Beyond Deepfakes: Synthetic Moving Images and the Future of History

Article · April 2024

CITATIONS
0

READS
27

1 author:

Williams Cole
The University of Arizona
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Beyond Deepfakes: Synthetic Moving Images and the Future of History

WR Cole

College of Applied Science and Technology (CAST)
University of Arizona
Phoenix, Arizona, United States of America

E-mail: wrossacole@gmail.com

Abstract: *This paper investigates the role of generative Artificial Intelligence (AI) tools in the production of synthetic moving images—specifically, how these images could be used in online disinformation campaigns and could profoundly affect historical footage archives. AI-manipulated content, especially moving images, will have an impact far beyond the current Information Warfare (IW) environment and will bleed into the unconsidered terrain of visual historical archives with unknown consequences. The paper will also consider IW scenarios in which new types of long-term disinformation campaigns may emerge and will conclude with potential verification and containment strategies.*

Keywords: *Deepfake*, *Synthetic Moving Images, Synthetic Media, Artificial Intelligence, Disinformation, Information Warfare, Historical Archives, Documentary Film*

Introduction

Synthetic media is real. In the current information environment, there are already debates around whether pieces of moving imagery should be verified as generated by AI tools or not. As these tools develop, the veracity of *every* piece of historical footage will demand a skeptical assessment: the Hindenburg burning, the Pearl Harbor attack, John F. Kennedy's assassination, the Apollo 11 moon landing, the Challenger space shuttle explosion, the Berlin Wall coming down, 9/11, Rodney King, George Floyd, January 6th, and the list goes on. Without a robust plan for verification and containment, those news and documentary clips that have defined collective memory, have recorded human achievement and misery, and have sparked social movements will be increasingly vulnerable to, and central in, disinformation campaigns that utilise altered historical records.

Due to the exponential pace of AI technology that allows almost anyone to generate synthetic visual versions of history, the implicit 'documentary proof' in most historical images will be compromised. While doctored or altered imagery—sometimes in the service of propaganda—has been present in societies since the beginning of the visual medium, the rapid technical development and open access to tools to do so are new. This raises profound questions: how can the proof in historical archival footage be protected and preserved? How can inaccurate synthetic versions of 'what happened' in history be contained, prevented, or contextualised? How will synthetic

historical footage be used in disinformation campaigns? And these developments may catalyze a profound epistemological crisis about how to verify visual history and what is 'real'.

As a documentary film professional with extensive experience in utilizing historical archival footage to create and enhance narratives, along with having a background in academic media and communications research, the author sees these developments from a unique vantage point. An interest in generative AI tools caused the author not only to see how these developments will affect production but also, specifically, the potential ramifications they have on the historical record.

This paper will outline the current state of generative AI technology in relation to creating synthetic moving images. It will then show how such imagery has manifested in the information environment so far, and then will go on to tackle the potential future of synthetic historical footage in the realm of disinformation. It will also argue that, without protocols in place, such material is likely to rapidly seep into commercial and governmental footage archives that act as repositories for collective history. Finally, it will consider potential measures of containment, verification, and authentication that need to be developed quickly to preserve accurate historical accounting.

It is important to first define some terms. The term 'deepfake' refers to a video, photo, or audio recording that seems real but has been manipulated with AI (GAO 2020). While mis-, dis-, and malinformation are still terms that are not universally agreed on, there are common concepts that define each category (Baines & Elliott 2020). If there is *no intention* to deceive but the information is false, such as forwarding a deepfake video that the sharer believes is real, then it is 'misinformation'; if the information is correct but used for deceptive ends, such as reposting a piece of old footage and saying it is from a current event and thereby reconfiguring the truth, it can be classified as 'malinformation'; and, finally, 'disinformation' involves *an intention* to deceive with false information. An intentional campaign can use all three methods (Baines & Elliott 2020; De Cock Buning 2018).

Because this paper draws from Information Warfare scholarship, it is also important to define that term. It is still a relatively poorly understood concept in the United States to the extent that it currently has no official U.S. government definition (Theohary 2018, 2023). Concepts of Information Warfare date back millennia, yet the pace of technology has radically expanded the information domain and how it can impact the operational environment (Bastian 2019). Generally, information warfare "is a struggle over the information and communications process, a struggle that began with the advent of human communication and conflict" (Lewis 1997). It includes multiple types of information manipulation and similar concepts—like active measures, hybrid warfare, gray zone warfare, soft power, and even public diplomacy, which enable a competitive advantage over an opponent (RAND 2023). Furthermore, these activities take place within an information environment that includes three layers: i) the physical infrastructure; ii) informational networks and systems; and iii) a cognitive layer that denotes the minds of people who transmit and respond to information (Theohary 2018). The manipulation of history, identity, and other effects of synthetic moving images fall into the cognitive layer and that is what this paper will focus on.

Where Are We? Artificial Intelligence and Synthetic Moving Images

Artificial Intelligence has been in development since the 1950s. Since 2010, breakthroughs have accelerated with machine learning advancements in neural networks and deep learning, powered

by exponential computer processing power (Muthukrishnan *et al.* 2020). Major achievements include, among others, image recognition systems and natural language processing (NLP), which is the basis for large language models (LLMs) like Chat GPT, the commercial AI service launched in late 2022. LLMs are complex AI models trained on huge amounts of text, resulting in the ability to generate human-like conversational interactions. All AI systems require massive, diverse sets of training data for the models to create outputs, predictions, and imagery (Rillig *et al.* 2023).

