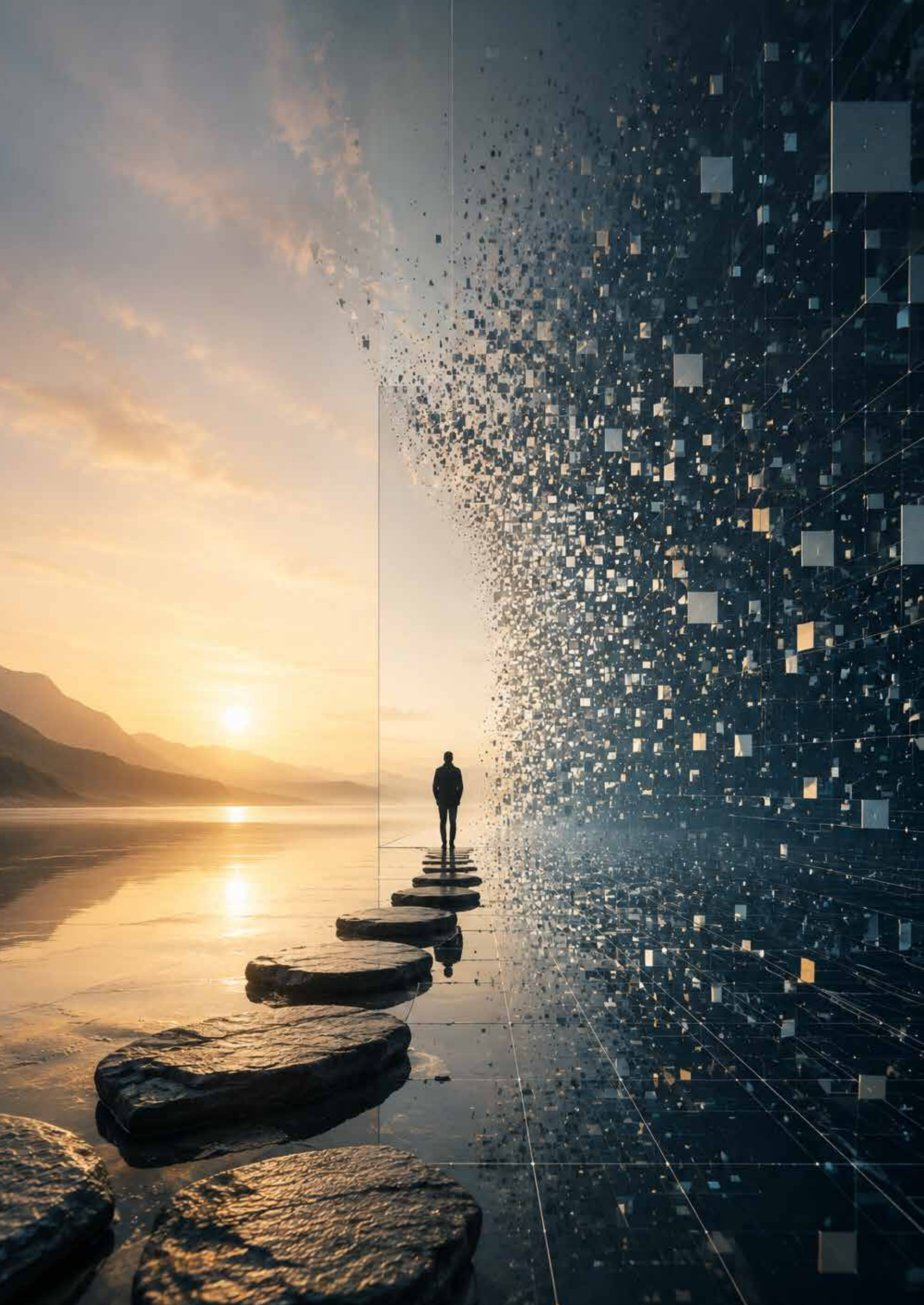




Realtà sintetiche

Video generativo tra creatività, limiti epistemici e verificabilità della comunicazione

testo di/text by Parisa Darv

Synthetic realities. Generative video: creativity, epistemic limits and the verifiability of communication

Introduction

In less than a decade, artificial intelligence has evolved from a niche technology to a general-purpose computing infrastructure. This paradigm shift is conventionally dated to the publication of *Attention Is All You Need* (Vaswani et al., 2017), which replaced recurrence and convolution with self-attention alone, making training massively parallelisable. Since then, research has followed a largely quantitative trajectory: increasing parameters, data and computational power. This programme has been empirically formalised in the scaling laws (Kaplan et al., 2020) and decisively refined through the notion of computational optimality (Hoffmann et al., 2022). Since 2022, the same technical framework has extended to a qualitatively different medium: video. The generation of coherent temporal sequences — with native audio from 2025 onwards — shifts the focus of research from text, which the reader interprets as discourse, to the moving image, which the viewer interprets as a trace of an event that has taken place. It is this semiotic difference, even before the technical one, that makes generative video a matter for the communication sciences and not merely for computer science. At the same time, generative video introduces a significant transformation in creative and design processes, as it no

Introduzione

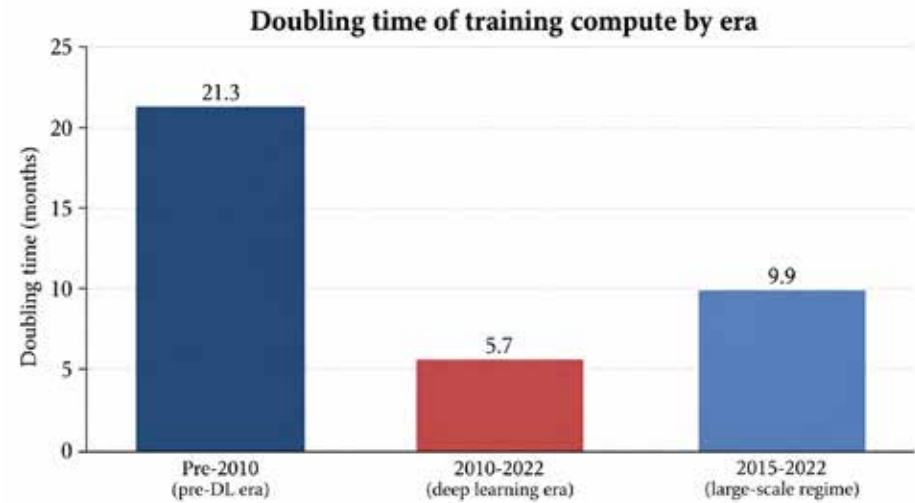
Nel giro di meno di un decennio, l'intelligenza artificiale è passata da tecnologia di nicchia a infrastruttura di calcolo generale. La discontinuità ha una data convenzionale: la pubblicazione di *Attention Is All You Need* (Vaswani et al., 2017), che sostituisce la ricorrenza e la convoluzione con la sola autoattenzione, rendendo l'addestramento massicciamente parallelizzabile. Da quel momento la ricerca ha seguito una traiettoria in larga parte quantitativa: aumentare i parametri, i dati e la potenza di calcolo. Questo programma ha ricevuto una formalizzazione empirica nelle leggi di scala (Kaplan et al., 2020) e una correzione decisiva nella nozione di ottimalità computazionale (Hoffmann et al., 2022). Dal 2022 il medesimo apparato tecnico si è esteso a una modalità qualitativamente diversa: il video. La generazione di sequenze temporali coerenti — con audio nativo a partire dal 2025 — sposta l'oggetto della ricerca dal testo, che il lettore interpreta come discorso, all'immagine in movimento, che lo spettatore interpreta come traccia di un evento accaduto. È questa differenza semiotica, prima ancora che tecnica, a rendere il video generativo un problema di scienze della comunicazione e non solo di informatica. Al tempo stesso, il video generativo introduce una trasformazione rilevante nei processi creativi e progettuali, poiché non opera più soltanto come strumento di rappresentazione di un'idea già definita, bensì come dispositivo capace di produrre rapidamente scenari, variazioni, simulazioni e alternative visive. La generazione audiovisiva entra così nella fase ideativa del progetto, modificando il rapporto tra intenzione autoriale, esplorazione delle possibilità e selezione critica degli esiti. La capacità di ottenere immagini in movimento altamente plausibili non coincide tuttavia con la loro validità progettuale: realismo percettivo, coerenza narrativa e correttezza fisica costituiscono dimensioni distinte, che richiedono forme specifiche di valutazione e responsabilità. Il presente contributo assume dunque una duplice domanda: quali sono le condizioni di validità del paradigma della scala quando applicato alla generazione video, e in che misura il quadro normativo emergente incide sulla produzione, sulla circolazione e sulla verificabilità dei contenuti sintetici? La metodologia è quella della rassegna critica di letteratura, integrata da analisi documentale delle fonti normative primarie (Regolamento (UE) 2024/1689; Legge 23 settembre 2025, n. 132) e da dati statistici ufficiali (Eurostat) e di settore (IAB). La selezione delle fonti privilegia contributi scientifici e documenti tecnico-istituzionali direttamente pertinenti all'evoluzione dei modelli generativi, con particolare attenzione alla letteratura pubblicata dal 2020 in poi, integrata dai riferimenti precedenti necessari a ricostruire i fondamenti teorici del fenomeno. I modelli e i benchmark considerati sono stati selezionati in base alla loro rilevanza nella transizione dalla generazione linguistica a quella audiovisiva, alla loro influenza nella letteratura recente e alla disponibilità di risultati comparabili sulla qualità, sulla coerenza tem-

