



unesco

Synthetic Content and its Implications for AI Policy

A Primer



Published in 2024 by the United Nations Educational, Scientific and Cultural Organization,
7, place de Fontenoy, 75352 Paris 07 SP, France

© UNESCO 2024

ISBN 978-92-3-100727-9

DOI: <https://doi.org/10.54678/URJF3412>



This publication is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (<https://www.unesco.org/en/open-access/cc-sa>).

Images marked with an asterisk (*) do not fall under the [CC-BY-SA](#) license and may not be used or reproduced without the prior permission of the copyright holders.

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

Authors: Mariagrazia Squicciarini, José Valdez Genao, Camilo Sarmiento

Cover photo: © metamorworks/Shutterstock.com*

Cover illustration: UNESCO/Jzan Tamielle Villanos

Graphic design: UNESCO

Printed by UNESCO

Printed in France

SHORT SUMMARY

Getting synthetic content right

In the age of generative Artificial Intelligence (AI), synthetic content has become ubiquitous. AI generated content encompassing texts, images, audios, and videos that look real, is often used to convey information and provide entertainment to different audiences. While these developments offer a range of opportunities, the proliferation of deepfakes, i.e. synthetic content depicting individuals performing actions or speaking, creates concerns.

Deepfakes are affecting all domains. Estimates suggest that in fintech they have increased 700% in 2023 and that deepfake-related fraud worldwide increased more than tenfold from 2022 to 2023. Deepfakes disproportionately target women and can cause harm, both psychological and reputational. A report from 2019 analyzing over 14,000 deepfake videos found that 96% of them were non-consensual intimate content and that, from the examined content on the top five deepfake pornography websites, all involved women.



96%
of **deepfake videos**
have non-consensual
intimate content

- While generative AI holds significant promises for societies and economies, it is paramount to address its challenges, for generative AI to be developed, deployed and used in a way that upholds human rights and fundamental freedoms, in line with UNESCO Recommendation on the Ethics of AI.
- Recognizing this need, stakeholders worldwide have started to devise and test mechanisms to make users of synthetic content able to distinguish what is real from what is not, and to design effective redressal mechanisms related to the harm that AI-generated synthetic content may cause.

If left unaddressed by policy, the possible misuse and abuse of synthetic content risks contributing to erode trust in individuals and institutions and to alter the perception of reality, ultimately affecting democracies and societies.



unesco

"Since wars begin in the minds of men and women it is in the minds of men and women that the defences of peace must be constructed"

Synthetic Content and its Implications for AI Policy

A Primer

Foreword



© UNESCO/Christelle ALIX

Synthetic content generated by advanced AI systems, i.e. data and text as well as highly realistic images, audios and videos, is taking the world by storm. The opportunities that such content creation at scale may trigger range from improved education and entertainment to firms' enhanced performance. Small and medium enterprises, for instance, may be able to afford virtual marketing specialists or finance experts that they would otherwise be unable to hire in person. Such opportunities may nevertheless come at the cost of jeopardized social and societal relationships, not only by making it harder for individuals to distinguish between what is artificial and what is real, but also by disrupting entire sectors—such as the creative industries—where human labor and intellectual property risk being displaced or inadequately recognized.

The ethical considerations that synthetic content triggers are complex and multifaceted. Issues related to authenticity, ownership, privacy, and the need to stop and redress possible misuse, to name but a few, are far from being addressed or resolved. AI-generated media challenge our perceptions of reality and can undermine the trustworthiness of individuals and institutions alike.

True, synthetic content can enable innovative governance practices and even help customize service provision and increase its quality, but it may also contribute to perpetuate or create biases, infringe on privacy, and disrupt social cohesion. Deepfakes, a specific form of synthetic media, exemplify the potential dangers of AI-generated content. While they can be used for benign purposes, such as language translation and virtual try-ons, they can nevertheless facilitate malicious activities like gender-based harassment, disinformation, and even cybercrime.

This calls for a reality check, and for ethical governance frameworks and models that make sure to uphold human rights, human dignities, and fundamental freedoms. The UNESCO Recommendation on the Ethics of Artificial Intelligence, adopted in 2021, can serve as a foundational guideline to navigate these complex issues in all UNESCO 194 member states.

The policy areas the Recommendation addresses and the solutions it proposes relate to all types of synthetic content-related issues, and can help speed and scale up effective policy responses, in an internationally harmonized and coordinated fashion.

However, no good solution can be found or designed in the absence of a clear understanding of the issues at stake, and the “primer” that the present note represents aims to shed light on key features and concerns. As we navigate the challenges and opportunities that synthetic content poses to economies and societies alike, it is imperative to foster a global dialogue that includes all stakeholders—governments, businesses, academia, and civil society. Together, we can ensure that AI technologies are developed, deployed and used in ways that enhance humanity, its wellbeing, and its inclusiveness. Halting technological progress or the solutions it offers is not the solution, as AI holds promises to address economic, environmental, and societal challenges. However, we need to recognize that risks and concerns may lead some to prioritize caution over opportunity, to the point that individuals and companies may choose to forgo the potential benefits of AI to avoid having to face its negative impacts.

Gabriela Ramos

Assistant Director-General
for Social and Human Sciences,
UNESCO

Acknowledgements

This report was written by Mariagrazia Squicciarini, Camilo Sarmiento, José Valdez Genao. We would like to thank UNESCO's Assistant Director-General for Social and Human Sciences, Gabriela Ramos, for her oversight and support. We are very grateful to her also for the feedback received on earlier versions of this paper and for steering our work as it advanced. We are also very grateful to participants in the "Policy Dialogue on AI Governance" that took place at UNESCO HQs on 28 June 2024, especially the panelists contributing to the discussion on "Synthetic Content". Many thanks go to Anne Neuberger, for providing critical context to the discussion, and to Lee Tiedrich, Rumman Chowdhury, Nataliya Kos'myna, Seydina Moussa Ndiaye, Julie Lavet, Bambang Riyanto Trilaksono, Bilel Jamoussi, Lucia Marucci, and Amir Banifatemi, for enriching the content of this report through their comments and the feedback received. Finally, we are grateful to Garance Sarlat, Mimouna Abderrahmane and Barbara Majorano for helping us go through the publication process smoothly, and to Ishita Bhatia for support in the very last stage of this publication. All errors remain our own.

Table of contents

ABSTRACT	11
-----------------	-----------

1. INTRODUCTION	12
------------------------	-----------

2. UNDERSTANDING SYNTHETIC CONTENT: BACKGROUND AND EXISTING LITERATURE	14
2.1 Definition and Context	15
2.2 Distinguishing Synthetic Content, from Synthetic Media, and Deepfakes	16

3. SYNTHETIC CONTENT: GENERATION AND OUTPUTS	18
3.1 Generating Synthetic Content	19
3.2 Classification of Outputs and Common Functions	20

4. POTENTIAL USES OF SYNTHETIC CONTENT: CONCERNS AND OPPORTUNITIES	22
4.1 Synthetic Data for Analytical Purposes	24
4.2 Synthetic Content and the Environment	25
4.3 Synthetic Content, Learning and Creativity	26
4.4 Deepfakes	28
4.5 Synthetic Biology	32

5. WHAT CAN POLICY DO?	33
5.1 Watermarks for AI-Generated Outputs	35
5.2 Content Moderation for Synthetic Media	35
5.3 Consent, Redressal, and Liability Regimes	36
5.4 Supporting Human Content and Fostering Wealth Redistribution	37
List of Figures	38
Bibliography	39

Characterizing Synthetic Content to Discuss the Possible Implications for AI Policy: A Primer

Abstract

The deployment of advanced Artificial Intelligence (AI) models, particularly generative AI, has sparked discussions regarding the creation and use of synthetic content - i.e. AI-generated or modified outputs, including text, images, sounds, and combinations thereof - and its impact on individuals, societies, and economies. This note explores the different ways in which synthetic content can be generated and used and proposes a taxonomy that encompasses synthetic media and deepfakes, among others. The taxonomy aims to systematize key characteristics, enhancing understanding and informing policy discussions. Key findings highlight both the potential benefits and concerns associated with synthetic content in fields like data analytics, environmental sustainability, education, creativity, and mis/disinformation and point to the need to frame them ethically, in line with the principles and values of UNESCO Recommendation on the Ethics of Artificial Intelligence. Finally, the note brings to the fore critical questions that policymakers and experts alike need to address to ensure that the development of AI technologies aligns with human rights, human dignity, and fundamental freedoms.

1. Introduction

The popularization of advanced Artificial Intelligence (AI) models and systems worldwide, including generative AI, is contributing to igniting the conversation about the impact that AI can have on individuals, societies, and economies alike, and about the need for governance approaches enabling the ethical development, deployment, and use of AI.

