



European
Writers'
Council

AI TOOL KIT FOR THE **BOOK** SECTOR.

Recommendations by the European Writers' Council (EWC) for writers and translators, publishers, booksellers, event organisers and further stakeholders of the book sector for bilateral and contractual agreements and technical requirements including opt-out procedures.



AI TOOL KIT FOR THE BOOK SECTOR

Recommendations by the European Writers' Council (EWC) for writers and translators, publishers, booksellers, event organisers and further stakeholders of the book sector for bilateral and contractual agreements and technical requirements.

This Tool Kit can be used by political decision-makers as well as AI providers who need to install transparent documentation of "training data" and remuneration obligations.

This AI Tool Kit is neither binding nor intended to replace or prescribe national agreements.

The European Writers' Council (EWC) is the world's largest representation of solely writers in the book sector and of all genres (fiction, non-fiction, academic, children's books, poetry, etc.). With 50 organisations and professional guilds from 32 countries of the EU, the EEA, and the European non-EU areas, the EWC represents 220,000 writers and translators. These individuals write and publish in 34 languages, also worldwide.

Writers, illustrators, cover designers, translators, audio book narrators as performers, as well as publishers, editors, and collective management organisations (CMOs, RROs) are directly and immediately affected by the consequences of so-called AI in the book sector, and in particular by the production and use of generative informatics. The EWC together with its expert task force has developed the following recommendations to help establish a set of fair practices within the book sector and between writers, their possible agents, translators, publishers, and, where applicable, book sellers or event organisers, and (generative) AI developers.

In this toolkit, you will find:

- Affected jurisdictions and the controversial text and data mining exception(s) (page 1)
- Not all AI is AI: defining the types of applications addressed by this Tool Kit (page 2)
- 10 recommendations for fair, practice-based business relations (page 4)
- Legal basis for and implementation of the 10 recommendations (page 6)

Note: *The recommendations are aimed at territories where Directive 2019/790 (EU) on Copyright in the Digital Single Market (CDSM Directive) applies and where the TDM Art. 4 exception regime was introduced accordingly; they also apply to translations published in the EU Member States. All other points are also applicable internationally.*

Please note that the interpretation of whether text and data mining (TDM) as laid down in CDSM Directive Art. 4 covers the further use of works for the development of generative AI remains legally controversial. The AI Act (of 12 July 2024) refers to the CDSM Directive in its recitals but does not formally clarify this aspect either. We expect corresponding litigation and/or clarification by the Commission during the evaluation of the CDSM Directive in the course of 2025-2027. The EWC's position is to recommend declaring the TDM reservation of rights ('opt-out') as a precautionary measure to avoid the utilisation of any material for generative AI (GenAI), but, at the same time, we foresee exploitation for the development of generative AI as a highly probable new exclusive right. In this sense, we are employing the opt-out as a protective shield until the legal situation, together with the question of TDM not covering GenAI development, has been formally clarified.

PART I

NOT ALL AI IS AI: DEFINING THE TYPES OF APPLICATIONS ADDRESSED BY THIS TOOL KIT

Already, numerous wrongful and damaging “AI business models” have cropped up in the book sector – with fake authors, fake books and also fake readers. It can be assumed that the fundamentals for large language models (LLM), such as GPT, Meta, StableLM or BERT, were generated from copyrighted book works and partly sourced from shadow libraries such as Library Genesis (LibGen), Z-Library (Bok), Sci-Hub and Bibliotik – piracy websites. Without legal regulation, generative technologies accelerate and enable the expansion of exploitation and the legitimisation of copyright infringement, accompanied by information and communication distortion, royalty fraud and collective licensing remuneration fraud.

At the same time, a close look and an assessment are needed to categorise and regulate the individual aspects of advanced informatics; since software can be “smart” without being AI-driven, not every application comes with the same risk.

For a start, the EWC classifies the following three systems:

- Assistive Informatics and Software – not considered as AI nor as a risk;
- Analysing Informatics – partly considered as AI and a potential risk;
- Generative Informatics – the AI category considered as a risk; as related to text, voice, image works: generative artificial intelligence, in short “GAI” or GenAI.

