt: "Next view: we see this scene from a lateral side view".

Input: Scena 1 Shot 3 First frame

Analisi: in questo caso, la tavola di storyboard è stata utilizzata esclusivamente come riferimento visivo e non come input diretto per la generazione. Il riferimento specifico è stato il first frame della scena 3, in quanto si desiderava ottenere un punto di vista differente mentre il bambino e la madre attraversavano la strada. L'approccio next view o next scene è molto diffuso sia nella generazione di immagini che di video per cercare di ottenere uno shot sequenziale coerente.

Iterazione 2

Prompt: "The child looks toward the left of the image. The teddy bear is

no longer in his arms. His torso is turned toward the left of the image, as if he has lost something".

Input: generazione ottenuta all'iterazione 1

Analisi: a seguito del prompt precedente, abbiamo ottenuto un'inquadratura di profilo della madre e del bambino mentre camminano. Il modello, tuttavia, non è riuscito a ruotare coerentemente anche l'ambiente circostante, evidenziando una difficoltà nel discernere correttamente lo spazio o nel generare viste alternative di un ambiente esteso. Al momento, si ritiene opportuno non concentrarsi sull'ambiente, ma lavorare selettivamente sulla posa dei personaggi. Nello specifico, si desidera ottenere il bambino con il torso ruotato all'indietro, coerentemente con la narrazione post-impatto in cui ha appena perso l'orsacchiotto alle sue spalle.

Iterazione 3

Prompt: "Extract the two subjects in the foreground and put them on a clean white background".

Input: generazione ottenuta all'iterazione 2

Analisi: passaggio preparativo in vista delle prossime lavorazioni. I personaggi sono stati estratti e posizionati su sfondo neutro.

Al fine di integrare anche l'ambiente è stato fatto un altro tentativo utilizzando la tavola di storyboard associata:

Iterazione 4

Prompt: "Re-render this image in a photorealistic way. The child looks back worriedly while crossing the road toward the right, holding his mother's hand. The crowd continues walking from left to right".

Input: tavola di storyboard corrispondente e turn-around del bambino in figura intera.

Analisi: per questo passaggio è stata utilizzata la stessa chat di Gemini impiegata per lo Shot 03; di conseguenza, il modello aveva già in memoria l'aspetto della madre. Sebbene non fosse un risultato esplicitamente atteso, la generazione è riuscita a discernere correttamente anche questo elemento. Lo storyboard è stato replicato decentemente; la posa del bambino risulta leggermente scorretta e in figura intera a causa delle reference di input fornite. Tuttavia, l'ambiente è coerente con il punto di vista ricercato.

Iterazione 5

Prompt: "Remove the people from this image".

Input: immagine generata all'iterazione 4

Analisi: rimuovendo i personaggi principali e la folla dalla generazione precedente, è stato possibile ottenere il solo ambiente disposto correttamente nello spazio.

Iterazione 6

Prompt: "Add the subjects in this background withouth changing the pose. If you need it's better to adapt the background then changing the characters".

Input: immagine generata all'iterazione 3 e immagine generata all'iterazione 5

Analisi: lo scopo è quello di ottenere i personaggi prima ottenuti integrati dentro l'ambiente corretto. Nel caso in cui al modello serva riadattare per ottenere una buona corrispondenza tra personaggi e ambiente, è preferibile che venga modificato l'ambiente.

Iterazione 7

Prompt: "Add the crowd walking from the left to the right and vice versas".

Input: immagine generata all'iterazione 6

Analisi: sebbene l'output non sia ancora ottimale, sia per quanto riguarda la posa che l'integrazione dei personaggi nell'ambiente, uno step successivo utile consiste nel ripristinare la folla di passanti all'interno della strada. L'obiettivo è ottenere una scena completa con tutti gli elementi presenti.

Iterazione 8

Prompt: "The camera is closer to the kid. Do not change the pose".

Input: immagine generata all'iterazione 7

Analisi: nonostante sia stato vantaggioso lavorare con il bambino in figura intera, è ora fondamentale ottenere la corretta inquadratura cinematografica. Con l'avvicinamento della camera, il modello ha apportato una leggera modifica alla posa del bambino, migliorandola. Sebbene non corrisponda esattamente allo storyboard, la nuova posa è ottimale per veicolare l'intenzione narrativa della scena.

Iterazione 9

Prompt: "The kid's face has to be this one".

Input: immagine generata all'iterazione 8 e turn-around del bambino in primo piano

Analisi: definita l'inquadratura, è fondamentale verificare la coerenza visiva tra i vari shot, assicurandosi che il volto del bambino corrisponda a quello originale.



Figura 4.12: Step di generazione per la prima Still della Scena 1 Shot 05

Sulla base della still finale ottenuta sono stati testati diversi prompt testuali in generazione video. Per citarne alcuni:

- *Prompt:* "Static camera: the child walks through the crowd holding his mother's hand while looking back";
- *Prompt:* "A young boy in a blue puffer jacket and beige beanie walks hand-in-hand with his mother through a bustling city street. The child repeatedly glances backward over his shoulder while maintaining forward movement, his curious expression contrasting with the mother's steady gaze ahead. The camera remains stationary as pedestrians flow around them, cold breath visible in the winter air";
- *Prompt:* "A young boy in a blue puffer jacket and beige beanie walks hand-in-hand with his mother through a bustling city street. The child repeatedly glances backward over his shoulder while maintaining forward movement, his curious eyes scanning the crowd behind them. The tracking camera follows their progress as pedestrians flow in both directions, cold winter breath visible in the air.";

Nessuno di queste iterazione, tuttavia restituiva un movimento che fosse adatto alla sequenza narrativa, facendo emergere la necessità di utilizzare anche un last frame. Ai fini della continuità dell'azione, si è rivelato più vantaggioso definire un punto di partenza antecedente piuttosto che generare un punto di arrivo successivo. Pertanto, basandosi sulla still precedente, è stato generato

un previous shot, con l'accorgimento di scambiare il first frame con il last frame in generazione video.

LAST FRAME

Iterazione 1:

Prompt: "Il bambino guarda in avanti verso il lato destro del frame mentre cammina tenendo per mano la madre".

Input: First frame Scena 1 Shot 05.

Analisi: dovendo ottenere un'azione precedente a quella della reference utilizzata, il bambino non può essere girato con il torso verso dietro, ma rivolto ancora parzialmente in avanti nella direzione di camminata. Lo scopo è far ruotare il corpo in animazione. In questo caso il prompt è stato formulato in italiano, puramente per comodità.

Iterazione 2:

Prompt: "Sposta il bambino e la madre più a sinistra mantenendo la loro posa".

Input: immagine generata all'iterazione 2.

Analisi: la precedente generazione aveva già modificato i passanti sullo sfondo, minimizzando la problematica di loop anomali durante la generazione video successiva. Di necessaria importanza riposizionare verso la destra del frame la madre e il bambino, così da consentire loro di effettuare alcuni passi in avanti prima della rotazione del torso. Per ottenere una cinematica più organica, sarebbe stato opportuno far percorrere ai personaggi una distanza maggiore prima del last frame. Di conseguenza, è stata valutata l'ipotesi di impiegare un first frame privo dei soggetti, con l'intento di introdurli dinamicamente in scena durante l'animazione. Tuttavia, l'assenza di un riferimento iniziale fa sì che il modello video tenda a perdere i riferimenti strutturali dei soggetti e generi artefatti di morphing durante la progressione della clip; questa soluzione è stata pertanto scartata.



Figura 4.13: Primo e ultimo frame Scena 1 Shot 05

VIDEO

- *Modello:* KlingAI 2.6
- *Prompt:* "The child walks through the crowd while holding his mother's hand, then looks back."
- *Input:* Last frame e First frame (invertiti rispetto all'ordine in cui sono stati generati).
- *Analisi:* un prompt semplice spesso si dimostra più efficace di prompt complessi. La generazione ottenuta ha restituito una prima, breve, camminata iniziale e un movimento successivo che segue la rotazione del corpo del bambino.

SCENA 1 - SHOT 07

FIRST FRAME

Iterazione 1:

Prompt: "Re-render this image in a photorealistic way. The teddy bear is on the ground, on the crosswalk. The crowd continues to walk around it. We see their feet. Use the atmosphere and style from the third image".

Input: tavola di storyboard corrispondente, asset dell'orsacchiotto in turn-around, First frame Scena 1 Shot 03

Analisi: in questo caso, la tavola di storyboard era molto semplice e priva di elementi specifici. È stata fornita per offrire al modello un riferimento visivo generale sull'azione da compiere, piuttosto che come una vera e propria guida per la composizione finale. Per quanto riguarda l'ambiente, è stata fornita la generazione finale dello Shot 03, così da consentire a Nano Bana di discernere autonomamente lo spazio, le condizioni di luce e i colori, riadattandoli per la scena da ricreare.

Iterazione 2:

Prompt: "Put the camera closer to the teddy bear. Create a more interesting diagonal to see it".

Input: immagine ottenuta dalla prima iterazione

Analisi: il risultato ottenuto in precedenza era tecnicamente valido, ma noioso e statico dal punto di vista artistico. Piuttosto che fornire un'ulteriore reference è stato chiesto al modello di restituire una risposta creativa.

Iterazione 3:

Prompt: "The crosswalk is directed in the wrong direction. They should follow the direction of the people walking. Change it".

Input: immagine ottenuta dall'iterazione 2.

Analisi: l'output restituito in precedenza presenta una composizione interessante e di impatto, con un'angolazione della camera non perpendicolare all'asse stradale. Si riscontra però un errore di semantica spaziale nella generazione della texture delle strisce pedonali. L'andamento delle linee è stato orientato a 90 gradi rispetto al fisiologico verso di percorrenza. Durante quest'iterazione si cerca dunque di risolvere il problema.