One AI system that has proved proficient in creating synthetic imagery is the generative adversarial network model (GAN). GANs use two neural networks working together, one to compose the image from training data and the other to 'discriminate' or change it, until it reaches the target of a prompt—often from text descriptions (Gonog & Zhou 2019). The other main category of image generation approach is Variational Auto Encoders (VAEs), although they are generally considered less effective (Coccomini *et al.* 2023). As these models evolve—trained on increasingly massive datasets—their ability to create synthetic images scales exponentially until they can virtually create any image. Moving images are essentially multiple frames of still images strung together and, as systems become more powerful, the continuum from a synthetic still image to a clip or piece of footage becomes easier and more realistic looking. Development is moving quickly (Aldausari *et al.* 2022) and the increasing power of current AI hardware training modules, such as NVIDIA's DGX H100—a system that runs five times as fast as the hardware that trained ChatGPT—points toward rapid development and capabilities (Witt 2023).

While special effects and Computer-Generated Imaging (CGI) have existed for decades, it has largely been limited to Hollywood, given the substantial cost and complexity required (Kertysova 2018). In comparison, new GAN-based AI tools, like OpenAI's Dall-e and Midjourney, cost relatively little for a consumer to use. While current models target creating still images and artistic imagery, established corporations like Meta and start-ups like Pika Labs are developing text to synthetic video generation (Metz 2023). Consensus in the generative AI technology world suggests that synthetic video generation will be available to the public soon (NVIDIA 2023; Witt 2023). Sora, a generative text-to-video GAN-based diffusion model that OpenAI previewed in early 2024, exemplifies how fast the technology is developing. Sora can create high-definition videos up to a minute long from text prompts while making great strides in issues like consistency and movement that plagued earlier models (Heaven 2024).

In addition, one of the most developed arenas in synthetic media is in voice generation, often an important element to creating realistic moving image clips. Companies like Eleven Labs not only allow fully synthetic voice generation but also the ability to 'clone' voice samples, thereby making the chosen voice say anything one desires via text prompts (Hu & Zhu 2023). This technology has developed significantly since WaveNet was introduced in 2016 (Google DeepMind 2016) and is used by phone scammers, for example. Bruce Reed, the Biden administration's AI strategy point person, has said that "voice cloning is one thing that keeps me up at night" (Veranasi 2023). In fact, a deepfake of Biden's voice during the 2024 New Hampshire Democratic primary urging voters not to cast ballots shocked disinformation experts causing one to state that "The political deepfake movement is here" (Murphy 2024). The point when publicly available tools can create synthetic moving image clips (with sound) of whatever one may desire is fast approaching.

AI and Disinformation—What Does the Future Look Like?

Disinformation, misinformation, and other forms of information manipulation have been part of

human society since the beginning of recorded history (Theohary 2018; Rid 2020). Forgeries, conveying false information to create advantage, and sowing distrust in institutions and individuals are typical information warfare tactics and strategies. False or misleading imagery in information campaigns date back to the advent of photography (Bravo 2017; Farid 2024) but, in the pre-digital age, 'faked' photos were usually a result of staged sets or manipulation in the photographic darkroom, a relatively involved and expensive process. In the digital era, software programs like Adobe Photoshop offered cheaper and more accessible ways to alter imagery. Other image manipulation methods include recontextualizing older imagery to promote an agenda, a common tactic present in information manipulation today.

The growth of social media and the digital environment in the last decades have created a sea-change in the ability for information manipulation to prosper: bad information travels faster and goes viral before fact-checkers and published corrections can get ahead of it; opaque algorithms prioritise engagement and clicks which, in turn, create incentives for ever more emotional and outrageous content; and filter bubbles reinforce perspectives and feed individuals increasingly extreme information with a particular perspective (Fitzer 2022). Moreover, recent studies have shown that AI-based technologies have quelled freedom of expression on the Internet and have boosted digital oppression (Ryan-Mosely 2023). A single AI program can now write as much text as all of humanity, and a relatively small image-generation model running on a laptop can compress all the pictures on the open web into a tool that generates images with creativity and precision (Suleyman & Bhaskar 2023). This environment has created a "disinformation crisis" that is largely unprecedented in its scope and rapidity (Sadiq & Saji 2022). And imagery, while only one component of the overall problem, is playing an increasingly central role due to the powerful ability of still and moving images to reach a cognitive and emotional part of human experience (Alpuim & Ehrenberg 2023). Imagery is a central component of narrative structure with the capacity to influence thoughts, emotions, and behavior—in effect, how humans often understand the world (Grall & Finn 2022). Regarding documentary and news footage, there persists an expectation of truth and reality, that it says something about the world and helps to make a coherent, emotional narrative of history (Bondebjerg 2014).

Newer generative AI models have supercharged both the ability to alter imagery and to create completely synthetic imagery. Manifestations of altered and synthetic imagery powered by AI technology surfaced soon after AI frameworks became accessible. For example, the term 'deepfake' first appeared on Reddit around 2017 by a user who utilised a machine learning algorithm, publicly available videos, and a home computer to swap celebrity faces onto pornographic performers (Cole 2018). This was followed by viral deepfakes of celebrities, politicians, and everyday people. Deepfake pornography, where users put faces scraped from social media onto bodies in pornographic scenes and use this for online harassment, has also proliferated (Reissman 2023). One high-profile example occurred in January 2024 when a pornographic deepfake of worldwide celebrity Taylor Swift racked up almost 50 million views before it was taken down (Saner 2024).

