

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/377187713>

Eastern Journal of European Studies Legal issues concerning Generative AI technologies

Article in Eastern Journal of European Studies · January 2024

DOI: 10.47743/ejes-2023-0203

CITATIONS

0

READS

258

3 authors, including:



[Carmen Ungureanu](#)

Universitatea Alexandru Ioan Cuza

40 PUBLICATIONS 36 CITATIONS

[SEE PROFILE](#)



[Aura Elena Amironesei](#)

Universitatea Alexandru Ioan Cuza

7 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)

Legal issues concerning Generative AI technologies

 Carmen Tamara Ungureanu^a,  Aura Elena Amironesei^a

^aAlexandru Ioan Cuza University of Iasi, Romania

Abstract

We are witnessing an accelerated technological evolution that has enabled the development of artificial intelligence in various fields, allowing it to gradually infiltrate the entire society. We intend to cover only a small subset of AI technologies in our paper, that of Generative Artificial Intelligence (GenAI). Our objectives are to shed light on the legal issues that GenAI can cause and to find solutions to them. We begin with a definition of GenAI in the much broader context of AI technologies. Answers to a few essential questions are to be found: 'How does GenAI work?', 'What could GenAI be used for?', 'What legal issues could arise from using a GenAI?'. To accomplish our goals, we first conduct a literature review to define artificial intelligence (AI) in general and GenAI in particular. Several lawsuits are chosen to illustrate the magnitude of the legal problems and to test the feasibility of possible solutions in both the national and EU legal systems. Then, we analyse GenAI's output, liability for its contents and for its use, altogether with examples of related contractual clauses.

Keywords: generative artificial intelligence, training data, civil liability, advertising

Introduction

ALIE (*Applied Lucent Intelligence Emulator*), the artificial intelligence that played a key role in the science fiction series *The 100* (an American post-apocalyptic science fiction drama series, available on Netflix), was designed to make life better for the humankind. However, as ALIE was not programmed to feel emotions or to value life, she tackled the overpopulation problem in a very effective manner, by wiping it out, which ultimately led to the destruction of the human race (almost) (Day, 2016). Fears about this possibility of AI evolving were early expressed (when AI was in its infancy): Elon Musk, 2018 - 'Mark my words – A.I. is far more dangerous than nukes' (Clifford, 2018); Stephen Hawking, 2014 - 'The development of full artificial intelligence could spell the end of the human race' (Cellan-Jones, 2014).

Now, we are witnessing an accelerated technological evolution that has enabled the development of artificial intelligence in various fields, allowing it to gradually infiltrate the entire society, sneaking into everyday life, almost unnoticed.

*✉ Carmen Tamara Ungureanu, Professor, PhD, Faculty of Law, Alexandru Ioan Cuza University of Iasi, Romania; email: carmen.ungureanu@uaic.ro.

AI is typically deployed in software applications that run on computers rather than in human form (Custers & Fosch-Villaronga, 2022), with a few exceptions, as Sophia, who is a Saudi Arab citizen, Nadine, Ameca, and others (Bellapu, 2023).

Examples of AI systems that are frequently used are multiple: AI that relies solely on software, such as smart assistants, image analysis software, search engines, foreign language translation applications (Google Translate, Reverso, etc.) voice and facial recognition systems, natural language processing (NLP) tools (Custers & Fosch-Villaronga, 2022, p. 7); AI embedded in hardware devices, such as robots, autonomous vehicles, drones or IoT (Internet of Things) applications (EESC, 2018, p. 51) etc.

As AI is a double-edge sword (with positive and negative consequences), to deal with the AI risks and the possibility of unwanted results, at the European Union level, in April 2021 the *Regulation establishing harmonized rules on artificial intelligence (Artificial Intelligence Act)* was proposed, and on September 28, 2022, the *Directive on the adaptation of non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*. These proposals are among the first international hard law norms.

We intend to cover only a small subset of AI technologies in our paper, that of Generative Artificial Intelligence (GenAI).

Our goals are to shed light on the legal issues that GenAI can give rise to and find solutions to them. We are going to start from the definition of the GenAI in the much broader context of AI technologies. Answers to a few essential questions are to be found: ‘How does GenAI work?’, ‘What could GenAI be used for?’, ‘What legal issues could arise from using a GenAI?’.

The methodology we will employ to achieve our objectives will consist of four components: a brief literature review, legal case law studies, case studies of a few GenAI providers' adhesion contracts, and a basic legislative evaluation. First, we will use a literature review to define artificial intelligence (AI) in general and GenAI in particular. Several case studies (still pending) in US courts will be used to illustrate the magnitude of the legal problems and to test the feasibility of possible solutions in both the national and EU legal systems. Another methodological approach will be case studies of existing contractual provisions, in order to identify the legal issues that might arise out of GenAI's output usage and to assess the mechanisms suitable for ensuring protection for both providers and users of GenAI. Throughout our attempt, we shall make references to several pieces of legislation either in force or in the proposal stage.

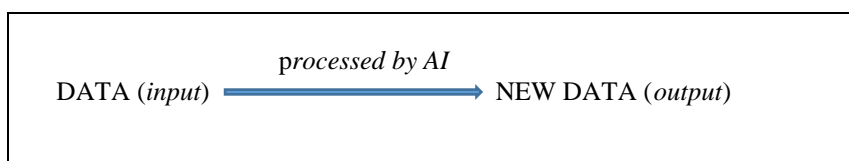
1. What does GenAI technology mean? How does it work?

AI is a difficult concept to define because it is a heterogeneous phenomenon with many facets. Considering AI as machines and *software* with specifically human skills and intelligence does not encompass the vast array of applications and

situations in which AI can be used (Bertolini, 2020, p. 9). There is no consensus (Bertolini, 2020, pp. 23-29) on the definition of AI (Häuselmann, 2022, p. 44).

AI has been viewed differently in technical and legal literature. For example, one author (Surden, 2019, p. 1307) who stated that he aims to explain AI in a way that lawyers can understand considered AI as a technology that automates tasks that normally require human intelligence. In a similar vein (Häuselmann, 2022, p. 46), AI refers to machines that can adapt (i.e., learn from experience) and automatically perform activities and tasks (i.e., make their own decisions), typically claiming human cognitive abilities. Other authors (Russell & Norvig, 2021, pp. 1-5) explained AI in terms of four characteristics: it thinks like a human, acts like a human, thinks rationally, and acts rationally. In another way (Paschen et al., 2020, p. 405), AI can be viewed as a computer system that acts rationally and solves problems based on the information/data it has access to/that is available to it. Specifically, AI can be explained using the *input-process-output* model. This means that AI receives data (input), processes it (process), and then provides other data (output).

Figure 1. Generative AI's basic functioning mechanism



Source: authors' representation

AI can be explained in two ways: narrowly and broadly. An AI technology, which is capable of solving a specific problem or performing a specific task, falls into the first category and refers to all existing AI systems: facial and image recognition systems (such as those used by Facebook and Google for automatic identification of people in photos), virtual assistants/smart assistants or chatbots (Siri, Alexa, ChatGPT), autonomous vehicles (cars, drones, ships etc.), etc. In a broad sense, AI is capable to fully mimic the functioning of the human brain (Häuselmann, 2022, p. 45) and is called AGI (Artificial General Intelligence) or Full Artificial Intelligence, as Stephen Hawking termed it. There is currently no AGI (luckily).

AI is an umbrella term, giving the impression that AI is a single entity (Bertolini, 2020, p. 15). In practice, however, AI comes in different forms (Häuselmann, 2022, pp. 47-67). Depending on the AI technologies results/outputs, they can be classified into AI systems that generate content (generative AI), which generate predictions, recommendations, or decisions. According to the proposed European regulation, Artificial Intelligence Act, the amended text from 14th of June 2023, 'artificial intelligence system' (AI system) means 'a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit

or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments' (Article 3, point 1).

In this paper, we will only discuss the AI technologies that generate content (GenAI). Depending on the quality of the data used to train it, GenAI can provide a wide range of new and creative content (texts, images, music, video, and other human-specific forms of expression) (Muller et al., 2022, pp. 1-7).

2. What could GenAI be used for?

The extended usage of GenAI is not a novelty to any industry. Multiple uses have already been predicted a few years ago and they continue to proliferate because, in general, the option of utilization is left to the user, not the provider of the AI services. Therefore, even though GenAI was initially developed for a specific goal, this may change when the content is exploited, as the user may decide the output is suitable for other objectives, unless the scope is contractually constrained, and the user cannot opt out.

The initial usage derives from the nature of GenAI content, which can be audio, text, graphic images (Garon, 2023, p. 13) and video. Further, the content delivered by GenAI is applied in multiple economic fields and industries, but also for domestic/personal purposes.

One of the most prominent economic applications of GenAI is in marketing (Garon, 2023, p. 23). First it can be used to generate advertising content such as advertisements, commercial texts, logos, slogans¹ and even social media content (Garon, 2023, p. 18). Second, it can be 'employed' for marketing strategies and branding (Garon, 2023, p. 18). Provided that GenAI is trained with a significant amount of data, it can 'learn' from previous advertising campaigns of numerous marketers and produce the best strategy. Third, AI-generated influencers are preferred over human ones, as the former are deemed to be more effective than humans due to generating a higher engagement (Klein, 2020). A misuse of GenAI in this field (of advertising) was pointed out though. It seems that many unreliable websites are created by GenAI where paid ads are served, apparently without the knowledge of the brands, which spend large amounts of money for advertising. The result, in a not very long term, would be an Internet entirely ruled by AI (Tate, 2023).