sotto/below: (in alto) Elaborazione dell'autore su Sevilla, Heim, Ho, Besiroglu, Hobbahn e Villalobos (2022), «Compute Trends Across Three Eras of Machine Learning», arXiv:2202.05924, Sez. 3; dataset aggiornato in Epoch AI, «Parameter, Compute and Data Trends in Machine Learning», <https://epochai.org/data> ; (sotto) Elaborazione dell'autore sui rapporti tecnici e sugli annunci ufficiali originali — Ho et al. (2022a); Singer et al. (2022); Blattmann et al. (2023); OpenAI (2024), «Video Generation Models as World Simulators»; Google DeepMind (2025a); Kuaishou Technology (2024). I valori indicano la durata massima dichiarata dai fornitori al momento dell'annuncio e non costituiscono una misura di qualità / (top) Author's adaptation of Sevilla, Heim, Ho, Besiroglu, Hobbahn and Villalobos (2022), 'Compute Trends Across Three Eras of Machine Learning', arXiv:2202.05924, Section 3; dataset updated in Epoch AI, 'Parameter, Compute and Data Trends in Machine Learning', <https://epochai.org/data> ; (below) Author's adaptation of the original technical reports and official announcements — Ho et al. (2022a); Singer et al. (2022); Blattmann et al. (2023); OpenAI (2024), 'Video Generation Models as World Simulators'; Google DeepMind (2025a); Kuaishou Technology (2024). The figures indicate the maximum duration stated by the providers at the time of the announcement and do not constitute a measure of quality

longer functions merely as a tool for representing a predefined idea but as a device capable of rapidly producing scenarios, variations, simulations, and visual alternatives. Audiovisual generation thus enters the conceptual phase of the project, altering the relationship between the author's intention, the exploration of possibilities and the critical selection of outcomes. However, the ability to produce highly plausible moving images does not necessarily equate to their design validity: perceptual realism, narrative coherence and physical accuracy constitute distinct dimensions, each requiring specific forms of evaluation and accountability. This paper therefore addresses a twofold question: what are the conditions for the validity of the scale paradigm when applied to video generation, and to what extent does the emerging regulatory framework affect the production, circulation and verifiability of synthetic content? The methodology employed is that of a critical literature review, supplemented by a documentary analysis of primary regulatory sources (Regulation (EU) 2024/1689; Law No. 132 of 23 September 2025) and official statistical data (Eurostat) and industry data (IAB). The selection of sources prioritises scientific contributions and technical-institutional documents directly relevant to the evolution of generative models, with particular attention to literature published from 2020 onwards, supplemented by earlier references necessary to reconstruct the theoretical foundations of the phenomenon. The models and benchmarks considered were selected based on their relevance to the transition from linguistic to audiovisual generation, their influence in recent literature, and the availability of comparable results on quality, temporal consistency, and physical plausibility. The analysis therefore brings together technical and scientific evidence, studies on communication and regulatory sources, and takes the verifiability of synthetic content as common ground for comparison.

The scaling paradigm and its correction

Kaplan et al. (2020) demonstrated that the loss of a linguistic model decreases as a power function of the number of parameters, the size of the dataset and the computational resources employed, with only a weak dependence on architectural details. This result has significant methodological implications: it makes performance predictable prior to training and transforms research into a resource allocation problem. However, Hoffmann et al. (2022) demonstrated that the prevailing allocation strategy at the time was suboptimal: the large models of that era were undertrained relative to their parameter counts. The Chinchilla model (70 billion parameters, 1.4 trillion tokens)

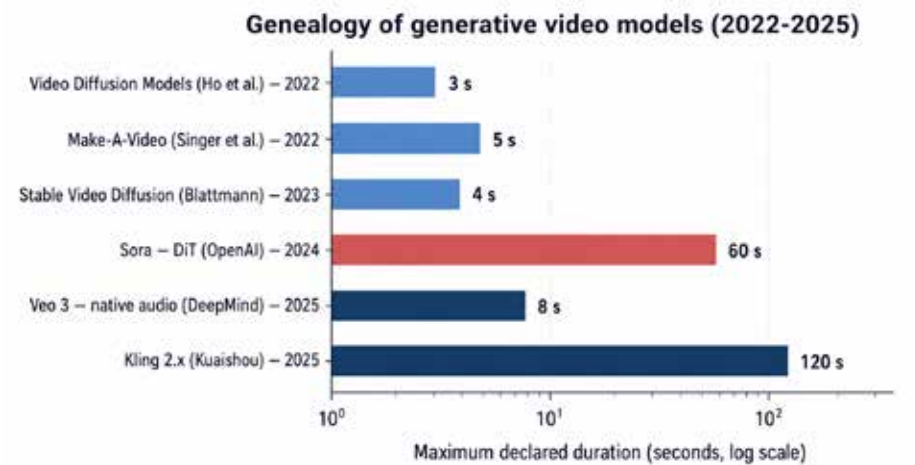


outperforms models four times its size whilst using the same amount of compute. As the authors summarise, 'we show that the size of the model must increase in tandem to increase the effective use of training compute' (Hoffmann et al., 2022, abstract). The practical consequence—shifting investment from parameters to data—has made high-quality data, rather than silicon, the limiting factor: a constraint that becomes even more acute in the video domain, given the scarcity of legally usable, annotated, high-resolution video corpora. The material dimension of this paradigm is documented by Sevilla et al. (2022), who identify three distinct eras. Prior to 2010, training compute power roughly doubled every 21 months, in line with Moore's Law; in the era of deep learning, the doubling time fell to approximately 5.7 months; from 2015 onwards, a large-scale regime, characteristic of cutting-edge laboratories, became established. A growth rate of this magnitude is not sustainable indefinitely: it entails rising energy, water, and capital costs. It creates a barrier to entry that concentrates research capacity among a few

private players — an asymmetry that Crawford (2021) interprets as a structural rather than accidental feature of contemporary AI. Video generation exacerbates the problem: computational cost increases with both temporal and spatial scales, and inference — not just training — becomes a significant cost per second produced. Since 2024, the focus has shifted towards test-time compute: reasoning-oriented models that generate extended chains of thought before responding sometimes yield greater gains, at the same cost, than a further increase in pre-training. In diffusion models, the corresponding factors are the number of denoising steps and the sampling resolution: two operational parameters that significantly alter perceived quality and which benchmarks rarely report, rendering many comparisons between video systems non-reproducible.