Iterazione 4:

Prompt: "Remove the teddybear and the people from the image".

Input: immagine ottenuta dall'iterazione 2.

Analisi: l'iterazione 3 ha replicato fedelmente l'output dell'iterazione 2, ignorando la nuova istruzione. Ci troviamo di fronte a uno scenario in cui il modello mostra difficoltà nell'interpretare correttamente la richiesta; la sfida operativa consiste nell'aggirare tali limitazioni ridefinendo la strategia comunicativa o sperimentando approcci alternativi. Analogamente a quanto fatto per lo Shot 05, disponendo di una composizione valida ma di un ambiente errato, si è deciso di separare i due elementi per procedere con una lavorazione più selettiva.

Iterazione 5:

Prompt: "The crosswalks in front must be in the same direction as the ones in the back".

Input: immagine ottenuta dall'iterazione 4.

Analisi: in assenza di altri elementi in scena, il modello è riuscito con precisione a sistemare il pattern delle strisce pedonali.

Iterazione 6:

Prompt: "Add the teddy bear and the walking crowd".

Input: immagine ottenuta dall'iterazione 5.

Analisi: provando a ricollocare l'orsacchiotto e la folla generati in precedenza (essendo nella stessa chat Gemini il modello ha un contesto) sul corretto ambiente, la generazione fallisce in ogni caso, rimodificando in

maniera anomala le strisce pedonali e vanifonda la miglioria appoarta nella precedente generazione.

In aggiunta ai prompt indicati, sono stati fatti altri tentativi, provando approcci sempre diversi come quello di scrivere manualmente sull'immagine per indicare al modello quali strisce specificatamente dovessero essere bianche a quali scure. Ogni risultato, però, continuava a presentare il problema.



Figura 4.14: Alcuni step di generazione falliti - Scena 1 Shot 07

Questo scenario rappresenta un caso studio significativo e spesso frequente, in cui i modelli generativi mostrano limiti nel progredire verso il risultato. Di conseguenza, nella realizzazione delle stills, il ricorso a tool di post-produzione come Adobe Photoshop diventa indispensabile.

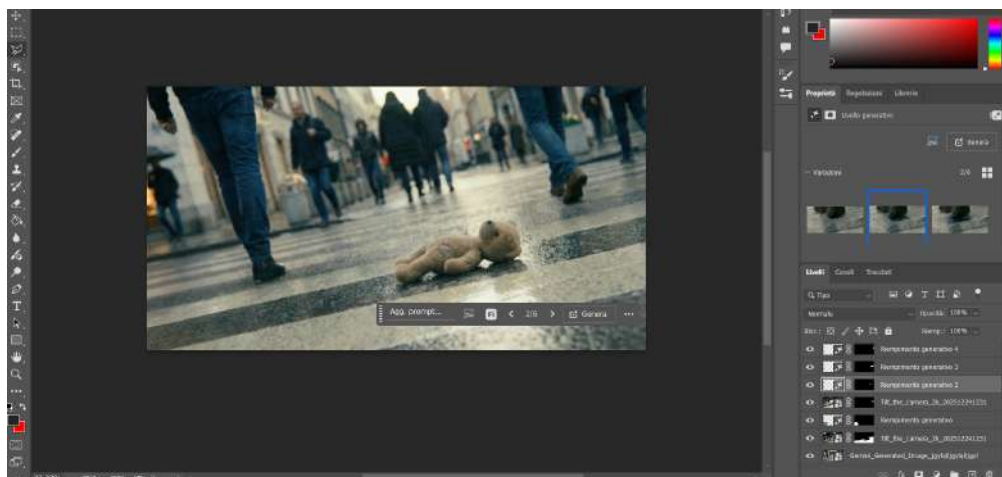


Figura 4.15: Lavorazioni in Adobe Photoshop - Scena 1 Shot 07

Gli output generativi più promettenti sono stati importati come livelli indipendenti, includendo il clean plate dell'ambiente derivante dall'iterazione

4, per procedere a un intervento intensivo di Digital Matte Painting (DMP) e alla rettifica manuale delle incongruenze. Nello specifico, la funzionalità 'Generative Fill' si è rivelata cruciale per il refining dei micro-dettagli e per garantire la coerenza visiva, permettendo di generare variazioni localizzate esclusivamente sulle aree selezionate (con strumenti di selezione come il 'lazo')

VIDEO

- *Modello*: KlingAI 2.6
- *Prompt 1*: "A static wide shot observes a weathered teddy bear abandoned at a pedestrian crossing while the crowd passes by"
- *Prompt 2*: "The camera slowly zooms in on the teddy bear as the crowd passes by".
- *Input*: First frame.
- *Analisi*: in questo caso, per la generazione richieste non è servito generare anche un last frame. Sono stati utilizzati due prompt, così da avere due alternative in fase di montaggio, come se fossero due take diversi di girato. Il primo è a camera statica, mentre il secondo presenta un lento push-in con la camera che ruota leggermente riportandosi in asse con l'asfalto. In fase di montaggio è stato poi prediletto il secondo.

SCENA 1 - SHOT 11

FIRST FRAME

Iterazione 1:

Prompt: "Add this woman to the background of the reference image in place of the character drawn in red. Integrate the woman according to the lighting and color conditions of the reference image".

Input: Come input è stato usato il turn-around della donna in figura intera e una porzione di strada generata in precedenza per un altro shot, a cui è stato aggiunto un disegno di un omino per indicare a larghe linee dove posizionare la donna

Analisi: lo scopo è ottenere un primo piano della donna che reagisce al bambino in pericolo. La strategia è quella di posizionare inizialmente la donna correttamente all'interno dello spazio, e poi andare a stringere il campo.

Iterazione 2:

Prompt: "Close up of this woman".

Input: generazione ottenuta all'iterazione 1.

Analisi: stringendo il campo e venendo da una figura intera, il modello tende a reinventare il volto della donna.

Iterazione 3:

Prompt: "The face of the woman should look like this one. She is panicking and staring slightly on her right. ".

Input: generazione ottenuta all'iterazione 2 e turn-around della donna in primo piano

Analisi: nonostante il volto fornito in input corrisponda a quello definito nella fase di character design, l'integrazione all'interno della scena ha generato un'identità visiva differente. È plausibile che le istruzioni aggiuntive fornite in questa iterazione — nello specifico la rotazione del capo e l'espressione preoccupata — abbiano condizionato l'elaborazione del modello. Ciononostante, il nuovo aspetto risulta credibile e coerente con il ruolo della madre. Trattandosi della prima generazione che ritrae il personaggio in questa specifica inquadratura, si è deciso di consolidare questa nuova estetica e di mantenerla per le produzioni future.

Iterazione 4:

Prompt: "The woman is looking almost in the centre of the camera".

Input: generazione ottenuta all'iterazione 3

Analisi: in questo caso si è deciso di avere un'altra versione da usare in generazione video in cui la donna guarda centralmente oltre la camera, da testare come possibile alternativa come first frame in sostituzione alla still prodotta durante l'iterazione 3, o eventualmente come last frame.

Iterazione 5:

Prompt: "Set diffused lighting, ensuring the light source is not positioned behind the woman".

Input: generazione ottenuta all'iterazione 3

Analisi: un problema dell'iterazione 3 è la sorgente di luce non coerente con le precedenti scene. Si è cercato quindi di cambiarne la posizione e renderla più diffusa. Una delle difficoltà di lavorare con un ambiente non 3D è appunto quella di poter riposizionare la sorgente a piacimento. Al fine di versionare in maniera più efficace, in questo caso è stato utilizzato Nano Banana su Flow. La stessa operazione è stata eseguita per la generazione ottenuta all'iterazione 4.



Figura 4.16: Step di generazione - Scena 1 Shot 11



Figura 4.17: Versioning su Flow - Scena 1 Shot 11

VIDEO

- *Modello:* KlingAI 2.6
- *Prompt 1:* "The woman looks to her right, then looks straight ahead"
- *Prompt 2:* "The woman stares at a fixed point with a terrified look, takes a breath, and then looks straight ahead".
- *Prompt 3:* "The terrified woman stares at a fixed point for a few seconds, motionless, then looks straight ahead with concern".

- *Input*: iterazione 3 e iterazione 4 dopo il fix della luce su Flow.
- *Analisi*: anche in questo caso è stato preferito provare diversi prompt così da avere qualche take in più in fase di montaggio.

4.2.2 Generazione della seconda sequenza narrativa

La seconda sequenza narrativa, ha rappresentato una sfida tecnica la cui complessità principale risiede nelle coreografie d'azione tra il pugile e i robot: un dinamismo elevato che l'intelligenza artificiale non sempre riesce a restituire con efficacia e fluidità naturale. I limiti tecnici imposti dalla durata fissa delle clip (tra i 5 e i 10 secondi) hanno condizionato pesantemente la resa delle scene. Anche adottando tecniche di first e last frame, è risultato difficile ottenere azioni rapide e di grande impatto, poiché il modello tendeva a spalmare il movimento sull'intero arco temporale, generando spesso un effetto slow motion non voluto che smorzava l'energia del combattimento (si specifica come gli aggiornamenti che permettono una gestione granulare della durata delle clip siano stati rilasciati solo in una fase avanzata del progetto). Proprio per la natura altamente creativa di questa sequenza, inoltre, è stato pensato un approccio più ibrido, dove lo storyboard passa in secondo piano e viene delegato al modello l'esplorazione di soluzioni visive autonome. Questo ha permesso di scoprire traiettorie e angolazioni inedite, nate dalla "spontaneità" dell'algoritmo. Un'altra complessità di questa sequenza è infine dovuta a un passaggio ponderato tra lo stile fotorealistico e quello cartoon, richiedendo un approccio Video-To-Video che concretizzasse il trasferimento di stile voluto (Style Transfer).

