



# CEPIC

Centre of the Picture Industry

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

### **Visual Content is facing its next 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 infringe individual personal rights. These harms include:

- Fake news
- Infringement of Third-Party Rights
- 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. 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 and 2) when creating visual data using artificial intelligence. **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 and producers in 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. These Guidelines investigate two processes:

- **Data Collection/ Sourcing and**
- **Data Production.**

CEPIC International Association of the Media Licensing Industry A.I.S.B.L

(Nbr 0786.351.284)

Quai aux Pierres de Taille 29, 1000 Brussels, Belgium

[ceplic@ceplic.org](mailto:ceplic@ceplic.org)



# CEPIC

Centre of the Picture Industry

## 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 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 terms of services of the website does not prohibit the collection of data. Copyright protected visual data cannot be considered public data unless put into public domain by the copyright owner.

**Synthetic Media/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, and most importantly without human interference after the initial data input.

**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.

**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
3. Accuracy
4. Security



# CEPIC

Centre of the Picture Industry

## DATA COLLECTION/ SOURCING

### 1. Lawfulness

The Collection of Data (including Public Data) available on the Internet, either for training purposes or to produce modified or Synthetic Media/Content should be done within the boundaries of existing national and international laws and treaties at the time of collection, including copyright, intellectual property, and data privacy legislation/GDPR, as well as of contractual terms of services of scraped websites. No Data should be used without required authorization to do so. Access, collection, and treatment of Data (including Public Data) should be done in a transparent easily accessible manner to ensure that Third Parties are able to exercise their rights, if required.

**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.*

*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 (Directive Copyright in the DSM 790/2019/ EU)*

### 2. Transparency

Access, collection, and treatment of Data (including Public Data) should be done in a transparent easily accessible manner to ensure that Third Parties are able to exercise their rights, if required



# CEPIC

Centre of the Picture Industry

**Specifically:**

*Records of all content collected (including Public Data) should be maintained, including the websites used for collection and any content ID that represents the Visual Data collected.*

*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.*

### 3. Necessity

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.*



# CEPIC

Centre of the Picture Industry

*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.*

#### 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

***Specifically:***

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

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

*Licensing of Data should be preferred.*

#### 5. Security

All personal and proprietary Data collected should be protected. This includes licensed content (i.e. content protected by copyright) which may be used under a license and shall be adequately protected against unauthorized access. 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...



# CEPIC

Centre of the Picture Industry

**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 should not be altered or violated through reverse engineering activities.*

*All appropriate measures should be taken to comply with any request to remove, and cease to use, any content.*

*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.*

## DATA CREATION (SYNTHETIC CONTENT)

### 1. Lawfulness

The production and use of Synthetic Media/Content should be carried out in compliance with applicable legislation with a special focus on copyright/intellectual property and data privacy laws.

**Specifically:**

*Synthetic media/Content Creators should never produce Synthetic media/ Content that is deceptive in a way that causes harms such as:*

- 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.*



# CEPIC

Centre of the Picture Industry

## 2. Transparency

Synthetic Media/Content production should respect transparency rules about the way they have been produced. Transparency is important to avoid a misuse of Data including a breach of data privacy and copyright infringement.

### ***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 Synthetic Media or Synthetic Content, for instance on producer's website.*

*Inform users that the media or content is synthetic, through labelling or cryptographic means, when the media created includes synthetic elements.*

*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

Production of Synthetic Media/Content should be done within the scope and purpose of a determined project and when Third Party Rights can be lawfully obtained



# CEPIC

Centre of the Picture Industry

***Specifically:***

*Restrict use to Data which has been collected in accordance with the requirements listed in the section above with regards to the **Data Collection/ Sourcing***

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

#### **4. Accuracy**

Synthetic Media/Content should be made in good faith and avoid creating biases in the process of creation.

***Specifically:***

*Synthetic Media/Content should avoid biases by properly and accurately selecting Data, notwithstanding whether the origin is editorial or synthetic.*

*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.*

*Bias should 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**

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





# CEPIC

Centre of the Picture Industry

***Specifically:***

*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 synthetic media is distributed on a platform.*

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

*Employees should be educated of the legal obligations in the creation of Synthetic Media/Content and conduct regular reviews of policies to ensure that policies continue to comply with best practices and changes. Policies should be enforced companywide.*

**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 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.



# CEPIC

Centre of the Picture Industry

## **ABOUT THE AI WORKING GROUP**

The CEPIC AI Working Group was created in May 2022 with the goal to establish, as a first step, Guidelines for an Ethical Use of AI and follow its inception in the picture industry.

**We have a diverse and inclusive international group from various sides of the debate.**

- The group is coordinated by **Valérie Théveniaud-Violette** - previously active on the **Copyright Enforcement working group** (See [here](#))
- The group further includes **Marisol Muniz, Nancy Wolff, Florian Koempel** for the legal side.
- **Yair Adato** and **Ravit Dotan** (BRIA), **Michael Osterrieder, Mark Milstein**, and **Keren Flavell** (VAIsual) for the developers and licensors
- **Paul Seheult** (CEPIC, BAPLA, PICSEL) will wear these many hats in the group and ensure the liaison

## **ABOUT CEPIC**

CEPIC represents hundreds of Picture Libraries and Agencies represent hundreds of thousands of photographers, videographers and increasingly synthographers, 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, Magnum Photos, and Alamy, fine arts libraries such the Bridgeman Images, historical archives such as Roger-Viollet and AKG Images, news photo syndication such as Le Figaro, news agencies such as Belga, TT or ANP as well as representatives for European trade associations AEAPAF, BAPLA, BLF, SAB, SAPHIR and SNAPIG. CEPIC's membership also includes several cutting-edge technology developers, some of them involved in manipulating images with AI driven tools, others in creating/ producing AI generated images.

CEPIC International Association of the Media Licensing Industry A.I.S.B.L

(Nbr 0786.351.284)

Quai aux Pierres de Taille 29, 1000 Brussels, Belgium

[ceplic@ceplic.org](mailto:ceplic@ceplic.org)