Advanced AI systems are known for their ability to produce output that looks real, and to be able to do so at speed and scale, based on user-specific prompts. Such output, which can consist of text, images, and sounds or combinations thereof, among others, is impacting sectors ranging from education to information and entertainment, at the same time as they impact and shape societal interactions. Steering such fast-paced technological developments ethically entails making sure that they uphold human rights, human dignity, and fundamental freedoms, in line with the UNESCO Recommendation on the Ethics of Artificial Intelligence, adopted in 2021 and applicable to all its 194 Member States. Doing so nevertheless requires a better understanding of the phenomenon at stake, its main components, and its possible implications for societies and economies.

The present note addresses such a need, by contributing to shed light on what “synthetic content” may mean and consist of, and by proposing a taxonomy aimed at systematizing its main characteristics, to facilitate understanding and inform the scientific and policy discussion alike. Aware that the novelty and rapid pace of these advancements may lead to different — and at times changing, diverging, or partially exclusive — interpretations and assessments, the proposed taxonomy is to be considered informative rather than conclusive.

In what follows, the note first explores existing literature to address the question of what synthetic content may mean or entail and to try and disentangle synthetic content from related concepts, such as synthetic media and deepfakes. It then explores the different ways in which synthetic content can be generated and proposes a first classification of synthetic outputs. It further discusses the potential types and uses of synthetic content, including synthetic biology, highlighting both concerns and opportunities. Finally, the note highlights the possible ethical implications of creating and using synthetic content, vis-à-vis the principles, values, and policies contained in the UNESCO Recommendation on the Ethics of AI. It ends with a number of policy considerations informing the conversation about how to put these powerful technologies at the service of individuals and societies.

The stakes are high. On the one hand, the ability to generate (high-quality) synthetic content can serve a wide array of purposes and contribute to the technological leadership of businesses and countries alike. On the other hand, the way synthetic content gets generated and the possible implications that its use may have on education, the world of work, or societal relationships, to name but a few, trigger profound questions. It calls for the need to assess possible tradeoffs, in a view to striking the “right” balance between economic and technological opportunities vis-a-vis societal needs and desiderata, as well as human rights and fundamental freedoms.

2. Understanding Synthetic Content: Background and Existing Literature

The term *synthetic content* is often used interchangeably with *synthetic media* and *deepfakes*, broadly referring to any output generated by AI algorithms.¹ However, as each term may hold different connotations and varying meanings, in the present note we define synthetic content as any AI-generated or modified output, whereas synthetic media represents one specific application of synthetic content. In turn, we hold deepfakes to be a particular type of synthetic media created by AI.

2.1 Definition and Context

In general, saying that something is *synthetic* leads to referring to products created through artificial means that attempt to display a natural origin or features. In dictionaries, products are typically defined as synthetic if they are produced through artificial rather than natural processes.² However, this definition presents challenges in relation to categorizing a product as synthetic or not, as there is no consensus on where the limits or threshold may lie that makes something artificial. For instance, does growing plants using hydroponic techniques, which involve artificial water-based environments, make the plants synthetic?

To address these limitations, our framework focuses on understanding AI-generated synthetic content by discussing the different interpretations that the term synthetic may have.

In the context of artificial intelligence, synthetic content corresponds to digital output generated or modified by algorithms, typically AI techniques, such as machine learning. Such output is characterized by its artificial generation. It differs from naturally occurring or manually created content, and closely resembles reality, to the point of making it often difficult to distinguish what is real from what is not.³

As mentioned, in practice, it is challenging to draw the line between synthetic and AI-powered content, and no agreement exists in relation to the threshold or amount of treatment that data need to undergo for the resulting content to be considered synthetic. The question in a nutshell is: How synthetic does a product/output need to be for it to be considered synthetic? In a parallel to the denomination of origin of products, how much production or transformation is needed to happen in a specific location for a product to be certified as being from that very place?

1 Jeenisha Shrugare, "Exploring Synthetic Media: How Ai Is Changing Art," XRDS: Crossroads, The ACM Magazine for Students 29, no. 4 (June 2023): 64–65, <https://doi.org/10.1145/3596936>.

2 Cambridge Business English Dictionary, s.v. "Synthetic," accessed 15 June 2024, <https://dictionary.cambridge.org/dictionary/english/synthetic>.

3 Jeenisha Shrugare, "Exploring Synthetic Media: How Ai Is Changing Art," XRDS: Crossroads, The ACM Magazine for Students 29, no. 4 (June 2023): 64–65, <https://doi.org/10.1145/3596936>.

2.2 Distinguishing Synthetic Content, from Synthetic Media, and Deepfakes

Synthetic content encompasses any AI-generated or modified output. Synthetic content may become **synthetic media** when used to convey information, entertainment, or messages to an audience, utilizing technologies and institutions for diffusion.⁴ This process, known as mediatization, transforms synthetic content into “AI Media.” For example, if a person generates a picture with an AI app blending faces with another person as a form of entertainment, that image would be synthetic content. If these people then decide to post this image to social media, it would become synthetic media, because of the use made in media.

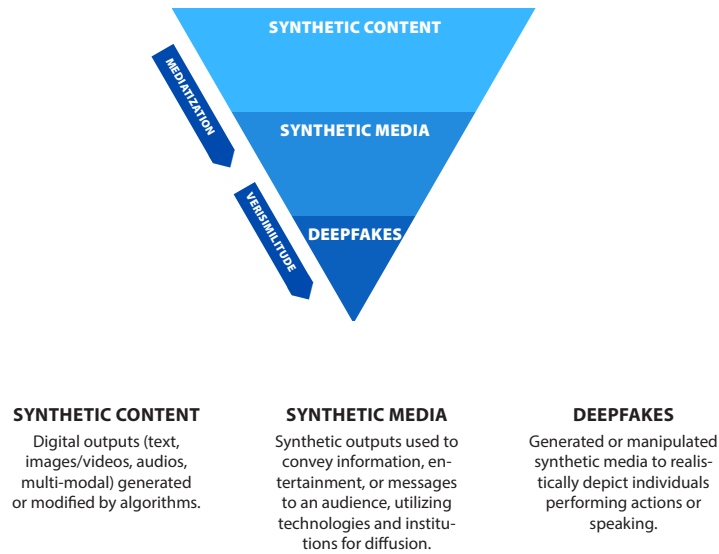
One of the most popular forms of AI media is **deepfakes**. Existing studies often use expressions such as synthetic media and deepfakes interchangeably, but synthetic content may be AI media without being classified as deepfakes, meaning these terms are conceptually different. For the sake of clarification, in our taxonomy, we refer to deepfakes as a specific type of synthetic media, involving the generation or manipulation of content to realistically depict individuals performing actions or speaking.⁵ Due to their realism, deepfakes can convince viewers that the content is real, even if it was synthetically produced.

Deepfakes are commonly associated with disinformation and harm. They may also hold potential for unharmed uses. Later we will explore the potential uses for these types of content while discussing the ethical concerns that they may trigger.

4 Katalin Feher, “Exploring AI Media. Definitions, Conceptual Model, Research Agenda,” *Journal of Media Business Studies*, April 13, 2024, 2–4, <https://doi.org/10.1080/16522354.2024.2340419>.

5 CPA, ed., “Parliamentary Handbook on Disinformation, AI and Synthetic Media,” Pam.int, January 4, 2024, https://issuu.com/theparliamentarian/docs/cpa_handbook_ai_disinformation_synthetic_media_onl.

Figure 1: Typifying Synthetic Content



Source: Authors' own compilation.

3. Synthetic Content: Generation and Outputs

Synthetic outputs can take the form of synthesized texts, images (including videos), audio, or combinations thereof.⁶ Generating these outputs generally involves machine learning, with the most advanced techniques utilizing deep learning. Popular machine learning models can easily modify existing data—creating outputs that closely resemble the input—or generate entirely new outputs.⁷

3.1 Generating Synthetic Content

Synthetic content can be generated in various ways. One approach involves creating outputs that resemble the characteristics of input data through modification. For example, a student might use a chatbot to refine an email she has written, instructing AI to make it sound more professional by changing the tone and structure of the original email. Although the resulting text is synthetic—an output modified by an AI model—it primarily involves the transformation of an original input.

AI can also generate entirely new synthetic outputs using advanced models. Advanced models, such as deep learning neural networks, can extract and learn complex patterns from data to then generate realistic content from scratch.⁸ For instance, continuing with the previous example, a chatbot trained to emulate human writing can compose an entirely new email, if given the appropriate prompt by the student.