SCENA 05 - SHOT 04

FIRST FRAME

Iterazione 1:

Prompt: "Siamo dentro questo mondo, inquadra dal basso verso l'alto la passerella che conduce al ring".

Input: come input è stato fornito l'asset del ring precedentemente realizzato.

Analisi: il testo è stato formulato in italiano per rapidità di esecuzione essendo all'interno di Gemini. Lo scopo è quello di ottenere una prospettiva del ring più in linea con l'inquadratura finale che si vuole ottenere.

Iterazione 2:

Prompt: "Aggiungi a questo pugile una vestaglia da ring rosso scuro, con il bordo oro".

Input: turn-around del pugile in figura intera.

Analisi: lo scopo è ottenere lo stesso pugile definito in fase di asset, che indossa una vestaglia nel momento del suo ingresso sul ring.

Iterazione 3:

Prompt: "Alza il cappuccio".

Input: generazione ottenuta all'iterazione 2

Analisi: il cappuccio deve essere alzato in preparazione allo shot voluto.

Iterazione 4:

Prompt: "Questo pugile si trova di spalle vicino la camera ad inizio passerella, rivolto verso il ring".

Input: generazioni ottenute all'iterazione 1 e 3

Analisi: lo scopo è posizionare il pugile all'interno dell'ambiente.

Iterazione 5:

Prompt: "Genera un'altro punto di vista del pugile sulla passerella che conduce al ring. E' visto davanti a 3/4, dal basso verso l'alto".

Input: generazioni ottenute all'iterazione 3 e 4

Analisi: la generazione ottenuta all'iterazione 4 è stata volutamente fatta di spalle così da avere un riferimento per diversi shot. Al fine di ottenere la specifica inquadratura voluta è necessario vedere il pugile sulla passerella dal davanti.

Iterazione 6:

Prompt: "Primo piano del pugile mantenendo la stessa angolazione di camera".

Input: generazione ottenuta all'iterazione 5

Analisi: avendo ottenuto un'angolazione di camera valida, necessitando per questo shot di un primo piano, si è proceduto ad avvicinare la camera verso il pugile.

Iterazione 7:

Prompt: "Primo piano strettissimo del volto del pugile mantenendo la stessa angolazione di camera".

Input: generazione ottenuta all'iterazione 6

Analisi: la generazione in uscita dall'iterazione 5 aveva la camera molto distante dal pugile, motivo per cui, al fine di ottenere un primo piano è stato necessario fare un passaggio in più di avvicinamento della camera.

Iterazione 8:

Prompt: "Generate nine different points of view of this boxer's close-up while he looks slightly upward.".

Input: generazione ottenuta all'iterazione 5

Analisi: Il primo piano ottenuto durante l'iterazione 7 presentava delle criticità a livello prospettico, a causa di una rotazione involontaria della

camera introdotta dal modello durante la fase di avvicinamento al soggetto. Invece di forzare complesse correzioni spaziali su quello specifico shot, si è ritenuto più efficace adottare un approccio retroattivo, piuttosto che procedere in maniera lineare: si è tornati all'output dell'iterazione 5 (a figura intera) utilizzandolo come base per generare nuove varianti tramite Flow. Il prompt adottato sfrutta una strategia consolidata che delega al modello l'autonomia compositiva dell'inquadratura, producendo così 9 punti di vista alternativi. Questa specifica dinamica procedurale rende la lavorazione di questa sequenza, pur trattandosi di una scena statica, un eccellente caso di studio. Poiché le operazioni si sono svolte su Flow, il prompt è stato formulato in lingua inglese. Per massimizzare le opzioni a disposizione, è stato eseguito un processo in batch da 4 generazioni; in questo modo, moltiplicando i 9 punti di vista, si è ottenuto un ventaglio complessivo di 36 alternative tra cui selezionare il primo piano definitivo.



Figura 4.18: Step di generazione - Scena 05 Shot 04

Iterazione 9:

Prompt: "Estrai l'immagine centrale a sinistra dal panel di immagini. Mantienila in 16:9, non generare niente di nuovo, estrai solo quello che hai già generato".

Input: generazione ottenuta all'iterazione 8

Analisi: dal batch di immagini è stata scelta quella contenente l'inquadratura voluta. Per estrarla è stata riportata su Gemini, chiedendo esplicitamente di selezionare univocamente quella. La richiesta di non generare niente di nuovo è dovuta al fatto che alcune volte il modello tende a reinterpretare e modificare l'immagine durante la fase di estrazione.

Iterazione 10:

Prompt: "Put the man from the second image in the same pose of the

man in the first image. Keep the same atmosphere, lighting, colors and composition from the first image".

Input: generazione ottenuta all'iterazione 9 e turn-around del pugile in primo piano

Analisi: dato che il workflow ha previsto l'ottenimento di un primo piano a partire da una figura intera, il viso del pugile non corrisponde più a quello determinato in fase di character design. Diventa necessario fare un'operazione di sostituzione del volto (Face Swap) mantenendo tutto il resto identico.

LAST FRAME

Iterazione 11:

Prompt: "Cambia l'espressione di quest'uomo. Strizza gli occhi per vedere meglio. Mantieni la stessa posa".

Input: generazione ottenuta all'iterazione 10

Analisi: al fine di ottenere un frame di chiusura per l'animazione, si è tornati su Gemini per modificare l'espressione del viso.



Figura 4.19: Face swap e cambio dell'espressione facciale - Scena 05 Shot 04

L'intero workflow dimostra l'efficacia di un approccio ibrido, che sfrutta la versatilità di Gemini nelle fasi di impostazione e le capacità di versioning di Flow. Molto spesso, partendo da un medesimo asset di base, risulta strategico impiegare i due ambienti operativi in parallelo. Questa sinergia permette di esplorare simultaneamente molteplici percorsi generativi, ottimizzando i tempi di esecuzione e massimizzando l'efficienza nel raggiungimento del risultato finale.

VIDEO

- *Modello:* Veo 3.1
- *Prompt:* "The man squints, looking ahead with curiosity".

- *Input*: immagine generata all'iterazione 10 come First Frame, immagine generata all'iterazione 11 come Last Frame.
- *Analisi*: l'impiego di un first frame e di un last frame fortemente simili, seppur con leggere variazioni, applicati a un'espressione facciale prevalentemente statica, rappresenta una precisa strategia produttiva. Sebbene l'output di 8 secondi generato dal modello Veo possa non mantenere una perfezione visiva costante nei fotogrammi intermedi, questo approccio garantisce un'elevata stabilità all'inizio e alla fine della clip. In questo modo si ottengono delle 'maniglie' (margini operativi) di alta qualità, fondamentali per effettuare successivi tagli (cut) di precisione in fase di montaggio sulla timeline.

SCENA 05 - SHOT 06

FIRST FRAME

Iterazione 1:

Prompt: "Mettili 3 di questi robot dentro il ring, rivolti verso la passerella".

Input: turn-around del robot in versione non minacciosa, output dell'iterazione 1 di Scena 05 - Shot 04.

Analisi: lo scopo è stato posizionare l'asset del robot all'interno del ring, con l'inquadratura corretta durante l'inizio del processo di generazione del precedente shot analizzato.

Al fine di ottenere un'inquadratura che riportasse nell'ambiente sia i robot che il pugile, correttamente integrati è stato fatto un lavoro di Digital Matte Painting su Adobe Photoshop, per rimpicciolire i robot, riposizionarli e aggiungere il pugile correttamente aggiunto durante l'iterazione 2 della Scena 05 - Shot 04.



Figura 4.20: Step di generazione preliminari - Scena 05 Shot 06

Questa still non è un punto d'arrivo fatto e compiuto, ma un passaggio preliminare necessario, come già specificato in precedenza, per intraprendere le lavorazioni per diversi shot di questa scena.

Iterazione 2:

Prompt: "Give me nine new angles of this scene".

Input: input ottenuto al precedente passaggio dopo le lavorazioni in Adobe Photoshop.

Analisi: si lascia al modello la delega di creare nuove angolazioni, utilizzando nuovamente Flow, così da avere un numero significativo di alternative. L'immagine voluta, così come fatto in precedenza, viene poi estratta dal panel di immagini.

Iterazione 3:

Prompt: "Generate nine different points of view of the robot in this scene".

Input: immagine selezionata all'iterazione 2.

Analisi: l'immagine precedente selezionata, in maniera iterativa, è stata sorgente di un nuovo panel di immagini. L'angolazione scelta è stata poi estratta dal panel.

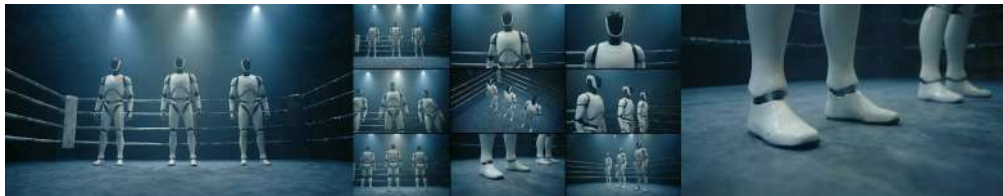


Figura 4.21: Step di generazione iterativi - Scena 05 Shot 06

Questa metodologia di lavoro non è soltanto tecnica ma anche artistica, un modo per esplorare possibilità e trovare ispirazione per nuove idee. In questo caso, l'intenzione di questa scena è quello di raffigurare la metamorfosi minacciosa dei robot a seguito dell'agire dell'effetto particellare. Essendo una transizione molto complessa da fare in maniera integrale, la precedente generazione ha suggerito come si potesse effettuare questo passaggio su un'inquadratura più stretta, come i piedi del robot, per poi riprendere con la loro camminata mentre l'effetto sparisce nella successiva inquadratura, semplificando la difficoltà tecnica di generazione e rappresentando comunque l'idea visiva ricercata.