The rise of the metaverse is nowadays opening doors for GenAI to enter a new field. Metaverses might be populated with AI-generated avatars, and these avatars are likely to have multiple roles.

Fashion industry also takes advantage of GenAI by using AI-models for their fashion campaigns and presenting the clothes virtually (Demopoulus, 2023) (through images or videos). Having an AI-model saves costs with the photographer, make-up artist, hair artist etc. Fashion shows might also be organised in the metaverse (Peter,

¹ See, for example, Copy.ai, <https://www.copy.ai/tools>.

2023), where AI models walk on the runway, wearing the digital twins (Fallmann, 2022) of the branded clothes. Moreover, GenAI may also predict and confirm trends in fashion (Banerjee & Mohapatra, 2021, pp. 165-180).

Other economic uses of GenAI that creates text are for customer support², generating business models (McKinsey, 2023), predictions, translations, emails drafting, video presentations, websites and the content for populating them. It can be integrated into different businesses as an API³, 'a set of functions and procedures allowing the creation of applications that access the features or data of an operating system, application, or other service'⁴, trained with the buyer's data and used internally for its business⁵.

Also in the journalistic field, GenAI can fill websites and newspapers (Garon, 2023, p. 59), by generating articles that are (or should be) slightly reviewed and adjusted before being published; at the same time, GenAI can also provide biographies (Meta.ai, n.d.a), relevant for writers. Game industry is as well among the 'beneficiaries' of GenAI (Gatto, 2023), as it is highly reliant on content (Generative AI, n.d.). GenAI can write computer code and generate videos, both being combined for games. GenAI is suitable for more other economic purposes, as it can develop new food recipes (thus being used in gastronomy), designing new products (Garon, 2023, p. 30) to meet customers' needs (regardless of the industry), drug research and development in the pharmaceutical industry, medical developments (Meta, n.d.), generated prototypes of cars and planes etc.

On the legal side, there are also multiple uses of AI. From replacing search instruments like Wikipedia and Google when it comes to legal provisions and official documents issued by authorities, to searching case-law, and even generating contracts and legal advice. However, as a recent case has proved (where a lawyer had used ChatGPT to produce a legal brief for a case in Federal District Court that was replete with fabricated legal citations and judicial opinions) (Weiser, 2023), legal advice provided by GenAI, without being verified by a human, could be disastrous. Currently, judges from different US courts have started requesting lawyers to certify that they did not use AI to draft legal documents without a human reviewing or even issuing orders asking to disclose the usage of GenAI tools (Merken, 2023).

Another economic goal is data augmentation. It is now considered that only tech giants that have access to large sets of data can truly develop artificial intelligence, as for its training there is required a huge amount of data. However, the final use of GenAI is data augmentation (Generative AI, n.d.), which refers to creating additional new data (Li et al., 2021) based on the input. As a result, GenAI

² See, for example, Ultimate.ai, <https://www.ultimate.ai/>.

³ For more information on APIs, see Getting Started with the Copy.ai Workflows API, <https://docs.copy.ai/docs>.

⁴ According to Oxford Languages.

⁵ Such an example is Gretel. <https://gretel.ai/>.

can multiply data and generate new data for its own training, thus helping for the AI system to enhance.

Regarding personal/domestic use of AI, GenAI can heighten creativity and enhance the analytic human abilities (Wilson, 2018). The downside is anyone could fall into the trap of fake news such as the new cat breed called ‘Serpens Catus’ (Hobbs, 2023), an information which was retrieved and reposted by many online news websites and social media pages (Jonsson & Tholander, 2022, p. 5) which in fact is a fictional one created by AI.

Furthermore, conversational AI, which is a type of GenAI, might have a personal purpose. Conversational AI is defined as a computer system capable ‘of engaging in natural language conversations with humans’ (Jaspet.ai chat, n.d.). This use is considered to be for entertainment purposes (McKinsey & Company, 2023). However, as already mentioned, behind the arousal of simply chatting with AI, conversational AI has as well economic purposes and it is not risk-free (El Atillah, 2023).

Finally, there might as well be a purely scientific and educational use of GenAI – the academic purpose. Students are taught to use it for inspiration in writing novels, papers and for doing research (UK Department for Education, 2023). However, it is worth mentioning that there are certain risks posed to academic integrity (Eke, 2023).

3. What legal issues could arise from using a GenAI?

At the heart of GenAI technology is data. AI, regardless of type, operates on and thanks to massive amounts of data. To detect the legal issues that the use of GenAI may cause, we will look at the model that this technology operates on *input-process-output*. Thus, we will first analyse whether the data with which the GenAI technology is ‘fed’ (the input) can raise questions regarding legality, and then, we will investigate whether the results obtained by processing the data (the output) can involve legal aspects, particularly civil liability. We will not address the *process* phase, as it comprises no legal aspects, but only technical ones.

3.1. Legal issues concerning the inputs

In the first part we will clarify the distinct types of data that can be used as training data for GenAI. Following that, we will discuss some recent litigation involving training data that is alleged to infringe intellectual property rights.

Training data categories

Where does the data that feeds and trains a GenAI technology come from? ChatGPT said:

‘OpenAI trained me on a diverse range of internet text, including websites, books, and social media. The training data was gathered from a range of sources and was filtered to ensure that it is high-quality and appropriate for a general audience. The data was pre-processed to clean and normalize it before it was used to train me. The specific sources of the training data are not publicly disclosed’⁶.

The new launched Meta’s AI model, Llama 2 ‘was pretrained on publicly available online data sources and over one million human annotations’, according to Meta (Meta.ai, n.d.b). FreeWilly1 and FreeWilly2, launched on 21st of July 2023 by Stability AI, the company behind the Stable Diffusion image generation AI, are both ‘based off of versions of Meta’s LLaMA and LLaMA 2 open-source models, but trained mainly on synthetic data’ (Franzen, 2023).

What kind of data is GenAI powered by? According to some authors (Sobel, 2020), who take intellectual property rights as a reference system, the training data of GenAI falls into three categories: data that is not subject to intellectual property rights, data that is exploited under different licenses and data protected by intellectual property rights but not authorized for use. If the *licit source* of the data is used as a criterion for classifying the training data, then they could also be divided into three categories, which overlap to some extent with those in the first classification: (a) open data, (b) data provided in a voluntary way by users, consumers, professionals in their online interactions and (c) data protected by various mechanisms (intellectual property rights, trade secrets, other means of protection). At these three categories of data, it should be added a fourth one, that of the *synthetic data*, the use of which is exponentially growing (d).

We will briefly analyse the data according to the last classification.

(a) The first data category includes two types of data: open data and data that was previously protected by intellectual property rights but has currently entered the public domain.

According to Directive (EU) 2019/1024 on open data and the re-use of public sector information (hereinafter, the ‘Open Data Directive’), the concept of open data is generally understood as designating data in open format that can be freely used, re-used and shared by anyone for any purpose, whether private or not, commercial or non-commercial (Recital 16), and originating from documents held by public sector bodies, bodies governed by public law and public undertakings as defined in the Directive (Carsaniga et al., 2022, p. 21). For example, geospatial, and citizen-generated data, as the data generated by a public transport app that informs passengers on the timetables, the actual position of the means of transportation, the estimated time of arrival; the geospatial data on properties that are linked to a

⁶ ChatGPT was accessed for the answer on 2nd February 2023.

position on earth provided by a satellite or by cell phones based on GPS location coordinates (Carsaniga et al., 2022, p. 21).

Articles 13 and 14 of the Open Data Directive provide for a list of high-value datasets held by public sector bodies and public undertakings, listed in Annex I of the Directive, as thematic categories: *geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, mobility*. On 20 January 2023, the European Commission has published ‘a list of high-value datasets that public sector bodies will have to make available for re-use, free of charge, within 16 months’ (European Commission, 2023). This list is provided in the *Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use*, the main goal of which is ‘to ensure that public data of highest socio-economic potential are made available for re-use with minimal legal and technical restriction and free of charge’ (Recital 2). These datasets are to be ‘made available for re-use under the conditions of the Creative Commons Public Domain Dedication (CC0) or, alternatively, the Creative Commons BY 4.0 license, or any equivalent or less restrictive open license’ (Article 4.3.).

CC0 means ‘no rights reserved’, which should be understood as anyone may freely build upon, enhance, and re-use the datasets for any purposes without restriction under copyright or database law (Creative Commons, n.d.). Creative Commons BY 4.0 license allows the users to share the datasets (copy and redistribute them in any medium or format) and to adapt them (remix, transform, and build upon them for any purpose, even commercially), having the obligation only to attribute the work/to acknowledge the paternity to the licensor (Creative Commons, Attribution 4.0 International, n.d.).

An *open license* as referred to in Article 4.3 (mentioned above) could be, as well, one of the *open-source licenses* that comply with the requirements of the Open Source Initiative. These types of licenses allow software to be freely used, modified, and shared (Open Source Initiative, n.d., b). For instance, the European Union Public License, version 1.2 (EUPL-1.2) (Open Source Initiative, n.d., a) is an open license. The licensee may use the original work in any way he/she wants, the licensor granting him/her a worldwide, royalty-free, non-exclusive, sub-licensable license. Nevertheless, the licensee has obligations, as well, among which that of acknowledging the paternity of the creator (the attribution obligation), of respecting the copyleft clause, the compatibility clause, of indicating the source code (Opensource.org, n.d.)