Deepfakes and AI-manipulated imagery are also increasingly used on the political front. For example, a deepfake video of Ukrainian President Volodymyr Zelenskyy 'surrendering' was traced back to a Russian disinformation initiative (Atlantic Council 2022); high-profile examples of viral synthetic still images include the Pentagon burning, which briefly panicked the New York Stock Exchange, and a series of AI-generated images of Donald Trump being arrested momentarily put AI capabilities in the spotlight. More recently, image-based disinformation in the Israel-Hamas

war has featured both synthetic images and re-contextualised images and audio (RumorGuard 2023). In addition to the powerful potential of OpenAI's Sora model, companies like Synthesia can create realistic looking avatars that have been used as news anchors to attempt to legitimise mis- and disinformation (Ryan-Mosely 2023). Likely, the rapid pace of synthetic imagery creation will fundamentally curtail the implicit documentary proof of any image, and how this development affects historical imagery may be profound.

The Potential Future

Consider the sheer amount of 'historical' imagery out there, not scripted or staged in the conventional sense of feature films and television series. Since the first moving image in 1888, humans have turned that recording technology on the world around them, producing millions of hours of material. And while a small amount has surely been faked or staged—and there are indeed valid epistemological debates around 'objectivity' and whether the camera lens can truly capture anything without bias (Godmilow & Shapiro 1997)—the vast majority of this footage is simply the camera trained on live action moments, be they pedestrian, beautiful, heartwarming, chaotic, or violent. Visual historical memory encompasses everything from iconic clips of momentous events to personal home movies, spanning the witnessing of both genocide and war, as well as capturing the triumphs of the human spirit.

Over the decades, footage archives have been formed by governmental, educational, and commercial entities to preserve the historical record and to create mechanisms to license footage to news and documentary productions. Globally, there is a lot of it: the overall state of visual archives worldwide could be defined as multi-faceted, complex, and generally unorganized by any shared codified standard. In the United States, for example, huge news and documentary archives are held by ABC News VideoSource, CNN ImageSource, NBC News Archives, and CBS News Archives. There are government-run entities like the National Archives and the Library of Congress, university-centered archives like those at UCLA and Indiana University, and there are hundreds of smaller archives—many of which have acquired one another, have shut down, or which have recycled various imagery (LOC 2022).

As Internet speeds became faster in the 2010s, new footage archives grew online, especially those like Pond5 and Shutterstock that offered quick and low-cost footage clips to anyone who wanted them and also allowed the uploading of licensable content by registered users. Many archives are represented by behemoths like Getty Images, which aggregates from multiple archives or buys smaller ones, markets the material, and then creates an online platform for licensing. In the contemporary landscape, the most accessible archives are largely those where governments have mandated protection of a nation's visual history or those who are motivated by commercial interests.

The variety of archives, perpetual resource constraints, and multiplicity of pre-digital formats (like U-matic ¾ inch, 1 inch, BetaSP, and MiniDV, for example) make it difficult to get an exact understanding of how much material still needs to be transferred into the digital sphere. 'Digitizing'—the act of transferring outdated physical media into a digital codec—is the present standard for material to be viewed, licensed, and theoretically used in a plethora of digital content throughout multiple conduits. Yet digitization also brings this material out of the physical capsules of provenance it was recorded on into the realm of ones and zeros, where veracity and authentication are still a goal rather than an inherent practice. Using AI-powered tools, moving

imagery in the digital realm is vulnerable to being re-created, altered, and regenerated. New, realistic looking 'archival' footage can be created, or existing historical footage can be manipulated with details and components that give it different meaning. This synthetic or altered footage can then potentially re-enter archives as rewritten versions of history.

Consequences and Long-Game Disinformation

There will be two separate but interrelated concerns because of these emerging technologies. First, the ability to essentially 'create history' comes with exciting but also potentially uncontrollable consequences. It is entirely possible that news and documentary film producers will embrace this technology to create accurate—or inaccurate—footage to illustrate narratives, thereby avoiding the hassle and expense of licensing footage from archives. This could also lead to a blossoming of storytelling as narratives of marginalised communities, never covered by news organizations, are 'brought to life' by the creation of realistic synthetic archival historical footage. But this may also act to further erode the sense of documentary proof that is currently present in content, where historical footage licensed from established archives is used to represent the past. This lack of implicit proof, in turn, may dissuade producers from using real archival footage as a valid storytelling method and, consequently, the synthetic generation of 'historical' footage may run amok.

The second concern is twofold: i) the need for widespread and effective tools to detect AI-generated content, and ii), perhaps more viable, the urgent need for standardised, vetted, and accessible methods where archives can authenticate the important historical footage they already have. In this way, archival footage that is an accurate (albeit not perfectly objective but directly recorded) representation of history will be protected. The 'infection' of synthetic historical material into established archives would have profound consequences and the rise of online footage portals—such as Wirestock, Adobe Stock, Pond5, and Shutterstock—where almost anyone can upload and tag footage to sell, may exacerbate this concern and could potentially add to mis- and disinformation.

Currently, generative AI image tools are increasingly being used in advertising, illustration, public relations, engineering, design, and other industries (GAO 2023). But they are poised to add a powerful element to information manipulation campaigns and strategies. A 2023 RAND study focusing on Chinese cognitive Information Warfare strategies outlined how these tools—in their ability to quickly generate photo-realistic images of people that do not exist—may play a role in a new wave of social media disinformation campaigns. For example, social media networks could be populated with 'friends' that are essentially supercharged bots—nonexistent individuals that can show themselves in synthetic photos with family, taking trips, camping, whatever—paired with text posts generated by LLMs that, at scale, could be used to spread disinformation in subtly manufactured ways (Marcellino *et al.* 2023).