Iterazione 4:

Prompt: "I piedi in figura devono appartenere a questo robot".

Input: input estratto dall'iterazione 3 e turn around del robot minaccioso in figura intera.

Analisi: si è proceduto a sostituire i piedi in preparazione alla transizione tra le due immagini. Le due still sono anche state controllate su Photoshop per accertarsi che combaciassero in posizione, così da avere poi un morphing il più fluido possibile.

Iterazione 5:

Prompt: "aggiungi l'effetto particellare presente sull'orsacchiotto ai piedi

del robot".

Input: immagine estratta all'iterazione 4, reference dell'orsacchiotto con l'effetto particellare applicato.

Analisi: in precedenza, l'effetto particellare era stato aggiunto sull'orsacchiotto per un'inquadratura della scena 3. Il prompt utilizzato era stato "aggiungi un effetto particellare di pixel trasparenti, colorati di viola, azzurro, verde acqua all'orsacchiotto". Il prompt era stato poi iterato fino all'ottenimento della resa visiva voluta. Anche sui piedi del robot, il prompt è stato iterato fino ad avere l'intensità corretta. Lo stesso passaggio è stato fatto per l'immagine selezionata all'iterazione 3.



Figura 4.22: Creazione del First Frame e del Last Frame - Scena 05 Shot 06

VIDEO

- *Modello:* Kling 2.6
- *Prompt:* "The pixelated particle effect entirely covers the legs, then, drifting slightly upward, it dissipates into the air, revealing the new legs".
- *Input:* First frame e Last Frame generati all'iterazione 5.
- *Analisi:* per questa generazione sono stati testati sia Veo 3.1 sia Kling. Soltanto quest'ultimo ha restituito un risultato efficace nella gestione della fisica delle particelle.

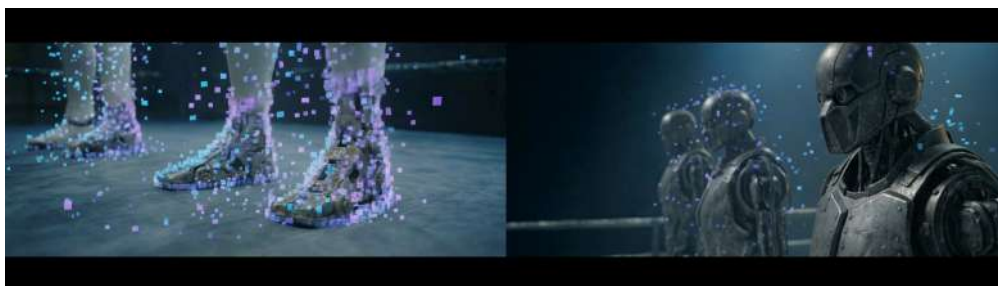


Figura 4.23: Sequenza narrativa - Scena 05 Shot 06/ Shot 07

SCENA 08 - SHOT 08

La Scena 08 è incentrata sul combattimento tra il pugile e il robot. In questa specifica sequenza narrativa, a differenza degli shot precedenti, il focus metodologico si sposta dalle immagini statiche (still) alla generazione video. L'approccio iniziale ha previsto l'impiego di immagini come first e last frame, sebbene in alcune fasi di test si sia provato a operare con il solo first frame. Tuttavia, questa seconda casistica ha evidenziato limiti strutturali notevoli: in assenza di un fotogramma finale di riferimento, le azioni rapide e la frequente presenza di occlusioni visive hanno comportato una marcata perdita di coerenza visiva, specialmente nei volti e nell'abbigliamento. Inoltre, i vincoli temporali imposti dai software (dai 5 ai 10 secondi su Kling e 8 secondi su Veo) rendevano complessa la resa fluida di movimenti così concitati, restituendo output non ottimali. Per ovviare a queste criticità, si è adottata una tecnica di campionamento: i frame più riusciti di queste prime clip, caratterizzati da pose dinamiche interessanti, sono stati isolati ed estratti (utilizzando DaVinci Resolve) per essere riutilizzati come nuovi first e last frame per le generazioni successive. Si è venuto così a creare un processo iterativo, in cui la direzione creativa umana lavora in sinergia con le alternative probabilistiche fornite dal modello.



Figura 4.24: First e last frame - Scena 08 Shot 08

VIDEO

- *Modello:* Kling 2.6
- *Prompt:* " ".
- *Input:* First frame e Last Frame.
- *Analisi:* in questo caso specifico, si è scelto di generare il video senza l'ausilio di un prompt testuale. A differenza di Veo, Kling permette infatti di creare animazioni basandosi esclusivamente su riferimenti visivi, come i frame di inizio e fine. Sebbene nell'economia dell'intero prodotto audiovisivo questa rappresenti un'eccezione, tale metodologia può costituire una

valida soluzione quando l'interpolazione tra i due stati risulta intuitiva per il modello. Qui, tuttavia, i tentativi di descrivere testualmente una scena particolarmente caotica e priva di una chiara direzione di movimento non avevano prodotto output soddisfacenti. Affidarsi all'interpretazione autonoma dell'IA è stato quindi un esperimento proficuo. Il risultato non solo si è rivelato coerente, ma ha integrato un movimento di macchina sofisticato che porta la clip ad essere una delle più efficaci dell'intera sequenza.

VIDEO-TO-VIDEO

- *Modello*: Kling Omni Video O1
- *Prompt*: Use Image as the first frame. Animate Image with the same motion as in the Video.
- *Input*: Clip generata al passo precedente e primo frame lavarato con Nano Banana per cambiare lo stile grafico da foto realistico a cartoon.
- *Analisi*: per la generazione delle clip in stile cartoon, con l'obiettivo di assicurarne la perfetta sincronia con le riprese fotorealistiche e consentire transizioni di tipo match cut in qualsiasi punto della sequenza, si è ricorsi a una tecnica video-to-video. Operando all'interno di Kling Omni, tramite la funzione Transformation, è possibile combinare un video di riferimento con un'immagine statica. La procedura ottimale prevede l'estrazione del primo fotogramma (first frame) dal filmato originario e la sua successiva rielaborazione stilistica. In fase di generazione, l'utente può scegliere se applicare lo stile globale al video oppure se animare l'immagine caricata sfruttando i dati di movimento del video di riferimento. Prediligendo questa seconda strategia, è stato possibile ottenere in output un duplicato esatto dell'azione originale per dinamica e timing, ma declinato in stile cartoon. Al fine di ottimizzare i tempi e i costi computazionali, il first frame non è stato estrapolato dalla clip grezza, bensì direttamente dalla porzione di video già ritagliata in timeline (cut definitivo), riducendo così la durata complessiva della nuova generazione.

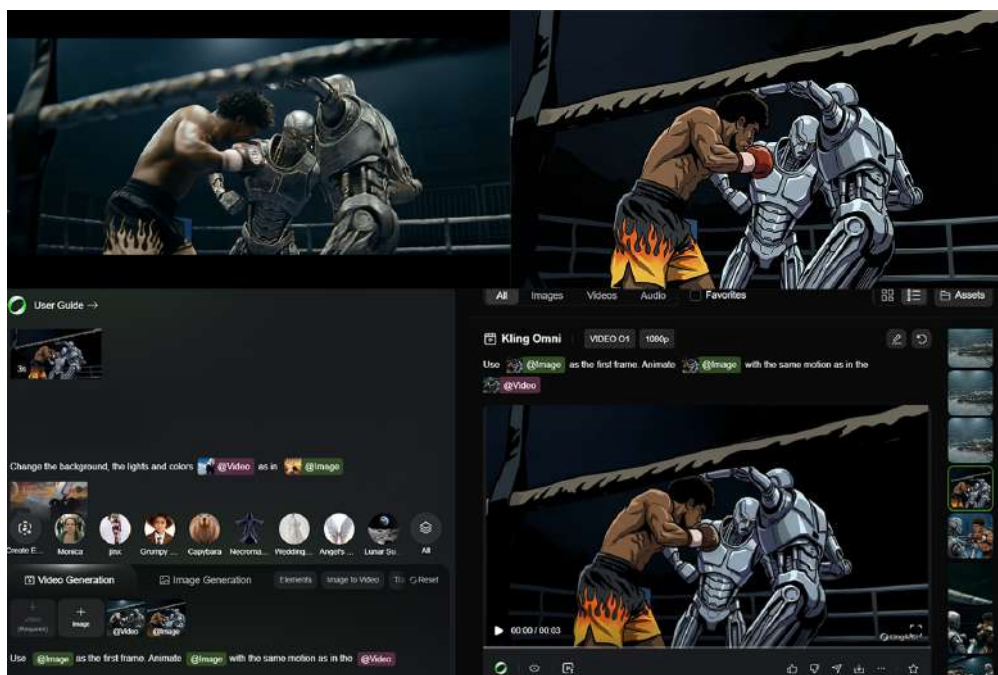


Figura 4.25: Video-To-Video - Scena 08 Shot 08

4.2.3 Generazione della terza sequenza narrativa

La terza sequenza narrativa è quella con meno criticità dal punto di vista tecnico, ma forse la più affascinante dal punto di vista visivo. La vera difficoltà risiede nell'accordo con il finale, dove sono previsti diversi matchcut che accompagneranno la sequenza dello spazio verso la scena finale.

SCENA 06 - SHOT 05

FIRST FRAME

Iterazione 1:

Prompt: "The astronaut's face, seen from behind in a three quarter view on the side of the image, looks through a very large porthole at planet Earth from space."

Input: turn-around dell'astronauta in primo piano.

Analisi: in questa fase, si è scelto di non fornire al modello ulteriori vincoli spaziali oltre al personaggio definito nel character design. Piuttosto che imporre un ambiente predefinito, è apparso metodologicamente più efficace lasciare al modello la libertà di generare il contesto, e valutare la costruzione della composizione del frame restituita in output.

