

Authorship Attribution in the Times of LLMs within the Framework of the CRediT Taxonomy

Paweł Kamocki, Andreas Witt

Leibniz Institute for the German Language
R5, 6-13, 68161 Mannheim, Germany
{kamocki, witt}@ids-mannheim.de

Abstract

This article examines the concept of authorship in the context of generative language models and other uses of Artificial Intelligence, and how this new ‘authorshipness’ can be represented in metadata. It analyses authorship under copyright law and proposes a metadata-based approach to disclosing the use of AI in publications, drawing on the widely adopted CRediT taxonomy developed by the National Information Standards Organization (NISO), and informed by guidance from the United States Copyright Office (USCO) and the International Association of Scientific, Technical and Medical Publishers (STM).

Keywords: metadata, authorship, AI

1. Introduction

AI tools have become an integral part of research. A recent report by Wiley (2025) indicates that 84% of researchers use AI tools in their work and that demand for expanded use is growing. Writing assistance is among the most common use cases: 74% of researchers report interest in using AI for this purpose, and 59% believe that AI already outperforms humans in this task. At the same time, researchers call for transparency: according to the same report, 66% consider it highly important for authors to disclose their use of AI in drafting and editing.

These developments are likely to drive changes in the concept of authorship in academia and beyond. Authorship is a foundational concept in copyright law and can be regarded as central to contemporary economic and cultural production.

This article proposes a mechanism to address the need to disclose the use of AI (Section 5), based on the well-established CRediT taxonomy. Before presenting this solution, Section 2 examines the concept of authorship and its implications in copyright law; Section 3 introduces the CRediT taxonomy; and Section 4 reviews recent developments concerning AI and authorship from both copyright law and publishing sector perspectives.

2. Authorship in Copyright Law

The author is placed at the centre of copyright law (called “author’s right” in many languages) as the initial holder of exclusive rights in a work.

In EU law, the sufficient and necessary condition for a work to be protected by copyright is its originality, understood as “author’s own intellectual creation”. According to the Court of Justice of the European Union (e.g., C-469/17 *Funke Medien*), this means that the work must reflect the author’s personality, which is the case when “the author was able to express his creative

abilities in the production of the work by making free and creative choices”. A logical consequence of this approach to originality is that only a work created by a human author can be protected by copyright, since only humans have a personality and are capable of making creative choices. The possibility of corporate authorship (like in a *work for hire*, where copyright is initially held by a legal entity that employs the human creator), although admitted in some (mostly common law) jurisdictions, remains an exception from the general principle of human authorship.

In many national jurisdictions copyright is transferrable, and in practice it is often transferred by the initial holder e.g. to a publisher. However, the author remains important throughout the lifecycle of a copyright-protected work. For example, the term of copyright protection (in most jurisdictions, life+70) is determined by the death of the author; the author also retains, even after the transfer of copyright, the right to claim authorship of his or her work (Article 6bis of the Berne Convention). *Author* metadata should be treated with great care also for another reason: the presumption of authorship of Article 15 of the Berne Convention. According to this provision, briefly put, in order to file an infringement case, it is sufficient that the claimant’s name appears on the work “in the usual manner.” Arguably, a “usual manner” to indicate the name of an author of a born-digital publication is in the metadata. Therefore, the person whose name appears as author in the metadata can sue for copyright infringement, even if the publication was in fact AI-generated and as such it is not protected by copyright. The burden of proving that the work was not in fact created by the person identified as the author would then rest on the defendant, and such proof is becoming increasingly difficult. This is one of the reasons why the AI Act (Article 50) requires that outputs of generative AI systems be clearly identifiable as such, but this requirement is particularly difficult to enforce in case of plain text outputs. Regardless of this legal obligation

(incumbent on the providers of AI systems, not on the users), disclosing the AI-generated nature of contents should be considered an ethical obligation of paramount importance (cf. Kamocki, Witt, 2022 and 2024).