From text to motion: a genealogy of generative video models

The starting point is Ho et al. (2022a), Video Diffusion Models, which extends the diffusion U-Net



with a factorised spatio-temporal architecture and joint image-video training. The authors state the objective in explicit terms: 'Generating temporally coherent high-fidelity video is an important milestone in generative modelling research' (Ho et al., 2022a). This line of research continues with Imagen Video (Ho et al., 2022b), which adopts a cascade of spatial and temporal super-resolution models, and with Make-A-Video (Singer et al., 2022), which decouples the learning of what — from text-image pairs — from the learning of how it moves, derived from unannotated videos: an elegant solution to the problem of text-video data scarcity. Blattmann et al. (2023) take a step towards industrial scale with Stable Video Diffusion, which transfers two-dimensional latent diffusion into the temporal domain by introducing temporal layers and formalising a three-stage data curation pipeline. The authors present the system as 'a latent video diffusion model for high-resolution, state-of-the-art text-to-video and image-to-video generation' (Blattmann et al., 2023, abstract). The decisive contribution is not architectural but methodological: it demonstrates that dataset curation, rather than dataset size, determines model quality. The architectural breakthrough is attributed to Peebles and Xie (2023), who replace the U-Net with a transformer operating on latent patches (DiT), demonstrating that sample quality scales with the number of floating-point operations; thus, scaling laws are introduced into the visual generative domain. Building on this, OpenAI (2024) developed Sora, trained jointly on videos and images with varying durations, resolutions, and aspect ratios, which were represented as spatiotemporal patches. The technical report states: 'We explore large-scale training of generative models on video data... Our largest model, Sora, is capable of generating a minute of high-fidelity video' (OpenAI, 2024). The very title of the paper — Video generation models as world simulators — puts forward a stronger argument, summarised by one of the authors' statements that 'AGI will be able to simulate the physical world, and Sora is a key step in that direction' (Brooks, cited in OpenAI, 2024). Three developments marked the following three years. Firstly, the integration of audio: Veo 3 (Google DeepMind, 2025) was the first state-of-the-art model to jointly generate synchronised images, dialogue and sound effects, thereby eliminating the final perceptual clue — lip-sync mismatch — that had previously made fakes recognisable. Secondly, international competition: Kling (Kuaishou, 2024) extends generation to sequences lasting nearly two minutes. Thirdly, the release of weights: Wan (Alibaba, 2025) and Open-Sora (Zheng et al., 2024) replicate DiT pipelines with downloadable

porale e sulla plausibilità fisica. L'analisi mette quindi in relazione evidenze tecnico-scientifiche, studi sulla comunicazione e fonti normative, assumendo la verificabilità dei contenuti sintetici quale terreno comune di confronto.

Il paradigma della scala e la sua correzione

Kaplan et al. (2020) hanno mostrato che la loss di un modello linguistico decresce come funzione di potenza del numero di parametri, della dimensione del dataset e del compute impiegato, con una dipendenza debole dai dettagli architetturali. Il risultato ha una portata metodologica notevole: rende le prestazioni prevedibili prima dell'addestramento e trasforma la ricerca in un problema di allocazione di risorse. Hoffmann et al. (2022) hanno però dimostrato che l'allocazione allora prevalente era subottimale: i grandi modelli dell'epoca erano sottoaddestrati rispetto al numero di parametri. Il modello Chinchilla (70 miliardi di parametri, 1,4 trilioni di token) supera modelli quattro volte più grandi a parità di compute. Come sintetizzano gli autori, «we show that the size of the model must increase in tandem to increase the effective use of training compute» (Hoffmann et al., 2022, abstract). La conseguenza pratica — spostare gli investimenti dai parametri ai dati — ha reso i dati di alta qualità, non il silicio, il fattore limitante: un vincolo che nel dominio video diventa ancora più stringente, data la scarsità di corpora video annotati e ad alta risoluzione legalmente utilizzabili. La dimensione materiale del paradigma è documentata da Sevilla et al. (2022), che ne distinguono tre ere. Prima del 2010, il compute di addestramento raddoppiava all'incirca ogni 21 mesi, in linea con la legge di Moore; nell'era del deep learning, il tempo di raddoppio si riduce a circa 5,7 mesi; dal 2015 si afferma un regime large-scale, proprio dei laboratori di frontiera. Un tasso di crescita di quest'ordine non è indefinitamente sostenibile: implica costi energetici, idrici e di capitale crescenti e introduce una barriera all'ingresso che concentra la capacità di ricerca in pochi attori privati — asimmetria che Crawford (2021) interpreta come tratto strutturale, e non accidentale, dell'IA contemporanea. La generazione video accentua il problema: il costo computazionale cresce con la dimensione temporale oltre che spaziale, e l'inferenza — non solo l'addestramento — diventa una voce di spesa rilevante per ogni secondo prodotto. Dal 2024 il baricentro si è spostato verso il *test-time compute*: modelli orientati al ragionamento che generano catene di pensiero estese prima di rispondere ottengono guadagni talvolta superiori, a parità di costo, rispetto a un ulteriore incremento del pre-addestramento. Nei modelli di diffusione i corrispettivi sono il numero di passi di denoising e la risoluzione di campionamento: due parametri d'uso che modificano sensibilmente la qualità percepita e che i benchmark raramente dichiarano, rendendo molte comparazioni fra sistemi video non riproducibili.

Dal testo al movimento: genealogia dei modelli generativi video

Il punto di partenza è Ho et al. (2022a), *Video Diffusion Models*, che estende l'*U-Net* di diffusione con un'architettura spazio-temporale fattorizzata e un addestramento congiunto immagine-video. Gli autori dichiarano l'obiettivo in termini espliciti: «Generating temporally coherent high fidelity video is an important milestone in generative modelling research» (Ho et al., 2022a). La linea prosegue con Imagen Video (Ho et al., 2022b), che adotta una cascata di modelli di super-risoluzione spaziale e temporale, e con *Make-A-Video* (Singer et al., 2022), che disaccoppia l'apprendimento del che cosa — dalle coppie testo-immagine — dall'apprendimento del come si muove, ricavato da video non annotati: una soluzione elegante al problema della scarsità di dati testo-video. Blattmann et al. (2023) compiono il passo verso la scala industriale con *Stable Video Diffusion*, che trasferisce la diffusione latente bidimensionale nel dominio temporale, inserendo strati temporali e formalizzando una pipeline di cura dei dati in tre stadi. Gli autori presentano il sistema come «a latent video diffusion model for high-resolution, state-of-the-art text-to-video and image-to-video generation» (Blattmann et al., 2023, abstract). Il contributo decisivo non è architetturale ma metodologico: dimostra che la cura dei dati, e non solo la sua dimensione, determina la qualità del modello. La svolta architetturale è dovuta a Peebles e Xie (2023), che sostituiscono l'*U-Net* con un *transformer* operante su *patch* latenti (DiT), mostrando che la qualità dei campioni scala con il numero di operazioni in virgola mobile: le leggi di scala vengono così importate nel dominio generativo visivo. Su questa base OpenAI (2024) costruisce Sora, addestrato congiuntamente su video e immagini di durata, risoluzione e proporzioni variabili, rappresentati come patch spazio-temporali. Il rapporto tecnico dichiara: «We explore large-scale training of generative models on video data... Our largest model, Sora, is capable of generating a minute of high fidelity video» (OpenAI, 2024). Il titolo stesso del documento — Video generation models as world simulators — avanza però una tesi più forte, riassunta dalla dichiarazione di uno degli autori secondo cui «AGI will be able to simulate the physical world, and Sora is a key step in that direction» (Brooks, cit.

sotto/below: Elaborazione dell'autore su Morphed, «AI Video Ads Statistics 2026: Spend, Performance & Production», aggiornamento 12 giugno 2026, e Interactive Advertising Bureau, «2025 Video Ad Spend & Strategy Full Report», New York, luglio 2025

(dato sugli inserzionisti). Scala logaritmica; le unità non sono omogenee e i valori non vanno confrontati direttamente fra loro / Author's analysis based on Morphed, 'AI Video Ads Statistics 2026: Spend, Performance & Production', updated 12 June 2026, and the

Interactive Advertising Bureau, '2025 Video Ad Spend & Strategy Full Report', New York, July 2025 (data on advertisers). Logarithmic scale; the units are not consistent and the values should not be compared directly with one another

models, shifting the capacity for synthetic production beyond the confines of large laboratories and, with it, any possibility of centralised upstream control.

Epistemic limits of video generation

The world simulator hypothesis has been subjected to systematic testing. Bansal et al. (2024) use VideoPhy to construct a testbed for elementary physical interactions (solid–solid, solid–liquid) and document recurring violations of fundamental principles, such as the conservation of mass and the laws of motion. Meng et al. (2024) extend the assessment to mechanics, optics, thermodynamics and material properties. Kang et al. (2024) provide the clearest evidence: the models generalise physical regularities only under conditions close to the training distribution, behaving as statistical interpolators rather than causal simulators. This leads to a clear epistemological conclusion: the perceptual fidelity of a frame does not indicate a model of the world, and visual consistency can coexist with causal inconsistency. This is precisely the visual counterpart to the argument put forward by Bender et al. (2021) regarding text. The dominant metric, the Fréchet Video Distance, exhibits a documented bias. Ge et al. (2024) show that 'FVD is biased towards per-frame quality [rather] than temporal consistency': temporally corrupted sequences obtain scores similar to those of coherent sequences when the quality of individual frames is preserved. A metric designed to capture the realism of motion turns out to be insensitive to that very realism. The community's response has been to decompose the evaluation: VBench (Huang et al., 2024) breaks the assessment down into sixteen dimensions validated against human judgements, whilst VBench-2.0 (Zheng et al., 2025) adds eighteen further dimensions focused on intrinsic fidelity—common sense, anatomy, and physical plausibility. The shift from an aggregated metric to a multidimensional framework mirrors, in the video domain, the criticism of metrics formulated by Schaeffer, Miranda and Koyejo (2023) regarding the so-called emergent abilities of language models: 'emergent abilities may be a mirage caused primarily by the researcher's choice of metric'. Kalai et al. (2025) attribute hallucinations to the incentives inherent in the training and evaluation process: 'language models hallucinate because the training and evaluation procedures reward guessing over acknowledging uncertainty' (abstract). This mechanism can be applied to video: when the objective function rewards perceptual plausibility, the model produces, without hesitation, a scene that is internally coherent yet factually false, without any sign of uncertainty.