While creating swarms of synthetic bots on social media to sway opinion is one thing, it is also crucial to imagine previously unthinkable scenarios of longer game disinformation that use moving image generation tools to rewrite history. For example, consider a scenario where ethnic, social, and political rivalries are exploited by using perceived historical injustices—seen in synthetic moving imagery—as a motivating tool. Unfortunately, the world is rife with ethnic and racial conflict with deep historical roots where volatile histories can be exploited. A short synthetic film—or one that mixes real and fake archival footage—which highlights atrocities that never happened, could be posted and shared, bypassing fact-checking media outlets and igniting emotion

and vengeance. These are scenarios where easy-to-create imagery could be easily weaponised since moving images hold the power of eliciting emotion quickly, especially if those images are geared to provoke outrage (Alpuim 2023).

Or, in a different scenario, additional 'footage' could be 'discovered' of an iconic and conspiracy-laden event, like the assassination of JFK or January 6th. State and non-state actors could use these tools to seed alternative clips of history around the Internet and even hack, implant, or legitimately upload 'historical' synthetic clips on archival footage sites. If downloaded and licensed, this footage could take on a life of its own in content of all kinds, perhaps eventually ending up in historical archives—official or not. These are just some examples of the possibilities, some of which are chilling in their long-term damage to historical truth and the potential to help sow discord and instability.

These possibilities will fit perfectly into the cognitive realm of information warfare strategies of state and non-state actors. China has made no secret of its interest in developing strategies to weaken perceptive abilities, to disseminate and shape information, and to alter and control decision making in new and creative ways (Jones *et al.* 2023). China's well documented interest in 'intelligentised' warfare explicitly includes the use of generative AI to help leverage cognitive dominance, a strategy some would call "5th Generation Warfare" (Schreckinger 2022). The use of synthetic historical video seeded throughout the Internet could be a powerful tool over months and years to help dominate long-term revisionist narratives and to re-orient historical documentation by a variety of motivated actors.

Russian information warfare strategies could also get a significant boost. If the objective is to alter the perception of information, to eliminate objective truth, to destabilise institutions, to destroy empirical knowledge, and to create fast and voluminous output, then synthetic moving imagery tools are a dream come true (Mullaney 2022). The infamous Russian "firehose of falsehood" (Paul & Matthews 2016) would also benefit from these tools, including chatbots that can generate better writing in English and other languages. Though less fully defined around cognitive strategies, U.S. intelligence agencies could utilise these tools as well, especially as conflicts in the Middle East and Ukraine strain current geopolitical alliances. With wide access to powerful generative tools and an action plan, smaller state and non-state actors could conceivably sow confusion and instability by manipulating and leveraging historical narratives.

Historical interpretations in the future could very well be mixed with or based on synthetic imagery; false or misleading cultural narratives could have more 'evidence' to use in social and political manipulation; and any cohesive public perception of historical events could become a thing of the past.

Ethical AI, Containment, and Verification

To mitigate this expansive problem, first and foremost, the overarching orientation of AI technology development must be guided with ethical frameworks and authentic audits that consider society and culture. AI Now, for example, has developed a more comprehensive Algorithmic Accountability protocol while AI for Good aims to develop standardized ethics via audits. These guidelines and regulations should incentivise 'human-first' development, where the models and algorithms are transparent and remedial and include guardrails around generative AI tools. The EU is actively developing these frameworks (EU 2023) while the U.S. is catching up to addressing AI's

transformational impact on society (White House 2023). Ethical AI development is an issue that is global in nature and there are also numerous burgeoning initiatives and resources that aim to bring together industry, government, academic, and civil society organizations to champion developing AI systems that are responsible, transparent, and inclusive (World Economic Forum 2024). For example, UNESCO's work on AI ethics and governance stems from the Recommendation on the Ethics of Artificial Intelligence that was adapted by 193 countries in 2021 and which set the first global normative framework to develop AI that would "benefit humanity" (UNESCO 2024).

More specifically, there are a variety of detection and authentication initiatives in the works, like aiornot.com, that claim to detect AI-generated imagery. And large tech companies like OpenAI, Alphabet, and Meta have also made voluntary commitments to implement measures like developing a "watermark" on all forms of AI generated content, so users will know that AI has been used to create it (Bartz & Hu 2023). And there are other watermark-oriented mechanisms being developed: Google DeepMind, for example, announced SynthID, a tool that would embed a digital watermark directly into the pixels of the image, making it imperceptible to the average viewer but detectable for verification that it is AI generated (Gowal & Kohli 2023); PhotoGuard, a tool developed by MIT researchers, employs pixel alterations with an image to disrupt an AI algorithm's ability to ingest an image and alter it (Brodsky 2023). See McFarland (2023) for more initiatives.

One umbrella approach led by the software company Adobe—and that a consortium of technology and media companies like the BBC and *The New York Times* has adapted—is the Coalition for Content Provenance and Authenticity (C2PA) standard*,* led by the Content Authenticity Initiative (CAI). This initiative would embed immutable metadata at the point of creation of an original image that would then transfer through any changes and edits, thereby allowing anyone to verify the source and authenticity of the image at any point (CAI 2023). New digital cameras are also being released with Content Credentials at the point of capture, a development that may spread through image-creation technology (Lyonn 2023). A related idea would use blockchain technology to cryptographically seal metadata into the image file where any tampering would break the digital signature, essentially mitigating authenticity (Ozair 2023).