Moreover, hybrid approaches exist that integrate both modification and generation techniques, leveraging the strengths of each to produce sophisticated synthetic content.⁹ For example, in the film industry, an AI model might first modify existing scripts to improve dialogue, pacing, and character development. Then, it can generate entirely new scenes or plotlines based on the refined script.

6 Lucas Whittaker et al., “‘All around Me Are Synthetic Faces’: The Mad World of Ai-Generated Media,” *IT Professional* 22, no. 5 (September 1, 2020): 90–99, <https://doi.org/10.1109/mitp.2020.2985492>.

7 CPA, ed., “Parliamentary Handbook on Disinformation, AI and Synthetic Media,” *Pam.int*, January 4, 2024, https://issuu.com/theparliamentarian/docs/cpa_handbook_ai_disinformation_synthetic_media_onl.

8 Neural networks are mathematical models inspired by the structure of biological neural networks and trained to simulate the neurons, similar to the human brain decision-making processes.

9 Khaled Emam, Lucy Mosquera, and Richard Hoptroff, *Practical Synthetic Data Generation* (O’Reilly Media, Inc, 2020).

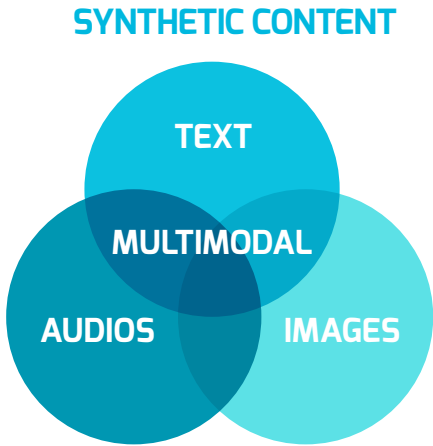
Figure 2: Common methods of synthetic content production

Common Methods of Synthetic Content Production	
Generative Adversarial Networks (GANs)	Utilizes a generator and a discriminator to produce data that becomes indistinguishable from real data through iterative training.
Variational Autoencoders (VAEs)	Detects patterns in datasets to generate new data with similar attributes, useful for complex data with various characteristics.
Transformer Models	Focused on language processing to generate or predict text sequences based on learned patterns.
Diffusion Models	Applied noise to input data and reverse the process to generate realistic images.

Source: Authors’ own compilation, adapted from Satwik Mishra, “Generative AI and Authenticity Balancing Innovation and Integrity in Media,” Center for Trustworthy Technology, 2024.¹⁰

3.2 Classification of Outputs and Common Functions

Figure 3: Types of synthetic outputs

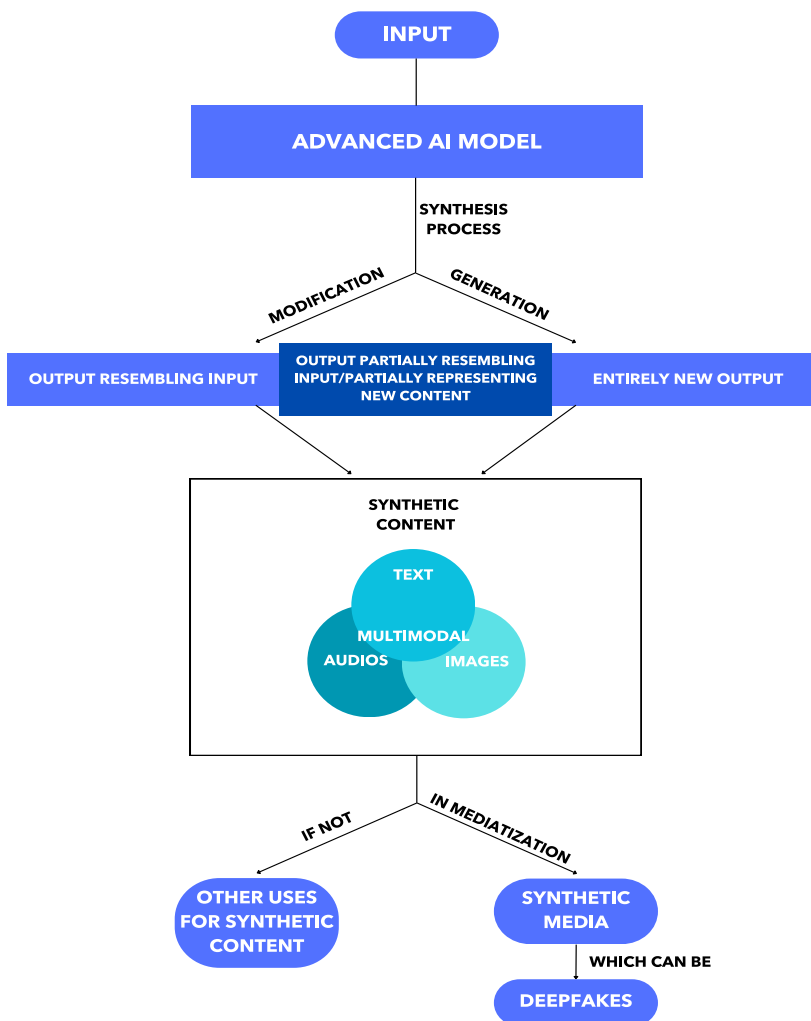


When the input given to an AI model undergoes a generative or modification process, the resulting output may be classified as synthetic content and can take different forms, from generating synthetic text to elaborating synthetic audio. These outputs need not be mutually exclusive and can be combined.

10 GANs utilize two neural networks trained in an iterative process where they compete. The generator’s objective is to produce data indistinguishable from real data, while the discriminator’s is to identify this data from real data.

Synthetic outputs may take various forms through the use of diverse techniques. Some of the more common are the following. Synthetic images, which include videos, involve creating new images, altering existing ones, applying new or different styles from one image to another, generating deepfakes, and creating virtual reality environments. Synthetic audio includes synthesized voices, generating speech from text, altering recordings, and composing new music. Synthetic text encompasses producing human-like writing, coding, translating, and converting speech to text. Multimodal synthetic content combines different types of synthetic outputs, such as synthetic videos which can be a combination of synthetic audio and images.

Figure 4: Exemplifying synthetic content generation and types



Source: Authors' own compilation

4. Potential Uses of Synthetic Content: Concerns and Opportunities

Synthetic content is applied across various domains, with its uses often overlapping and integrating. While it is possible to classify these applications, such classifications are not mutually exclusive, as many applications often coexist.

Figure 5: Possible uses and related concerns and opportunities of synthetic content

	USED OF SYNTHETIC CONTENT	OPPORTUNITIES	CHALLENGES
ANALYTICAL PURPOSES	<ul style="list-style-type: none"> • Synthetic data • Synthetic datasets • Advanced systems 	<ul style="list-style-type: none"> • Input data privacy • Addresses privacy concerns about data treatment 	<ul style="list-style-type: none"> • Learning on biased data and datasets • Privacy concerns
SUSTAINABILITY	<ul style="list-style-type: none"> • Environment 	<ul style="list-style-type: none"> • Supporting mechanisms for green transitioning 	<ul style="list-style-type: none"> • High electricity usage • Environmental footprints • Water consumption
CREATIVE REASONING TASKS	<ul style="list-style-type: none"> • Education • Training • Synthetic biology • Art • Industry assistance • Music • Literature • Video games • Academia • Virtual and augmented reality 	<ul style="list-style-type: none"> • Acquiring knowledge • Practical simulations • Design and analysis and predictions of biological systems • Innovative art • Automated decision making and assistance • Support in academia and research 	<ul style="list-style-type: none"> • Academic authorship • Skill endowment • Inaccurate outputs • Biosecurity and Bioethics • Intellectual property • Job insecurity
DEEPEAKE MEDIA	<ul style="list-style-type: none"> • Deepfakes in media, videos, audio 	<ul style="list-style-type: none"> • Realistic audio translations • Business tool • Entertainment • Dubbing 	<ul style="list-style-type: none"> • Gender based attacks • Disinformation • Media manipulation and extortion • Fraud

Source: Authors' own compilation.

In what follows, we delve deeper into some of the possible uses of synthetic content outlined in Figure 5, to discuss key opportunities and challenges of relevance for policy discussion.

4.1 Synthetic Data for Analytical Purposes

AI models are capable of generating synthetic data and datasets, which then can be utilized as input content. Similarly, AI models can be used to generate synthetic algorithms such as new neural network structures. Synthetic data and algorithms can bring important benefits but also trigger notable risks.

In research projects where datasets lack adequate representation of e.g. certain populations or languages, synthetic data can help fill these gaps by providing data points with characteristics similar to the original ones.¹¹ This can make datasets more representative and help reduce biases. Additionally, utilizing synthetic data can help address privacy concerns by reducing the need to use real individuals' data.