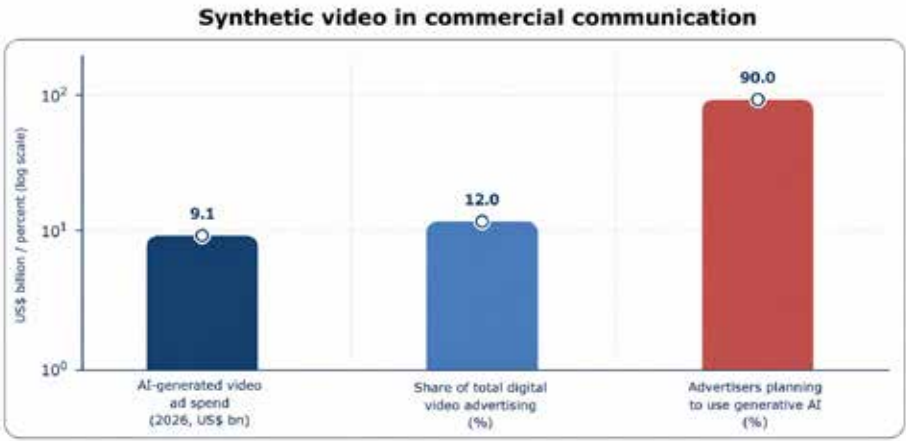
in OpenAI, 2024). Il triennio successivo è segnato da tre sviluppi. Primo, l'integrazione dell'audio: Veo 3 (Google DeepMind, 2025) è il primo modello di frontiera a generare congiuntamente immagine, dialogo ed effetti sonori sincronizzati, eliminando l'ultimo indizio percettivo — l'asincronia labiale — che rendeva riconoscibili i falsi. Secondo, la competizione internazionale: Kling (Kuaishou, 2024) porta la generazione a sequenze di durata prossima ai due minuti. Terzo, l'apertura dei pesi: Wan (Alibaba, 2025) e Open-Sora (Zheng et al., 2024) riproducono pipeline DiT con modelli scaricabili, spostando la capacità di produzione sintetica fuori dal perimetro dei grandi laboratori e, con essa, ogni possibilità di controllo centralizzato a monte.

Limiti epistemici della generazione video

La tesi del simulatore di mondo è stata sottoposta a una verifica sistematica. Bansal et al. (2024) costruiscono con VideoPhy un banco di prova per interazioni fisiche elementari (solido-solido, solido-liquido) e documentano violazioni ricorrenti di principi basilari, quali la conservazione della massa e le leggi del moto. Meng et al. (2024) estendono la valutazione alla meccanica, all'ottica, alla termologia e alle proprietà dei materiali. Kang et al. (2024) forniscono l'evidenza più netta: i modelli generalizzano le regolarità fisiche soltanto in condizioni prossime alla distribuzione di addestramento, comportandosi da interpolatori statistici anziché da simulatori causali. Ne discende una tesi epistemologica precisa: la fedeltà percettiva di un fotogramma non è indice di un modello del mondo, e la coerenza visiva può coesistere con l'incoerenza causale. È esattamente la versione visiva dell'argomento di Bender et al. (2021) sul testo. La metrica dominante, la Fréchet Video Distance, presenta una distorsione documentata. Ge et al. (2024) mostrano che «FVD is biased towards per-frame quality [rather] than temporal consistency»: sequenze temporalmente corrotte ottengono punteggi simili a quelle coerenti quando la qualità dei singoli fotogrammi è preservata. Una metrica pensata per catturare il realismo del movimento risulta insensibile proprio a tale realismo. La risposta della comunità è la scomposizione della valutazione: VBench (Huang et al., 2024) articola il giudizio in sedici dimensioni validate contro giudizi umani, e VBench-2.0 (Zheng et al., 2025) ne aggiunge diciotto orientate alla fedeltà intrinseca — senso comune, anatomia, plausibilità fisica. Il passaggio da una metrica aggregata a una griglia multidimensionale ricalca, nel dominio video, la critica metrica formulata da Schaeffer, Miranda e Koyejo (2023) per le cosiddette abilità emergenti dei modelli linguistici: «emergent abilities may be a mirage caused primarily by the researcher's choice of metric». Kalai et al. (2025) riconducono le allucinazioni agli incentivi del processo di addestramento e valutazione: «language models hallucinate because the training and evaluation procedures reward guessing over acknowledging uncertainty» (abstract). Il meccanismo è trasferibile al video: quando la funzione obiettivo premia la plausibilità percettiva, il modello produce, senza esitazione, una scena internamente coerente e materialmente falsa, senza alcun segnale di incertezza. Nel testo l'allucinazione è un enunciato verificabile; nel video è un evento apparentemente testimoniato. La differenza non è di grado ma di statuto probatorio.

Video sintetico e comunicazione

L'adozione più rapida non avviene nel giornalismo ma nella pubblicità. Le stime di settore indicano per il 2026 una rapida crescita degli investimenti pubblicitari destinati a contenuti video prodotti o assistiti mediante IA generativa, mentre l'Interactive Advertising Bureau (2025) riporta che la quasi totalità degli inserzionisti prevede di impiegare l'IA generativa nella produzione di creatività video. Il



In text, a hallucination is a verifiable statement; in video, it is an apparently witnessed event. The difference is not one of degree but of evidential status.

Synthetic video and communication

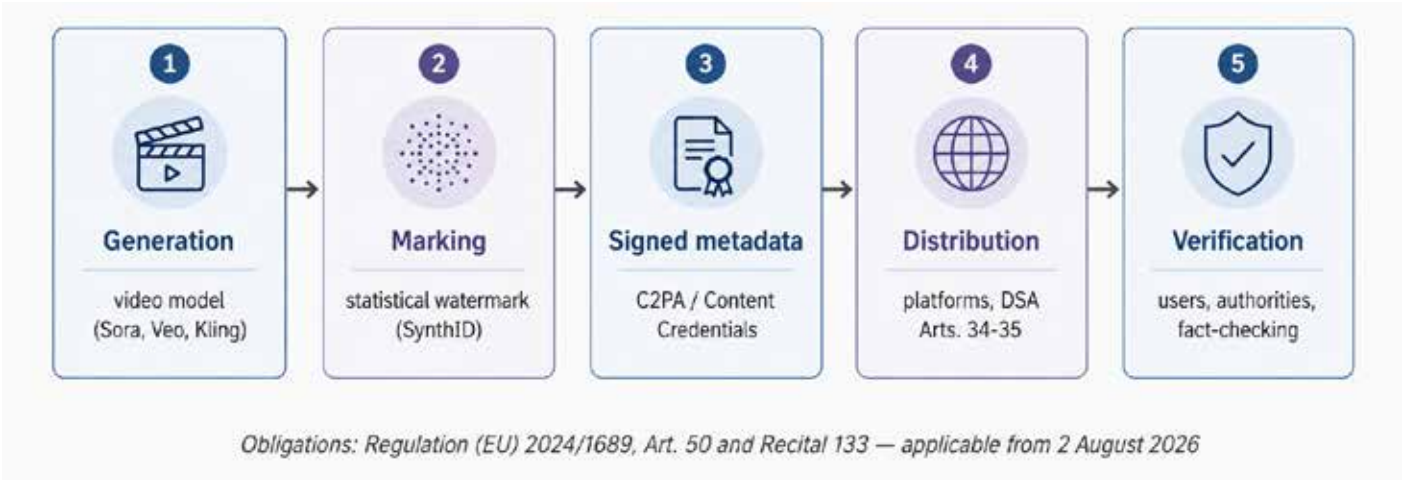
The fastest adoption is not taking place in journalism but in advertising. Industry estimates point to rapid growth in advertising spend on video content produced or assisted by generative AI by 2026, whilst the Interactive Advertising Bureau (2025) reports that almost all advertisers plan to use generative AI in the production of video creative. The key figure is not the volume but the reduction in marginal cost: the production of a set of advertising clips drops from tens of thousands of dollars to a few tens, with immediate effects on the structure of the creative services market and on employment in intermediate professions. Three findings from the literature precisely define the problem. First, detectability: Köbis, Doležalová, and Soraperra (2021) document, in a pre-registered study, that people cannot distinguish deepfakes from authentic content and, at the same time, overestimate their own ability to do so — a combination of incompetence and overconfidence that undermines countermeasures based on individual vigilance. Groh et al. (2022) clarify the picture: 'ordinary humans perform in the range of the leading machine learning model', and crowds informed by the model outperform both humans and the machine acting in isolation. The design implication is that verification should be conceived as a hybrid socio-technical system, neither an individual task nor an automatic classifier—secondly, the liar's dividend. Chesney and Citron (2019) observe that 'harmful lies are nothing new. However, the ability to distort reality has taken an exponential leap forward with "deepfake" technology', and show that the main harm lies not in the circulation of the fake but in the possibility, for those filmed, of dismissing an authentic recording as a fake. The social cost therefore falls on the genuine evidence, not just on the false evidence. Thirdly, the epistemic function of the recording. Rini (2020) argues that 'recordings have so far provided an epistemic backstop to our testimonial practices': the mere possibility of subsequent audiovisual verification regulates public testimony. Fallis (2021) formalises this consequence in terms of information theory: deepfakes reduce the informational content of the video, regardless of whether a single fake is actually disseminated. The threat is structural, not episodic. That of saturation compounds the problem of intentional falsification. The term AI slop — named Merriam-Webster's Word of the Year for 2025 — describes the large-scale produc-