Iterazione 2:

Prompt: "Reverse shot: from outside the porthole, we see a three-quarter extreme close-up of the astronaut looking nostalgically past the camera. A beautiful blue reflected light shines on him, coming from the Earth, which is not visible in this shot".

Input: generazione ottenuta all'iterazione 1

Analisi: l'inquadratura precedentemente generata riprende l'astronauta mentre guarda la terra nostalgico. Ai fini della sequenza narrativa, si tenta di avere un'altra angolazione dove l'astronauta viene ripreso da fuori la navicella.

Iterazione 3:

Prompt: "The camera should be positioned closer to the man's face, located inside the porthole".

Input: generazione ottenuta all'iterazione 2

Analisi: la precedente generazione non convinceva, anche se l'angolazione di camera sembrava essere corretta. Si è quindi proceduto a generare uno shot simile con la camera dentro la navicella spaziale



Figura 4.26: Still - Scena 06 Shot 05

VIDEO

- *Modello:* Veo 3.1
- *Prompt:* "The astronaut looks nostalgically at a fixed point in front of him and lets out a sigh, while making natural micro-movements with his body".

- *Input*: First frame, ripetuto come Last Frame.
- *Analisi*: in assenza di dinamiche di movimento particolarmente complesse, una prassi metodologica consolidata consiste nel far coincidere il first frame con il last frame. In tal modo, il soggetto compie l'azione per poi ricongiungersi alla posa di partenza. Si tratta di un espediente tecnico volto a mitigare il rischio di perdita di coerenza visiva durante il processo generativo. Fornendo al modello un doppio ancoraggio temporale, si ottimizza la stabilità morfologica del personaggio.

SCENA 06 - SHOT 01/02/03

STILL

Iterazione 1:

Prompt: "Shuttle in partenza al crepuscolo, blue hour".

Analisi: si vuole immortalare il momento in cui lo shuttle sta per essere lanciato nello spazio.

Iterazione 2:

Prompt: "Lo shuttle deve ancora partire".

Input: immagine generata all'iterazione 1.

Analisi: con questa semplice modifica si otterrà un first frame da usare in combinazione con la generazione precedente, usata da last frame, per ottenere in animazione l'accensione dei motori e la partenza dello shuttle.

Iterazione 3:

Prompt: "Genera la vista dall'alto di questa scena. La camera è posizionata sopra lo shuttle e guarda in giù".

Input: immagine generata all'iterazione 2.

Analisi: lo shuttle, non ancora partito, viene visto da un'altra angolazione.

VIDEO

Scena 06 Shot 01

- *Modello*: Veo 3.1
- *Prompt 1*: "The camera slowly pushes in toward the shuttle".
- *Prompt 2*: "The camera slowly zooms in toward the shuttle, almost in slow motion".
- *Prompt 3*: "The camera slowly approaches the shuttle, maintaining a certain distance".
- *Input*: still ottenuta all'iterazione 3.

- *Analisi*: essendo l'unica animazione il movimento di camera e non essendoci particolari azioni, sono stati testati più prompt, così da avere più take eventuali in fase di montaggio.

Scena 06 Shot 02

- *Modello*: Veo 3.1 usato su Gemini
- *Prompt*: "Voglio vedere il propulsore del razzo visto da sotto. Deve occupare l'intera inquadratura, voglio proprio vedere il tubo di scappamento, da cui esce il fuoco".
- *Input*: still ottenuta all'iterazione 3.
- *Analisi*: questa scena rappresentando un caso interessante. Essendo un fegatello riempitivo ottenuto da una delle ultime batch di generazione, è stato possibile testare la funzionalità introdotta in chat su Gemini di poter generare video. Il risultato ottenuto è stato soddisfacente.

Scena 06 Shot 03

- *Modello*: Veo 3.1
- *Prompt*: "The shuttle is ready for takeoff; the engines ignite, smoke begins to billow out, and then the shuttle takes flight."
- *Input*: still ottenuta all'iterazione 2 come First Frame, still ottenuta all'iterazione 3 come Last Frame.
- *Analisi*: animazione dello shuttle che viene lanciato verso lo spazio.



Figura 4.27: Sequenza narrativa Shot 06 - Shot 01/02/03

SCENA 03 - SHOT 01/02

STILL

Iterazione 1:

Prompt: "Siamo dentro il mondo della prima immagine, voglio vedere gli

astronauti della seconda immagine camminare verso lo shuttle visti di spalle".

Input: still di partenza dello Scena 06 shot 01 e turn-around dell'astronauta in figura intera.

Analisi: lo scopo è raffigurare due astronauti di spalle che camminano verso lo shuttle, pronti per il viaggio.

VIDEO

- *Modello:* Kling 3.0 Multi-Shot.
- *Prompt 1:* "The two astronauts walk toward the rocket as the camera tracks them, keeping a constant distance".
- *Prompt 2:* "Extreme close-up of an astronaut's hand gliding along the handrail as he walks".
- *Input:* Still prima ottenuta come First Frame.
- *Analisi:* in questo caso, i due prompt testuali non sono due alternative, ma usando la modalità Multi-Shot di Kling 3.0 è possibile concatenare all'interno di una stessa generazione, due generazioni successive, utilizzando un'unica immagine di reference. Sarà il modello a dedurre il contesto in cui generare la seconda clip. Un'altra funzionalità significativa è quella di poter scegliere in maniera granulare la durata delle clip. In questo caso è stato assegnato un tempo di 4s per l'animazione dovuta al primo prompt, e 2s per quella dovuta al secondo prompt.



Figura 4.28: Generazione video in Multi-Shot Shot 03 - Shot 01/02

4.2.4 Generazione della sequenza finale

Per chiudere la narrazione, una sequenza di match cut segna la transizione dal viaggio spaziale alla vignetta di un albo a fumetti, tenuto tra le mani dal bambino introdotto nell'incipit. Questo espediente di montaggio svela la reale natura diegetica degli eventi: l'intera sequenza a cui lo spettatore ha assistito viene retroattivamente ricontestualizzata come la trasposizione visiva dell'immaginazione infantile, innescata dalla lettura delle tavole.

Per realizzare tecnicamente questa sequenza sono utilizzati diversi approcci. La scena del viaggio spaziale, che ritrae lo shuttle in transito attraverso diversi scenari, è stata costruita in modo sequenziale. A partire dalla generazione di un primo fotogramma statico (still), è stata elaborata la prima clip video. Una volta importata in timeline, la clip è stata tagliata nel punto esatto previsto per il match cut con l'ambientazione successiva. L'ultimo fotogramma utile è stato quindi estratto ed elaborato tramite Nano Banana. Questa operazione ha permesso di mantenere inalterata la posizione e la morfologia della navicella rispetto al frame finale della clip precedente. Gli interventi di modifica si sono concentrati esclusivamente sull'ambientazione e sulla luce incidente sul veicolo, al fine di garantirne una perfetta integrazione cromatica e luminosa nel nuovo scenario. Generando la clip video successiva a partire da questa nuova immagine, è stato possibile ottenere una continuazione fluida dell'azione. Viene così garantita una rigorosa continuità visiva grazie alla costanza dello shuttle, mutando al contempo in modo dinamico l'ambiente circostante.

In preparazione alla transizione all'albo a fumetti, dall'ultima clip è stato estratto il fotogramma iniziale. Questa immagine è stata successivamente rielaborata all'interno di Nano Banana per conferirle un'estetica cartoon. Una volta definita la nuova identità visiva sul singolo frame, il mutamento stilistico è stato applicato in modo coerente all'intero segmento video avvalendosi del modello Kling 3.0 Omni. Utilizzando quindi la clip video e l'immagine cartoon come input, il prompt utilizzato è stato: "Change the style of Video in the cartoon style of Image".

Per ottenere la still dell'albo a fumetti, è stato utilizzato il seguente prompt:

Prompt: "Genera la pagina di un fumetto tenuto aperto dalle mani di un bambino di 6 anni. POV, vediamo solo le mani che tengono il fumetto. Le mani sono pulite".

Le vignette sono state poi sostituite manualmente su Adobe Photoshop, così che ogni vignetta rappresentasse le ultime clip viste prima di arrivare al finale.

La transizione finale da clip video ad albo a fumetti è stata ottenuta utilizzando la versione dello shuttle cartoon, prima generata, come First Frame e quella dell'albo a fumetti, modificata su Photoshop, come Last Frame. Il prompt utilizzato è stato:

Prompt: "The camera pulls back, revealing it to be a panel within a comic".

Per questa generazione sono testati Veo 3.1 su Flow, Veo 3.1 su Gemini e Kling 2.6, restituendo risultati tra loro diversi. Il peggiore è quello di Flow, che con questo prompt non è riuscito a creare un movimento continuo. Gemini ha aggiunto fluidità ma non restituendo un match cut perfetto con il first frame. Kling ha consentito di avere il miglior compromesso e ottenere la transizione cercata.



Figura 4.29: Da fotorealistico a fumetto

Infine è stato realizzato lo shot in cui il bambino legge il fumetto.

STILL

Iterazione 1:

Prompt: "Togli il giubotto e il berretto".

Input: turn-around del bambino in figura intera.

Analisi: lo scopo è di mantenere coerenza visiva con l'asset del bambino definito in precedenza.

Iterazione 2:

Prompt: "Voglio il controcampo della prima immagine. Il bambino è seduto sul divano, visto di tre quarti mentre legge il fumetto all'interno del soggiorno".

Input: immagine del fumetto tenuto in mano dal bambino e generazione ottenuta all'iterazione 1.