However, generating synthetic data also poses risks, as the models used to create this data are trained on existing datasets. If these datasets are not representative themselves, there is a risk of perpetuating or exacerbating biases.¹² Algorithms trained on biased data will produce biased synthetic data, which in turn will train new algorithms with similar characteristics, thus amplifying biases and generating new ones.

This concern is directly addressed by the principle of Fairness and Nondiscrimination, outlined in paragraphs 28, 29, and 30 of the [UNESCO Recommendation on the Ethics of AI](#).¹³ The Recommendation emphasizes the importance of safeguarding fairness and nondiscrimination to promote social justice through AI systems. It prioritizes the inclusion of all members of society, with particular emphasis on people with disabilities, women and children, and other marginalized groups, considering their specific needs and language requirements. Moreover, it highlights the necessity to address the digital divide and prevent AI systems from reinforcing or perpetuating biases and stereotypes.¹⁴ By aligning synthetic content generation with these principles, it would be possible to mitigate the risks associated with biased datasets.

When it comes to privacy, with the “right” prompt AI systems can inadvertently reveal data from their training datasets, including personal information about individuals, at times collected from the open internet. If this information was never intended to be processed or made available, this may trigger significant privacy risks and potential misuse of sensitive data. Paragraphs 32, 33, and 34 of the UNESCO Recommendation on the Ethics of AI address such privacy and data protection issues, emphasizing the importance of defending human dignity, autonomy, and

11 Khaled Emam, Lucy Mosquera, and Richard Hoptroff, *Practical Synthetic Data Generation* (O'Reilly Media, Inc, 2020).

12 Dominic-Madori Davis, “Women in AI: Anika Collier Navaroli Is Working to Shift the Power Imbalance,” *TechCrunch*, June 21, 2024, <https://techcrunch.com/2024/06/23/women-in-ai-anika-collier-navaroli-is-working-to-shift-the-power-imbalance/>.

13 UNESCO, “Recommendation on the Ethics of Artificial Intelligence” (Paris: United Nations Educational, Scientific and Cultural Organization, 2021).

14 UNESCO, “Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence.” UNESDOC Digital Library, 2023.

agency by enforcing national and international laws in relation to data collection, use, sharing, storage, and deletion. The Recommendation highlights the need to adopt robust data protection frameworks and governance mechanisms to safeguard personal information.¹⁵

4.2 Synthetic Content and the Environment

The intersection between producing synthetic content and the environment is important. AI-enabled technologies hold the potential for positive climate action, as they can enhance efficiency in data centers and model ecological systems to support green transitioning. However, the deployment of these technologies has also increasingly raised sustainability concerns.

The environmental impact of AI is another concern that, while being seldom taken into account, is conversely very much underlined in the UNESCO Recommendation on the Ethics of Artificial Intelligence. The production of synthetic content entails substantial use of electricity, as well as non-negligible water consumption, in addition to contributing to generating carbon emissions. The processes of producing synthetic content, particularly through deep learning and large language models, require vast computational power, leading to considerable environmental footprints. This increase in energy usage directly contributes to greenhouse gas emissions, which are linked to climate change.¹⁶

One study estimates that the carbon footprint of training a single big language model is equal to around 300,000 kg of carbon dioxide emissions. This is of the order of 125 round-trip flights between New York and Beijing.¹⁷ Similarly, another recent study found that training a large language model emits 25 times more carbon than a single air traveler on a one-way trip from New York to San Francisco.¹⁸ While these two studies diverge in the order of magnitude of the problem, they both point towards the same direction and underline the need to look at the carbon footprint of generative AI and the creation of synthetic content.

Moreover, water usage associated with AI development has surged in recent years. AI computing hardware and infrastructure use water directly for cooling and indirectly for electricity generation, as well as for the production of components like semiconductors. Reports indicate that between

15 UNESCO, "Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence." UNESDOC Digital Library, 2023.

16 Alokya Kanungo, "The Real Environmental Impact of AI," Earth.Org, March 5, 2024, <https://earth.org/the-green-dilemma-can-ai-fulfil-its-potential-without-harming-the-environment/>.

17 Fernando Sevilla Martínez, Raúl Parada, Jordi Casas-Roma, "CO2 impact on convolutional network model training for autonomous driving through behavioral cloning," *Advanced Engineering Informatics*, Volume 56, April 2023, <https://doi.org/10.1016/j.aei.2023.101968>.

18 Nestor Maslej et al., "The AI Index 2023 Annual Report," Stanford University | AI Index, April 2023, <https://aiindex.stanford.edu/report/>.

2021 and 2022, water usage increased by 20% at Google and 34% at Microsoft.^{19, 20} In Microsoft's case, water usage reached nearly 1.7 billion gallons, equivalent to more than 2,500 Olympic-sized swimming pools.

It becomes paramount to bear in mind and address the possible sustainable development-related concerns related to leveraging AI technologies, also when creating synthetic content. Policy Area 5 on the environment and ecosystems of the UNESCO Recommendation on the Ethics of AI urges Member States and companies to take responsibility for AI's direct and indirect environmental impacts. This includes adopting practices that minimize energy consumption and promote sustainability throughout the lifecycle of AI technologies.

4.3 Synthetic Content, Learning and Creativity

Synthetic outputs hold significant promises for education and training. From language learning to medical practice simulations, knowledge that was once costly to obtain is now readily accessible through synthetic content. AI tutors can assist students in enhancing their math skills and aid academics in generating comprehensive literature reviews.²¹

Figure 6: The Next Rembrandt, a famous AI-generated artwork in Amsterdam



Source: ING Group, "The Next Rembrandt," [Flickr \[CC BY 2.0\]](#) (December 3, 2021)*.

19 Google, ed., "2023 Environmental Report - Google Sustainability," Sustainability, 2023, <https://sustainability.google/reports/google-2023-environmental-report/>.

20 Microsoft, ed., "2022 Environmental Sustainability Report," Microsoft.com, 2023, <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW15mgm>.

21 Maria Radeva, "Guest Post: The Benefits and Risks of Chatgpt for Education," The Learning Scientists, February 15, 2024, <https://www.learningscientists.org/blog/2024/2/15-1>.

However, the generative capabilities of these models have also raised concerns in the educational sphere. The ease with which chatbots can produce high-quality essays raises questions about learning processes, skill endowment, fairness in grading, and authorship in academic work, as students may rely on AI to generate assignments with minimal effort.

Synthetic content also impacts art. Advanced AI models can generate new paintings, reproduce pictures in the style of renowned artists, create realistic music covers in the voices of famous singers, and even write scripts for films and television. In 2016, millions in Amsterdam and online were astounded by the realism of a 3D painting created using deep learning models in the style of the Dutch painter Rembrandt. The final result, shown in Figure 6, was so convincing that many could not distinguish it from his original work.

The use of AI in art has raised critical questions about ownership rights, copyright, and the impact on human creativity.²² Such is the case with an artist who produced a synthetic image that won a Sony Award. At present, it remains unclear who owns AI-generated content, whether the creator of the model, the algorithm itself (if at all possible or acceptable from a legal point of view), or the user who commissions or utilizes the content.

Many artists have raised concerns about the AI models generating graphics having been trained on their artworks without consent, about intellectual property rights being abused, and human creativity being devalued. A related question is one of legal personality, i.e. whether these tools can be considered authors or co-authors. Addressing this concern, Paragraph 68 of the UNESCO Recommendation states unequivocally that “when developing regulatory frameworks, Member States should, in particular, take into account that ultimate responsibility and accountability must always lie with natural or legal persons and that AI systems should not be given legal personality themselves,”²³ ruling out the possibility of AI being considered authors or co-authors. This reinforces the concept that AI shall assist and augment rather than replace human creativity, and that the legal and moral rights associated with creative works shall remain with human creators.

These impacts are resonating across various industries, also raising concerns about job security. In 2023, Hollywood was paralyzed by a strike, one of the main demands being to stop the rise of scripting algorithms and synthetic images in the industry. Workers voiced their concerns about AI models generating scripts and images of high quality at a faster pace, leading to job losses.²⁴

22 Marion Lagardère, “L’artiste Boris Eldagsen révèle que sa photo, récompensée par un Sony Award, est l’œuvre d’une IA et veut ouvrir le débat,” *Franceinfo*, April 19, 2023, https://www.francetvinfo.fr/replay-radio/l-etoile-du-jour/l-artiste-boris-eldagsen-revele-que-sa-photo-recompensee-par-un-sony-award-est-l-uvre-d-une-ia-et-veut-ouvrir-le-debat_5752295.html

23 UNESCO, “Recommendation on the Ethics of Artificial Intelligence” (Paris: United Nations Educational, Scientific and Cultural Organization, 2021).