dato rilevante non è il volume ma l'abbattimento del costo marginale: la produzione di un set di clip pubblicitarie passa da decine di migliaia di dollari a poche decine, con effetti immediati sulla struttura del mercato dei servizi creativi e sull'occupazione delle professioni intermedie. Tre risultati della letteratura definiscono il problema con precisione. Primo, la rilevabilità: Köbis, Doležalová e Soraperra (2021) documentano in uno studio preregistrato che le persone non distinguono i deepfake dai contenuti autentici e, al contempo, sopravvalutano la propria capacità di farlo — una combinazione di incompetenza e sovrastima che vanifica le strategie di contrasto fondate sull'attenzione individuale. Groh et al. (2022) precisano il quadro: «ordinary humans perform in the range of the leading machine learning model», e le folle informate dal modello superano sia gli umani sia la macchina isolatamente. L'implicazione progettuale è che la verifica vada concepita come un sistema socio-tecnico ibrido, non come un compito individuale né come un classificatore automatico. Secondo, il dividendo del mentitore. Chesney e Citron (2019) osservano che «harmful lies are nothing new. But the ability to distort reality has taken an exponential leap forward with 'deep fake' technology», e mostrano che il danno principale non è la circolazione del falso ma la possibilità, per chi è ripreso, di liquidare come sintetica una registrazione autentica. Il costo sociale ricade dunque sulla prova vera, non solo su quella falsa. Terzo, la funzione epistemica della registrazione. Rini (2020) argomenta che «recordings have so far provided an epistemic backstop to our testimonial practices»: la sola possibilità di una verifica audiovisiva successiva disciplina la testimonianza pubblica. Fallis (2021) formalizza la conseguenza in termini di teoria dell'informazione: i deepfake riducono il contenuto informativo del video in quanto tale, indipendentemente dal fatto che un singolo falso venga effettivamente diffuso. La minaccia è strutturale, non episodica. Al problema della falsificazione intenzionale si affianca quello della saturazione. Il termine *AI slop* — definito parola dell'anno 2025 da Merriam-Webster — descrive la produzione su larga scala di contenuti sintetici di bassa qualità come strategia per conquistare l'attenzione. L'effetto non è un inganno su un fatto specifico, ma la riduzione del rapporto segnale-rumore dell'intero ecosistema informativo e l'aumento del costo cognitivo della verifica per il pubblico. Sul piano teorico, il fenomeno riporta al centro la questione dell'economia dell'attenzione: la scarsità non è più la capacità di produrre contenuti, ma quella di attribuire loro fiducia.

Governance: dalla trasparenza dichiarata alla provenienza verificabile

L'*AI Act*, in vigore dal 1° agosto 2024, adotta un'architettura basata sul rischio: pratiche vietate (art. 5), sistemi ad alto rischio (allegato III), obblighi di trasparenza (art. 50) e rischio minimo. Per la comunicazione sintetica la norma cardine è l'articolo 50, che impone ai fornitori di sistemi generativi di marcare i contenuti in formato leggibile dalla macchina come artificialmente generati o manipolati, e ai deployer di deep fake di dichiararne la natura artificiale, con eccezioni per finalità di prevenzione e accertamento dei reati e con modalità adatte — non invasive — per le opere manifestamente artistiche, creative, satiriche o di finzione. Il considerando 133 specifica lo strumentario tecnico ammesso: filigrane, identificazione tramite metadati, metodi crittografici per la prova della provenienza e dell'autenticità, registri, impronte digitali. Nel luglio 2026 la Commissione ha pubblicato le linee guida attuative sull'articolo 50 e ha riconosciuto il *Code of Practice on Transparency of AI-generated Content* quale strumento idoneo a facilitarne la conformità. La conformità all'articolo 50 poggia su due famiglie di strumenti: le filigrane statistiche impresse nel segnale (SynthID per immagine, audio, testo e video; il precedente metodologico è la marcatura a liste verdi e rosse proposta da Kirchenbauer et al. 2023 per il testo) e i metadati firmati crittograficamente secondo lo standard C2PA, diffusi commercialmente come Content Credentials. Entrambe le famiglie hanno limiti documentati. I metadati sono volontari e vengono rimossi da qualunque ricodifica, ritaglio, conversione di formato o ricarica su una piattaforma non conforme; le filigrane statistiche resistono meglio, ma degradano sotto compressione aggressiva e sotto attacchi mirati. La stessa OpenAI avverte che «provenance signals can be helpful indicators of origin, but they are not a guarantee that content is accurate, unedited, legally owned». Ne consegue una tesi centrale del presente lavoro: la provenienza è un'infrastruttura di attestazione dell'origine, non di verifica della verità, e la sua efficacia dipende meno dalla robustezza crittografica che dall'adozione coordinata lungo l'intera catena — generazione, distribuzione, visualizzazione — con obblighi speculari a carico delle piattaforme ai sensi degli articoli 34-35 del *Digital Services Act*. L'Italia è il primo Stato membro dotato di una legge nazionale organica: la Legge 23 settembre 2025, n. 132, pubblicata in Gazzetta Ufficiale n. 236 del 10 ottobre 2025. Il testo fissa principi settoriali (sanità, giustizia, lavoro, pubblica amministrazione), delega al Governo l'adozione dei decreti attuativi e individua due autorità nazionali: l'AgID, con funzioni di notifica e di promozione dell'innovazione, e l'ACN, con funzioni di vigilanza del mercato e di sicurezza. Sul terreno della comunicazione sintetica la

sotto/below: (in alto) Elaborazione dell'autore su Coalition for Content Provenance and Authenticity, «C2PA Technical Specification» v2.x; Google DeepMind, «SynthID»; Regolamento (UE) 2024/1689, art. 50 e considerando 133; Regolamento (UE) 2022/2065 (Digital Services Act), artt. 34-35; (sotto) Elaborazione dell'autore su Regolamento (UE) 2024/1689, artt. 4, 5, 50, 51-56 e 113, e su Commissione europea, «Guidelines on the Implementation of the Transparency Obligations... under Article 50 of the AI Act», C(2026) 5054 final, 20 luglio 2026 / (above) Author's adaptation of Coalition for Content Provenance and Authenticity, 'C2PA Technical Specification' v2.x; Google DeepMind, 'SynthID'; Regulation (EU) 2024/1689, Article 50 and recital 133; Regulation (EU) 2022/2065 (Digital Services Act), Articles 34-35; (below) Author's adaptation of Regulation (EU) 2024/1689, Articles 4, 5, 50, 51-56 and 113, and of the European Commission, 'Guidelines on the Implementation of the Transparency Obligations... under Article 50 of the AI Act', C(2026) 5054 final, 20 July 2026



tion of low-quality synthetic content as a strategy to capture attention. The effect is not deception regarding a specific fact, but rather a reduction in the signal-to-noise ratio of the entire information ecosystem and an increase in the cognitive cost of verification for the public. On a theoretical level, the phenomenon brings the issue of the attention economy back to the fore: scarcity no longer lies in the ability to produce content but in the ability to earn trust for it.