Copyright law does not, on the other hand, recognise “Contributor” as an autonomous concept: in works created by multiple individuals (synchronously or asynchronously), each of them is granted the status of an author, as long as his or her contribution is original (i.e., in other words, as long as he or she left his or her “personal stamp” in the work). This is of practical significance, e.g. for the term of copyright protection, which for works of joint authorship is determined by the death of the last surviving co-author. Hypothetically, if a much younger assistant is to be considered a co-author of a work, the work is likely to remain in copyright for a much longer period. However, contributors who cannot be considered co-authors of a work, i.e. those whose contribution was not original, are not protected by copyright law in any way. Classically, a PhD supervisor who contributes ideas and provides guidance to a student, but does not participate in the drafting process, is not regarded as a co-author and therefore does not hold any copyright in the thesis. The same holds, e.g. for a technical group that contributes the data analyses in a life science article – despite the fact that the findings presented in such a paper depend on the data analysis.

This may contrast with the established academic practice described in the following section.

3. Authors/Contributors in Academic Contexts according to the CRediT system

The academic community tends, for good reason, to recognize a wide range of contributions, also with ‘faux’ authorship. In a widely quoted example, one physics paper had 5,154 authors (Aad et al., 2015). Obviously, such “authorship” often does not meet the standards of copyright law discussed in the previous section.

This phenomenon is partly addressed by the *Contributor* metadata term (present e.g. in the Dublin Core Metadata Terms, DCMI 2020), defined as “an entity responsible for making contributions to the resource” and distinct from the *Creator* term, defined as “an entity responsible for making the resource” (DCMI 2020; note that the Dublin Core Metadata Terms diplomatically avoid the term Author altogether).

In order to allow the community to distinguish between the various contributions to a published work, the CRediT (Contributor Roles Taxonomy, <https://credit.niso.org>) was introduced by the National Information

Standards Organisation (NISO) (Hosseini et al., 2026). Approved in 2022 as an ANSI/NISO standard, the Taxonomy recognises the following roles, which can be attributed to every author of a published paper :

- **Conceptualisation** (formulation or evolution of overarching research goals and aims),
- **Data Curation** (management activities to prepare data for initial use and later re-use),
- **Formal Analysis** (application of formal techniques to analyse or synthesise the data),
- **Funding acquisition**
- **Investigation** (conducting a research and investigation process, specifically performing experiments or data collection),
- **Methodology** (development or design of methodology, creation of models),
- **Project Administration** (management and coordination of the research activity planning and execution),
- **Resources** (provision of materials, samples, instrumentation, computing resources or other analysis tools),
- **Software** (programming, implementation of existing code, testing of existing code),
- **Supervision** (oversight and leadership, including mentorship),
- **Validation** (verification of replication/reproducibility of results),
- **Visualisation** (preparation of the published work, specifically visualisation/data presentation),
- **Writing – original draft** (drafting original text, including substantive translation),
- **Writing – review & editing** (specifically critical review, commentary or revision, including pre- or post-publication stages).

In the era of generative AI, most of these contributions (arguably, all of them, apart from Funding acquisition and Supervision) can be made with AI assistance. Nevertheless, this article focuses on the roles that are decisive for authorship as it is understood in copyright law (cf. Section 2 above), that is Writing, both the “original draft” and “review & editing”. Incidentally, these are also the roles (alongside Visualisation) where there is the greatest demand for disclosure of the use of AI tools according to the abovementioned report (Wiley, 2025).

Metadata are the right place for such a disclosure. Before the specific proposal for how to incorporate this information within the CRediT taxonomy is made in Section 5, Section 4 explores the various ways in which AI can be used in the writing process.

4. AI and Authorship: Copyright and Publishing Sector Perspectives

Widespread use of AI in the writing process is very likely to have a disruptive effect on copyright law and the publishing sector.

The issue of copyrightability of AI-assisted works is one of the main challenges copyright law will have to face in the near future. On the one hand, it is rather undisputed that AI-generated works should not be protected by copyright (cf. above in Section 2); on the other hand, some degree of AI-assistance in the creative process should not bar copyrightability of the output.