24 Brian Merchant, “Your Boss Wants AI to Replace You. the Writers’ Strike Shows How to Fight Back,” *Los Angeles Times*, May 11, 2023, <https://www.latimes.com/business/technology/story/2023-05-11/column-the-writers-strike-is-only-the-beginning-a-rebellion-against-ai-is-underway>. Writer and activist Molly Crabapple stated “Another place that used to hire human artists has filled the spot with schlock from [AI image generator] Midjourney. If illustrators want to remain illustrators in two years, they have to fight now.” <https://unesdoc.unesco.org/ark:/48223/pf0000380455>.

Other industries are facing similar concerns, as synthetic content holds the potential to provide automated services with high expertise in fields such as law, accounting, and finance.²⁵ Policy Area 10 on Economy and Labour of the UNESCO Recommendation stresses the importance of investing in reskilling and upskilling workers, to prepare them and have them adapt to an ever-evolving job landscape. This includes providing the education and tools needed to integrate AI effectively in the workplace and for people to work with AI, including skills such as learning how to learn, communication, critical thinking, teamwork, empathy, and the ability to transfer knowledge across domains.²⁶ Having a workforce that is adaptable and equipped with both technical and soft skills is necessary to ensure that AI's benefits, including synthetic content creation, are distributed equitably, and to mitigate the potential negative impacts on employment that AI may have.

4.4 Deepfakes

When considering the potential uses of synthetic content, it becomes important to discuss deepfakes. They are among the most well-known types of synthetic media and are often associated with disinformation and manipulated videos. While deepfake videos are the most commonly debated, photo and audio deepfakes also exist and may trigger non-negligible ethical concerns.

However, deepfakes need not be harmful. For example, these techniques can generate audio translations of educational or entertainment content into languages that are not typically dubbed. Businesses can utilize it to create synthetic images of customers trying on cosmetics, new hairstyles, or clothes virtually, without causing any harm to the individuals directly involved.²⁷

Deepfake audios have also become a form of entertainment, exemplified by an AI cover featuring Drake and The Weeknd that went viral on social media.²⁸ Popular apps allow users to swap or blend their faces with others. Additionally, deepfake audios and videos can be used to lip-sync translations in movies, matching the original actors' voices and lip movements to the audio.

25 Marco Iansiti and Karim R Lakhani, "The Age of AI," essay, in *Competing in the Age of AI : Strategy and Leadership When Algorithms and Networks Run the World*. (Cambridge, Massachusetts: Harvard Business Review, 2020), 1–17.

26 UNESCO, "Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence." UNESDOC Digital Library, 2023.

27 Jan Kietzmann et al., "Deepfakes: Trick or Treat?" *Business Horizons* 63, no. 2 (March 2020): 135–46, <https://doi.org/10.1016/j.bushor.2019.11.006>.

28 Mark Savage, "AI-Generated Drake and The Weeknd Song Goes Viral," BBC News, April 17, 2023, <https://www.bbc.com/news/entertainment-arts-65298834>.

Figure 7: AI-generated cover of a song featuring Drake and The Weeknd

Source: AI Covers, "Drake AI - 'heart on My Sleeve' (Ft. The Weeknd AI)," YouTube, September 29, 2023, <https://www.youtube.com/watch?v=qPjgkKZ0P2M>.

Recent political events have brought attention to new unharmed — yet ethically questionable usage of deepfakes. The sophistication of current synthetic outputs brings to the forefront of new political campaigns what some are starting to define as "softfakes", meaning the use of deepfake images, videos, and audio that are modified to portray political candidates more appealingly.²⁹ In most cases, the purpose of softfakes seems to be to humanize candidates, to make them more related to the population they are targeting, thus augmenting their political capital. Recent examples are the digital resurrection of a deceased former president in Indonesia endorsing a running candidate, the use of AI-generated articles by politicians, or using deepfake audio to make candidates appear to be speaking many languages or singing.

Despite the above, deepfakes have become mostly known for their harmful uses. The phenomenon gained widespread attention when celebrities became targets, appearing in videos and images depicting them in explicit sexual acts.³⁰ Also, politicians were featured in viral videos, saying things that seemed unusual, which were later revealed to be synthetic content.³¹

Deepfake synthetic content has become a powerful tool to spread disinformation and cause harm. For example, a viral image on Twitter in 2023 depicted an explosion at the United States Pentagon, causing a slight drop in financial stocks.³² A year before, a video showed Ukrainian

29 Rumman Chowdhury, "AI-Fuelled Election Campaigns Are Here - Where Are the Rules?," Nature News, April 9, 2024, <https://www.nature.com/articles/d41586-024-00995-9>.

30 AFP, "In Age of AI, Women Battle Rise of Deepfake Porn," The Economic Times, 2023, <https://economictimes.indiatimes.com/tech/technology/in-age-of-ai-women-battle-rise-of-deepfake-porn/articleshow/102069206.cms?from=mdr>.

31 Daniel I. Weiner, Lawrence Norden, "Regulating AI Deepfakes and Synthetic Media in the Political Arena," Brennan Center for Justice, December 05, 2023, <https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena>.

32 Luke Hurst, "Fake Image of Pentagon Explosion Causes Real Dip on Wall Street," Euronews, May 23, 2023, <https://www.euronews.com/next/2023/05/23/fake-news-about-an-explosion-at-the-pentagon-spreads-on-verified-accounts-on-twitter>.

President Zelensky urging Ukrainians to surrender.³³ Another notable case involved a financial attack on a German company, where scammers used a deepfake voice to make a phone call impersonating the CEO, leading a manager to approve a transaction of over 240,000 euros.³⁴

The ability to create highly realistic yet entirely fabricated videos, audio, and images poses significant risks to public trust, individual reputations, and the integrity of information ecosystems. This misuse of AI technology necessitates robust strategies to combat the spread of false information and protect societies from its detrimental effects.

In response to these challenges, the UNESCO Recommendation on the Ethics of AI, particularly Policy Area 9 on Communications and Information, emphasizes the need for Member States to improve access to accurate information and knowledge, respect and promote freedom of expression, and encourage a diversity of viewpoints. It also underscores the importance of digital literacy skills.³⁵ Specifically, paragraph 114 of the Recommendation calls on Member States to “invest in and promote digital and media and information literacy skills to strengthen critical thinking and competencies needed to understand the use and implication of AI systems, in order to mitigate and counter disinformation, misinformation, and hate speech.”³⁶ By fostering these skills, we can empower individuals to critically evaluate the content they encounter, recognize deepfakes and other forms of synthetic content, and navigate the digital landscape more safely and effectively.

4.4.1 Gender-based Deepfakes

Women have been disproportionately targeted by deepfakes. In 2020, a publication done for the United States Department of Homeland Security reported over 100,000 computer-generated fake nude images of women created without their consent or knowledge.³⁷ An industry report from 2019 analyzing over 14,000 deepfake videos found that 96% of them were non-consensual intimate content and that, from the examined content on the top five deepfake pornography websites, all involved women.³⁸

This form of synthetic content can be used for violence and revenge purposes, depicting women in actions they never consented to. Celebrities like Taylor Swift have gone viral due to deepfakes, receiving also high-level institutional support, but other women and girls are equally vulnerable.

33 Jane Wakefield, “Deepfake Presidents Used in Russia-Ukraine War,” BBC News, March 18, 2022, <https://www.bbc.com/news/technology-60780142>.

34 Jesse Damiani, “A Voice Deepfake Was Used to Scam a CEO out of \$243,000,” Forbes, February 20, 2024, <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/>.

35 UNESCO, “Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence,” UNESDOC Digital Library, 2023.

36 UNESCO, “Recommendation on the Ethics of Artificial Intelligence” (Paris: United Nations Educational, Scientific and Cultural Organization, 2021).

37 Tina Brooks et al., “Increasing Threat of Deepfake Identities,” Department of Homeland Security, 2022, https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

38 Beatriz Kira, “Deepfakes, the Weaponisation of AI against Women and Possible Solutions,” Verfassungsblog, June 3, 2024, <https://verfassungsblog.de/deepfakes-ncid-ai-regulation/>.

A notable case is that of Noelle Martin, an 18-year-old Australian student, who discovered deepfake pornography of herself on the internet after searching her name online.³⁹ The malicious use of deepfakes against individuals, especially women and girls, cannot be overlooked by policymakers.