Governance: from declared transparency to verifiable provenance

The AI Act, which comes into force on 1 August 2024, adopts a risk-based framework: prohibited practices (Article 5), high-risk systems (Annex III), transparency obligations (Article 50) and minimal risk. For this summary, the key provision is Article 50, which requires providers of generative systems to mark content as artificially generated or manipulated in a machine-readable format. It requires those deploying deepfakes to declare their artificial nature, with exceptions for the prevention and investigation of criminal offences, and with adapted — non-invasive — methods for manifestly artistic, creative, satirical, or fictional works. Recital 133 specifies the permitted technical tools: watermarks, identification via metadata, cryptographic methods for proving provenance and authenticity, registers, and digital fingerprints. In July 2026, the Commission published the implementing guidelines on Article 50 and recognised the Code of Practice on Transparency of AI-generated Content as a suitable tool to facilitate compliance. Compliance with Article 50 relies on two families of tools: statistical watermarks embedded in the signal (SynthID for images, audio, text and video; the methodological precedent is the 'green and red lists' marking proposed by Kirchenbauer et al. 2023 for text) and cryptographically signed meta-

data in accordance with the C2PA standard, commercially available as Content Credentials. Both families have documented limitations. Metadata is voluntary and is removed by any re-encoding, cropping, format conversion or re-upload to a non-compliant platform; statistical watermarks are more resilient, but degrade under aggressive compression and targeted attacks. OpenAI itself warns that 'provenance signals can be helpful indicators of origin, but they are not a guarantee that content is accurate, unedited, legally owned'. This leads to a central thesis of this paper: provenance is an infrastructure for attesting to origin, not for verifying truth, and its effectiveness depends less on cryptographic robustness than on coordinated adoption throughout the entire chain — generation, distribution, display — with corresponding obligations on platforms pursuant to Articles 34–35 of the Digital Services Act. Italy is the first Member State to have a comprehensive national law: Law No. 132 of 23 September 2025, published in the Gazzetta Ufficiale No. 236 of 10 October 2025. The text sets out sector-specific principles (health, justice, labour, public administration), delegates the adoption of implementing decrees to the Government, and identifies two national authorities: AgID, responsible for notification and the promotion of innovation, and ACN, responsible for market supervision and security. In terms of concise

communication, the most significant development relates to criminal law: the introduction into the Criminal Code of the offence of unlawful dissemination of content generated or altered using artificial intelligence systems (Article 612-quater of the Criminal Code), which punishes the dissemination without consent of images, videos or audio that are generated or manipulated and likely to mislead as to their authenticity, with aggravating circumstances in sensitive contexts. The Italian model thus combines the administrative and preventive approach of the AI Act with individual criminal protection, but entrusts oversight to two separate authorities, which the implementing decrees must coordinate.

Adoption and absorption capacity

Eurostat data show rapid but uneven adoption: the proportion of EU-27 businesses using at least one AI technology rises from 13.5 per cent (2024) to 20.0 per cent (2025); Italy's figure rises from 8.2 per cent to 16.4 per cent, remaining below the European average and far behind the frontrunners Denmark (42.0 per cent) and Finland (37.8 per cent). The gap cannot be attributed solely to technological endowment: the average size of enterprises, digital skills and sectoral structure also play a role. At the microeconomic level, experimental studies point to a positive but heterogeneous ef-



fect: Noy and Zhang (2023) find productivity gains and a reduction in qualitative variation in professional writing; Brynjolfsson, Li and Raymond (2025) observe benefits concentrated among less experienced workers; Dell'Acqua et al. (2023) introduce the notion of a jagged frontier, according to which AI assistance improves tasks within the frontier of the model's capabilities and worsens those outside it. At the macroeconomic level, Acemoglu (2024) estimates a modest increase in total factor productivity, of the order of 0.7 per cent over ten years. Applied to the creative industries, the jagged frontier suggests a testable prediction: high gains in the production of standardised, short-form content, and losses in production that requires narrative continuity, causal coherence and editorial responsibility.

Critical discussion

The first issue concerns the relationship between perceptual fidelity and understanding of the world. Laws of scale promise predictability of loss, not of capabilities; in video, they promise frame realism, not correct dynamics. Benchmarks of physical common sense show that the latter does not follow from the former. Until the community adopts continuous, calibrated and multidimensional measures, claims about 'world simulators' will remain epistemically fragile. A second issue arises in the relationship between abundance and trust. The reduction in the cost of audiovisual production goes hand in hand with democratisation and inflation. The informational value of a recording depended on the cost of falsifying it; once that cost was eliminated, trust had to be rebuilt through institutional and infrastructural means — provenance, editorial responsibility, hybrid verification — rather than through perception. The liar's dividend shows that inaction carries an asymmetric cost, which falls on authentic evidence. A further critical issue concerns the relationship between formal transparency and substantive verifiability. Article 50 mandates labelling, but this labelling is fragile. It can be removed voluntarily, and its obligations apply to compliant providers, not to malicious actors or to locally executable open-source models. The risk is a regulatory regime that burdens legitimate operators without affecting illicit use. The plausible response is not regulatory in the strict sense but systemic: to make the absence of provenance credentials a negative signal interpretable by the public, which requires universal adoption throughout the distribution chain. Finally, this reflection extends to the relationship between generation and design. The introduction of generative video into creative processes does not merely accelerate the production of representations, but alters

novità di maggior rilievo è penalistica: l'introduzione nel codice penale del delitto di illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale (art. 612-quater c.p.), che punisce con la reclusione — nella forbice indicata dalla dottrina fra uno e cinque anni — la diffusione senza consenso di immagini, video o audio generati o manipolati e idonei a indurre in inganno sulla loro genuinità, con aggravanti in contesti sensibili. Il modello italiano combina così la logica amministrativa e preventiva dell'AI Act con una tutela penale individuale, ma affida a due autorità distinte una vigilanza che i decreti attuativi dovranno coordinare.

Diffusione e capacità di assorbimento

I dati Eurostat mostrano una diffusione rapida ma diseguale: la quota di imprese UE-27 che utilizza almeno una tecnologia di IA passa dal 13,5% (2024) al 20,0% (2025); l'Italia cresce dall'8,2% al 16,4%, restando sotto la media europea e distante dalle capofila Danimarca (42,0%) e Finlandia (37,8%). Il divario non è riducibile alla sola dotazione tecnologica: pesano la dimensione media d'impresa, le competenze digitali e la struttura settoriale. Sul piano microeconomico, gli studi sperimentali convergono su un effetto positivo ma eterogeneo: Noy e Zhang (2023) rilevano guadagni di produttività e una riduzione della dispersione qualitativa nella scrittura professionale; Brynjolfsson, Li e Raymond (2025) osservano benefici concentrati sui lavoratori meno esperti; Dell'Acqua et al. (2023) introducono la nozione di frontiera frastagliata, secondo cui l'assistenza dell'IA migliora i compiti interni alla frontiera delle capacità del modello e peggiora quelli esterni. Sul piano macroeconomico, Acemoglu (2024) stima un incremento della produttività totale dei fattori modesto, dell'ordine dello 0,7% in dieci anni. Applicata alle industrie creative, la frontiera frastagliata suggerisce una previsione verificabile: guadagni elevati nella produzione di contenuti standardizzati e brevi, perdite nella produzione che richiede continuità narrativa, coerenza causale e responsabilità editoriale.