Yet any kind of magic bullet is out of reach: detection tools are complicated by low-resolution images and an overreliance on technical traits rather than contextual clues (Thompson & Hsu 2023; Knibbs 2023); CAI requires wide compliance and effective use largely in Adobe software and, given wide array of archival imagery that is not yet verified, it is difficult to see how it might be a 'go to' for authentication (Wilson 2023); watermarks that would identify AI generated images have already been added to real images in an attempt to discredit them. Deepfake detection models and methods, while showing some promise, are facing challenges in real-world situations where they need to have a more generalised detection ability and the problem that they often work for a limited amount of time needs to be addressed (Coccomini *et al.* 2023). Blockchain technology requires massive power and storage capabilities, especially with video. And something as simple as a screen recording before use of imagery in disinformation campaigns, for example, could bypass any embedded provenance or authentication.

Another idea for containment may be a strategy where large commercial models limit certain kinds of synthetic moving images, such as they do with violence and pornography. The control of the output is theoretically in the hands of those running the model so 'bad actors' could be suppressed by content moderation. For example, models might limit the generation of outright 'historical'

footage of certain kinds like that based on the aesthetics of certain eras like WWII. OpenAI, for example, claims that its groundbreaking text-to-video model Sora will have robust text classifiers that will check and reject text input prompts that are in violation of their usage policies (OpenAI 2024). Yet this is still a challenging proposition, as the restriction of certain prompts and outputs in the AI models would not only affect commercial viability but would also affect free speech issues. In addition, "prompt hacking" and related manipulations regarding text inputs are already a concern with Large Language Models (Mittal 2023).

The most effective limit may be those inherent in the technology itself. Generative AI models using neural networks with billions of parameters cost a lot of money to run and use a lot of energy. It is estimated that a LLM like ChatGPT, for example, requires hundreds of thousands of dollars per day to run and that cost will only grow as the parameters of the model grow (McQuate 2023). The testing and vetting of powerful models before they are released to the public may be an option that helps contain some of the worst-case scenarios. OpenAI, for example, claims to be adversarially testing Sora with experts in areas like misinformation, hateful content, and bias before it releases it to the public (OpenAI 2024).

Yet, the larger epistemological problem at this point may be the 'poisoning of the pool', whereby the very idea that synthetic imagery is possible has compromised trust in *all* imagery (Hsu & Thompson 2023). Unless verification tools and standards catch up, this may be the biggest sea-change of all: the widespread undermining of a psychological perception of 'truth'. There is also the possibility that deepfakes and synthetic video will not enter the more consequential domains of disinformation but, rather, will remain as memes, jokes, pornography, and the like and that the "doomsayers" and "alarmists" may be just that (Immerwahr 2023). But the technology is developing daily, and it is still unclear what the capabilities will be.

Perhaps then, an immediate preliminary guardrail would be to understand what footage archives are doing, if anything, to 'protect' historical footage in the face of these developments. A survey of the landscape of institutional and commercial archives should take place to develop a baseline understanding—including potential policy steps around protection—while technical solutions are developed that could authenticate, verify, and contain important historical archival footage. If developed in an accessible, affordable, and user-friendly way, this 'protection' strategy could offer a long-term solution for archives that would take a proactive approach to separate the authentic from the coming barrage of inauthentic and manipulated imagery. For example, Starling Lab is developing advanced cryptography and Web3 concepts utilizing blockchain technology to capture, store, and verify original material in ways where it would be immutable and decentralized.

In this scenario, commercial and governmental archives, for example, must prioritise certain historical footage that would then be imported and given a cryptographic stamp. It would then be exported to a decentralised immutable online ledger. Material could also be validated via provenance to the original source material (such as film, tape, or cassette) as authentic and traceable historical footage. While the technology to do this is not fully formed, it would give content-makers, such as documentary filmmakers, archival producers, and others looking for vetted historical footage a verified option that could be manifested as an official label or via a third-party audit. The archives of the future will demand footage that is 'authenticated' historical representation. And it is not only content creators that will need this: national educational curriculum; evidentiary corroboration in legal processes; the historical integrity of nationhood—the stability of large parts of the overall

social project may be compromised if there is not some kind of protocol developed sooner than later.

With so much material not even digitised, the resources, time, and storage to do this are not insignificant. But there must be momentum to ensure that footage archives—although unorganised, underfunded, and unstable—do not remain open to such transformative and consequential vulnerabilities.

Conclusion

AI-generated synthetic moving images will soon become easy, cheap, and ubiquitous. The idea that 'seeing is believing' is over has wide and profound repercussions for the information landscape. Complex changes that the 'AI revolution' will bring to societies all over the world are just beginning to seep into channels of influence and understanding. That one can no longer give recorded imagery an automatic seal of documentary proof will affect the information landscape, the legal protocols that underpin liberal democracies, and many other aspects of politics, industry, and society. The potential uses of synthetic imagery in Information Warfare and disinformation are far-reaching in a time when all imagery cannot be easily vetted either by third parties or individuals using tools or critical faculties. It will require training and change over generations to shed an inherent believability in what is seen, and to transform that reflex into an intrinsic skepticism to defend against manipulation.

