



CEPIC

Centre of the Picture Industry

AI Ethical Guidelines for Responsible Re-Use and Production of VISUAL CONTENT

I. Preamble

Visual Content is facing a revolution.

Artificial intelligence has the potential to foster innovation, cultural development, growth, and prosperity thereby benefiting society at large including creators. However, if not used responsibly, AI generated products may cause unintentional harms, stifle innovation and growth and seriously infringe individual personal rights. These harms include:

- Fake news
- Infringement of Third-Party Rights, including copyright
- Bias and Diversity issues
- Hampering sustainable innovation by providing a competitive edge to irresponsible companies

It is important that companies who provide AI generated products respect core principles, so the benefits of AI products outdo the harms that may be caused.

Our primary concern relates to collection/ sourcing of datasets. Data are like natural resources we exploit; any exploitation needs to be done in a sustainable way to preserve the future of the natural resources, i.e. the visual works used by artificial intelligence for training. These resources must be supported to ensure they can renew and continue to sustain future AI training and accuracy. Thirty years of development on the Internet have shown that proper licensing of visual data is the best way forward to create a sustainable digital environment in which creators may thrive. A better, responsible, and sustainable AI will allow creators to create more, produce more AI.

These Guidelines are written with the goal to minimise the risk of unintentional harms when:

- 1) collecting/ sourcing existing visual data to train algorithms, commonly known as DATASCRAPING**
- and**
- 2) creating or producing new visual data using artificial intelligence (GenAI)**

The entire value chain should be accountable, from using existing data to train the algorithm to its publication. This document aims to answer the questions of investors in and producers of AI-generated visual content as well as its users.

These Guidelines focus on Visual Media and recognise that the same rules may not be applied equally to all Creative Content and to Editorial Content.

II. Definitions

- **Data in this document is defined as information derived from visual media, including its metadata (i.e. data about data), whether embedded, linked attached to the main file or in any other form.**
- **Data Collection** is defined as the process of sourcing visual data (input data) in a systematic manner from numerous different databases to train algorithms for data mining purposes that will be used either to alter other visual media or to produce new data (output data including Synthetic Content). Visual Data are usually collected on open and freely accessible websites.
- **Data Production** is defined as the generation of visual media content which includes already existing visual media content.
Public data is defined as information that an individual user with a personal computer, tablet or mobile device can access without restriction, and where the data is accessible to all. This does not mean that it is part of the Public Domain. Visual data protected by copyright may be freely accessible. This does not mean that they are also free to reproduce and distribute.
- **Generative visual content** is defined as any piece of media or content which had been either entirely generated by, or altered through, a trained neural network (AI) to such a significant degree that the original nature of the expression of the media is noticeably different-
- **Third Party Rights** include all Intellectual Property Rights (copyright & authors rights, trademark, patent, neighbouring rights), privacy rights including bio-metric rights and personality/image rights.

III. Best Practices or Principles

All along the process of AI generation, from collection/ sourcing of data to production of fully synthetic media the same overriding principles listed here should apply:

1. Lawfulness
1. Transparency
2. Necessity or minimization
3. Accuracy
4. Security

1. Lawfulness

The collection of data available on the Internet, whether for the purpose of Learning algorithms or to produce visual, modified, or generative content, must be done in compliance with national and international laws and treaties in force at the time of collection, particularly with regard to copyright, intellectual property, and data, including data privacy legislation/the RGDP, as well as the contractual conditions of the services of the scrapped websites.

No data is to be used without consent.

Specifically:

Images should only be collected from sites that have lawfully obtained the visual content, i.e. the relevant licenses. They should only be displayed publicly, distributed in any way or, for that matter, sold with the required authorization to do so.

Organizations disseminating artificial intelligence models and datasets, whether open source or not, to train learning models for algorithms capable of generating digital images in particular, must indicate the legal basis on which they have obtained the data: either through an exploitation license/or an exploration license or machine learning license, or on the basis of a copyright exception that must be specified.