In this paper, we focus on the legal, administrative, and technical aspects of so-called “generative AI” (GenAI) and related practices.

- We define the legal situation regarding input (contractual and technical routines) and output (labeling issues and transparency requirements).
- The paper examines automated text robotics (example: GPT), automated translation machines (example: DeepL), generative image production (example: Midjourney), synthetic cloning of human voices or otherwise AI-generated voices.

This is done taking into account the AI Act on the one hand, and the EU and non-European legal frameworks on text and data mining or intellectual property issues that still require clarification on the other hand: such as the definition of “machine learning”, which we consider to be closer to “algorithm programming” and which includes reproduction processes in preparation for GenAI development, such as scraping, temporary conversion of .pdf, .mobi, or .epub files into .xml; continuous copying in order to create a corpus (corpora) of words; the deposition of source files for reproducibility or verifiability purposes; copying, storing, and contextual breaking up and reproducing individual expression within artificial large language models incl. aspects of proximity and style imitation; image diffusion and synthetic voice cloning.

Please note: Analytical or assistive technology and software, such as semantic and proofreading analysis (example: Word Editor), image refinement (Photoshop), database management, filing, converting, citation indexes, storyboard software, text summarisation for metadata generation, sound mixing studio editing or automated inventory processes including CAT tools used by translators are *not* covered by this tool kit.

PART II

10 RECOMMENDATIONS FOR FAIR, PRACTICE-BASED BUSINESS RELATIONS

1. **Authorisation of exploitation** for text and data mining or GenAI development: Informed consent and written permission by authors, visual artists and audio book performers lay the groundwork for respecting their intellectual property rights as for any use of their text, visual, or audio works for (a) TDM, (b) scraping, (c) any related steps within algorithm programming for the development of generative AI. Therefore, contractual and communication routines for TDM opt-outs are to be developed, and individual (via publishers) or collective licences (via collective management organisations / CMOs) shall be obtained by any AI developers. Authors who want to protect their already published or to-be-published works from TDM scraping and any exploitation for GenAI, shall require their publishers in written form to declare an opt-out.
2. **Remuneration:** Authors, artists, and performers who give their informed and uncoerced permission for their works or performances to be exploited for text and data mining or for the programming of generative AI shall receive appropriate and proportional remuneration, in line with licensing models that are limited in time and scope. Licensing models should be transparent as to their purpose and as to how the works will be used, and observe clear reporting obligations over these uses in order to ensure adequate and recurring remuneration. Whether this is administered through individual or rather collective licences via CMOs and RROs is a question of national legislation and of negotiations among all national stakeholders.
3. **Transparency on input:** Scrapers and crawlers, as well as corpora builders and, ultimately, AI developers, are required by the AI Act to provide a sufficiently detailed summary of legally accessed titles, authors, and sources, as well as their acquisition methods of protected works including any relevant IP data. For this purpose, the book sector could develop and implement suitable and harmonised standards. At present, the following technical solutions for declaring opt-outs or giving licensing information are in use or development: metadata, the International Standard Content Code (ISCC) plus an attached rights declaration (e.g., but not limited to, using software by Liccium or Creators' Credentials), Digital Object Identifier (DOI), ISBN (International Standard Book Number), and the W3C (World Wide Web Consortium) TDMRep within ONIX. These systems not only allow the declaration of opt-outs, but also facilitate the tracking of works licensed by (Gen)AI developers, including online and other sources accessed.
4. **Transparency on output:** Every automated text including those obtained by machine translation, any AI-generated visual product, as well as any synthetic audio product should upon publication be labelled as AI-generated output. Upon consideration, 100-percent human works could also be labelled, following the concept of "Trusted Shops" models.
5. **Clear communication and respecting the moral rights of authors** to the integrity of their work. Publishers and other contractual counterparts shall seek author's approval before using generative AI in relation to their works and establish a mutual understanding on the utilisation of different kinds of software and advanced informatics, such as, but not limited to, synthetic voices in audio books, machine translation, generative cover illustration, and any other adaptation of the work by generative AI. Authors should have the right to choose to use human work and refuse AI-generated covers, AI audiobook adaptations or AI translations of their work without being disadvantaged and without negative consequences, such as lower royalties. Publishers should also be able to be sure that they know whether a work contains AI-generated components.