While the generation of synthetic imagery may be a boon for filmmakers and creatives striving to create compelling visual narratives of events never recorded by conventional news and documentary cameras, this potentially positive scenario still begs the question: how will visual history be verified and authenticated? How will synthetic visual history affect how societies unfold? While various initiatives in synthetic media detection tools hold promise, there is the danger that an 'arms race' will develop—especially regarding disinformation—between synthetic imagery generation and those invested in figuring out what is real and what is fake, a process with no real resolution. The development of robust verification protocols, especially for sensitive historical footage that is important for long-term and immutable authentication, will be crucial to develop. Above all, widespread education about the capabilities of synthetic image generation tools coupled with early implementation of critical media literacy skills on a global scale will be of increasing importance (Kertysova 2018).

Given the rapid development of AI-fueled synthetic image generation tools, these initiatives will not be a choice but a necessity. Synthetic imagery is real; it will be the new normal, and we must be ready.

References

Aldausari, N, Sowmya, A, Marcus, N & Mohammadi, G 2022, 'Video generative adversarial networks: A review', *ACM Digital Library*, vol 55, no. 2, pp. 1-25, viewed 10 March 2024, <https://dl.acm.org/doi/full/10.1145/3487891>.

Alpuim, M & Ehrenberg, K 2023, 'Why images are so powerful–and what matters when choosing them', 3 August, viewed 10 March 2024, <https://www.bonn-institute.org/en/news/psychology-in-journalism-5>.

Atlantic Council (Digital Forensic Research Lab) 2022, 'Russian war report: Hacked news program and deepfake video spread false Zelenskyy claims', 16 March, viewed 10 March 2024, <https://www.atlanticcouncil.org/blogs/new-atlanticist/russian-war-report-hacked-news-program-and-deepfake-video-spread-false-zelenskyy-claims/>.

Baines, D & Elliott, R 2020, 'Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic', ResearchGate, 4 April, pp. 8-12, viewed 25 January 2024, <https://www.researchgate.net/publication/341130695_Defining_misinformation_disinformation_and_malinformation_An_urgent_need_for_clarity_during_the_COVID-19_infodemic>.

Bartz, D & Hu, K 2023, 'OpenAI, others pledge to watermark AI content for safety', *Reuters*, 21 July, viewed 10 March 2024, <https://www.reuters.com/technology/openai-google-others-pledge-watermark-ai-content-safety-white-house-2023-07-21/>.

Bastian, ND 2019, 'Information warfare and its 18th and 19th century roots', *The Cyber Defense Review*, November 15, viewed 20 February 2024, <https://cyberdefensereview.army.mil/CDR-Content/Articles/Article-View/Article/2017786/information-warfare-and-its-18th-and-19th-century-roots/>.

Bondebjerg, I 2014, 'Documentary and cognitive theory: Narrative, emotion and memory', *Media and Communication*, vol. 2, no. 1, pp. 13-22, viewed 15 January 2024, <https://www.cogitatiopress.com/mediaandcommunication/article/view/17>.

Bravo, R 2017, 'Image manipulation practices throughout the history and evolution of photography', *Proceedings of the 23rd International Symposium on Electronic Arts*, ISEA2017, eds. J Arango, A Bubarno, F Londono & G Meji, Bogota, Columbia, viewed 15 February 2024, <https://www.researchgate.net/publication/333672206_Image_Manipulation_Practices_Through_the_History_and_Evolution_of_Photography>.

Brodsky, S 2023, 'How a new tool by MIT researchers counters AI image manipulation threats', *Lifewire*, 27 July, viewed 10 March 2024, <https://www.lifewire.com/mit-photoguard-fights-ai-image-manipulation-7566258>.

CAI 2023, 'How it works', Content Authenticity Initiative website, viewed 10 March, 2024, <https://contentauthenticity.org/how-it-works>.

Coccomini, D, Caldelli, R, Falchi, F & Gennaro, C 2023, 'On the generalization of deep learning models in video deepfake detection', *Journal of Imaging*, vol. 9, no. 5, p. 4, viewed 10 March 2024, <https://www.mdpi.com/2313-433X/9/5/89>.

Cole, S 2018, 'We are truly fucked: Everyone is making AI-generated porn now', *Vice Motherboard*, 28 January, viewed 9 March 2024, <https://www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley>.

De Cock Buning, M 2018, 'A multi-dimensional approach to disinformation: A report of the independent high-level group on fake news and online disinformation', *European Union*

Parliament News, 19 December, viewed 10 March 2024, < https://op.europa.eu/en/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1>

EU (European Union) 2023, 'EU AI act: First regulation on Artificial Intelligence', *European Parliament News*, 19 December, viewed 10 January 2024, <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>.

Farid, H 2024, 'Photo tampering throughout history', Hany Farid professional website, viewed 10 March 2024, <https://faculty.cc.gatech.edu/~beki/cs4001/history.pdf>.

Fitzer, G 2022, 'Combatting the social media disinformation crisis: Why reforming Section 230 is not the answer but legislation mandating transparency is', *Social Science Research Network (SSRN)*, 21 July, viewed 10 March 2024, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4162179>.

Gonog, L & Zhou, Y 2019, 'A review: Generative Adversarial Networks', *14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 16 September, Xi'an, China, viewed 5 March 2024, <https://ieeexplore.ieee.org/abstract/document/8833686>.

GAO 2020, US Government Accountability Office, GAO-20-379SP, *Science and Tech Spotlight: Deepfakes*, February, viewed 5 March 2024, <https://www.gao.gov/products/gao-20-379sp>.

—2023, US Government Accountability Office, GAO-23-106782, *Science & Tech Spotlight: Generative AI*, 13 June, viewed 5 March 2024, <https://www.gao.gov/products/gao-23-106782>.