Terms of service of website need to be respected. If the terms of service do not permit scraping of copyright protected data, this should be considered as sufficient to qualify as opt out from the application of the text and data mining exception under European Union law (Art. 3 or 4 of the Directive Copyright in the DSM 790/2019/ EU)

In the event of unauthorised use or non-compliance with the conditions of use, the unlawfully retrieved content/data must be removed at the request of the rights holders.

Producers of visual content modified or generated using artificial intelligence algorithms are subject to additional obligations.

Specifically:

Creators of generative visual content should never produce content:

- *which is misleading and/or causes harm to third parties*
- *Defaming any person by representing that a person acted or behaved in a way they did not, or saying things that they did not say, etc.*
- *Creating “Fake News” by representing the dissemination of events that did not occur or falsification of these events to alter public opinion.*

2. Transparency

Transparency is essential because it ensures that the entire production chain, from data collection to generative creation, operates smoothly and is held accountable.

Without an obligation of transparency, all the principles listed in this document, including the first principle of lawfulness, cannot be applied. Without transparency, it is not possible to remunerate rights holders who wish to do so, nor can they exercise their right to opt-out.

Data access, collection and processing must therefore be organized transparently and simply to ensure that third parties are able to exercise their rights if necessary.

The producer of manipulated visual content, and a fortiori generative content, must respect the rules of transparency regarding the way in which the content was created.

Specifically:

Records of all content collected (including data collected on publicly accessible sites) should be maintained, including that represents the Visual Data collected, an

Records of the visual content collected (including data collected on publicly accessible sites) should be maintained, including the websites used for collection and any content ID or any reference enabling the visual data collected and the rights attached to it to be identified.

If a user account is created for the purpose of Data Collection, this email account should be monitored by the company or individual scraping the Data.

For transparency purposes accurate source information of all content, such as technology providers’ bandwidth, IP address purchased and used should be retained.

Transparency is extremely important to avoid misuse of visual content, such as falsification of information, violation of data confidentiality and copyright infringement.

Visual content, whether modified or generative, must therefore be marked in an appropriate, timely, clear, and visible manner, indicating as far as possible the name of the natural or legal person who generated or manipulated it.

Producers of visual content modified or generated using artificial intelligence algorithms are subject to additional obligations.

Specifically:

Ensure transparent and informed consent of persons photographed complying with GDPR; appropriate model release forms are available. When working with persons, make sure that consent is obtained through legally drafted model releases.

Inform users about capabilities, limitations, and potential risks of manipulated or generative visual content, for instance on producer's website.

Inform users that the media or content is generative or includes modified elements, through labelling, metadata, or cryptographic means.

Support and/or participate in all organisations/ initiatives aiming at authenticating media content through metadata or other means.

Companies should document their content creation process to show that humans are involved in the process and to avoid any claims that AI created the entire content.

3. Necessity or minimisation

Lawful access, collection, and any treatment of Data (including Public Data) should be limited to what is directly relevant and necessary to accomplish the specific project. Data should only be retained for as long as it is necessary to fulfil that purpose. In case of licensed content the Data shall only be retained for as long as provided under the license granted.

Specifically:

Data should only be collected for the purpose of training algorithms following a risk analysis to ensure that the Data collected is necessary for the project and will not cause any of the harms identified in the introduction of this Guidelines.

All collected Data not necessary to train an algorithm should be deleted, with records kept identifying the source.

If collected Data is stored it should only be stored for the time frame strictly necessary to carry out content analysis and be deleted after the use is no longer required.

The amount of Data collected over time from any single source both in terms of quality and in quantity should remain minimal and "insubstantial".

The frequency of websites visits should not materially increase the volume of traffic of any website.

Restrict use to Data that is necessary to the production. Unused image files should be deleted.

4. Accuracy

Personal data shall be compliant with all privacy laws; they should also be accurate. Accuracy of all data used to train the system is important as otherwise it can perpetuate or even amplify biases that exist in the trained Data.