Analisi: l'immagine dove viene visto il fumetto lascia intravedere parti della stanza in cui si trova il bambino. Si vuole adesso ottenere un totale di quella stanza, mostrando il personaggio.

Iterazione 3:

Prompt: "Il bambino della prima immagine deve assomigliare in viso a quello della seconda immagine".

Input: immagine ottenuta all'iterazione 2 e turn-around del bambino in primo

piano.

Analisi: si cerca di riottenere coerenza sul viso del bambino.

Iterazione 4:

Prompt: "Non mi piace il taglio di capelli, puoi farne uno simile ma più naturale, mantenendo i capelli lisci il ciuffo che nell'immagine con il cappello usciva dal cappello".

Input: immagine ottenuta all'iterazione 3.

Analisi: i capelli sembravano troppo appiattiti, è stato fatto un lavoro per renderli più mossi e naturali.

Per concludere, è stato fornito a Gemini una descrizione della storia narrata e chiesto di generare una copertina per un fumetto che la narrasse. La copertina è stata poi aggiunta come modifica all'iterazione 4.

VIDEO

- *Modello:* Kling 2.6
- *Prompt:* "The child reads and flips through the comic book with great concentration".
- *Input:* Still ottenuta dopo la modifica all'iterazione 4.
- *Analisi:* per quanto riguarda l'animazione del bambino nell'atto di sfogliare il fumetto, l'obiettivo è stato quello di ottenere un movimento il più possibile semplice e naturale. Al fine di mantenere l'attenzione visiva focalizzata su questa specifica azione, si è scelto di mantenere l'inquadratura statica, omettendo intenzionalmente qualsiasi movimento di camera.

Capitolo 5

Post-produzione e montaggio

5.1 Montaggio

Sebbene una fase di montaggio preliminare fosse già stata avviata durante la produzione, agendo fondamentale come strumento di feedback per valutare la fluidità narrativa e l'eventuale necessità di nuove generazioni, al termine del processo creativo si è proceduto a un affinamento definitivo dell'opera. L'editing finale, effettuato all'interno di DaVinci Resolve, ha portato alla strutturazione di una timeline composta da 78 inquadrature, successivamente convertita nel formato cinematografico 2.35:1.

5.2 Altre lavorazioni

sono stati effettuati interventi di color correction, resi necessari dalla tendenza dei modelli generativi a produrre clip con un'eccessiva saturazione cromatica. In questa fase, l'obiettivo primario è stato la normalizzazione dei contributi video per garantire una coerenza visiva costante tra i diversi shot. Al termine della correzione primaria, è stata applicata una fase di color grading volta a conferire all'opera un look cinematografico più marcato, arricchito dall'integrazione della grana tipica delle pellicole (film grain) e di una leggera vignettatura.

A completamento del comparto visivo, sono stati eseguiti mirati interventi di finishing per risolvere specifiche anomalie insorte durante la generazione. Un esempio significativo ha riguardato la correzione del colore degli occhi del bambino che, in una specifica clip, tendeva erroneamente all'azzurro: l'incongruenza è stata risolta mediante la creazione di una maschera sulla pupilla e il relativo tracciamento (tracking) del movimento, permettendo di scurire l'area in modo selettivo e naturale.

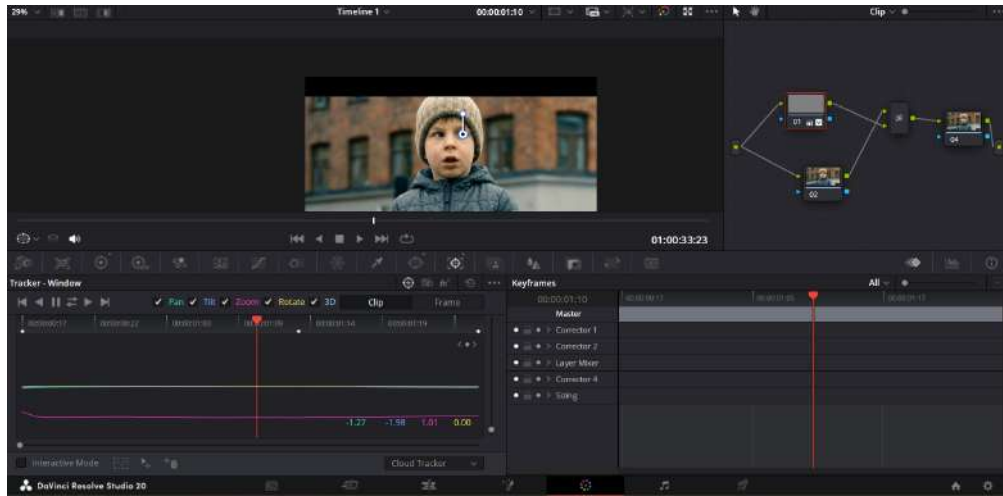


Figura 5.1: Intervento sulla pupilla del bambino

Infine, il processo si è concluso con la fase di sound design. Il Voice Over (V.O.) è stato sintetizzato attraverso la piattaforma ElevenLabs, mentre per la componente sonora sono stati testati diversi tool generativi come Suno, per la stesura della colonna sonora, e Adobe Firefly per la generazione di specifici effetti sonori (SFX).

Capitolo 6

Conclusioni

6.1 Risultati, difficoltà e prospettivi futuri

Il presente progetto di tesi ha permesso di esplorare analiticamente le potenzialità e i limiti qualitativi offerti dall'Intelligenza Artificiale generativa nel panorama tecnologico attuale. L'IA, oltre a rivelarsi un potente strumento integrabile in molteplici workflow, manifesta la concreta possibilità di affermarsi come un nuovo medium a sé stante, ponendosi come una valida alternativa alla cinematografia tradizionale o ai prodotti d'animazione convenzionali.

Per quanto concerne la generazione di immagini, il livello di controllo raggiunto è già notevole e garantisce una grande versatilità nel raffinamento delle stills per ottenere il risultato estetico desiderato. Tuttavia, un ostacolo significativo è rappresentato dai rigidi protocolli di sicurezza dei modelli cloud-based, che spesso incappano in fenomeni di over-blocking, censurando contenuti di natura puramente artistica. Sebbene tali filtri siano comprensibili e necessari, l'introduzione futura di profili "creator" certificati, che tengano conto delle finalità autoriali della generazione, rappresenterebbe un'evoluzione auspicabile. In alternativa, l'adozione di modelli in locale permette di aggirare tali limitazioni, sebbene l'elevata capacità computazionale richiesta costituisca tuttora una barriera d'accesso per chi non dispone di hardware di fascia alta. Un limite rimane la mancanza di riferimenti spaziali. Sebbene esistano strumenti per ottenere da riprese 2D, geometrie in 3D (gaussian splatting) siamo ancora indietro da questo punto di vista, determinando consequenzialmente complicazioni nell'andare a muovere la camera all'interno di un ambiente o modificare la direzione delle sorgenti di luce.

Un ulteriore limite strutturale risiede nella difficoltà da parte dei modelli di comprendere, giustamente, i riferimenti spaziali. Sebbene si stiano affermando tecnologie promettenti per l'estrazione di geometrie tridimensionali da sorgenti 2D, come il Gaussian Splatting, l'integrazione di una reale consapevolezza volumetrica

appare ancora acerba. Tale carenza determina inevitabili complicazioni nel controllo della posizione della camera all'interno degli ambienti e nella manipolazione accurata delle sorgenti luminose. Senza una base geometrica sottostante, la gestione delle parallassi e del relighting rimane vincolata a un'approssimazione bidimensionale, limitando la precisione necessaria per produzioni cinematografiche di alto profilo.

Sul fronte della produzione video, i limiti appaiono più marcati: la tecnologia non consente ancora un controllo assoluto sulla coerenza temporale e spaziale. Nonostante l'introduzione di strumenti come il Video-to-Video e la gestione granulare della durata delle clip abbiano ampliato le possibilità di intervento, restano necessari significativi passi avanti nell'acting dei personaggi, nella loro espressività e nella sincronizzazione del parlato. In conclusione, sebbene il progresso sia costante, siamo ancora distanti dal grado di precisione e controllo che un regista esercita sulla scena e sulle azioni degli attori in un set tradizionale.

Permangono, inoltre, alcuni vincoli tecnici legati alla qualità del segnale video generato. In primo luogo, la gamma dinamica risulta ancora limitata: la maggior parte dei modelli attuali produce output con una profondità di colore a 8-bit, una specifica sensibilmente inferiore agli standard HDR e ai campionamenti ad alta dinamica garantiti dalle cineprese professionali. Questo divario si traduce in una minore flessibilità durante la fase di color grading e in una resa meno dettagliata nelle zone di massima luce e ombra. Un ulteriore limite, sebbene di minor impatto stilistico, riguarda il frame rate, spesso vincolato a uno standard fisso di 24 fps. Infine, la risoluzione nativa costituisce tuttora una criticità: gli output si attestano generalmente sui formati HD o Full HD, rendendo indispensabile l'impiego di upscaler esterni per il raggiungimento di standard qualitativi in 4K o superiori, con il rischio di introdurre artefatti digitali non sempre prevedibili.

L'attuale stato dell'arte, tuttavia, induce a un marcato ottimismo. La rapidità con cui i modelli sono stati perfezionati nell'ultimo periodo, unitamente all'impatto dirompente delle recenti innovazioni, testimonia un'evoluzione tecnologica accelerata. Tale dinamismo suggerisce che molti degli attuali limiti tecnici siano destinati a essere superati in tempi brevi, consolidando ulteriormente il ruolo dell'Intelligenza Artificiale come pilastro fondamentale della nuova economia creativa.