6. **Respecting the writers', translators', or artists' and performers' own choice** of working methods: Authors, translators, performers, or illustrators should not be forced to employ any GenAI or to work from AI-generated text including machine translation or GenAI images.
7. **A clear licensing and opt-out information chain:** Publishers need to declare to third parties including platforms, aggregators, libraries, or trade distributors, whether these have the right(s) to sublicense, reproduce and/or otherwise use the work in any manner for purposes of text and data mining or for the programming of generative technologies (GenAI) – or not. The rights reservation protocol (“TDM / AI / GenAI opt-out”) should be communicated for each title file, e.g., in meta data or in other digital rights management procedures such as, but not limited to, the ISCC+ rights declaration or TDM Reservation Protocol (TDMRep)+ONIX.
8. **When applicable, personal websites of authors, artists, and performers could,** and official websites of publishers and retailers should declare their TDM reservation under Art. 4, CDSM Directive 2019/790 and make clear they do not grant permission for TDM nor for scraping and exploitation for machine learning or algorithm programming of generative AI. This is possible and complies with requirements when stated in the general terms and conditions (T&C) or in the imprint, but especially when stated in a machine-readable way, for example in the robots.txt of a website URL or via the TDM Reservation Protocol (TDMRep). Publishers are required to raise a text and data mining opt-out flag on their company websites for each title or for the portfolio in its entirety, and the same applies to retailers and book sellers on online websites. Beyond websites, i.e. for the work itself, other technical and machine-readable standards can be used, such as the asset-based ISCC+ rights declaration, or the file-based meta data within the work.
9. **Check the T&C of your software:** Any publishing or aggregators' companies as well as agents, editors or translators should check the software they use to make sure that it does not claim within its T&C to be allowed to scrape, use, copy, store any content(s) for developing, improving or enhancing any AI including generative AI; this is also to be applied to platforms, social media and portals, where video recordings from reading events or panel discussions are to be published.
10. **Everyone should be aware of their own ethical responsibilities.** Most larger GenAI systems in existence today are alleged to have been built on copyright infringement. The works and investments of authors and publishers have been used without their knowledge, authorisation, remuneration or transparency, partly through piracy sites and well before the non-retroactive TDM exceptions of the CDSM Directive 2019/790 came into force. A concerted effort to denounce and monetarily compensate for this damage, as well as to push for said GenAI systems to be shut down, if necessary, should be an aspiration of the sector to secure its future. The oft-repeated misinterpretation of machine “learning” being the same as human reading, therefore constituting a “right”, is wrong. National entities, in cooperation with for instance, but not limited to, CMOs and stakeholders, should assess the economic damage already done by the uncontrolled development of GenAI. **European and national class actions, together with internationally applicable regulations that specifically hold AI developers accountable, are necessary.** Working together for an “AI and IP traffic code” and for regulations for a fair, ethical, and regulated future will protect the power of human innovation and creativity and prevent the disruption of knowledge and culture.

We are at the beginning of a discussion in the book sector that will shape future generations as it is continued. May these recommendations by the EWC serve as an initial impetus.

PART III

LEGAL BASIS FOR AND IMPLEMENTATION OF THE 10 RECOMMENDATIONS

1. On Contractual Matters

Overview of the legal basis for the recommendations:

- 1.1. Authors' intellectual property rights (copyright, authors' right) as related to commercial text and data mining (TDM) and to the new hitherto forms of exploitation: scraping, copying, storing, and other forms of exploitation for machine "learning" aka algorithm programming for generative AI.
- 1.2. Authors' moral rights: integrity of the work incl. translation, cover art, and audio narrating
- 1.3. Practical agreements on labelling and assessment of remuneration entitlement

1.1. Authors' intellectual property rights: (I) Right to opt out from TDM and (II) Reservation rights for all processes involved in algorithm programming with the purpose of developing (Gen)AI :