Discussione critica

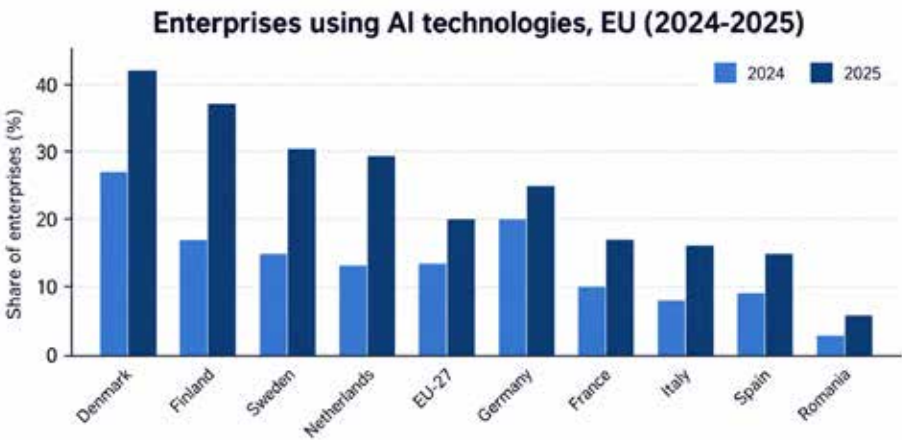
Il primo nodo riguarda il rapporto tra la fedeltà percettiva e la comprensione del mondo. Le leggi di scala promettono prevedibilità della loss, non delle capacità; nel video promettono realismo del fotogramma, non correttezza della dinamica. I *benchmark* di senso comune fisico mostrano che la seconda non segue dalla prima. Finché la comunità non adotterà misure continue, calibrate e multidimensionali, le affermazioni sui «simulatori del mondo» resteranno epistemicamente fragili. Un secondo nodo emerge nel rapporto tra abbondanza e fiducia. L'abbattimento del costo della produzione audiovisiva è insieme alla democratizzazione e all'inflazione. Il valore informativo di una registrazione dipendeva dal suo costo di falsificazione; azzerato quel costo, la fiducia deve essere ricostruita per via istituzionale e infrastrutturale — provenienza, responsabilità editoriale, verifica ibrida — anziché per via percettiva. Il dividendo del mentitore mostra che l'inazione ha un costo asimmetrico, che ricade sulle prove autentiche. Un ulteriore elemento critico riguarda il rapporto tra trasparenza formale e verificabilità sostanziale. L'articolo 50 impone la marcatura, ma la marcatura è fragile e volontariamente rimovibile, e i suoi obblighi si applicano ai fornitori conformi, non agli attori malevoli né ai modelli a pesi aperti eseguibili localmente. Il rischio è un regime che grava sugli operatori leciti senza incidere sull'uso illecito. La risposta plausibile non è normativa in senso stretto ma sistemica: rendere l'assenza di credentials di provenienza un segnale negativo interpretabile dal pubblico, il che richiede l'adozione universale lungo la catena di distribuzione. La riflessione si estende, infine, al rapporto tra generazione e progettazione. L'introduzione del video generativo nei processi creativi non si limita ad accelerare la produzione di rappresentazioni, ma modifica il momento stesso in cui le alternative progettuali vengono formulate, visualizzate e selezionate. La possibilità di ottenere in pochi secondi molteplici versioni di uno scenario sposta una parte del lavoro dalla costruzione diretta dell'immagine alla definizione delle condizioni generative, alla valutazione degli output e alla loro correzione progressiva. In questo quadro, prompt, iterazione e selezione assumono una funzione progettuale, ma non trasferiscono automaticamente al modello l'intenzionalità dell'autore. Il sistema genera configurazioni plausibili sulla base delle regolarità statistiche apprese, mentre resta al progettista il compito di stabilirne la pertinenza, la coerenza e il valore rispetto agli obiettivi del progetto. Tale distinzione diventa particolarmente rilevante quando il video sintetico viene utilizzato per anticipare spazi, prodotti, materiali, interazioni o comportamenti che non esistono ancora. In questi casi il realismo della rappresentazione può produrre un effetto di validazione anticipata: ciò che appare visivamente convincente rischia di essere interpretato come tecnicamente possibile, materialmente coerente o progettualmente risolto. La generazione audiovisiva richiede quindi una nuova alfabetizzazione progettuale, capace di distinguere tra plausibilità percettiva e fattibilità, tra variazione formale e innovazione, tra simulazione narrativa e verifica. L'IA generativa

sotto/below: Eurostat, «Use of Artificial Intelligence in Enterprises», Statistics Explained, 2026 (indagine comunitaria sull'uso delle TIC nelle imprese) / Eurostat, 'Use of Artificial Intelligence in Enterprises', Statistics Explained, 2026 (Community survey on the use of ICT in enterprises)

the very moment at which design alternatives are formulated, visualised and selected. The ability to obtain multiple versions of a scenario in a matter of seconds shifts part of the work from the direct construction of the image to the definition of generative conditions, the evaluation of outputs and their progressive refinement. In this context, prompts, iteration, and selection serve a design function but do not automatically transfer the designer's intent to the model. The system generates plausible configurations based on learned statistical regularities, whilst it remains the designer's task to determine their relevance, coherence and value in relation to the project's objectives. This distinction becomes particularly relevant when synthetic video is used to anticipate spaces, products, materials, interactions or behaviours that do not yet exist. In such cases, the realism of the representation can produce an effect of anticipatory validation: what appears visually convincing risks being interpreted as technically feasible, materially coherent or design-wise resolved. Audiovisual generation therefore requires a new form of design literacy, capable of distinguishing among perceptual plausibility and feasibility, formal variation and innovation, and narrative simulation and verification. Generative AI can expand the scope of possibilities, but creative value does not equate to the number of alternatives produced; it depends on the critical ability to guide, select, and transform them into intentional decisions. From this perspective, even the provenance of content takes on a distinctly design-related dimension. Making the synthetic origin of a visualisation recognisable is not merely a regulatory requirement but also part of the project's communicative responsibility. The more the boundary between document, simulation, and design anticipation becomes perceptually indistinct, the more transparency about the conditions of the image's production becomes integral to its communicative integrity.

Research agenda

The research agenda emerging from the analysis centres, first and foremost, on the need to develop calibrated, multidimensional metrics, replacing FVD with grids validated against human judgments and introducing a declaration of the computational budget for inference—denoising and resolution steps—as a condition for reproducibility. At the same time, it appears necessary to extend benchmarks such as VideoPhy and VBench-2.0 with counterfactual experiments to distinguish statistical interpolation from causal representation. A further area concerns the robustness of provenance, through a systematic adversarial evaluation of watermarks and metadata subjected to



può ampliare lo spazio delle possibilità esplorabili, ma il valore creativo non coincide con il numero di alternative prodotte: dipende dalla capacità critica di orientarle, selezionarle e trasformarle in decisioni intenzionali. Da questa prospettiva, anche la provenienza dei contenuti assume una dimensione propriamente progettuale. Rendere riconoscibile l'origine sintetica di una visualizzazione non costituisce soltanto un adempimento normativo, ma anche una componente della responsabilità comunicativa del progetto. Quanto più il confine tra documento, simulazione e anticipazione progettuale diventa percettivamente indistinto, tanto più la trasparenza sulle condizioni di produzione dell'immagine diventa parte integrante della sua correttezza comunicativa.

Agenda di ricerca

L'agenda di ricerca che emerge dall'analisi si articola innanzitutto sulla necessità di sviluppare metriche calibrate e multidimensionali, sostituendo FVD con griglie validate rispetto ai giudizi umani e introducendo la dichiarazione del budget di calcolo in inferenza — passi di denoising e di risoluzione — come condizione di riproducibilità. Parallelamente, appare necessario estendere *benchmark* quali VideoPhy e VBench-2.0 mediante interventi controfattuali, così da distinguere l'interpolazione statistica dalla rappresentazione causale. Un ulteriore ambito riguarda la robustezza della provenienza, attraverso una valutazione avversariale sistematica di filigrane e metadati sottoposti a ricodifica, ritaglio, ripresa da schermo e attacchi mirati, accompagnata da metriche standard di sopravvivenza del segnale. Sul piano comunicativo saranno inoltre necessari studi longitudinali sull'erosione del valore probatorio del video nella percezione del pubblico e sull'incidenza empirica del dividendo del mentitore nei contesti elettorali e giudiziari. A ciò si aggiunge l'esigenza di tradurre gli obblighi previsti dagli artt. 50-56 in test riproducibili, rendendo la conformità un oggetto scientifico e non esclusivamente documentale. Infine, occorrono studi panel sulle imprese italiane della comunicazione che mettano in relazione l'adozione, le competenze e la produttività, superando i limiti dei disegni sperimentali di breve periodo.