Bibliografia

- [1] Jeffrey A. Okun e Susan Zwerman. *The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures*. Third. Routledge, 2025 (cit. a p. 3).
- [2] Commissione Europea. *Entra in vigore il regolamento sull'IA - Articolo*. Ago. 2024. URL: https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_it (cit. a p. 8).
- [3] Commissione Europea. *Ai Pact*. URL: <https://digital-strategy.ec.europa.eu/en/policies/ai-pact> (cit. a p. 8).
- [4] Commissione Europea. *Entrata in vigore della legge europea sull'intelligenza artificiale - Comunicato stampa*. Ago. 2024. URL: https://ec.europa.eu/commission/presscorner/detail/it/ip_24_4123 (cit. a p. 8).
- [5] Commissione Europea. *Legge sull'IA*. URL: <https://digital-strategy.ec.europa.eu/it/policies/regulatory-framework-ai> (cit. a p. 8).
- [6] Commissione Europea. *Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio*. Giu. 2024. URL: <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32024R1689> (cit. a p. 9).
- [7] Michael Chen. *Chinese AI Companies and EU/U.S. Frontier Safety Regulation*. Gen. 2026. URL: https://www.researchgate.net/publication/400376104_Chinese_AI_Companies_and_EUUS_Frontier_Safety_Regulation9 (cit. a p. 10).
- [8] *LEGGE 23 settembre 2025, n. 132*. URL: <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:2025-09-23;132> (cit. a p. 10).
- [9] Studio legale Enrico Candiani. *IA e Diritto d'autore: Tutto sulla Nuova Legge Italiana (L. 132/2025)*. Nov. 2025. URL: https://www.youtube.com/watch?v=Dm_o7vvJGiI (cit. a p. 10).
- [10] Ciao Internet - Matteo Flora. *Legge 132/2025 su AI-Act: ora il copyright vale davvero per l'AI. Con Lucia Maggi e Anna Rosa Così*. Ott. 2025. URL: <https://www.youtube.com/watch?v=QWw6VQUFU9Q> (cit. a p. 10).

- [11] Coalition for Content Provenance and Authenticity. *Advancing digital content transparency and authenticity*. URL: <https://c2pa.org/> (cit. a p. 11).
- [12] Content Authenticity Initiative. *Restoring trust and transparency in the age of AI*. URL: <https://contentauthenticity.org/> (cit. a p. 12).
- [13] J. Scott Evans, Adobe. *Protecting Creators in the Age of Generative AI*. Giu. 2023. URL: <https://copyrightalliance.org/protecting-creators-generative-ai/> (cit. a p. 12).
- [14] Content Credentials. *An evolution in understanding online content*. URL: <https://contentcredentials.org/> (cit. a p. 12).
- [15] Andy Parsons, Sr. Director, CAI. *Durable Content Credentials*. Apr. 2024. URL: <https://contentauthenticity.org/blog/durable-content-credentials/> (cit. a p. 13).
- [16] Adobe Help Center. *Come addestrare i modelli personalizzati di Firefly*. Gen. 2026. URL: <https://helpx.adobe.com/it/firefly/web/work-with-enterprise-features/train-custom-models/train-firefly-custom-models.html> (cit. a p. 14).
- [17] Adobe Help Center. *Modelli partner nei prodotti Adobe*. Mar. 2026. URL: <https://helpx.adobe.com/it/creative-cloud/apps/generative-ai/non-adobe-models-in-adobe-products.html> (cit. a p. 15).
- [18] Adobe. *Firefly Legal FAQs – Enterprise Customers*. Mag. 2024. URL: <https://www.adobe.com/cc-shared/assets/pdf/enterprise/firefly-legal-faqs-enterprise-customers-2024-06-11.pdf> (cit. a p. 15).
- [19] Bloomberg. *Adobe’s ‘Ethical’ Firefly AI Was Trained on Midjourney images*. URL: <https://www.bloomberg.com/news/articles/2024-04-12/adobe-s-ai-firefly-used-ai-generated-images-from-rivals-for-training> (cit. a p. 15).
- [20] Markus Brinsa. *Adobe Firefly vs Midjourney - The Training Data Showdown and Its Legal Stakes*. Giu. 2025. URL: <https://www.linkedin.com/pulse/adobe-firefly-vs-midjourney-training-data-showdown-its-brinsa-w5vmc> (cit. a p. 15).
- [21] Google. *Gestisci i crediti AI in Flow*. 2026. URL: <https://support.google.com/flow/answer/16526234?hl=en> (cit. a p. 18).
- [22] Google DeepMind. *We want to build AI responsibly to benefit humanity*. URL: <https://deepmind.google/responsibility-and-safety/> (cit. a p. 19).
- [23] Google. *Veo3 Model Card*. Mag. 2025. URL: <https://storage.googleapis.com/deepmind-media/Model-Cards/Veo-3-Model-Card.pdf> (cit. a p. 19).

- [24] Google. *Veo: a text-to-video generation system*. URL: <https://storage.googleapis.com/deepmind-media/veo/Veo-3-Tech-Report.pdf> (cit. a p. 19).
- [25] Google. *Responsible AI Progress Report*. Feb. 2025. URL: <https://ai.google/static/documents/ai-responsibility-update-published-february-2025.pdf> (cit. a p. 19).
- [26] Don - Admiral. *How to Opt-Out of AI Training Bots by Google Bard and OpenAI ChatGPT*. Nov. 2023. URL: <https://blog.getadmiral.com/how-publishers-can-opt-out-of-ai-training-bots-by-google-bard-and-openai-chatgpt> (cit. a p. 19).
- [27] Digwatch. *Google admits using opted-out content for AI training*. Mag. 2025. URL: <https://dig.watch/updates/google-admits-using-opted-out-content-for-ai-training> (cit. a p. 19).
- [28] Zach Vallese - CNBC. *Creators say they didn't know Google uses YouTube to train AI*. Giu. 2025. URL: <https://www.cnbc.com/2025/06/19/google-youtube-ai-training-veo-3.html> (cit. a p. 19).
- [29] Matt Growcoot - PetaPixel. *YouTubers Surprised That Google Uses Their Videos to Train AI Models*. Giu. 2025. URL: <https://petapixel.com/2025/06/20/youtubers-surprised-that-google-uses-their-videos-to-train-its-ai-models-veo-3/> (cit. a p. 19).
- [30] Google DeepMind. *A tool to watermark and identify content generated through AI*. URL: <https://deepmind.google/models/synthid/> (cit. a p. 19).
- [31] Laurie Richardson. *How we're increasing transparency for gen AI content with the C2PA*. Set. 2024. URL: <https://blog.google/innovation-and-ai/products/google-gen-ai-content-transparency-c2pa/> (cit. a p. 19).
- [32] Naina Raisinghani. *Nano Banana 2: Combining Pro capabilities with lightning-fast speed*. Feb. 2026. URL: <https://blog.google/innovation-and-ai/technology/ai/nano-banana-2/> (cit. a p. 19).
- [33] Google Cloud. *Governance dei dati e AI generativa*. URL: <https://docs.cloud.google.com/generative-ai-app-builder/docs/data-governance?hl=it> (cit. a p. 19).
- [34] Google Cloud. *Service Specific Terms*. URL: <https://cloud.google.com/terms/service-terms> (cit. a p. 19).
- [35] Google Cloud. *Shared fate: Protecting customers with generative AI indemnification*. Ott. 2023. URL: <https://cloud.google.com/blog/products/ai-machine-learning/protecting-customers-with-generative-ai-indemnification> (cit. a p. 19).

- [36] NYU|TISCH. *2026 Post Next 50*. Feb. 2026. URL: <https://tisch.nyu.edu/alumni/alumni-news/post-next-50.html> (cit. a p. 20).
- [37] Reuters. *AI company Runway valued at 1.5billioninlatestfunding*. Giu. 2025. URL: <https://www.reuters.com/technology/ai-company-runway-valued-15-bln-latest-funding-source-2023-06-29/> (cit. a p. 20).
- [38] Runway. *Introducing Runway Gen-4.5: A new frontier for video generation*. Dic. 2025. URL: <https://runwayml.com/research/introducing-runway-gen-4.5> (cit. a p. 20).
- [39] Runway. *Introducing Runway Aleph*. Lug. 2025. URL: <https://runwayml.com/research/introducing-runway-aleph> (cit. a p. 21).
- [40] Runway. *Runway Partners with Lionsgate*. Set. 2024. URL: <https://runwayml.com/news/runway-partners-with-lionsgate> (cit. a p. 22).
- [41] Matt Growcott - PetaPixel. *Movie Studio Lionsgate is Struggling to Make AI-Generated Films With Runway*. Set. 2025. URL: <https://petapixel.com/2025/09/23/movie-studio-lionsgate-is-struggling-to-make-ai-generated-films-with-runway/> (cit. a p. 22).
- [42] Roger Cheng, Jeremy Fuster. *What Happened to Lionsgate's Splashy Plan to Make AI Movies With Runway? It's Complicated | Exclusive*. Set. 2025. URL: <https://www.thewrap.com/lionsgate-runway-ai-deal-ip-model-concerns/> (cit. a p. 22).
- [43] Runway. *Security*. URL: <https://www.runway.team/security> (cit. a p. 22).
- [44] Runway. *Privacy Policy*. Set. 2025. URL: <https://runwayml.com/privacy-policy/> (cit. a p. 22).
- [45] Runway. *Runway Security*. Mag. 2025. URL: <https://runwayml.com/data-security> (cit. a p. 22).
- [46] Deepti Ghadiyaram - Runway. *Foundations for Safe Generative Media*. Ott. 2024. URL: <https://runwayml.com/research/foundations-for-safe-generative-media> (cit. a p. 22).
- [47] Samantha Cole - 404Media. *AI Video Generator Runway Trained on Thousands of YouTube Videos Without Permission*. Lug. 2024. URL: <https://www.404media.co/runway-ai-image-generator-training-data-youtube/> (cit. a p. 22).
- [48] Mike Kaput. *Runway Trains on YouTube Videos Without Permission*. Lug. 2024. URL: <https://www.marketingaiinstitute.com/blog/runway-youtube> (cit. a p. 22).