According to Articles 9.1, 9.2, and 9.3 of the Berne Convention, every author, artist, and audio performer has the right to decide how their work is published, distributed, copied, and used, unless national or transnational laws impose restrictions in form of exceptions, limitations, or other binding agreements. This includes the decision whether the work may be copied and used (a) for TDM for non-commercial and for commercial general purposes, (b) for the practices of scraping, copying, and storing, and (c) for machine programming ("training") for general purposes as well as for developing generative informatics (text, image, voice) and GenAI. The EWC regards the uses named under (b) and (c) as new types of uses previously not within the remit of author contracts or legislative exceptions. Exploitation of the works needs authors' written consent.

In the EU, the exceptions contained in Art. 3 (non-commercial TDM) and Art. 4 (commercial TDM) Directive 2019/790 on Copyright in the Single Market, came into force on 7 June 2021. Art. 4 allows text and data mining for commercial purposes; the only way authors and publishers can object to this non-remunerated exploitation is by exercising their right to declare a rights reservation in machine-readable or otherwise sufficient manner, the so-called "TDM opt-out" or "TDM rights reservation".

There are still many unresolved issues needing to be clarified, such as: how to deal with works existing before 2021, whether these also need a machine-readable opt-out, how proof should be made of their obvious use by AI companies, how they should be tracked, and, when applicable, remunerated. As long as it is not clarified under current EU law by e.g. the Commission or the Courts whether algorithm programming of economic substitutions of authors' works (GenAI) is at all covered by the exception of Art. 4 CDSM Directive 2019/790 and its national implementations, **the contractually declared or otherwise agreed-upon opt-out from TDM for commercial purposes by the author** (writer, translator, illustrator, performer) is the parachute with which authors and their publishers can reserve their rights until clarification is provided. Also, an opt-out entitles to licensing if the author so wishes, which can then be carried out by the publisher or CMOs.

→ For existing contracts, an addendum covering TDM rights reservation is recommended. However, amending old contracts (of which there are up to 13.8-20 million in Europe alone) will require high administrative and personnel efforts. Here, publishers and authors need to find procedures for works already on the market, including whether, for example, publishers will always declare an opt-out, except for those cases where an author expressly authorises TDM under Art. 4.

→ In parallel, authors should require their publishers in written form to declare the opt-out on any published as well as on their upcoming works.

New contracts including foreign rights / translation agreements shall include a clause or another form of written confirmation to this effect, coordinating the opt-out. For instance:

- The right to commercial text and data mining is not transferred with this contract (TDM reservation right under Art. 4, CDSM Directive 2019/790 (EU)). The author requests that the publisher inform them about this TDM opt-out declaration in a binding way.
- The author explicitly reserves the right to text and data mining, scraping, copying and storage//storing, as well as for algorithm programming for general purposes including but not limited to the purpose of producing any generative AI. The author requests that the publisher inform them about this TDM opt-out declaration in a binding way.
- The publisher will apply all necessary measures to communicate the commercial TDM rights reservation in an appropriate and effective manner, including but not limited to, within the meta data and ONIX, with machine-readable indications on websites or the website imprint, and ensures that licensees and/or retailers also communicate any associated reservations of rights.

With this paper trail, publishers can act in a compliant manner and label to-be-published books (e-books, digitised audio books, as well as printed books) with this TDM reservation in compliance with the current Art. 4 CDSM Directive 2019/790 requirements (within metadata and ONIX, Imprint, ISCC identifier + rights declaration, or via TDMRep).