A recent report by UNESCO informs that 58% of young women and girls have experienced online harassment.⁴⁰ In such a dangerous online ecosystem, deepfakes and synthetic content can only exacerbate technology-facilitated gender-based violence (TFGBV). TFGBV leads to manipulation and intimidates women, including those with an active public life. For instance, thousands of female journalists and politicians around the world reported receiving TFGBV attacks, and many times, these attacks have managed to erode victims' public life, reputation, and positions. The UNESCO report further finds that, in addition to deepfakes, synthetic content can be utilized to automate harassment campaigns against women. Such was the case of Filipino-American journalist Maria Ressa, who has been the target of an overwhelming harassment campaign on social media. She mentions how at one point, only in one hour, she received more than 90 hate messages on Facebook. This coordinated attack used similar speech - i.e. common derogatory terms and slurs, including racist, homophobic, and sexist comments, or death and rape threats. After further investigation, it was revealed that many of the attacks were fueled by fake accounts and bots.⁴¹

It is clear that, based on the above, such deepfakes can cause significant harm, whether psychological or reputational, and can impact individuals and societies. The UNESCO Recommendation on the Ethics of AI addresses this issue in Policy Area 6 on Gender. It calls on Member States and other stakeholders, including developers, to "ensure that the potential of AI systems to advance the achievement of gender equality is realized".⁴² This entails actively preventing AI technologies from exacerbating existing gender gaps and instead using them to eliminate these disparities.

39 Justin Sherman, "'completely Horrifying, Dehumanizing, Degrading': One Woman's Fight against Deepfake Porn," CBS News, October 14, 2021, <https://www.cbsnews.com/news/deepfake-porn-woman-fights-online-abuse-cbsn-originals/>.

40 Rumman Chowdhury and Dhanya Lakshmi, "'Your Opinion Doesn't Matter, Anyway': Exposing Technology-Facilitated Gender-Based Violence in an Era of Generative AI," Unesdoc.unesco.org, 2023, <https://unesdoc.unesco.org/ark:/48223/pf0000387483>.

41 Julie Posetti et al., "Maria Ressa: Fighting an Onslaught of Online Violence. A Big Data Analysis.," International Center for Journalists, March 2021, https://www.icfj.org/sites/default/files/2021-03/Maria%20Ressa-%20Fighting%20an%20Onslaught%20of%20Online%20Violence_0.pdf

42 UNESCO, "[Recommendation on the Ethics of Artificial Intelligence](#)" (Paris: United Nations Educational, Scientific and Cultural Organization, 2021).

4.5 Synthetic Biology

Synthetic biology also benefits from AI's ability to generate synthetic content. Algorithms can now design new synthetic biological products, such as proteins or entire genomes, which can subsequently be engineered.⁴³ The use of AI systems enables the integration and analysis of extensive datasets, uncovering new patterns and insights in synthetic biology. By enhancing our understanding of biological systems, practices like complex gene editing have become feasible. Another breakthrough is the ability to predict biological behaviors and interactions, significantly reducing the reliance on traditional trial-and-error experimental designs. This may pave the way for the development of new vaccines, drugs, and diagnostic tools.⁴⁴

Despite these advances, challenges persist. The synthetic content generated in this field may lack accuracy, as predictive algorithms depend on existing datasets that might contain flaws. Additionally, ethical concerns in both synthetic biology and AI usage must be addressed, including biosafety, data usage, and the potential modification of biological materials for unethical purposes.⁴⁵

43 Hector García Martín, Stanislav Mazurenko, and Huimin Zhao, "Special Issue on Artificial Intelligence for Synthetic Biology," *ACS Synthetic Biology* 13, no. 2 (January 12, 2024): 408–10, <https://doi.org/10.1021/acssynbio.3c00760>.

44 Houbing Song et al., "Guest Editorial Ai-Driven Synthetic Biology for Human Wellbeing," *IEEE Journal of Biomedical and Health Informatics* 26, no. 10 (October 2022): 5042–43, <https://doi.org/10.1109/jbhi.2022.3204369>.

45 Arif Amaan, Garg Prekshi, and Srivastava Prachi, "Unlocking the Transformative Power of Synthetic Biology," *Archives of Biotechnology and Biomedicine* 8, no. 1 (April 18, 2024): 009–016, <https://doi.org/10.29328/journal.abb.1001039>.

5. What Can Policy Do?

As with many technologies in the past, advanced AI systems and the generation and use of synthetic content have once again outpaced the development of regulatory frameworks. This has led to concerns about the ethical, societal, and policy implications of such technologies. The proposition of some to have a moratorium putting the development of these technologies at standby for some time risks being ineffective and even harmful, as moratoriums have never worked, even in the context of less pervasive technologies.⁴⁶ In addition, this would entail putting at a halt the development, deployment and use of the societally and economically enhancing developments that AI can bring, to contain AI systems' possible drawbacks.

In response to the proposed moratorium, UNESCO's Director-General Audrey Azoulay, emphasized that the biggest challenge is guaranteeing stronger ethical rules for AI, and urged all governments to implement UNESCO's Recommendation on the Ethics of Artificial Intelligence immediately. This global normative framework provides all necessary safeguards and mechanisms to ensure that AI systems respect the rule of law, that harm is prevented and that there is accountability and redressal, in case of harm. She further underlined that self-regulation does not represent an alternative, as it risks being inefficient, ineffective and uncoordinated, and thus unable to ensure that pervasive and fast evolving technologies such as AI are developed, deployed and used ethically, for the good of people and the planet.⁴⁷

It becomes clear that leveraging the positive impacts of AI, including generative AI, for our societies, requires targeted policy interventions that can guide the development and use of synthetic content in a manner that upholds human rights, respects human dignities and fundamental freedoms, and promotes societal well-being, in an inclusive and non-discriminatory fashion. Moreover, as AI is a set of technologies that do not know or respect geographical borders — unless deliberately imposed — managing synthetic content in an ethical fashion calls for coordinated, interoperable governance mechanisms able to steer the development, deployment, and use of these technologies for good. If left unregulated, synthetic content may contribute to eroding trust in individuals and institutions alike, and affect the perception of reality, thus impacting societies at large. Addressing such shortcomings becomes important for all stakeholders, including governments, businesses, and civil society, and the conversation about possible guardrails needs to happen on the basis of universally agreed standards, for policies to be maximally interoperable and effective.

In what follows, we discuss a number of options that economic actors around the world are considering to make it possible to distinguish real from artificial content and to contain and redress the possible harm that AI-generated content can trigger. The list is not considered exhaustive, as its only aim is to shed light on some of the possible concerns and related approaches.

⁴⁶ Future of Life Institute, "Pause Giant AI Experiments: An Open Letter," Future of Life Institute, March 22, 2023, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

⁴⁷ Clare O'Hagan, "Artificial Intelligence: UNESCO Calls on All Governments to Implement Global Ethical Framework Without Delay," UNESCO, March 30, 2023, <https://www.unesco.org/en/articles/artificial-intelligence-unesco-calls-all-governments-implement-global-ethical-framework-without>.

5.1 Watermarks for AI-Generated Outputs

Watermarks have been proposed as a possible solution to help identify synthetic content, through clear and indelible signs that help distinguish AI-generated content from human-created content.⁴⁸ This approach entails applying unique, invisible markers to AI-generated content to identify it as such. These markers, created during the AI model's training phase, are detectable through specialized algorithms and are traceable back to the original AI content-generation model. This technology can help authenticate digital files and combat misinformation, deepfakes, and other deceptive media, by providing a reliable method to verify content authenticity and provenance.⁴⁹ This would help ensure transparency and accountability. The main proponents of watermarking do not see it as a solution to the challenges presented by synthetic content, but rather as a mechanism to help people discern the origin of the content they are seeing.

While some watermarking tools are already being commercialized, and some countries are considering legislating in relation to this, watermarking is also raising concerns, as no agreement exists on the best methods to implement effective watermarking.⁵⁰ Moreover, implementing watermarking tools that are accurate or unbiased against e.g. English texts produced by non-native speakers, may not be trivial. It is reported that AI-text detectors exhibit a bias by tending to wrongly classify non-native writing as AI-generated.⁵¹ Another challenge is that individuals and entities willing to do so could still manipulate, remove, or alter watermarks, that is, watermarks are a mechanism that is not exempt from attacks.⁵²

5.2 Content Moderation for Synthetic Media

In terms of content moderation for social media, platforms like YouTube or Meta have developed systems for moderating, identifying, labeling, and managing synthetic content. However, such interventions are still in their early days, and the accuracy rate of such content moderation is still unknown. Yet, content moderation is crucial to prevent the spread of misinformation and to protect users from harmful synthetic media.