Conclusioni

Il paradigma della scala ha prodotto, in meno di un decennio, la trasformazione tecnologica più rapida della storia dell'informatica e, nel giro di quattro anni, ha esteso la sintesi generativa al dominio audiovisivo. I limiti epistemici — insensibilità delle metriche al movimento, incoerenza causale delle scene, inevitabilità statistica delle allucinazioni — indicano che il progresso ulteriore dipenderà dalla qualità della valutazione tanto quanto dalla quantità di calcolo. I limiti comunicativi mostrano che l'effetto socialmente rilevante non è tanto il singolo falso quanto l'erosione della registrazione come sostegno epistemico della testimonianza pubblica. I limiti normativi, con l'articolo 50 dell'AI Act e la Legge 132/2025, spostano il criterio di successo dalla prestazione media alla verificabilità dell'origine. Sul piano creativo e progettuale, la stessa trasformazione impone di distinguere la capacità generativa da quella di progetto: la moltiplicazione delle alternative e l'aumento del realismo non sostituiscono l'intenzionalità, la selezione critica e la responsabilità dell'autore, ma ne ridefiniscono gli strumenti e il campo d'azione. La domanda decisiva per il prossimo decennio non è quanto realistici possano diventare i video generati, ma quanto affidabilmente sapremo attribuirne l'origine, misurarne i limiti e governarne la circolazione.

recoding, cropping, screen capture and targeted attacks, accompanied by standard signal survival metrics. In terms of communication, longitudinal studies will also be required on the erosion of video's probative value in the public's perception and on the empirical impact of the liar's dividend in electoral and judicial contexts. Added to this is the need to translate the obligations set out in Articles 50–56 into reproducible tests, making compliance a scientific rather than a purely documentary matter. Finally, panel studies are needed on Italian media companies to establish a link between adoption, skills, and productivity, thereby overcoming the limitations of short-term experimental designs.

Conclusions

In less than a decade, the scale paradigm has brought about the most rapid technological transformation in the history of computing and, within four years, has extended generative synthesis to the audiovisual domain. Epistemic limitations — metrics' insensitivity to movement, causal inconsistency in scenes, and the statistical inevitability of hallucinations — suggest that further progress will depend as much on the quality of evaluation as on the quantity of computation. Communicative limitations show that the socially significant effect is not so much the individual fake as the erosion of the recording as an epistemic foundation for public testimony. Normative limitations, with Article 50 of the AI Act and Law 132/2025, shift the criterion for success from average performance to the verifiability of origin. On a creative and design level, this very transformation requires us to distinguish between generative capacity and design capacity: the proliferation of alternatives and the increase in realism do not replace the author's intentionality, critical selection and responsibility, but rather redefine their tools and scope of action. The crucial question for the coming decade is not how realistic generated videos can become, but how reliably we will be able to attribute their origin, measure their limitations and regulate their circulation.

References

- Acemoglu, D. (2024). The simple macroeconomics of AI (NBER Working Paper No. 32487). National Bureau of Economic Research. <https://doi.org/10.3386/w32487>

- Alibaba (Wan-Video Team). (2025). Wan: Open and advanced large-scale video generative models [Technical report].

- Bansal, H., Lin, Z., Xie, T., Zong, Z., Yarom, M., Bitton, Y., Jiang, C., Sun, Y., Chang, K.-W., & Grover, A. (2024). VideoPhy: Evaluating physical commonsense for video generation. *arXiv*. <https://arxiv.org/abs/2406.03520>

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>

- Blattmann, A., Dockhorn, T., Kulal, S., Mendelevitch, D., Kilian, M., Lorenz, D., Levi, Y., English, Z., Voleti, V., Letts, A., Jampani, V., & Rombach, R. (2023). Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv*. <https://arxiv.org/abs/2311.15127>

- Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. *The Quarterly Journal of Economics*, 140(2), 889–942. <https://doi.org/10.1093/qje/qjae044>

- Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107, 1753–1820.

- Coalition for Content Provenance and Authenticity. (2024). C2PA technical specification (Version 2.x). Linux Foundation Joint Development Foundation.

- Commissione europea. (2026). Guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of Regulation (EU) 2024/1689 (C(2026) 5054 final).

- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

- Dell'Acqua, F., McFowland, E., Ill, Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraye, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality (Harvard Business School Working Paper No. 24-013). Harvard Business School. <https://doi.org/10.2139/ssrn.4573321>

- Eurostat. (2026). Use of artificial intelligence in enterprises. *Statistics Explained*. Publications Office of the European Union.

- Fallis, D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology*, 34(4), 623–643. <https://doi.org/10.1007/s13347-020-00419-2>

- Ge, S., Mahapatra, A., Parmar, G., Zhu, J.-Y., & Huang, J.-B. (2024). On the content bias in Fréchet Video Distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *arXiv*. <https://arxiv.org/abs/2404.12391>

- Google DeepMind. (2025, May 20). Fuel your creativity with new generative media models and tools [Blog post].

- Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, 119(1), e2110013119. <https://doi.org/10.1073/pnas.2110013119>

- Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021). Measuring massive multitask language understanding. In *International Conference on Learning Representations*. <https://arxiv.org/abs/2009.03300>

- Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., & Fleet, D. J. (2022a). Video diffusion models. *Advances in Neural Information Processing Systems*, 35. <https://arxiv.org/abs/2204.03458>

- Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D. P., Poole, B., Norouzi, M., Fleet, D. J., & Salimans, T. (2022b). Imagen Video: High definition video generation with diffusion models. *arXiv*. <https://arxiv.org/abs/2210.02303>

- Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. D., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G. V., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., . . . Sifre, L. (2022). Training compute-optimal large language models. *arXiv*. <https://arxiv.org/abs/2203.15556>

- Huang, Z., He, Y., Yu, J., Zhang, F., Si, C., Jiang, Y., Zhang, Y., Wu, T., Jin, Q., Chanpaisit, N., Wang, Y., Chen, X., Wang, L., Lin, D., Qiao, Y., & Liu, Z. (2024). VBench: Comprehensive benchmark suite for video generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 21807–21818). <https://doi.org/10.1109/CVPR52733.2024.02060>

- Interactive Advertising Bureau. (2025). 2025 video ad spend & strategy full report.

- Italia. (2025). Legge 23 settembre 2025, n. 132: Disposizioni e deleghe al Governo in materia di intelligenza artificiale. *Gazzetta Ufficiale della Repubblica Italiana*, Serie Generale, n. 236, 10 ottobre 2025.

- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 248. <https://doi.org/10.1145/3571730>

- Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). Why language models hallucinate. *arXiv*. <https://arxiv.org/abs/2509.04664>

- Kang, B., Yue, Y., Lu, R., Lin, Z., Zhao, Y., Wang, K., Huang, G., & Feng, J. (2024). How far is video generation from world model: A physical law perspective. *arXiv*. <https://arxiv.org/abs/2411.02385>

- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv*. <https://arxiv.org/abs/2001.08361>

- Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023). A watermark for large language models. In *Proceedings of the 40th International Conference on Machine Learning (Vol. 202, pp. 17061–17084)*. *Proceedings of Machine Learning Research*.

- Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience*, 24(11), 103364. <https://doi.org/10.1016/j.isci.2021.103364>

- Kuaishou Technology. (2024). Kling: Text-to-video generation model [Product announcement].

- Meng, F., Liao, J., Tan, X., Shao, W., Lu, Q., Zhang, K., Cheng, Y., & Li, D. (2024). Towards world simulator: Crafting physical common-sense-based benchmark for video generation. *arXiv*. <https://arxiv.org/abs/2410.05363>

- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>

- OpenAI. (2024, February 15). Video generation models as world simulators [Technical report].

- Parlamento europeo & Consiglio dell'Unione europea. (2022). Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (Digital Services Act). *Gazzetta ufficiale dell'Unione europea*, L 277, 1–102.

- Parlamento europeo & Consiglio dell'Unione europea. (2024). Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act). *Gazzetta ufficiale dell'Unione europea*.

- Peebles, W., & Xie, S. (2023). Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. <https://arxiv.org/abs/2212.09748>

- Rini, R. (2020). Deepfakes and the epistemic backstop. *Philosophers' Imprint*, 20(24), 1–16.

- Schaeffer, R., Miranda, B., & Koyejo, S. (2023). Are emergent abilities of large language models a mirage? *Advances in Neural Information Processing Systems*, 36. <https://arxiv.org/abs/2304.15004>

- Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., & Villalobos, P. (2022). Compute trends across three eras of machine learning. *arXiv*. <https://arxiv.org/abs/2202.05924>

- Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O., Parikh, D., Gupta, S., & Taigman, Y. (2022). Make-A-Video: Text-to-video generation without text-video data. *arXiv*. <https://arxiv.org/abs/2209.14792>

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

- Zheng, D., Huang, Z., Liu, H., Zou, K., He, Y., Zhang, F., Zhang, Y., & He, J. (2025). VBench-2.0: Advancing video generation benchmark suite for intrinsic faithfulness. *arXiv*. <https://arxiv.org/abs/2503.21755>

- Zheng, Z., Peng, X., Yang, T., Shen, C., Li, S., Liu, H., Zhou, Y., Li, T., & You, Y. (2024). Open-Sora: Democratizing efficient video production for all. *arXiv*. <https://arxiv.org/abs/2412.20404>