To complete the picture about TDM rights reservation, please note the following:

- Translators also must comply with a reservation of rights to text and data mining or to the programming of generative AIs, and should not upload texts to machine translation software without the knowledge and written consent of the author or rights holder;
- Booksellers or other intermediaries such as librarians should also not feed texts or files into a GPT, e.g., to have summaries generated, as this would be in violation to an opt-out;
- Libraries giving access to digitised works under the Ulmer v. TU Darmstadt ruling (C-117/13, 2014) are not entitled to sublicense or give access to these for commercial TDM;
- Within a cloud-based publishing system, it must be communicated towards authors whether and in which steps a publisher enters texts or visual art into GenAI systems e.g., to extract keywords, summaries, or other knowledge from them, as this could also be ruled out by the authors' reservations of rights. Here, all parties involved must exchange information openly and transparently and come to an agreement and mutual understanding on the publishing workflow;

- While Art. 8 (2019/790 (EU)) entitles certain entities to digitise out-of-commerce works, this same article does not sign over the right to give access to these works for commercial TDM or algorithm coding (GenAI development). Entitled entities shall ensure that a machine-readable opt-out is applied (e.g., but not limited to, to archives digitised and made public online) to such works; they may not prevent rights holders from exercising their reservation of rights.

1.2. Authors' moral rights

Moral rights of authors include the right to having the integrity of their work respected – this is called the “integrity right”. In practice, this has been observed for decades, for example, when an author checks proofs and approves these before they go to print, or gives their written consent to an abridged version of their work for audiobooks or digest editions. Also, authors did not, ever, imagine their works and individual expressions scattered into millions of pieces for the development of GenAI systems. The unexpected digital use of their work for any other purpose than the original one of conveying art is in violation of the authors' integrity right.

In the field of GenAI, the author's moral right on the integrity of their work is extended to:

- Audio editions
- Translations
- Equipment, especially covers and illustrations
- Transforming and presenting their work.

The moral right shall also include the right to attribution (i.e., the right to have the authors' name on the work and no other). GenAI software shall have no right to appear in the imprint in the same way as a human author.

Agreements on audio book production

- Authors and publishers should agree that the writer has the right to refuse licensed audiobook editions of their works using artificial and/or synthetic voices. In principle, the writer must give their permission in a written form and must not fear any disadvantages if they reject GenAI voices or insist on human narrators when a licensed audio book is made. Some publishers and audiobook producers may have different views, e.g., based on economic considerations. It is to be hoped that the common interest in the appreciation of human work and cultural skills will continue to prevail.
- Please note: The requirements of the European Accessibility Act (EAA, Directive 2019/882), coming into force in 2025, allow text-to-speech adaptations of e-books, which will be carried out by GenAI voices on devices. This presumably cannot be contractually excluded, but it can be restricted by specific definition, such as “AI voice output is only permitted in the context of legally authorised text-to-speech under Directive 2019/882. Any further use of the text set to audio in this way for TDM or algorithm coding is not permitted.” Politically, it would be desirable to come to an agreement with legislators that any text-to-speech AI e-book edition shall not compete with original audiobook editions by human narrators.

- If authors record a book themselves, it must be stipulated in a separate contractual agreement that recording for voice cloning is not permitted except with their explicit consent, which shall include remuneration.
- Accordingly, the audiobook publisher should respond to the author's wishes to opt out of audio TDM and the use of the narrator's performance for synthetic voice replication and apply a machine-readable opt-out to these effects.

Agreements on translations in case of transfer of foreign rights

- To preserve the integrity of the work in translation as well as to reserve the right of use for TDM and for any GenAI developing or enhancing, an author has the right to refuse to have their work translated in whole or in part by machine translation, and to exercise their right of approval over machine translation incl. to reject pre-MT and post-editing by a translator. This is particularly important if the author wants to exercise their legitimate right *not* to have their texts used for GenAI developing, which already happens when manuscripts are fed into machine translation software. Unfortunately, authors may experience that the refusal of machine translation might lead to the effect that a publisher decides to not pursue a translation of the work. We hope that the publishing industry will remain true to its USP and the value of human work.
- At the same time, translators also have copyright and moral rights, and can refuse to post-edit a pre-machine-translated text. Here, agreements and clear communication in the author-publisher-translator triangle are important, especially with respect to any responsibility for the resulting translation and to labelling requirements. Unfortunately, authors may experience that any refusal of the use of GenAI by the author and/or the translator might result in the decision by the publisher to not pursue a particular exploitation of a work.
- Translators also have the right to reserve their rights for TDM and exploitation for developing GenAI as far as their language version is concerned. Both author and translator should, at best, agree that the final work may not be used for TDM, scraping-copying-storage, or machine programming and any production of generative informatics, and require the publisher to declare the opt-out within the contract or in other binding written form. The foreign publishing house shall apply the necessary measures to declare the machine-readable rights reservation, and to ensure that all sublicensees and further distributors are informed of and respect and apply the rights reservation protocol.
- Both author and translator should not be afraid to reject the use of machine translation, and such a rejection should not result in any disadvantages to them.