Where public sector bodies/undertakings have *sui generis* rights over databases they are not allowed to exercise them to prevent the re-use of documents/data or to restrict their re-use (Article 1(6) of Open Data Directive).

Among the purposes of re-using open data is that of the AI development (Margoni & Kretschmer, 2022, p. 699). Creative Commons licensed works, for

instance, can be used in the development of new technologies, such as training of artificial intelligence software (Creative Commons FAQ, n.d.).

The same category includes data/datasets regarding creations, which have been protected by intellectual property rights, but which have entered the public domain. The consequence is that being in the public domain they no longer benefit from legal protection and can be used by anyone without restrictions. The rule is that legal protection is limited in time, the terms being specific to each individual intellectual property right (Dominte, 2021, pp. 76, 109-110, 172, 214-216) and jurisdiction. Regarding copyright, by way of exception, *moral copyright* is unlimited in time (in continental law). Thus, for example, the right to authorship of the work/the paternity, to its integrity, does not expire. Author's patrimonial rights, however, enter the public domain, according to Romanian legislation, after 70 years from the author's death (as a rule) (Dominte, 2021, pp. 282-283); from this date the work can be used by any person freely, including as training data for AI. Being in the public domain, the work can be reproduced, distributed, retransmitted, derivative works can be made, without restrictions (Articles 10, 12 and 13 from the Romanian Law no. 8/1996 on copyright and related rights).

(b) The second category of training data is data voluntarily provided by users, consumers, professionals in their online interactions. These data/datasets are data that they licence, by expressing the consent stipulated in the Terms & Conditions (T&C) / Terms of Use /Terms of Service/Terms, which no one reads, except when a dispute arises. For example, Instagram in its Terms of Use specifies that: 'We do not claim ownership of your content, but you grant us a license to use it' (Instagram, n.d.); TikTok has a similar provision (TikTok, n.d.) and GitHub (a code hosting platform, where the developers work together to build, scale, and deliver secure software) as well (GitHub Docs, n.d.).

Another source of training data for AI could be data provided in an altruistic way, or the so-called *data altruism*. As Article 2.16 from the Data Governance Act (DGA) reads:

“data altruism” means the voluntary sharing of data on the basis of the consent of data subjects to process personal data pertaining to them, or permissions of data holders to allow the use of their non-personal data without seeking or receiving a reward that goes beyond compensation related to the costs that they incur where they make their data available for objectives of general interest as provided for in national law, where applicable, such as healthcare, combating climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy making or scientific research purposes in the general interest’.

Finally, personal data could be provided as input/training data. All the above from this category could be personal data or mixt data. Whether provided by a user or by the developer of the AI system, this automatically falls under data privacy protection. When it comes to the developer, certain filters could be put in place for assuring that no personal data processing will be performed. However, the content provided by the user when using conversational AI might not be filtered easily (Business & Human Rights Centre, 2023) and neither the collection of personal data prevented (Ericsson, 2022). In the EU, when personal data is processed, the General Data Protection Regulation ('GDPR') is applicable. If the processing is not lawful, then the provider of the AI might be held liable. Moreover, if fed with personal data, GenAI can generate personal data in the output (such as real names, addresses), hence make them publicly known. A proposed solution would be to ensure that the GenAI is 'fed' only with anonymised data and apply data filters that detect and remove personal information from the training data.

As it is incredibly burdensome for the GenAI provider to implement effective mechanisms for checking whether the personal data provided by a user has been lawfully collected, the contractual clauses set forth between the provider and the user are meant to shift the liability from the former to the latter. Examples of such clauses are comprised in Synthesia's terms of service (point 10 - Software License and Content). A second safe measure taken is the Privacy Policy, which regulates a notice and take down system, together with the discretionary right of the provider to erase the unlawful collected data: 'where content includes *personal information* about *private individuals* this will be further regulated by our *Privacy Policy*, DPA, or other individual agreement' (Synthesia, n.d.).

In what concerns the intentional training of GenAI with personal data, the provider is liable when infringing GDPR's provisions. Even if this regulation is already well known in all industries, worldwide, breaches still happen. Recently, an AI software developer has been fined by the Italian data protection authority for unlawful processing of biometric data, used to train facial recognition AI (European Data Protection Board, 2022). The authority also ordered the erasure of the data, leading to a loss of training data. However, a question mark remains in relation to the effectiveness of such an erasure as the data has been already used for the training, thus AI 'learnt' from it and even developed/enhanced its performance due to it.

(c) Regarding the third category of training data, those protected by various mechanisms (intellectual property rights, trade secrets, etc.), for their use, permissions must be obtained from their right holders.

In EU, for instance, DGA provides rules that potential AI developers must comply with. These rules apply to *certain categories of data held by public sector bodies* [art. 1.1. (a)], which are protected on grounds of: *commercial confidentiality, including business, professional and company secrets; statistical confidentiality; the protection of intellectual property rights of third parties; or the protection of personal data* (Article 3.1. DGA).

If protected data are used without permission, then the AI provider is liable.

(d) Synthetic data. ‘Artificially annotated information generated by computer algorithms or simulations’ (Lucini, 2021) is what synthetic data is.

Synthetic data is necessary to deal with real data's unavailability, scarcity, or high cost; to find a replacement for real data's poor quality, which could lead to inaccurate or imprecise outputs due to misunderstanding; to prevent privacy concerns (for instance, in sensitive applications like medical imaging or medical record aggregation); and to comply with copyright laws (Lu et al., 2021). Therefore, GenAI creates synthetic data for other AI models or for its own training. Due to the fact that it is simpler, quicker, and less expensive to do so, the practice of training AI using synthetic data is expanding quickly (Alemohammad et al., 2023). Recent studies, however, have come to the conclusion that ‘repeating this process for generation after generation of models forms an autophagous (self-consuming) loop’; ‘without enough fresh real data in each generation of an autophagous loop, future generative models are doomed to have their quality (precision), or diversity (recall) progressively decrease’ (Alemohammad et al., 2023). By drawing comparisons to mad cow disease, they gave this ailment the name Model Autophagy Disorder (MAD).

But, at least, there is no apparent legal issue, except, maybe, the outputs which generate misinformation and disinformation, discrimination, to name just a few.

Disputes on training data protected by intellectual property rights rules

A potential legal issue is the intellectual property rights to the data used to train GenAI. In fact, class actions on this basis have already been filed in the United States and the United Kingdom (Goldman, 2023).

Last year and early this year two collective actions were initiated in the USA, both at the United States District Court for the Northern District of California (Goldman, 2023). In the first lawsuit (*Doe v. GitHub*, 2022), filed on November 10, 2022, a group of developers sued GitHub, OpenAI, and Microsoft (which owns GitHub and part of OpenAI) alleging that Copilot, an AI technology implemented by GitHub and OpenAI, uses copyrighted code among the data that Copilot is trained with, which the plaintiffs have posted on GitHub under various open-source licenses. On the GitHub platform, the answer to the FAQ question ‘What data has GitHub Copilot been trained on?’ is: ‘GitHub Copilot is powered by Codex, a generative pretrained AI model created by OpenAI. It has been trained on natural language text and source code from publicly available sources, including code in public repositories on GitHub’ (Github.com, n.d.).

In the second US lawsuit (*Andersen v. Stability AI et al.*) (Vincent, 2023) filed on January 13, 2023, by three visual artists (Sarah Andersen, Kelly McKernan, and Karla Ortiz), the plaintiffs are suing artistic GenAIs, specifically, Stable AI (the developer behind Stable Diffusion), MidJourney (a popular image generator) and Deviant Art (developer of the DreamUp app). They claim that among the data on

which GenAI was trained were their creations protected by intellectual property rights, which were used to create derivative works without their permission. They have asked for ‘injunctive relief, declaratory judgment, and compensatory damages for alleged direct and vicarious copyright infringement, violation of Plaintiffs’ statutory and common law rights of publicity and violation of California’s Unfair Competition Law’ (California Northern District Court, 2023).

The third lawsuit (*Getty Images v. Stable Diffusion*) was filed in the United Kingdom in the High Court of Justice in London on January 17, 2023 by Getty Images against Stability AI, alleging that the defendant, Stability AI, ‘unlawfully copied and processed millions of images protected by copyright and the associated metadata owned or represented by Getty Images absent a license to benefit Stability AI’s commercial interests and to the detriment of the content creators’ (gettyimages.com, 2023).

With these first three cases against AI companies, the confrontations on intellectual property field are just getting started. On July 7, 2023 the comedian and author Sarah Silverman along with another two authors have sued OpenAI and Meta, alleging that their AI models (Chat GPT and Llama) were trained on copyrighted books without permission; they claimed that the books were illegally obtained from so called ‘shadow library’ websites such as Bibliotik, Library Genesis, Z-Library, and others, where their works are ‘available in bulk via torrent systems’ (Davis, 2023). In September 19 in a federal court in New York, *Authors Guild* (representing about 14,000 US authors), George R.R. Martin (the author of the *Game of Thrones*) and sixteen others have filed a lawsuit against OpenAI over a similar accusation of ‘systematic theft on a mass scale’ (Italie, 2023). They contend that ChatGPT could generate derivative works without permission and compensation (Mok, 2023).