Moderation of synthetic content is still subject to the internal policies of each media company, thus policy discussions could lead to standardization across industries, to make sure that the labeling of synthetic content is consistent across the internet. In line with UNESCO's guidelines

48 Tate Ryan-Mosley, "The inside Scoop on Watermarking and Content Authentication," MIT Technology Review, November 6, 2023, <https://www.technologyreview.com/2023/11/06/1082996/the-inside-scoop-on-watermarking-and-content-authentication/>.

49 Alessandra Sala, "AI Watermarking: A Watershed for Multimedia Authenticity," ITU, May 27, 2024, <https://www.itu.int/hub/2024/05/ai-watermarking-a-watershed-for-multimedia-authenticity/>.

50 JaeYoung Hwang, and SangHoon Oh, "A Brief Survey of Watermarks in Generative AI." The 15th International Conference on ICT Convergence, 2023. <https://ictc.org/media?key=site/ictc2023a/abs/P1-21.pdf>. Countries that have worked on watermarking regulations include China, the United States, the European Union and South Korea.

51 Weixin Liang et al., "GPT Detectors Are Biased against Non-Native English Writers," *Patterns* 4, no. 7 (July 2023): 100779, <https://doi.org/10.1016/j.patter.2023.100779>.

52 Tambiama Madiaga, "Generative AI and Watermarking - European Parliament," European Parliament, December 2023, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI\(2023\)757583_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf).

for the Governance of Digital Platforms, content moderation systems should be reliable, effective, and scalable, ensuring accuracy and non-discrimination. They should align with international human rights standards, considering context, linguistic nuances, and cultural particularities. Platforms should further provide moderation in all major languages of their operating regions and act promptly to remove harmful content while preserving it for law enforcement, if needed. Transparency in moderation decisions and appeal processes for users are also considered as essential, as is fairness towards human content moderators, and the need to ensure that they are adequately trained, supported, and protected from harm.⁵³

It is also recommended to moderate and label AI-generated content that can harm users or produce misinformation. Similarly, users of synthetic content should be required to clearly identify when the content has been generated or modified by AI. This transparency is critical for maintaining trust and ensuring that consumers are aware of the origins of the content they engage with.

It should be noted that experts have raised concerns about the potential ineffectiveness of AI-generated labels if they are overused in social media. Future research could explore such concerns with empirical evidence, to make sure that policy interventions can maximize their effectiveness.

5.3 Consent, Redressal, and Liability Regimes

The use of images of individuals, especially to create deepfakes or other synthetic media, should be governed by stringent consent laws, with possible misuse that may be deterred also through sanctioning, with the aim to protect individuals' privacy and dignity. Rigorous enforcement mechanisms may be key to guaranteeing the effectiveness of such interventions.

Governments can leverage existing laws and rules or issue new ones, as needed, to make sure that citizens are protected against any harm that synthetic content may produce. In line with what is stated in the UNESCO Recommendation on the Ethics of AI, it would be important to establish redressal mechanisms, whereby harm is stopped when happening and those being hurt are compensated. It would also be important to discuss establishing liability regimes for both creators and distributors of synthetic content and ensure that there is protection, accountability, and redressal for any harm caused by such content, be it reputational, psychological, or financial.

⁵³ UNESCO, "Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multi-Stakeholder Approach," UNESDOC Digital Library, 2023, <https://unesdoc.unesco.org/ark:/48223/pf0000387339>.

5.4 Supporting Human Content and Fostering Wealth Redistribution

With the rise of synthetic content, it becomes key to develop strategies to support and enhance the value of human work and creativity. This can include funding for creative industries, support for the development of new skills, and mechanisms to ensure that human creators across industries are not only deprived of the content they create but are also fairly compensated for their work. Similarly, education and training can help to leverage the use of AI for human benefit. Providing workers with skills that allow them to take advantage of the technology.

Moreover, as AI systems are largely trained on open data — without necessarily having the direct consent of those whose data are used to generate such profits — the issue of how to best redistribute the wealth generated by AI technologies becomes even more pertinent in the case of generative AI. The question of how to (re)distribute the returns generated through and by AI ranges from ensuring that the wealth produced by AI technologies is shared with those very humans the technology may make fully or partially redundant, to rewarding humans for having used their data, often unwillingly. In light of the economic benefits that AI and synthetic content are generating through the use of data, it would be important to have a policy conversation about the mechanisms that can ensure a fair redistribution of the wealth created with and through AI.

List of Figures

Figure 1	Authors’ own compilation, Typifying Synthetic Content, 2024.
Figure 2	Table by author, Common methods of synthetic content production. Adapted from Satwik Mishra, “Generative AI and Authenticity Balancing Innovation and Integrity in Media,” Center for Trustworthy Technology, 2024
Figure 3	Authors’ own compilation, Types of Synthetic Outputs, 2024.
Figure 4	Authors’ own compilation, Characterizing synthetic content, 2024.
Figure 5	Authors’ own compilation, Uses, concerns, and opportunities of synthetic content, 2024.
Figure 6	ING Group, “The Next Rembrandt,” Flickr [December 3, 2021], https://web.archive.org/web/20211203183102/https://www.flickr.com/photos/inggroup/25681989673 .
Figure 7	Ai-generated cover of a song featuring Drake and The Weeknd.AI Covers, “Drake AI - ‘heart on My Sleeve’ (Ft. The Weeknd AI),” YouTube, September 29, 2023, https://www.youtube.com/watch?v=qPjgkKZ0P2M

Bibliography

- AFP. "In Age of AI, Women Battle Rise of Deepfake Porn." *The Economic Times*, 2023. <https://economictimes.indiatimes.com/tech/technology/in-age-of-ai-women-battle-rise-of-deepfake-porn/articleshow/102069206.cms?from=mdr>
- Amaan, Arif, Garg Prekshi, and Srivastava Prachi. "Unlocking the Transformative Power of Synthetic Biology." *Archives of Biotechnology and Biomedicine* 8, no. 1 (April 18, 2024): 009–016. <https://doi.org/10.29328/journal.abb.1001039>
- Brooks, Tina, Princess G, Jesse Heatley, Jeremy J, Scott Kim, Samantha M, Sara Parks, et al. "Increasing Threat of Deepfake Identities." Department of Homeland Security, 2022. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf
- Cambridge Business English Dictionary, s.v. "Synthetic," accessed 2024, <https://dictionary.cambridge.org/dictionary/english/synthetic>
- Chowdhury, Rumman, and Dhanya Lakshmi. "'Your Opinion Doesn't Matter, Anyway': Exposing Technology-Facilitated Gender-Based Violence in an Era of Generative AI." *Unesdoc.unesco.org*, 2023. <https://unesdoc.unesco.org/ark:/48223/pf0000387483>
- Chowdhury, Rumman. "AI-Fuelled Election Campaigns Are Here - Where Are the Rules?" *Nature News*, April 9, 2024. <https://www.nature.com/articles/d41586-024-00995-9>
- CPA, ed. "Parliamentary Handbook on Disinformation, AI and Synthetic Media." Pam.int, January 4, 2024. https://issuu.com/theparliamentarian/docs/cpa_handbook_ai_disinformation_synthetic_media_onl
- Damiani, Jesse. "A Voice Deepfake Was Used to Scam a CEO out of \$243,000." *Forbes*, February 20, 2024. <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/>
- Weiner, Daniel I., Lawrence Norden, "Regulating AI Deepfakes and Synthetic Media in the Political Arena," Brennan Center for Justice, December 05, 2023, <https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena>
- Davis, Dominic-Madori. "Women in AI: Anika Collier Navaroli Is Working to Shift the Power Imbalance." *TechCrunch*, June 21, 2024. <https://techcrunch.com/2024/06/23/women-in-ai-anika-collier-navaroli-is-working-to-shift-the-power-imbalance/>
- Emam, Khaled, Lucy Mosquera, and Richard Hoptroff. *Practical Synthetic Data Generation*. O'Reilly Media, Inc, 2020
- Feher, Katalin. "Exploring AI Media. Definitions, Conceptual Model, Research Agenda." *Journal of Media Business Studies*, April 13, 2024, 2–4. <https://doi.org/10.1080/16522354.2024.2340419>
- Future of Life Institute. "Pause Giant AI Experiments: An Open Letter." Future of Life Institute, March 22, 2023. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- García Martín, Hector, Stanislav Mazurenko, and Huimin Zhao. "Special Issue on Artificial Intelligence for Synthetic Biology." *ACS Synthetic Biology* 13, no. 2 (January 12, 2024): 408–10. <https://doi.org/10.1021/acssynbio.3c00760>
- Google, ed. "2023 Environmental Report - Google Sustainability." Sustainability, 2023. <https://sustainability.google/reports/google-2023-environmental-report/>
- Hurst, Luke. "Fake Image of Pentagon Explosion Causes Real Dip on Wall Street." *Euronews*, May 23, 2023. <https://www.euronews.com/next/2023/05/23/fake-news-about-an-explosion-at-the-pentagon-spreads-on-verified-accounts-on-twitter>