Agreements on cover art and other design (graphics, illustrations, pictures)

- Publishers should not, without the author's written consent, use covers, graphics, illustrations, or other design that has been created exclusively or to a significant extent by generative image (re)production (GenAI, text to image) to equip their work.
- Neither cover designers, illustrators, or other visual artists nor authors should have to fear any disadvantages, e.g. lower payments or royalties, if they reject the use of GenAI.

- In general, parties must be transparent as to the extent to which advanced informatics is or has been used as an *assistive* tool, such as image enhancement of human-created artwork, or automatically generated ALT text for accessible e-book formats. When producing ALT text, publishers should ensure that the images they load into AI description software are not exploited by the software developer, e.g., for GenAI training.
- Clauses to these effects can already be included in new contracts.

Respecting the writers', translators', or visual artists' own choice of working

Writers, translators, visual artists, and performers should not be required or coerced to use any generative AI or to work from AI-generated text or GenAI images against their wishes.

1.3. Self-Declarations by authors, and publishers' requirements of labelling

GenAI products are not human creations and, therefore, not protected by intellectual property rights. Accordingly, machine text or images have no legal right on remuneration and cannot be licensed to an exploiter. Machine output can be copied and reused by anyone.

This also applies to an author who offers a manuscript; if it is GenAI-produced, they do not have the right to grant licences of exploitation or to receive remuneration. Accordingly, agents and editors need to know whether the author has made use of GenAI. In this way, the transparency chain can be built all the way up to the reader and the public: knowing whether a to-be-published or an already published work – book, audio book, or visual, for example – has been fully made by humans or whether it was generated (fully or in part) by software will be relevant when money comes into play, e.g. in the form of remuneration from collective management organisations (for copies in copy shops, print book loan via PLR, equipment levies, but also performers' rights), where royalties are to be shared, but also as to the labelling obligations according to the AI Act. Institutions that award prizes or scholarships must also be sure that they continue to honour human achievement. Furthermore, machine outputs must not benefit from reduced VAT, fixed book prices or other subsidies dedicated to human works and protected cultural assets.

Whether the writer or translator must provide a corresponding self-declaration: strictly speaking, the usual clause in today's contracts stating that the writer or translator is the (sole) originator of the work, in accordance with national copyright law covering the level of creation (= 100% their own human work), is sufficient. However, publishers, out of legitimate concerns about, e.g., copyright infringements, plagiarism, their own obligation to label AI products, and wanting to ensure their rights to sublicence, may like to have the author sign a self-declaration concerning any (non-)generative AI use.

Labelling of any published work, whether fully GenAI-produced or in part, is necessary for all liability reasons: firstly, for clearly determining which authors and other rightsholders should be remunerated. Secondly, for knowing which intellectual property rights are engaged, and thirdly, for any allocations of liability (for infringement, plagiarism, violation of personal rights, disinformation). In this context, it is equally important to equip AI products with human-readable labels. On the one hand, to allow the reader to make a fully informed decision about what to spend their money on. On the other hand, so that the privileges enjoyed by books as a cultural asset in

many member states, such as reduced VAT, publishing subsidies or grants and prizes, will not be applied to machine products. Likewise, book aggregators and distributors are insisting that AI products be labelled. However, it is to be expected that there will be different views in the publishing industry on the requirement of labelling all generative technologies used.

The EWC follows the principle that only full, reliable transparency and trust in human endeavour will make our sector future-proof and trustworthy for readers.