Until the courts reach a solution, we are making a few brief assessments regarding their merits. In the US, the reproduction of a work protected by intellectual property rights is considered legal, if it meets the requirements of the *fair use doctrine* (Franceschelli & Musolesi, 2022, e17-1-e17-18). In American law, the court assesses whether a fair use has occurred depending on: the nature and purpose of the use; the nature of the protected work; the amount and importance of the part of the protected work used in the derivative work; the effect of unauthorized use on a potential market or value of the protected work (Inalton, 2020, pp. 15-22).

In terms of the amount and importance of the part used from the protected work in the derivative work, GenAI, in principle, uses the protected works in their entirety, not parts of them, because they enter the data on which GenAI is trained, taking the ideas, principles, correlations between the data (Franceschelli & Musolesi, 2022, e17-5). However, copyright does not protect ideas, theories, concepts, as provided for in most legislations (Margoni & Kretschmer, 2022), including the Romanian Copyright Law (Article 9 letter a) of Romanian Law no. 8/1996 on copyright and related rights), but rather their original expression (Dominte, 2021, p. 240). Data mining techniques do not use the protected works themselves, but only

the information they contain, and the information is not copyright protected; this may lead to the idea that copyright cannot be infringed (Franceschelli & Musolesi, 2022, e17-5). In the same vein, Professor Matthew Sag sees the GenAI as a student learning from its training data, rather than copying it ‘like a scribe in a monastery’ (Testimony before the U.S. Senate Committee on the Judiciary Subcommittee on Intellectual Property, 2023, p. 4). He argues that in the past courts have ruled that ‘non-expressive uses’ of technologies such as reverse engineering, search engines, and plagiarism detection software are fair use, distinguishing between protectable original expression and unprotectable facts, ideas and so on. In the case of a GenAI (or LLM - *Large Language Model*- as he refers to it), the output would determine whether the training data is a non-expressive usage.

‘If an LLM is trained properly and operated with appropriate safeguards, its outputs will not resemble its inputs in a way that would trigger copyright liability. Training such an LLM on copyrighted works would thus be justified under the fair use doctrine’ (Testimony before the U.S. Senate Committee on the Judiciary Subcommittee on Intellectual Property, 2023, p. 3).

However, if we look at things from another angle, *the output* of the processing of ideas from protected works may contain ideas *expressed* by the copyright holders (Sobel, 2020, pp. 19-21), because GenAI could memorise parts from the training data and therefore be in violation of the copyright. For now, it seems that the GenAI dealing with text-to-image transformation is more prone to memorisation (Sag, 2023, pp. 129-140).

3.2. Legal issues concerning the outputs

The output could raise legal concerns in different areas, which we cannot entirely cover here. To name just a few of these legal issues (combined with social ones), we could mention: discrimination, perpetuation of stereotypes and social biases, usage of toxic language, providing false or misleading information, disinformation, illicit advice, criminal content, privacy issues, plagiarism, intellectual property rights violation etc. We will further examine only some of the intellectual property rights aspects in relation to the content generated (briefly) and liability for the output.

Is the GenAI an author, enjoying the copyright?

Most recent discussions about GenAI revolve around the issue of qualifying the output as a work of art and establishing who has the copyright. We will not go into detail about this here; we will just sketch down a few ideas.

Authorship has been protected since a long time ago for encouraging creation and protecting the creators (Dee, 2018, p. 33). The title of author is reserved to humans, individuals (Dee, 2018, p. 36). This is also the case in Romania, where only a natural person can be an author. The US Copyright Office expressed the same opinion when it denied protection under copyright of an AI generated work (U.S. Copyright Office, 2023), arguing that the images generated were ‘not the product of human authorship’. Consequently, GenAI cannot be considered author when it solely created the output. However, there is a precedent in Canada where a work has been registered with two co-authors: a human being author and an AI (JDSUPRA, 2023), the work being created under direct guidance of a human (Dee, 2018, p. 35).

If deciding who has the copyright of the output poses difficulties, then it should at least be determined who has the ownership of the content. This is an issue sorted out by the contractual terms for now. For example, in the terms of use of ChatGPT, it is stipulated that the rights over the output are assigned to the user, which implies that initially the rights are with the provider (OpenAI, n.d., a). This means that the GenAI provider may be entitled to intellectual property rights of the output, such as databases, trade secrets, patent rights, and so on.

We cannot answer yet the question in the title of this section in a clear-cut manner, but we are confident that in a near future the practice will clarify the issue.

Liability for the output

The output generated by GenAI is called content and after providing it, the ‘job’ of GenAI is ended. From this point, it is up to the user whether the content will be utilized and for which purpose. Consequently, in respect to the output, there are two main types of liability both contractual or/and non-contractual: for the output itself (for the content) and for the exploitation of the output.

Civil contractual liability for the output (’s use)

GenAI is being utilized based on contracts (‘Terms and Conditions’, ‘Terms of Service’, etc.) and include clauses from traditional software licensing agreements (O’Leary and Armfield, 2020, pp. 249-272). Under the contract, the user is entitled to receive content generated by GenAI. However, this content might not be perfect (inaccurate, incorrect, incomplete etc.) and the service does not have a 100% availability. In these cases, the questions are whether the provider can be held liable for the output itself and for the unavailability of the service, following the contractual relationship.

As there is a contract in place, for answering the above questions there must be conducted an analysis on whether these issues are regulated by the contract and if there is any breach of contract. If the outcome is positive, then breaching the

contract implies contractual liability (Garon, 2023, p. 39). Nevertheless, it does not exclude the applicability of non-contractual liability.

Most rules for providing the content, using it and the related liability are defined in the contract between the provider of GenAI services and the user (Helberger & Diakopoulos, 2023, p. 5). Even though there is still considered to be a dispute on the liability when it comes to using the content and harming another individual (Dentons, 2023), we deem that this kind of debate can be sorted out firstly based on the contractual terms.

What is particular about GenAI is that the liability following its use is not as debatable as the one for employing decisional artificial intelligence. The main reason is that while the decisional AI leads to decisions, which might affect rights and freedoms of the human beings, the generative AI simply generates. It is up to the user of the service whether the generated content (output) is exploited in any way and how.

Four issues must be addressed in the contractual liability for output: liability for its use (a), liability disclaimers (b), limitation of liability (c) and circumventing any potential sole liability of the initial developer of a GenAI for its output (d).

(a) Liability for use of output.

Users are granted a license for the use of output ('Company grants you a limited, non-exclusive, non-transferable *license* (...)') (Rephrase.ai's Terms of Service, n.d.), which is subject to limitations enshrined in the terms and conditions of the platforms that provide GenAI services. Consequently, the users are allowed to exploit the output only for certain objectives, while others are disallowed⁷. Open AI platform, for instance, has even a separate document just for regulating the usage policies, through which it prohibits the use of their models, tools, and services for (i) 'illegal activity'; (ii) 'generation of hateful, harassing, or violent content'; (iii) 'content that expresses, incites, or promotes hate based on identity'; (OpenAI, n.d., b).

When GenAI is contractually confined to a particular purpose, then the user might be liable when the content/output shifts to any other goal (Gretel.). So, if the user disregards the grounds contractually prohibited, he/she could be held liable for breaching the contractual clauses. In addition, if by using the output a damage occurs, then the user will also be responsible and perhaps he/she will be denied the access to the GenAI services – 'We may immediately *discontinue your access* to the Platform in the event of breach of the Acceptable Use Policy' (Synthesia, n.d.).

Also, the user is prohibited to sublicense the use of the output - 'You have no right to sublicense the license rights granted herein' (Rephrase.ai's Terms of service, n.d.). Therefore, only the person owning the account and concluding the contract with the platform is allowed to receive and use the output. Consequently, the user

⁷ Examples of prohibited uses can be found in Synthesia.io's terms of service: 'In any way that violates any applicable national or international law or regulation; For the purpose of exploiting, harming, or attempting to exploit or harm minors in any way by exposing them to inappropriate content or otherwise; In any way that infringes upon the rights of others (...)' (<https://www.synthesia.io/terms/terms-of-service>).

might be liable when the account is shared or when the content is transmitted and exploited by third parties, regardless of the purpose.

If the user is held liable for the unlawful content, the provider of the GenAI cannot or if it will, then it will be able to redress against the user. Similarly, if the user breaches the contract by using the output for a prohibited purpose, then even if the provider would be initially held accountable for the content, it could redress against the user based on the contractual provisions. A relevant example is Synthesia.ai:

'To the extent permitted by applicable law, you will defend, indemnify and hold harmless Synthesia (...) from and against any and all losses, damages, costs, expenses (...) and liabilities arising from, connected with or relating to your use of the service or breach of these terms' (Synthesia, n.d.).

Currently, the most effective mechanism for limiting or preventing liability for use of output is having it humanly filtered, which is a technique known as 'human in the loop' (Mosqueira-Rey, 2023, pp. 3005-3054). These techniques generally imply having in place both a pre-moderation (Hacker et al., 2023, p. 3) and a moderation procedure during which humans detect and remove any inappropriate or illegal content (input or output). Even if human mistakes are not to be excluded, such a procedure will reduce the risks of using the output, as it is being 'moderated'. However, such mechanisms imply more resources invested from the provider's side. Meta is also recommending a similar approach involving human input, named 'Reinforcement Learning from Human Feedback (RLHF)' (Meta.ai, 2023, p. 11), which implies designing a system for assessing the quality and risks/safety of the output and then providing the GenAI with the results of the assessments, in numbers, so it can further learn to reduce the risky content. A more convenient solution would be simply excluding any obligation for a certain quality and requirements of the content, together with disclaimers and direct exclusion of warranties in relation to the output.