Google DeepMind 2016, 'WaveNet: A generative model for raw audio', Google DeepMind website, 8 September, viewed 5 March 2024, <https://deepmind.google/discover/blog/wavenet-a-generative-model-for-raw-audio/>.

Gowal, S & Kohli, P 2023, 'Identifying AI-generated images with SynthID', Google DeepMind website, 29 August, viewed 5 March 2024, <https://deepmind.google/discover/blog/identifying-ai-generated-images-with-synthid/>.

Grall, C & Finn, E 2022, 'Leveraging the power of media to drive cognition: A media-informed approach to naturalistic neuroscience', *Social Cognitive and Affective Neuroscience*, 3 June, vol. 17, no. 6, pp. 598-608, viewed 10 March 2024, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9164202/>.

Godmilow, J & Shapiro, A 1997, 'How real is the reality in documentary film?', *History and Theory*, vol. 36, no. 4, pp. 80-101, viewed 10 March 2024, <https://www.jstor.org/stable/2505576>.

Heaven, WD 2024, 'OpenAI teases an amazing new generative model called Sora', *MIT Technology Review*, 15 February, viewed 10 March 2024, <https://www.technologyreview.com/2024/02/15/1088401/openai-amazing-new-generative-ai-video-model-sora/>.

Hsu, T & Thompson S 2023, 'A.I. muddles Israel-Hamas war in unexpected ways', *The New York Times*, 28 October, viewed 10 March 2024, <https://www.nytimes.com/2023/10/28/business/

media/ai-muddies-israel-hamas-war-in-unexpected-way.html?smid=nytcore-ios-share&referringSource=articleShare>.

Hu, W & Zhu, X 2023, 'A real-time voice cloning system with multiple algorithms for speech quality improvement', *National Library of Medicine*, 3 April, vol. 18, no. 4, pp. 1-14, viewed 11 March 2024, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10069766/>.

Immerwahr, D 2023, 'What the doomsayers get wrong about deepfakes', *The New Yorker*, 13 November, viewed 10 March 2024, <https://www.newyorker.com/magazine/2023/11/20/a-history-of-fake-things-on-the-internet-walter-j-scheirer-book-review#:~:text=The%20fakes%20aren%27t%20deep,our%20media%20as%20counterfeited%20evidence>.

Jones, S, Harding, E, Doxsee, C, Harrington, J & McCabe, R 2023, 'Competing without fighting: China's strategy of political warfare', *Center for Strategic & International Studies (CSIS)*, 2 August, viewed 11 March 2024, <https://www.csis.org/analysis/chinas-strategy-political-warfare>.

Kertysova, K 2018, 'Artificial Intelligence and disinformation: How AI changes the way disinformation is produced, disseminated and can be countered', *Security and Human Rights*, 12 December, vol. 29, nos. 1-4, pp. 55-81, viewed 11 March 2024, <https://www.scienceopen.com/document?vid=f0151b0c-b6aa-42c8-ba99-471321efa5fe>.

Knibbs, K 2023, 'Researchers tested AI watermarks—and broke all of them', *Wired*, 3 October, viewed 10 March 2024, <https://www.wired.com/story/artificial-intelligence-watermarking-issues/>.

Lewis, B 1997, 'Information Warfare and the intelligence community', *The Final Report of the Snyder Commission*, eds. E Cheng & D Snyder, Woodrow Wilson School Policy Conference 401a: Intelligence Reform in the Post-Cold War Era, viewed 11 March 2024, <https://irp.fas.org/eprint/snyder/>.

Library of Congress (LOC) 2022, *Resources for moving image related subjects & materials*, The Library of Congress Moving Image Research Center, viewed 5 March 2024, <https://www.loc.gov/rr/mopic/onlinesources.html>.

Lyonn, S 2023, 'Leica launces world's first camera with content credentials', Content Authentication Initiative (CAI), 26 October, viewed 5 March 2024, <https://contentauthenticity.org/blog/leica-launches-worlds-first-camera-with-content-credentials>.

Marcellino, W, Beauchamp-Mustafaga, N, Kerrigan, A, Chao, L & Smith J 2023, 'The rise of generative AI and the coming era of social media manipulation 3.0', Santa Monica, CA, US, RAND Corporation, PE-A2679-1, viewed 11 March 2024, <https://www.rand.org/pubs/perspectives/PEA2679-1.html>.

McFarland, A 2023, '5 best detector tools & techniques', *Unite.AI*, November, viewed 5 March 2024, <https://www.unite.ai/best-deepfake-detector-tools-and-techniques/>.

McQuate, S 2023, 'Q&A: UW researcher discusses just how much energy ChatGPT used', *UW News*, July, viewed 11 March 2024, <https://www.washington.edu/news/2023/07/27/how-much-energy-does-chatgpt-use/>.

Metz, C 2023, 'Instant videos could represent the next leap in A.I. technology', *The New York Times*, 4 April, viewed 11 March 2024, <https://www.nytimes.com/2023/04/04/technology/runway-ai-videos.html>.

Mittal, A 2023, 'Prompt hacking and misuse of LLMs,', *Unite.AI*, 19 October, viewed 11 March 2024, <https://www.unite.ai/prompt-hacking-and-misuse-of-llm/>.

Mullaney, S 2022, 'Everything flows: Russian information warfare forms and tactics', *The Cyber Defense Review*, 14 November, pp. 193-211, viewed 5 March 2023, <https://cyberdefensereview.army.mil/Portals/6/Documents/2022_fall/12_Mullaney.pdf?ver=AUTQuqWAV0vDAeGvWN-dYw%3d%3d>.