Accordingly, bilateral contractual and communication practices between authors and publishers must be established:

- Already in existing contracts, the author states that they have created the work in full of their own creative powers, in accordance with the respective national provisions on authors rights and copyright and the necessary level of protected creation, e.g., Germany: §2(2) UrhG, “Schöpfungshöhe”, and own the full rights to transferring exploitation rights. Strictly speaking, this standard clause already excludes the work containing any generative AI.
- In the future, publishers might prefer a more specific self-declaration by the author to not have used automatically generated text, image, or MT translation in the work to be published, or to indicate any employment of generative technologies. The issue of whether some “small” quantity of GenAI should be permissible is volatile and far from settled; in the U.S., for example, “subordinate extent of GenAI” is considered to be a maximum of 5 % of the total for a certain work to still be accepted as a copyright-protected human work. Whether this matter can be resolved by specifying percentages and/or by specifying the assistive or analytical AI applications used will keep associations as well as the entire book sector occupied for some time. As a matter of principle, there should be mutual understanding and transparency on all sides.
- The utilisation of *assisting or analysing* software applications (for example, automatic citation indexing, automated synonym or rephrasing suggestions, photoshop as a purely assistive software, Word Editor), or having been inspired by regarding an AI image or reading AI-generated “poems”, e.g., shall be no subject of declarations. The EWC is of the opinion that assistive technologies that do not automatically or mechanically generate a “work” or parts of the work are exempt from the obligation to declare. Generatively produced “works” or parts of works, on the other hand, must be indicated without exception.

Fair play among colleagues: if you use ChatGPT, Bert or other generative text, image, or speech robots, you may infringe copyrights of your fellow authors and performers.

It is alleged that the foundation models of large language models were built using collections of over 4 million copyright-protected works. 194,000 titles from the Pile, Books1, Books2, and Books3 corpora have already been identified; their sources: bit torrent piracy sites. GenAI software copies and ‘memorises’ word chains and individual expressions from existing works, and often produces output with proximities to the originals, or even entire paragraphs copying an original word by word. Everyone who uses these actual applications risks infringing the copyright of authors whose works have been exploited and are now being reproduced, whether obviously or not, in this process, especially as these applications do not provide traceability of the works that were exploited to generate the text delivered.

Content control software employed by large e-book distributors examines each book before allowing the upload, and more and more anti-plagiarism software is being used, with the goals to (a) eliminate AI products and (b) detect copyright infringement. It is recommended to not use any GenAI for texts that are to be published.

2. Let's talk about TDM and opt-out techniques: From B2B ("Business to Business") to the General Terms & Conditions of your Software

2.1. Everyone using software in the book sector: Is your software scraping you?

In 2023 and 2024, nearly all software manufacturers extended their T&Cs. This concerns text, image, and management software, collaboration tools, online storage, cloud providers, mail providers, social media, etc. Now, the new T&Cs contain clauses that allow them to copy, store, reproduce and use any text, image, and further content for the development or optimisation of AI, including GenAI. Even though this is not allowed under EU copyright, data protection and privacy law, opting out is often made difficult if not downright impossible. When consent is denied, the full functionality of the software is restricted by the manufacturer. It is the same for Microsoft, Adobe, Apple, Google, Meta and other companies: either you give us access, or you no longer have all services or functionalities available. Accordingly, it is up to everyone, from authors to agents to editors and publishers, to check the software they use so as not to inadvertently open a loophole that allows access to works, work data and other sensitive business information.

2.2. Website developers of publishers, book trade, and authors: Be clear in opting out of TDM and prohibiting any scraping and exploiting of works to develop (generative) AI

- **TDM reservation in written format:** {your company or author name} expressly reserves the right to use the entire website content for commercial text and data mining within the scope of {your national legislation on TDM}. Similarly, {your company or author name} expressly reserve(s) all rights to grant scraping and machine learning for purposes such as, but not limited to, AI and generative AI development. Anyone wishing to obtain a licence to use this material, please contact {your email}.
→ Please note: It is very likely that crawlers will not be able to "understand" human language, but will only be able to read codes, cryptograms, or other machine-readable options, for instance:
- **Opt-out via robots.txt by manual coding:** <https://www.iubenda.com/en/help/137640-block-openai-crawlers>
- **The W3C TDM rights reservation protocol** within EPUB and pdf files of a book, and every URL where a book title is listed: <https://www.w3.org/2022/tdmrep/>
- **Opt-Out via ISCC (International Standard Content Code identifier) plus a softbinding, asset-based rights declaration.** The EWC's recommendation to the book sector is using the ISCC + rights declaration. This code can be used for works in all formats (print, digital, audio, image), and can be combined with a declaration that carries all essential information