The accuracy of all data used in learning algorithms is essential, as otherwise any biases can be perpetuated or even amplified in the final product.

Specifically:

Every reasonable step must be taken to ensure that Personal Data that is inaccurate, incorrect, or misleading, is erased or corrected.

The AI algorithm used to produce new content should not be trained on ethically suspicious Data and should be obtained legally. With these requirements, licensing is preferred.

Collected Data should avoid biases in the selection of Data that should be made in good faith.

Bias should therefore be addressed by best effort in using an unbiased dataset containing as many and diverse persons in various contexts, and by being as inclusive as possible.

5. Security

Security is important to avoid any copyright infringement, privacy breach and also to allow any misuse of the system to intentionally introduce bias in the system...

All personal and proprietary Data collected in a text and data mining process should therefore be protected. This includes licensed content (i.e. content under copyright) which may be used and shall be adequately protected against unauthorized access).

Specifically:

All technical and organisational measures must be taken to protect the collected Data against possible attacks or hacks.

Integrity and security of websites used for data collection and data mining should not be altered or violated through reverse engineering activities.

All measures to protect the collected Data must be protected against possible cyber-attacks or hacks.

It should not be possible to reverse engineer the AI-generated images to reveal any underlying individual images.

Prompt adjustments should be made whenever a complaint is received regarding labels, downranking, removal, or any other useful interventions, when intentionally misleading and harmful generative content is distributed on a platform.

Publicly available internal compliance policies should be introduced and adhere to strict use restrictions as outlined in these Guidelines.

Introduce publicly available internal compliance policies which shall be in line with these Guidelines. Educate employees of the legal obligations in Data Collection and conduct regular reviews of policies to ensure compliance with best practices and all applicable laws.

Licensing of Data should be preferred.

V. Conclusion:

CEPIC is convinced that Artificial Intelligence in the visual sector brings about great opportunities to visual content creators and producers.

For this contribution to be only constructive, it is necessary and sufficient to respect the basic principles described above.

These already exist in current legislation (RGDP and Intellectual Property) and can easily be applied to new usages.

CEPIC will be a strong advocate of these basic principles.

The **AI Ethical Guidelines for Responsible Re-Use and Production of VISUAL CONTENT** will be regularly up-dated so that examples reflect changing practices and technology.

20.10.2023

ABOUT THE CEPIC AI WORKING GROUP

CEPIC's AI Working Group was created in May 2022 with the aim of establishing a best practice guide for the ethical use of AI and monitoring its beginnings in the imaging industry. It has been composed with an international, diverse and inclusive approach, bringing together participants from different parts of the debate. Participants are drawn from visual content creation professionals, legal experts and AI specialists.

This working group is permanent and evolving.

More information on the composition and evolution of the AI working group can be found on the CEPIC website www.cepic.org

ABOUT CEPIC

CEPIC represents hundreds of Picture Libraries and Agencies representing hundreds of thousands of photographers whose core business is the direct licensing of visual content off-line and online. Acting as right holders, Picture Libraries and Agencies license digital assets for all kinds of commercial uses, to newspapers, magazines, advertising, broadcasters, etc. CEPIC members are continuously adaptive towards innovative technology solutions for the growth in digital enterprises and have developed sophisticated digital platforms to both market digital content online and provide digital access to images. Amongst CEPIC members are global players such as **Getty Images**, **Shutterstock**, **Adobe**, visual content marketplaces such as **Alamy** and **Dreamstime**, fine arts libraries such **Bridgeman Images** or **AKG Images**, historical archives such as **Roger-Viollet**, news agencies such as **Belga** (Belgium), **TT** (Sweden), **ANP** (Netherlands) or the **AFP** (France), more recently AI focused image libraries such as **VAlsual**, photographers associations such as the **Union des Photographes Professionnels**, the **UPP**, as well as representatives for European trade associations **AEAPAF**, **BAPLA**, **BLF**, **FNAPPI**, **SAB**, **SAPHIR** and **SNAPIG**.