Muthukrishnan, N, Maleki, F, Ovens, K, Reinhold, C, Forghani, B & Forghani, R 2020, 'Brief history of Artificial Intelligence', *Neuroimaging Clinics of North America*, 30 November, vol. 30, no.4, pp. 393-9.

NVIDIA 2023, 'NVIDIA brings Generative AI to world's enterprises with cloud services for creating large language and visual models', Press Release, 21 March, viewed 11 March 2024, <https://nvidianews.nvidia.com/news/nvidia-brings-generative-ai-to-worlds-enterprises-with-cloud-services-for-creating-large-language-and-visual-models>.

OpenAI 2024, 'Creating video from text, Sora is an AI model that can create realistic and imaginative scenes from text instructions', OpenAI website, viewed 10 March 2024, <https://openai.com/sora>.

Ozair, M 2023, 'Manifesting responsible AI: Blockchain meets Artificial Intelligence', NASDAQ, 26 September, viewed 5 March 2024, <https://www.nasdaq.com/articles/manifesting-responsible-ai-blockchain-meets-artificial-intelligence>.

Paul, C & Matthews, M 2016, 'The Russian firehose of falsehood propaganda model: Why it might work and options to counter it', RAND Corporation, Santa Monica, CA, US, viewed 11 March 2024, <https://www.rand.org/pubs/perspectives/PE198.html>.

RAND 2023, 'Information Operations, RAND Corporation', Santa Monica, CA, US, viewed 11 March 2024, <https://www.rand.org/topics/information-operations.html>.

Reissman, H 2023, 'What is deepfake porn and why is it thriving in the age of AI?', Annenberg School of Communication, 13 July, viewed 10 March 2024, <https://www.asc.upenn.edu/news-events/news/what-deepfake-porn-and-why-it-thriving-age-ai>.

Rid, T 2020, *Active Measures: The Secret History of Disinformation and Political Warfare*, Farrar, Strauss & Giroux, New York, NY, US.

Rillig, M, Agersrtand, M, Bi, M, Gould, K & Sauerland, U 2023, 'Risks and benefits of large language models for the environment', *Environmental Science and Technology*, vol. 57, no. 9, pp. 3464-6, viewed 10 March 2024, <https://pubs.acs.org/doi/full/10.1021/acs.est.3c01106>.

Ryan-Mosely, T 2023, 'How generative AI is boosting the spread of disinformation and propaganda', *MIT Technology Review*, 4 October, viewed 11 March 2024, <https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/>.

RumorGuard 2023, 'AI-generated content distorts events in Gaza', 26 October, News Literacy Project, 26 October, viewed 11 March 2024, <https://www.rumorguard.org/post/ai-generated-content-distorts-events-in-gaza>.

Sadiq, M & Saji, M 2022, 'The disaster of misinformation: A review of research in social media', *National Library of Medicine*, vol. 13, no. 4, pp. 271-85, viewed 10 March 2024, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8853081/>.

Saner, E 2024, 'Inside the Taylor Swift deepfake scandal: It's men telling a powerful woman to get back in her box', *The Guardian*, 31 January, viewed 11 March, 2024, <https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box>.

Schreckinger, B 2022, 'Introducing the "5G" war', *Politico*, 15 December, viewed 5 March 2024, <https://www.politico.com/newsletters/digital-future-daily/2022/12/15/the-other-5g-00074195>.

Suleyman, M & Bhaskar, M 2023, *The Coming Wave: Technology, Power and the 21st Century's Greatest Dilemma*, Crown, New York, NY, US.

Theohary, C 2018, 'Information Warfare: Issues for Congress', Congressional Research Service, 7 March, pp. 1-2, viewed 10 March 2024, <https://sgp.fas.org/crs/natsec/R45142.pdf>.

—2023, 'Defense primer: Operations in the information environment', Congressional Research Service, 14 December, viewed 10 March 2024, <https://crsreports.congress.gov/product/pdf/IF/IF10771/10>.

Thompson, S & Hsu, T 2023, 'How easy is it to fool A.I. detection tools?', *The New York Times*, 28 June, viewed 11 March 2024, <https://www.nytimes.com/interactive/2023/06/28/technology/ai-detection-midjourney-stable-diffusion-dalle.html>.

UNESCO 2024, 'Ethics in Artificial Intelligence', UNESCO website, viewed 11 March, 2024, <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>.

Veranasi, L 2023, 'Biden's AI chief says 'voice cloning' is what keeps him up at night', *Business Insider*, 5 November, viewed 10 March 2024, <https://www.businessinsider.com/voice-cloning-technology-worries-biden-ai-bruce-reed-elevenlabs-scammers-2023-11>.

White House 2023, 'Executive Order on the safe, secure and trustworthy development and use of Artificial Intelligence', 30 October, viewed 11 March 2024, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

Wilson, M 2023, 'A super group of companies reveals an icon to combat misinformation. It may just confuse things more', *Fast Company*, 10 October, viewed 10 March 2024, <https://www.fastcompany.com/90964745/adobe-released-a-super-icon-to-combat-misinformation-will-it-work>.

Witt, S 2023, 'How Jenson Huang's NVIDIA is powering the A.I. revolution, *The New Yorker*, 27 November, viewed 5 March 2024, <https://www.newyorker.com/magazine/2023/12/04/how-jensen-huangs-nvidia-is-powering-the-ai-revolution>.

World Economic Forum 2024, 'AI governance alliance', World Economic Forum website, viewed 11 March 2024, <https://initiatives.weforum.org/ai-governance-alliance/home>.