such as rights reservations (or even licensing information), work data, author details, etc. in an irreplaceable form, as the declaration is not implemented in the work, but an external asset, and which is recorded in a data base of ISCC-registered works. In this way, not only can opt-outs be effectively declared, but AI developers are also able to use the ISCC code to document when they legally license works and to easily compile the title lists for proof of use in compliance with the AI Act transparency requirements. For illustrators, the combination with the software by Creators' Credentials is also a good way of proving the original (and human) provenance of an image.

- **Further information:**

- <https://iscc.codes>
- <https://iscc.io/>
- <https://www.youtube.com/watch?v=S1vK8LMK0f4>
- <https://docs.tdmai.org/>

Please note: The EWC will organise Webinars for its members on the ISCC and TDM / AI / GenAI rights reservation for book works as well as routines for a robots.txt opt-out for websites in late 2024/early 2025.

2.3. Publishers: **Close the backdoor of sub-licensing or ignoring your TDM opt-out.**

Publishers need to declare to third parties including platforms, aggregators (incl. libraries for e-lending) or distributors (incl. print-on-demand services) that these do NOT have the right(s) to sublicense or to reproduce and/or otherwise use the work in any manner for purposes of TDM under Art. 4 (2019/790 (EU)) for the development of artificial intelligence technologies generating text, images, or voices. The TDMRep may be useful for this purpose. Awareness must also be raised within the book sector that booksellers, librarians, literary critics, translators, scouts, agents, (peer) reviewers, etc., should not enter protected texts into software systems such as ChatGPT or Llama to generate summaries, keywords, or other information, as this already constitutes a copying and storage process for developing or refining generative AI, which the author has objected to.

2.4. Publishers: **Make your meta data fit in an era of scraping and crawling.**

Publishers are obliged to make the opt-out obvious by way of machine-readable meta data and ONIX, T&C on the website, with ISCC+rights declaration, TDMRep, or in the imprint – although this last option might in no way be sufficient to signal crawlers the rights reservation, but may only be useful when read by the human eye. Also, a declaration protocol must be agreed upon by publishers and online retailers and book trade portals, who are also obliged to indicate to any robot, crawler, and scraper that the opt-out applies to any book works online – for example through the TDMRep which is recognised by the ONIX for books. Therefore, a harmonised standard for bilateral publisher-bookseller information flow, the “opt-out chain”, shall be established, as well as the interfaces of online bookshops updated.

Two sources to learn more:

- <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/>
- <https://iscc.codes>

2.5. Event organisers: **Use caution when streaming or video-recording a reading, panel, or lecture.**

Whether it's a reading, a symposium, or a book fair discussion at a round table: events are often recorded on video and later posted on websites, YouTube channels or social media. YouTube started to sublicense videos to AI developers like Open AI, allowing them to transcribe speech to text for feeding into GenAI models. This is neither licensed nor remunerated and constitutes a considerable copyright infringement. Take the earliest opportunity to research//check the terms and conditions of the respective platform, to find out whether any of your content will be reused for the development of any type of AI – including voice, likeness, and the informative content of the lecture. Establish bilateral agreements with the artist(s) and author(s) as to whether they agree to any recording and republishing. In turn, author(s) and artist(s) should have the opportunity to opt out of the use of their likeness, voice, and presented content.

Related resources

- A model clause by the US Authors' Guild
- Practical considerations by the UK Society of Authors
- Analysis by EWC: <https://europeanwriterscouncil.eu/gai-is-based-on-theft/>
- An EWC dictionary on advanced informatics
- EWC's 10 principles to regulate (Gen)AI
- AI-Training is Copyright Infringement: Technological and Legal Tandem-Study