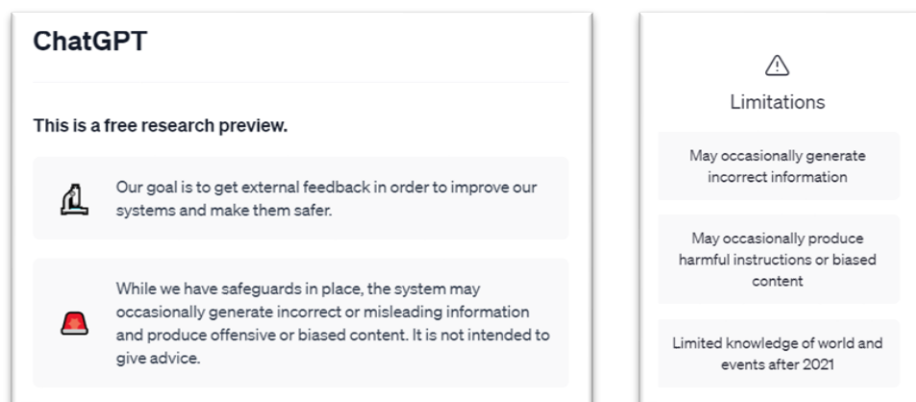
(b) Liability for the output. Disclaimers

As already mentioned, the output might not be accurate, and sometimes not even appropriate. If GenAI is fed with inaccurate, biased, or discriminatory data, then from this input AI will 'learn' and generate a similar output. The GenAI providers inform the users in advance about this possibility.

GenAI providers can opt to take no responsibility for the quality and standards of the output. Together with the acceptance of the 'as is' or 'as available' clauses, the user is agreeing to the content in the form delivered by the GenAI, accepting thus that it might be inaccurate, unreliable, erroneous, interrupted. The 'as is' or 'as available' clauses are a contractual shield against complaints. Thus, a 'take it or leave it' approach is applicable. Multiple GenAI providers include such clauses for their provision of services, relatable examples being Synthesia.io and Rephrase.ai. In

some instances, the user is fully liable for its content, as stipulated in the terms and conditions of ChatGPT: ‘you are responsible for Content, including for ensuring that it does not violate any applicable law or these Terms’ (OpenAI, n.d., a).

Figure 2. Disclaimers displayed by ChatGPT



Source: <https://chat.openai.com/> (retrieved on 06.05.2023)

In most cases, providers state the probability of an illicit or incorrect input – a proper example can be found in the Terms of Service of Rephrase.ai: ‘Company makes no warranties, expressed or implied, and hereby disclaims and negates all other warranties, including without limitation, implied warranties or conditions of merchantability, (...) *non-infringement of intellectual property or other violation of rights*’. This is done with the help of disclaimers, used for the protection of the provider against liability; these disclaimers are intricately linked to the ‘as is’ or ‘as available’ clauses (for example, terms of service of Rephrase.ai: ‘The service rendered on company's platform are provided *‘as is’* and *‘as available’*’), and imply a lack of warranties (‘ (...) *do not warrant* that: (a) the service will be secure or available (...); (b) any defects or errors will be corrected; (...) Company *does not warrant* that the user will be able to always use the platform (...)’ (Rephrase, n.d.).

Moreover, the user is made aware that the content might be harmful, and upon his/her consent to use it, the provider of GenAI cannot be held liable. If an individual requires correct and non-harmful content, he/she must not access the services of that platform and search for a GenAI that ensures a set of minimum standards of the output. Consequently, any use of GenAI it is, theoretically, at his/her risk – ‘Our Platform allows you (...) to create or generate graphics, videos, or other material (referred to as ‘User Generated Content’). *You are responsible* for Your Content and User Generated Content, including its legality, reliability, and appropriateness’ and ‘We take no responsibility and assume no liability for content you post or create on or through the Platform’ (Rephrase.ai, n.d.).

(c) *Limitation of liability*

Contracts for GenAI concluded with regular users include boilerplate clauses for limiting or excluding the liability of the GenAI provider. There are multiple examples of such clauses. For instance, on Gretel.ai: ‘To the maximum extent permitted by law, Gretel.ai will not be liable for any incidental, special, exemplary or consequential damages (...), whether based on warranty, contract, tort (including negligence), product liability or any other legal theory (...)’ (Gretel.ai, n.d.).

The enforceability of these limitations is debatable, being considered, for example, that even if GenAI providers such as Microsoft, Google and OpenAI have included warnings on the inaccuracy of the output (Fried, 2023), this does not lead to a full exoneration. When it is legally impossible to exclude liability, this is limited to a certain modest level:

‘To the maximum extent permitted by law, Gretel.ai’s total liability (...) will not exceed the amounts you have paid to gretel.ai (...). If you have not had any payment obligations to gretel.ai, (...) total liability exceeds one hundred dollars (\$100)’ (Gretel.ai, n.d.).

However, provided that most users are consumers, the limitations of liability might not be fully enforceable in every jurisdiction. Also, even if the parties are allowed to agree on the *lex contractus*, the choice of laws has limitations when it comes to consumers. For example, in the EU, *lex contractus* cannot be chosen in detriment of consumers, when it would affect their rights and/or deprive them of the protection provided by their habitual residence law. These legal requirements and limitations are also acknowledged by GenAI providers and mentioned in the contract:

‘The foregoing does not affect any warranties which cannot be excluded or limited under applicable law. In particular, if you are a consumer and have your habitual residence (...) the European Economic Area, applicable consumer laws may not allow some of the exclusions and limitations set out above (...)’ (Synthesia, n.d.).

There are also jurisdictions in which liability cannot be limited when it arises from gross negligence or intent (e.g., Romanian Civil Code - Article 1355 (1) (Baiaș, 2022)). In addition, in Romania, under Article 1355 (3) of the Civil Code, liability for bodily or mental injury cannot be contractually limited, meaning that if the AI provides harmful advice to a human, which leads to physical injuries, then the provider of the AI might be held liable. This type of liability would be for being ‘reckless or negligent in the creation of the system’ (Fried, 2023).

(d) *Circumventing any potential sole liability of the initial developer of a GenAI for its output*

After multiple concerns being raised in relation to the liability for the output provided by GenAI, in particular a few related to the impossibility of entirely avoiding the liability (Ariyaratne, 2023), Meta came up with a new GenAI model, which *theoretically* takes GenAI off Meta's hands and puts it 'in the hands of people globally' (Llama 2 Responsible Use Guide, 2023, p. 4).

For not having the liability centred around one business developer of the GenAI, Meta advanced on the market an open source GenAI, available to be freely downloaded, used and further developed by any interested person that requests the software. This open source GenAI was initially called Llama, its newest version being Llama 2. This GenAI is no longer exclusively controlled and developed by the business that released it, but by persons all over the world, including other businesses and/or associations.

Through the Llama project, Meta is managing to 'diffuse' the liability for the output of the GenAI it initially developed. Meta is not opening only the GenAI technology, but also the responsibility of its developments, hence also the liability for the output. The users of Llama 2 might be developers of GenAI at the same time. The responsibility is smoothly shifted towards the user of Llama 2, who is being given access to a 'Responsible Use Guide' (Llama 2 Responsible Use Guide, 2023) that operates with concepts as 'responsible generative AI' and 'AI safety' and under which the developers should focus on 'building responsibility' (Llama 2 Responsible Use Guide, 2023, p. 3). Also, the guide makes it clear that the decision on the usage (and related consequences) is solely with the user/developer of Llama 2, who should decide to implement best practices depending on the jurisdiction where her/his products 'will be deployed and should follow his/her company's internal legal and risk management processes' (Llama 2 Responsible Use Guide, 2023, p. 3). Furthermore, the Responsible Use Guide holds the user/developer accountable for 'assessing risks' associated with the GenAI use they found and for 'applying best practices to ensure safety' (Llama 2 Responsible Use Guide, 2023, p. 7).

As an open source, the GenAI can be improved, or its issues can be tackled. There is a system in place for reporting any risks or malfunctions the users of Llama 2 experience. Consequently, the challenge to build a responsible GenAI addresses to all the users, who cannot only complain about the risks posed by GenAI but can now take measures. Users are no longer given a passive role; it is not a take-it-or-leave-it GenAI.

Non-contractual liability for the output

As to the non-contractual liability for GenAI's output, a few jurisdictions are in the 'work in progress' phase. For now, the EU advanced hard law proposals, while the US has soft law provisions at the federal level and a brand new executive order (The White House, 2023). China came out with the first in force hard law in August 2023.

We are going to point out only a few aspects on (potential) solutions provided by the three jurisdictions.

(a) EU proposed regulations on AI

The proposed Artificial Intelligence Act (European Commission, 2021) and AI Liability Directive (European Commission, 2022), when adopted, will also govern non-contractual liability.

The AI Liability Directive applies only to non-contractual fault-based law claims for damages, focusing on the high-risk AI, as it is classified by the Artificial Intelligence Act, but also taking into account all the other AI systems. When it states in Article 4 that national courts must presume ‘the causal link between the fault of the defendant and the output produced by the AI system or the failure of the AI system to produce an output’, the Directive refers directly to the output as the element for which the provider is liable. As it does not distinguish between the types of the AI systems, it also covers the output of GenAI.