- Hwang, JaeYoung, and SangHoon Oh. "A Brief Survey of Watermarks in Generative AI." The 15th International Conference on ICT Convergence, 2023. <https://ictc.org/media?key=site/ictc2023a/abs/P1-21.pdf>
- Iansiti, Marco, and Karim R Lakhani. "The Age of AI." Chapter. In *Competing in the Age of AI : Strategy and Leadership When Algorithms and Networks Run the World.*, 1–17. Cambridge, Massachusetts: Harvard Business Review, 2020
- Kanungo, Alokya. "The Real Environmental Impact of AI." Earth.Org, March 5, 2024. <https://earth.org/the-green-dilemma-can-ai-fulfil-its-potential-without-harming-the-environment/>
- Kietzmann, Jan, Linda W. Lee, Ian P. McCarthy, and Tim C. Kietzmann. "Deepfakes: Trick or Treat?" *Business Horizons* 63, no. 2 (March 2020): 135–46. <https://doi.org/10.1016/j.bushor.2019.11.006>
- Kira, Beatriz. "Deepfakes, the Weaponisation of AI against Women and Possible Solutions." *Verfassungsblog*, June 3, 2024. <https://verfassungsblog.de/deepfakes-ncid-ai-regulation/>
- Lagardère, Marion. "L'artiste Boris Eldagsen révèle que sa photo, récompensée par un Sony Award, est l'oeuvre d'une IA et veut ouvrir le débat." *Franceinfo*, April 19, 2023. https://www.francetvinfo.fr/replay-radio/l-etoile-du-jour/l-artiste-boris-eldagsen-revele-que-sa-photo-recompensee-par-un-sony-award-est-l-uvre-d-une-ia-et-veut-ouvrir-le-debat_5752295.html
- Liang, Weixin et al., "GPT Detectors Are Biased against Non-Native English Writers," *Patterns* 4, no. 7 (July 2023): 100779, <https://doi.org/10.1016/j.patter.2023.100779>
- Madiega, Tambiama. "Generative AI and Watermarking - European Parliament. European Parliament, December 2023. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI\(2023\)757583_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf)
- Martínez, Fernando Sevilla, Raúl Parada, Jordi Casas-Roma, "CO₂ impact on convolutional network model training for autonomous driving through behavioral cloning," *Advanced Engineering Informatics*, Volume 56, April 2023, <https://doi.org/10.1016/j.aei.2023.101968>
- Maslej, Nestor, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, et al. "The AI Index 2023 Annual Report." Stanford University | AI Index, April 2023. <https://aiindex.stanford.edu/report/>
- Merchant, Brian. "Your Boss Wants AI to Replace You. The Writers' Strike Shows How to Fight Back." *Los Angeles Times*, May 11, 2023. <https://www.latimes.com/business/technology/story/2023-05-11/column-the-writers-strike-is-only-the-beginning-a-rebellion-against-ai-is-underway>
- Microsoft, ed. "2022 Environmental Sustainability Report." Microsoft.com, 2023. <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW15mgm>
- Mishra, Satwik. "Generative AI and Authenticity Balancing Innovation and Integrity in Media." Center for Trustworthy Technology, 2024. <https://c4tt.org/wp-content/uploads/2024/01/Generative-AI-and-Authenticity.pdf>
- OECD. "What Is the Environmental Footprint of Artificial Intelligence?" OECD events, 2022. <https://www.oecd-events.org/cop27/session/f174ec37-5145-ed11-819a-00224880a4d8/what-is-the-environmental-footprint-of-artificial-intelligence->
- O'Hagan, Clare. "Artificial Intelligence: UNESCO Calls on All Governments to Implement Global Ethical Framework Without Delay." UNESCO, March 30, 2023. <https://www.unesco.org/en/articles/artificial-intelligence-unesco-calls-all-governments-implement-global-ethical-framework-without>
- Posetti, Julie, Diana Maynard, Kalina Bontcheva, Don K Hapal, and Dylan Salcedo. "Maria Ressa: Fighting an Onslaught of Online Violence. A Big Data Analysis." International Center for Journalists, March 2021. https://www.icfj.org/sites/default/files/2021-03/Maria_Ressa_-_Fighting_an_Onslaught_of_Online_Violence_0.pdf

- Radeva, Maria. "Guest Post: The Benefits and Risks of Chatgpt for Education." The Learning Scientists, February 15, 2024. <https://www.learningscientists.org/blog/2024/2/15-1>
- Ryan-Mosley, Tate. "The inside Scoop on Watermarking and Content Authentication." MIT Technology Review, November 6, 2023. <https://www.technologyreview.com/2023/11/06/1082996/the-inside-scoop-on-watermarking-and-content-authentication/>.
- Sala, Alessandra. "Ai Watermarking: A Watershed for Multimedia Authenticity." ITU, May 27, 2024. <https://www.itu.int/hub/2024/05/ai-watermarking-a-watershed-for-multimedia-authenticity/>
- Savage, Mark. "AI-Generated Drake and The Weeknd Song Goes Viral." BBC News, April 17, 2023. <https://www.bbc.com/news/entertainment-arts-65298834>
- Sherman, Justin. "'completely Horrifying, Dehumanizing, Degrading': One Woman's Fight against Deepfake Porn." CBS News, October 14, 2021. <https://www.cbsnews.com/news/deepfake-porn-woman-fights-online-abuse-cbsn-originals/>
- Shrungare, Jeenisha. "Exploring Synthetic Media: How Ai Is Changing Art." XRDS: Crossroads, The ACM Magazine for Students 29, no. 4 (June 2023): 64–65. <https://doi.org/10.1145/3596936>
- Song, Houbing, Yuan Zhang, Jose Neuman Souza, and Jianqiang Li. "Guest Editorial Ai-Driven Synthetic Biology for Human Wellbeing." *IEEE Journal of Biomedical and Health Informatics* 26, no. 10 (October 2022): 5042–43. <https://doi.org/10.1109/jbhi.2022.3204369>
- Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. <https://doi.org/10.18653/v1/p19-1355>
- UNESCO. "Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence." UNESDOC Digital Library, 2023. <https://unesdoc.unesco.org/ark:/48223/pf0000385629>
- UNESCO. "Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multi-Stakeholder Approach." UNESDOC Digital Library, 2023. <https://unesdoc.unesco.org/ark:/48223/pf0000387339>
- UNESCO. "Recommendation on the Ethics of Artificial Intelligence." Paris: United Nations Educational, Scientific and Cultural Organization, 2021 <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
- Wakefield, Jane. "Deepfake Presidents Used in Russia-Ukraine War." BBC News, March 18, 2022. <https://www.bbc.com/news/technology-60780142>
- Whittaker, Lucas, Tim C. Kietzmann, Jan Kietzmann, and Amir Dabirian. "'All around Me Are Synthetic Faces': The Mad World of Ai-Generated Media." *IT Professional* 22, no. 5 (September 1, 2020): 90–99. <https://doi.org/10.1109/mitp.2020.2985492>

Synthetic Content and its Implications for AI Policy:

A Primer

This publication fosters a better understanding of what synthetic content is and does and how the creation and use of synthetic content can impact individuals, societies, economies. It further discusses the possible ethical guardrails to be applied to generative AI and the creation of use of synthetic content, and highlights the role that policy makers and other relevant stakeholders can play to make sure that advanced AI systems and the content they generate do not harm people or impinge upon the well-being of individuals and societies alike. The ability to produce synthetic content can drive innovation in fields such as education, creativity, and data analysis, enabling tailored services and expanding access to knowledge. However, while enabling significant opportunities, synthetic content may challenge authenticity, privacy, trust and even democratic processes, among others. Moreover, a great concern is the rise of deepfakes, which are particularly harmful when used maliciously. Deepfakes often target women, leading to severe reputational and psychological impacts. The report also hints at the rapid increase in deepfake-related fraud, signaling broader social vulnerabilities. Key conclusions highlight the importance of devising and implementing solutions (e.g. watermarking; content moderation) making users able to distinguish what is real from what is synthetic. It is further underscored the value added of pursuing coordinated, globally applicable approaches, such as those referred to in the UNESCO Recommendation on the Ethics of AI.



unesco

United Nations
Educational, Scientific
and Cultural Organization

Social and Human Sciences Sector

7, place de Fontenoy
75352 Paris 07 SP France



https:

Follow us



9 789231 007279