- [49] Emma Roth. *Runway's AI video generator trained on thousands of scraped YouTube videos*. Lug. 2024. URL: <https://www.theverge.com/2024/7/25/24206120/runway-ai-video-generator-scraped-youtube-videos-report> (cit. a p. 22).
- [50] Mogin Law LLP. *Artists Sue AI Companies for Copyright Infringement*. Giu. 2025. URL: <https://moginlawllp.com/artists-sue-ai-companies-for-copyright-infringement/> (cit. a p. 22).
- [51] CANVAS - Business Model. *What is the Brief History of Luma AI Company?* Giu. 2025. URL: <https://canvasbusinessmodel.com/blogs/brief-history/luma-ai-brief-history> (cit. a p. 22).
- [52] Luma. *Luma is Partnering with HUMAIN to Accelerate the Arrival of Multimodal AGI*. Mag. 2025. URL: <https://lumalabs.ai/blog/news/luma-joins-forces-with-humain> (cit. a p. 22).
- [53] Luma. *Ray3 Highlights*. URL: <https://lumalabs.ai/ray> (cit. a p. 23).
- [54] Luma. *Introducing Camera Motion Concepts*. Mar. 2025. URL: <https://lumalabs.ai/blog/news/camera-motion-concepts> (cit. a p. 23).
- [55] Luma. *Introducing Camera Angle Concepts*. Apr. 2025. URL: <https://lumalabs.ai/blog/news/camera-angle-concepts> (cit. a p. 23).
- [56] Luma. *Introducing Ray3 Modify*. Dic. 2025. URL: <https://lumalabs.ai/blog/news/ray3-modify> (cit. a p. 23).
- [57] Luma. *Ray3.14 is here: Native 1080p, 3x cheaper and 4x faster*. Gen. 2026. URL: https://lumalabs.ai/blog/news/ray3_14 (cit. a p. 23).
- [58] Luma. *Introducing Modify Video*. Giu. 2025. URL: <https://lumalabs.ai/blog/news/introducing-modify-video> (cit. a p. 23).
- [59] Luma. *Introducing Modify with Instructions*. Ago. 2025. URL: <https://lumalabs.ai/blog/news/modify-with-instructions#limitations> (cit. a p. 23).
- [60] Post di Luma su X. *Monster Camp*. URL: <https://x.com/LumaLabsAI/status/1800921393321934915> (cit. a p. 24).
- [61] Inno Flores. *Luma's AI Video Tool Dream Machine Under Scrutiny After Altered Version of Disney Character Appears in Its Trailer*. Giu. 2024. URL: <https://www.techtimes.com/articles/305824/20240619/luma-ai-video-tool-dream-machine-under-scrutiny-altered-version-disney.htm> (cit. a p. 24).
- [62] Luma. *Terms of Use*. Feb. 2025. URL: <https://lumalabs.ai/learning-hub/terms-of-use> (cit. a p. 24).

- [63] OpenAI. *DALL-E 3*. URL: <https://openai.com/it-IT/index/dall-e-3/> (cit. a p. 24).
- [64] OpenAI. *Generating videos on Sora*. URL: <https://help.openai.com/en/articles/9957612-generating-videos-on-sora> (cit. a p. 25).
- [65] OpenAI. *Sora 2 è qui*. Set. 2025. URL: <https://openai.com/it-IT/index/sora-2/> (cit. alle pp. 25, 26).
- [66] OpenAI. *Getting started with the Sora app*. URL: <https://help.openai.com/en/articles/12456897-getting-started-with-the-sora-app> (cit. a p. 26).
- [67] OpenAI. *Un lancio responsabile per Sora*. Set. 2025. URL: <https://openai.com/it-IT/index/launching-sora-responsibly/> (cit. a p. 26).
- [68] OpenAI. *Creazione di contenuti su Sora in linea con le nostre politiche*. Dic. 2025. URL: <https://openai.com/it-IT/policies/creating-sora-videos-in-line-with-our-policies/> (cit. a p. 26).
- [69] OpenAI. *La Walt Disney Company e OpenAI raggiungono un accordo storico per portare su Sora gli amati personaggi dei marchi Disney*. Dic. 2025. URL: <https://openai.com/it-IT/index/disney-sora-agreement/> (cit. a p. 27).
- [70] Ashley Capoot - CNBC. *AI Sam Altman and the Sora copyright gamble: 'I hope Nintendo doesn't sue us'*. Ott. 2025. URL: <https://www.cnbc.com/2025/10/04/sora-openai-video-app.html> (cit. a p. 27).
- [71] Kuaishou. *Kuaishou Unveils Proprietary Video Generation Model 'Kling; Testing Now Available*. Giu. 2024. URL: <https://ir.kuaishou.com/news-releases/news-release-details/kuaishou-unveils-proprietary-video-generation-model-kling/> (cit. a p. 27).
- [72] KlingAI. *KLING VIDEO O1 User Guide*. Dic. 2025. URL: <https://app.klingai.com/global/quickstart/klingai-video-o1-user-guide> (cit. a p. 29).
- [73] KlingAI. *Kling VIDEO 3.0 Omni Model User Guide*. Feb. 2026. URL: <https://app.klingai.com/global/quickstart/klingai-video-3-omni-model-user-guide> (cit. a p. 29).
- [74] AnimateAi.Pro. *Kling AI NSFW Policy Explained: Why Professional Creators Prefer Commercial-Grade Safe Models*. Mar. 2026. URL: <https://animateai.pro/blog/kling-ai-nsfw-policy-explained-why-professional-creators-prefer-commercial-grade-safe-models/> (cit. a p. 29).

- [75] Marina Londei - data4biz. *Un modello di IA cinese censura argomenti politicamente sensibili*. Lug. 2024. URL: <https://data4biz.com/articles/un-modello-di-ia-cinese-censura-argomenti-politicamente-sensibili> (cit. a p. 29).
- [76] KlingAI. *Kling AI Terms of Service*. Dic. 2025. URL: <https://app.klingai.com/global/docs/user-policy> (cit. a p. 30).
- [77] stability.ai. *Stable Diffusion Public Release*. Ago. 2022. URL: <https://stability.ai/news/stable-diffusion-public-release> (cit. a p. 30).
- [78] stability.ai. *James Cameron, Academy Award-Winning Filmmaker, Joins Stability AI Board of Directors*. Set. 2024. URL: <https://stability.ai/news/james-cameron-joins-stability-ai-board-of-directors> (cit. a p. 30).
- [79] stability.ai. *Stability AI and EA Partner to Empower Artists, Designers, and Developers to Reimagine Game Development*. Ott. 2025. URL: <https://stability.ai/news/stability-ai-and-ea-partner-to-reimagine-game-development> (cit. a p. 30).
- [80] stability.ai. *Universal Music Group and Stability AI Announce Strategic Alliance to Co-Develop Professional AI Music Creation Tools*. Ott. 2025. URL: <https://stability.ai/news/universal-music-group-and-stability-ai-announce-strategic-alliance> (cit. a p. 30).
- [81] stability.ai. *Warner Music Group and Stability AI Join Forces To Build The Next Generation Of Responsible AI Tools For Music Creation*. Nov. 2025. URL: <https://stability.ai/news/warner-music-group-and-stability-ai-join-forces-to-build-next-gen-tools> (cit. a p. 30).
- [82] J.R. Johnivan. *Stability AI Claims Getty's Copyright Lawsuit Stifles Freedom of Expression*. Giu. 2025. URL: <https://www.eweek.com/news/getty-lawsuit-stability-ai-copyright-uk-trial/> (cit. a p. 31).
- [83] Varuni Parnavitane, Ph.D. *Getty Images vs. Stability AI: The UK Court Battle That Could Reshape AI and Copyright Law*. Lug. 2025. URL: <https://www.finnegan.com/en/insights/articles/getty-images-vs-stability-ai-the-uk-court-battle-that-could-reshape-ai-and-copyright-law.html> (cit. a p. 31).
- [84] Maria Vittoria Pessina. *Concessa autorizzazione ad appellare nella causa Getty Images vs Stability AI*. Gen. 2026. URL: <https://dirittoaldigitale.com/2026/01/26/getty-images/> (cit. a p. 31).

- [85] Robert Booth - The Guardian. *AI firm wins high court ruling after photo agency's copyright claim*. Nov. 2025. URL: <https://www.theguardian.com/media/2025/nov/04/stability-ai-high-court-getty-images-copyright> (cit. a p. 31).
- [86] Edoardo Frasso. *Warner Bros. fa causa a Midjourney per violazione del copyright*. Set. 2025. URL: <https://ainews.it/warner-bros-fa-causa-a-midjourney-per-violazione-del-copyright/> (cit. a p. 32).
- [87] Blake Brittain. *Stability AI, Midjourney should face artists' copyright case, judge says*. Mag. 2024. URL: <https://www.reuters.com/legal/litigation/stability-ai-midjourney-should-face-artists-copyright-case-judge-says-2024-05-08/> (cit. a p. 32).
- [88] Zach Schor, JD '26. *Andersen v. Stability AI: The Landmark Case Unpacking the Copyright Risks of AI Image Generators*. Dic. 2024. URL: <https://jipel.law.nyu.edu/andersen-v-stability-ai-the-landmark-case-unpacking-the-copyright-risks-of-ai-image-generators/> (cit. a p. 32).
- [89] Angela Yang. *ByteDance responds to copyright infringement concerns with Seedance 2.0*. Feb. 2026. URL: <https://www.nbcnews.com/tech/tech-news/seedance-2-bytedance-copyright-infringement-concerns-hollywood-rcna259173> (cit. a p. 33).