As regards the Artificial Intelligence Act, its rules take a risk-based approach, establishing obligations for providers and users based on the level of risk generated by AI. Noncompliance with its provisions may subject the provider or the user to liability. A breach of compliance would be, for example, the release on the market of a GenAI that includes in its output:

‘(...) purposefully manipulative or deceptive techniques, with the objective to or the effect of materially distorting a person’s or a group of persons’ behaviour by appreciably impairing the person’s ability to make an informed decision, thereby causing the person to take a decision that that person would not have otherwise taken in a manner that causes or is likely to cause that person, another person or group of persons significant harm’ (Article 5 (a) of Artificial Intelligence Act as amended by the European Parliament on 14 June 2023 (European Parliament, 2023)).

We are not going to go deeper into the proposed legislation, as it is subject to changes, but for now we can outline that there are proposed four categories of AI, as follows: unacceptable AI (extremely risky), high-risk AI, limited-risk AI (under which ChatGPT would fall), and low-risk AI. The first category is prohibited from being traded, for the next two categories the obligations imposed are proportionate to the category of risk, and for the last category there are currently no specific obligations proposed. Qualifying AI as high-risk under the Artificial Intelligence Act is dependent on the purpose of its usage. However, it is up to the user sometimes to decide on the use of the content, thus the user determines the category of the risk (Helberger & Diakopoulos, 2023, p. 2).

(b) US regulatory approach on AI

In October 2022 the White House Office of Science and Technology Policy developed a document called the *Blueprint for an AI Bill of Rights*, which contains a

set of five guiding principles for the creation and use of AI systems. These principles have a status of soft law and are aimed at the protection of US citizens. So, the Americans:

'should be protected from unsafe or ineffective systems, should not face discrimination by algorithms and systems should be used and designed in an equitable way, should be protected from abusive data practices via built-in protections and should have agency over how their data is used, should know that an automated system is being used and understand how and why it contributes to outcomes that impact them, should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems they encounter' (The White House, 2022).

The *Blueprint for an AI Bill of Rights* is addressed to companies and government authorities, who are advised to implement the five principles in their policy regarding the development and use of AI. Although it can be considered a legislative act that takes an important step towards protecting the fundamental rights of US citizens (Park, 2023, p. 27), having no binding power could determine the private sector to simply ignore it (Hine & Floridi, 2023, p. 285).

One year later, on October 30, 2023, the Biden administration adopted the *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (The White House, 2023). This order, even if it has binding force, does not have the 'power' of legislation passed by the US congress (Glover, 2023). It imposes certain rules on federal agencies regarding the monitoring of AI risks and it is based on eight principles, listed in section 2 of the order: *Safety and Security, Promoting Responsible Innovation, Competition and Collaboration, Supporting American Workers, Protecting Equity and Civil Rights, Maintaining Consumer Protections, Protecting Privacy and Civil Liberties, Use Responsible AI to Help the Government Work Better, Create an International Framework for Responsible AI* (Glover, 2023).

The time will reveal more about the current US endeavor.

(c) China's regulatory perspectives on GenAI's output

China, the largest world producer of AI research (Sheehan, 2023), has published earlier this year for public debate a proposal for a regulation on AI - 'Measures for the Management of Generative Artificial Intelligence Services' (hereinafter, the 'AI Measures Act'). Then, it took some steps further and decided to be the pioneer (Wu, 2023) in regulating GenAI by adopting a set of interim measures - the 'Interim Measures for the Administration of Generative Artificial Intelligence Services' (hereinafter, the 'AI Interim Measures Act'), which entered into force on the 15th of August 2023. At first sight, it is clear that, although initially proposing rigid and final requirements, China ended up passing an interim act revealing thus

the current impracticability of a final and immutable regulation for a constantly evolving technology.

The initial AI Measures Act was significantly limiting the sets of data that can be used for trainings, as any output had to be compliant with the act. Considering that one of the most important arguments in deciding the ‘best’ GenAI is the size of the data sets used for training, such requirements could have resulted in a less global competitive GenAI. Nevertheless, this issue was solved in the AI Interim Measures Act by limiting the scope of the measures and narrowing it to the GenAI services provided only to the public.

According to Article 4 of the AI Interim Measures Act, the provision of GenAI products or services shall comply with a set of requirements. The providers of GenAI have been granted the role of content moderators, having the obligation to moderate (Wu, 2023) the output when it is not compliant. We deem that Article 4 of the AI Measures Act is partially shifting the liability from the user (as we noticed it was the case in the EU and US according to the contractual terms and conditions) to the provider of the GenAI, who is responsible to ensure from the very beginning that the GenAI will not provide any content breaching the law. Also, the provider of GenAI is obliged to: protect personal information (Article 9); mark the output generated by AI accordingly (Article 12); take down or stop generating immediately any illegal content (Article 14).

In case the output is in breach of the AI Interim Measures Act because of programming, wrongful pre-training or incorrect collected data, then there will be sanctions applicable to the provider (Article 21). Consequently, the recent Chinese regulation is focusing the entire liability for the output on the provider of the GenAI.

As a conclusion to the three jurisdictions’ possible regulatory solutions briefly presented above, the US stated in the Biden administration’s order of October 30, 2023 that the US should take the lead and ‘engage with international allies and partners in developing a framework to manage AI’s risks, unlock AI’s potential for good, and promote common approaches to shared challenges’ (Section 2, h). Meanwhile, the EU, in the Article 2 of the proposed Artificial Intelligence Act establishes the extraterritorial effects of European rules, which will also affect American companies that provide GenAI services. Chinese legislation appears to be limited to the national level.

Conclusions

AI has a disruptive effect not only on technology (Mann, 2023), but also on the legal field, regardless of being decisional or generative.

At the heart of GenAI technology is data. To detect the legal issues that the use of GenAI may cause, we looked at the model that this technology operates on: *input-process-output*. We classified the input (training data) using the criterion of its licit source into: open data, data provided in a voluntary way by users, consumers,

professionals in their online interactions, data protected by various mechanisms (intellectual property rights, trade secrets, other means of protection), *synthetic data*.

We found that the training data is already the ‘protagonist’ in a few lawsuits, mainly in US. We did not find any information on similar litigation in EU. The US courts are yet to reach a solution, but it seems that they apply the *fair use doctrine*. We shall see the decisions that are still to be rendered.

Starting with the data that serves as input for GenAI and going beyond to the use of its output, this new technology raises multiple legal issues that are laborious to be covered by legal frameworks. Technology has a faster pace of development, its usages are daily evolving and multiplying, while law is left behind struggling with its strings of formalisms and procedures. At the moment, the most suitable legal instrument for regulating the use of GenAI is the contract, as it can be literally overnight amended by the parties and adapted to any development, technological or legal. However, there still exists the concept of the Black Box AI (Rouse, 2023), which would initially result in a ‘black hole’ for the legal research on GenAI, until finding the correct formula for deciding on liability on the output it generates.

The first and most stringent legal issues that have been sketched until now are the ones related to the data that serves as input and the content generated as output, together with its usage. For some of them the current legal framework does provide solutions, but for others it can be argued that for now there is a legislative vacuum, and the solution is provided by the contract in place. However, both USA and EU are working, even together (Smalley, 2023), for regulating and making lawful the use of GenAI. As GenAI is already a global technology, with borderless exploitation, the legal response should reflect these characteristics, involving the cooperation between multiple jurisdictions. However, there might also be countries like China, trying to regulate AI on their own.

As to the issue of qualifying the GenAI output as a work of art and establishing who has the copyright, the answer seems quite clear for the moment: only a human being could be considered an author. Regarding the GenAI providers, they may be entitled to the intellectual property rights of the output, such as databases, trade secrets, patent rights, and so on.

The liability for the output was addressed discussing a few contractual and non-contractual liability aspects. We made a few remarks on four issues of the contractual liability for output: liability for its use, liability disclaimers, limitation of liability and circumventing any potential sole liability of the initial developer of a GenAI for its output.

In relation to the last aspect, noticing the legal issues raised by GenAI, as well as the risks posed concerning privacy, copyright, and human rights, GenAI developers and/or providers started to focus on shifting their liability and came with the solution of the open source GenAI models designed for further and extensive research, a particular example being Llama 2, which made possible the development of the GenAI Free Willy 2 (Bastian, 2023). In this case, the final GenAI is not

entirely developed by the initial developer, and the liability for any legal issues it might create will be bore by the person/entity that ‘added on top’ of the open-source software basis and created the final GenAI that was used.

Concerning the non-contractual liability, a few jurisdictions are in the ‘work in progress’ phase. For now, the EU has hard law proposals, US has soft law provisions at the federal level and an *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* from October 30, 2023. China came out with the first in force hard law in August 2023.

Finally, yet important, this paper is meant to underline the main core of GenAI: data. Once again, a new technology is developed and gravitates around data, opening a new door for regulating usage and control of data, while providing the opportunity to enforce the already existing regulations in a new field, for new sets of data.

Acknowledgement: This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS – UEFISCDI, project number PN-III-P4-PCE-2021-1878, within PNCDI III ‘Institutions, digitalization and regional development in the European Union’.

References

- Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A. I., Babaei, H., LeJeune, D., Siahkoohi, A. G., & Baraniuk, R. (2023). *Self-Consuming Generative Models Go MAD*. arXiv:2307.01850. <https://doi.org/10.48550/arXiv.2307.01850>
- Ariyaratne, H. (2023). *ChatGPT and Intermediary Liability: Why Section 230 Does Not and Should Not Protect Generative Algorithms*. <http://dx.doi.org/10.2139/ssrn.4422583>
- Baias F.-A. (Ed.) (2022). *Romanian Civil Code*. Editura Monitorul Oficial.
- Banerjee, S. S., & Mohapatra, S. (2021). Developing a framework of artificial intelligence for fashion forecasting and validating with a case study. *International Journal of Enterprise Network Management*, 12(2), 165 – 180. <https://doi.org/10.1504/IJENM.2021.116438>
- Bastian, M. (2023, July 22). Stability AI's new open source LLM “FreeWilly” outperforms Meta's Llama 2. *The Decoder*. <https://the-decoder.com/stability-ais-new-open-source-llm-freewilly-outperforms-metas-llama-2/>
- Bellapu, A. (2023, January 8). Top 10 Humanoid Robots Set to Grow 10 Times by 2023. *Analytics Insight*. <https://www.analyticsinsight.net/top-10-humanoid-robots-set-to-grow-10-times-by-2023/>
- Bertolini, A. (2020). Artificial Intelligence and Civil Liability. *Policy Department for Citizens' Rights and Constitutional Affairs*, 1-132. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU\(2020\)621926_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf)

- Business & Human Rights Centre (2023, January 23). *OpenAI and Sama hired underpaid Workers in Kenya to filter toxic content for ChatGPT*. <https://www.business-humanrights.org/en/latest-news/openai-and-sama-hired-underpaid-workers-in-kenia-to-filter-toxic-content-for-chatgpt/>
- California Northern District Court (2023). *Andersen et al. v. Stability AI Ltd. et al.* 3:23-CV-00201. <https://unicourt.com/case/pc-db5-andersen-et-al-v-stability-ai-ltd-et-al-1380299>
- Carsaniga, G., Eline, N. Lincklaen, A., Jochem, D., Van Assen, M., & Cecconi, G. (2022). *Open Data Maturity Report 2022*. https://data.europa.eu/sites/default/files/data.europa.eu_landscaping_insight_report_n8_2022_2.pdf
- Cellan-Jones, R. (2014, December 2). Stephen Hawking warns artificial intelligence could end mankind. *BBC*. <https://www.bbc.com/news/technology-30290540>
- Clifford, C. (2018, May 13). Elon Musk: 'Mark my words — A.I. is far more dangerous than nukes'. *CNBC*. <https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html>
- Creative Commons CC BY 4.0. (n.d.). Attribution 4.0 International (CC BY 4.0). <https://creativecommons.org/licenses/by/4.0/>
- Creative Commons, CC0. (n.d.). 'No Rights Reserved'. <https://creativecommons.org/share-your-work/public-domain/cc0/>
- Creative Commons. (n.d.). Frequently Asked Questions. <https://creativecommons.org/faq/#artificial-intelligence-and-cc-licenses>
- Custers, B. & Fosch-Villaronga E. (2022). *Humanizing Machines: Introduction and Overview*. In: B. Custers & E. Fosch-Villaronga (Eds.), *Law and Artificial Intelligence Regulating AI and Applying AI in Legal Practice* (pp. 3-28). T.M.C. ASSER PRESS. <http://dx.doi.org/10.2139/ssrn.4331754>
- Davis, W. (2023, July 9). Sarah Silverman is suing OpenAI and Meta for copyright infringement. *The Verge*. <https://www.theverge.com/2023/7/9/23788741/sarah-silverman-openai-meta-chatgpt-llama-copyright-infringement-chatbots-artificial-intelligence-ai>
- Day, C. (2016, May 5). 'The 100' Takes Brutality to New Levels with ALIE, But Where Does It Go from Here?. *Collider*. <https://collider.com/100-season-3-alie/>
- Dee, C. M. A. (2018). Examining Copyright Protection of AI-Generated Art, *Delphi - Interdisciplinary Review of Emerging Technologies*, 1(1), 31-37. <https://doi.org/10.21552/delphi/2018/1/11>
- Demopoulos, A. (2023, April 3). Computer-generated inclusivity: fashion turns to 'diverse' AI models. *The Guardian*. <https://www.theguardian.com/fashion/2023/apr/03/ai-virtual-models-fashion-brands>
- Dentons (2023, January 20). *Part 3: Generative AI – Navigating commercial and civil liability*. https://www.dentons.com/en/insights/articles/2023/january/20/part-three-generative-ai-navigating-commercial-and-civil-liability?utm_source=email&utm_medium=email&utm_campaign=vuture

- Doe v. Github (2022). Case 3:22-cv-06823-KAW. United States District Court, Northern District of California, San Francisco Division. <https://www.documentcloud.org/documents/23323955-doe-v-github-class-action-complaint>
- Dominte, N. R. (2021). *Dreptul proprietății intelectuale. Protecție juridică* [Intellectual property law. Legal protection]. Solomon Publishing House
- EESC. (2018, December 6). *Opinion of the European Economic and Social Committee on 'Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions — Artificial Intelligence for Europe'*. OJ C 440/51. 51-56 <https://eur-lex.europa.eu/legal-content/RO/TXT/?uri=CELEX:>
- Eke, D. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity?. *Journal of Responsible Technology*, 13, 1-4. <https://doi.org/10.1016/j.jrt.2023.100060>
- El Atillah, I. (2023, March 31). Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change. *Euronews*. <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate>
- Ericsson. (2022). *Ericsson, AI and privacy: Everything you need to know about trust and technology*. <https://www.ericsson.com/en/blog/2022/8/ai-and-privacy-everything-you-need-to-know>
- European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts*. COM/2021/206 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>
- European Commission. (2022). *Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*. COM/2022/496 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496>
- European Commission. (2023). *Commission defines high-value datasets to be made available for re-use*. <https://digital-strategy.ec.europa.eu/en/news/commission-defines-high-value-datasets-be-made-available-re-use>
- European Data Protection Board. (2022). *Facial recognition: Italian SA fines Clearview AI EUR 20 million*. https://edpb.europa.eu/news/national-news/2022/facial-recognition-italian-sa-fines-clearview-ai-eur-20-million_en
- European Parliament. (2023, June 14). *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*. COM (2021) 0206 – C9-0146/2021 – 2021/0106 (COD). https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html
- Fallmann, D. (2022, May 3). Using Digital Twins and Preparing for The Metaverse, *Forbes*. <https://www.forbes.com/sites/forbestechcouncil/2022/05/03/using-digital-twins-and-preparing-for-the-metaverse/>

- Franceschelli, G. & Musolesi, M. (2022). Copyright in generative deep learning. *Data & Policy*, 4, E17. <https://doi.org/10.1017/dap.2022.10>
- Franzen, C. (2023, July 24). Stability AI unveils new FreeWilly language models trained using minimal — and highly synthetic. *Venture Beat*. <https://venturebeat.com/ai/stability-ai-unveils-new-freewilly-language-models-trained-using-minimal-and-highly-synthetic-data>
- Fried, I. (2023, February 24). Generative AI is a legal minefield. *Axios*. <https://www.axios.com/2023/02/24/chatgpt-generative-ai-legal-minefield>
- Garon, J. (2023). *A Practical Introduction to Generative AI, Synthetic Media, and the Messages Found in the Latest Medium*. <https://dx.doi.org/10.2139/ssrn.4388437>
- Gatto, J. (2023). How Generative AI Generates Legal Issues in the Games Industry. *Sheppard Mullin*. <https://www.lawoftheledger.com/wp-content/uploads/sites/15/2023/03/AI-and-Games-Article-0223.pdf>
- Generative AI. (n.d.). *The Power of Generative AI*. <https://generativeai.net/>
- gettyimages.com. (2023). Getty Images Statement. News.room. <https://newsroom.gettyimages.com/en/getty-images/getty-images-statement>
- GitHub Docs. (n.d.). *GitHub Terms of Service*. <https://docs.github.com/en/site-policy/github-terms/github-terms-of-service>
- Github.com. (nd). *Copilot*. <https://github.com/features/copilot>
- Glover, E. (2023, October 31). What You Need to Know About the Biden-Harris Executive Order on AI. *BuiltIn*. <https://builtin.com/artificial-intelligence/ai-executive-order>
- Goldman, J. (2023, January 18). Do AI generators infringe? Three new lawsuits consider this mega question. *Lexology*. <https://www.lexology.com/library/detail.aspx?g=c85167ff-90fe-41eb-863e-b2601ac9d058>
- Gretel.ai (n.d.). *Gretel.ai's subscription services agreement*. <https://gretel.ai/terms>
- Gretel.ai (n.d.). *Terms of use*. <https://gretel.ai/acceptable-use>
- Hacker, P., Engel, A., & Mauer, M. (2023). Regulating ChatGPT and other Large Generative AI Models. *arXiv:2302.02337*. <https://doi.org/10.48550/arXiv.2302.02337>
- Häuselmann, A. (2022). *Disciplines of AI: An Overview of Approaches and Techniques*. In: B. Custers & E. Fosch-Villaronga (Eds.), *Law and Artificial Intelligence Regulating AI and Applying AI in Legal Practice* (pp. 43-70). T.M.C. ASSER PRESS
- Helberger, N., & Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review*. 12(1). 2-6. <https://doi.org/10.14763/2023.1.1682>
- Hine, E., Floridi, L. (2023). The Blueprint for an AI Bill of Rights: In Search of Enaction, at Risk of Inaction. *Minds and Machines*, 33, 285-292. <https://doi.org/10.1007/s11023-023-09625-1>
- Hobbs, J. (2023, March 15). 'Amazon snake cat' photo goes viral and mystifies internet. *New York Post*. <https://nypost.com/2023/03/15/amazon-snake-cat-photo-goes-viral-and-mystifies-internet/>

- Inaltong, N. U. (2020). The Evolution of The Notion of Fair Use in Copyright Law: Where Are We Today. *SSRN*, 15-22. <http://dx.doi.org/10.2139/ssrn.3928852>
- Instagram. (n.d.). *Terms of Use*. https://help.instagram.com/581066165581870/?cms_id=581066165581870
- Italie, H. (2023, September 21). ‘Game of Thrones’ creator and other authors sue ChatGPT-maker OpenAI for copyright infringement. *AP New*. <https://apnews.com/article/openai-lawsuit-authors-grisham-george-rr-martin-37f9073ab67ab25b7e6b2975b2a63bfe>
- JDSUPRA. (2023, August 17). *The Generative AI Revolution: Key Legal Considerations for the Fashion & Retail Industry*. <https://www.jdsupra.com/legalnews/the-generative-ai-revolution-key-legal-3041938/>
- Jonsson, M., & Tholander, J. (2022). Cracking the code: Co-coding with AI in creative programming education. *C&C '22: Creativity and Cognition*, 5-14. <https://doi.org/10.1145/3527927.3532801>
- Klein, M. (2020, November 17). The Problematic Fakery of Lil Miquela Explained—An Exploration of Virtual Influencers and Realness. *Forbes*. <https://www.forbes.com/sites/mattklein/2020/11/17/the-problematic-fakery-of-lil-miquela-explained-an-exploration-of-virtual-influencers-and-realness/>
- Li, Y., Wang, X., Miao, Z., & Tan, W. (2021). Data Augmentation for ML-driven Data Preparation and Integration. *Proceedings of the VLDB Endowment*, 14(12), 3182-3185. <https://doi.org/10.14778/3476311.3476403>
- Lu, Y., Shen, M., Wang, H., & Wei, W. (2021). Machine Learning for Synthetic Data Generation: A Review. *Journal of Latex Class Files*, 15(8), 1-17. <https://doi.org/10.48550/arXiv.2302.04062>
- Lucini, F. (2021). The real deal about synthetic data. *MIT Sloan Management Review*, 63(1), 1–4. <https://sloanreview.mit.edu/article/the-real-deal-about-synthetic-data/>
- Mann, J. (2023, March 13). AI considered among the top three disruptive technologies. *Ship Technology*. <https://www.ship-technology.com/features/ai-considered-among-the-top-three-disruptive-technologies/>
- Margoni, T., & Kretschmer, M. (2022). A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology. *GRUR International*, 71(8), 685–700. <https://doi.org/10.1093/grurint/ikac054>
- McKinsey & Company. (2023, January 19). *What is generative AI?*. <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai>
- Measures for the Management of Generative Artificial Intelligence Services. (2023). *Popular Republic of China*. http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm
- Merken, S. (2023, June 9). Another US judge says lawyers must disclose AI use. *Reuters*. <https://www.reuters.com/legal/transactional/another-us-judge-says-lawyers-must-disclose-ai-use-2023-06-08/>

- Meta. (n.d.) *Accelerating MRI Scans with AI*. <https://ai.meta.com/research/impact/fastmri/>
- Meta.ai. (2023), *Llama 2 Responsible Use Guide*.
<https://github.com/facebookresearch/llama/blob/main/Responsible-Use-Guide.pdf>
- Meta.ai. (n.d.a). *Generating Biographies. Meta AI develops a novel dataset and model to help bring more representation to Wikipedia*. <https://ai.meta.com/research/impact/generating-biographies/>
- Meta.ai. (n.d.b). *Introducing Llama 2*. <https://ai.meta.com/llama/>
- Mok, A. (2023, September 2021). 'Game of Thrones' author George R.R. Martin and other writers sue ChatGPT's maker claiming copyright infringement: 'Generative AI threatens to decimate the author profession'. *Insider*.
<https://www.businessinsider.com/game-of-thrones-author-sues-chatgpt-openai-copyright-infringement-lawsuit-2023-9>
- Mosqueira-Rey, E., Hernández-Pereira, E., & Alonso-Ríos, D. (2023), Human-in-the-loop machine learning: a state of the art. *Artif. Intell. Rev.*, 56, 3005–3054.
<https://doi.org/10.1007/s10462-022-10246-w>
- Muller, M., Chilton, L. B., Kantosalo, A., Martin, C. P., & Walsh, G. (2022). GenAICHI: Generative AI and HCI. In: S. Bardosa, C. Lampe, C. Appert & D. A. Shamma (Eds.). *CHI Conference on Human Factors in Computing* (pp. 1-7). Association for Computing Machinery New York. United States.
<https://doi.org/10.1145/3491101.3503719>
- O'Leary, T. P. & Armfield T. (2020). Adapting to the Digital Transformation. *Alberta Law Review Society*, 58(2), 249-272. <https://www.canlii.org/en/commentary/doc/2020CanLIIDocs3296>
- Open Source Initiative (n.d., a). *European Union Public License*. Version 1.2.
<https://opensource.org/license/eupl-1-2/>
- Open Source Initiative (n.d., b). *OSI Approved Licenses* (italic).
<https://opensource.org/licenses/>
- OpenAI (n.d., a). *Terms of use*. <https://openai.com/policies/terms-of-use>
- OpenAI (n.d., b). *Usage Policies*. <https://openai.com/policies/usage-policies>
- Park, E. (2023). The AI Bill of Rights: A Step in the Right Direction. *Orange County Lawyer Magazine*, 65(2). <https://ssrn.com/abstract=4351423>
- Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons*, 63, 403-414. <https://doi.org/10.1016/j.bushor.2020.01.003>
- Peter, J. (2023, April 5). Metaverse Fashion Week had big brands but few people. *The Verge*.
<https://www.theverge.com/23668846/decentraland-metaverse-fashion-week-2023>
- Rephrase.ai (n.d.). *Terms of Service*. <https://www.rephrase.ai/terms-of-service>
- Rouse, M. (2023, June 16). What Does Black Box AI Mean?. *Techopedia*.
<https://www.techopedia.com/definition/>

- Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. 4th edition. Pearson
- Sag, M. (2023, August 10). Copyright Safety for Generative AI. *Houston Law Review*, 61 (2), 104-148. <http://dx.doi.org/10.2139/ssrn.4438593>
- Sag, M. (2023, July 12). *Testimony before the U.S. Senate Committee on the Judiciary Subcommittee on Intellectual Property Hearing on 'Artificial Intelligence and Intellectual Property – Part II: Copyright and Artificial Intelligence'*. https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_sag.pdf
- Sheehan, M. (2023, July 10). China's AI Regulations and How They Get Made. *Carnegie Endowment for International Peace*. <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>
- Smalley, S. (2023, January 28). U.S. and EU to launch first-of-its-kind AI agreement. *Reuters*. <https://www.reuters.com/technology/white-house-european-commission-launch-first-of-its-kind-ai-agreement-2023-01-27/>
- Sobel, B. (2020). A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning. In: R. Hilty, J.-A. Lee & K.-J. Liu (Eds.). *Artificial Intelligence and Intellectual Property*. Oxford University Press (pp. 1-36). <http://dx.doi.org/10.2139/ssrn.3677548>
- Sobel, B. L. W. (2018). Artificial Intelligence's Fair Use Crisis. *The Columbia Journal of Law & the Arts*, 41(1), 45–97. <https://doi.org/10.7916/jla.v41i1.2036>
- Surden, H. (2019). Artificial Intelligence and Law: An Overview. *Georgia State University Law Review*, 35(4), 1305-1337. <https://readingroom.law.gsu.edu/gsulr/vol35/iss4/8>
- Synthesia. (n.d.). *Terms of Service*. <https://www.synthesia.io/terms/terms-of-service>
- Tate, R.-M. (2023, June 26). Junk websites filled with AI-generated text are pulling in money from programmatic ads. *MIT Technology Review*. <https://www.technologyreview.com/2023/06/26/1075504/junk-websites-filled-with-ai-generated-text-are-pulling-in-money-from-programmatic-ads/>
- The White House (2022). *Applying the Blueprint for an AI Bill of Rights*. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/#applying>
- The White House (2023, October 30). *FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*. <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>
- TikTok (n.d.). *Terms of Service*. <https://www.tiktok.com/legal/page/eea/terms-of-service/en>
- U.S. Copyright Office (2023). *Zarya of the Dawn Letter* (Registration # VAu001480196). <https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>
- UK Department for Education (2023). *Generative artificial intelligence in education. Departmental statement*. <https://assets.publishing.service.gov.uk/government/>

uploads/system/uploads/attachment_data/file/1146540/Generative_artificial_intelligence_in_education_.pdf

- Vincent, J. (2023, January 16). AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit. *The Verge*. <https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart>
- Weiser, B. & Schweber N. (2023, June 8). The ChatGPT Lawyer Explains Himself. *The New York Times*. <https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html>
- Wilson, H. J. & Daugherty, P. R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces. *Harvard Business Review*. <https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces>
- Wu, Y. (2023, July 23). How to Interpret China's First Effort to Regulate Generative AI Measures. *China Briefing*. <https://www.china-briefing.com/news/how-to-interpret-chinas-first-effort-to-regulate-generative-ai-measures/>