In Part 2 of its Report on Copyright and Artificial Intelligence, The United States Copyright Office (USCO, 2025) presented the conclusions from a large public consultation on copyrightability of AI outputs. The Report distinguishes between five types of human interactions with generative AI systems in the creative process:

- **Assistive uses** such as e.g. error correction or “brainstorming”. These uses do not affect copyrightability of the resulting works;
- **Prompting**; according to USCO, prompting alone, even very detailed and repeated, “[does] not provide sufficient human control to make users of an AI system the authors of the output”. USCO, however, leaves open the possibility that this could change with technological progress, if AI tools give users greater control over the final shape of the outputs;
- **Expressive Inputs** that are intended to be perceptible in the output (e.g. where AI is used for translation); according to USCO, in such cases the user retains at least partial authorship of the output which, however, does not extend to the AI-generated components, in a manner analogous to a derivative work;
- **Modifying or arranging AI-generated content**. In cases where the modifications are sufficiently creative, human authors can claim copyright in the final result, which, however, does not extend to individual AI-generated elements. Since this particular use of AI is of little relevance for the writing process (as opposed to, e.g., data compilation), it will not be elaborated upon in this article;
- **Inclusion of AI-Generated Content** in a larger human-authored work (e.g., special effects in a movie) should not affect copyrightability of the larger work.

Also in 2025, the International Association of Scientific, Technical & Medical Publishers (STM) issued its Recommendations for a Classification of AI Use in Academic Manuscript Preparation (2025). The document contains 9 categories of AI uses:

1. *Refinement*, correction, editing and formatting the manuscript to improve

clarity of language, e.g. via the use of spell checkers, grammar checkers and similar tools – according to STM, this is the only use of AI that should not be necessary to disclose;

2. *Writing* or drafting (parts of) manuscript content (either from prompts, or by asking to substantially expand or rewrite the input);
3. *Translation* of manuscript text for the purpose of publishing (as distinct from translation of source texts in the research process);
4. Refining or *formatting of data* reported in the manuscript;
5. *Generation*, refinement, correction, editing or formatting of images, diagrams or other *figures* for illustrative purposes only;
6. *Generation*, refinement, correction, editing or formatting of *visualisations* of research *data* or results;
7. *Refinement* or formatting of *code* reported in the submitted manuscript;
8. Assisting with *gathering references*;
9. Presentation of any kind of content generated by AI tools as though it were original research data/results from non-machine sources.

According to STM, the use described in point ‘9’ should be prohibited. All the other uses are allowed, but should be disclosed (apart from use 1, for which disclosure is not mandatory).

Due to its level of detail, the STM classification can be tentatively mapped onto the classification proposed by USCO, as shown in Table 1.

USCO	STM
Assistive use	Refinement (1), Gathering references (8), Formatting of data (4), Refinement of code (7)
Prompting	Writing (2) [from prompt]
Expressive Inputs	Writing (2) [from expressive input], Translation (3) [of manuscript]
Inclusion of AI-Generated Content	Generation of figures (5), Generation of data visualisations (6)

Table 1: The relation between the categories of AI uses proposed by USCO and STM

5. Disclosure of AI Use in *Author Metadata*: Proposal for Extension of CRediT

Author metadata is an appropriate place to disclose the use of AI in the writing process. This disclosure should on the one hand ensure a high level of transparency and by doing so promote trust; on the other hand, the adopted solution should not create unnecessary burden or stigma that could discourage users from disclosing the use of AI or, worse, prevent legitimate uses of AI.

The authors of this article propose to achieve this by extending the existing CRediT taxonomy (Section 3). Contributions consisting of Writing (be it “original draft” or “review & editing”) should come with an additional information concerning the use of AI. For the sake of systematisation, the authors of this article propose to distinguish between five types of AI use (Table 2), based on the USCO and STM classifications presented in Section 4.

AI use category	Examples
Assistive use	Idea generation, topic discovery, argument development and critique, gap identification, summarising and suggesting sources
Editing	Grammar and syntax correction, spellchecking, flow improvement, style and tone adaptation
Generation of minor elements	Generation of introduction/conclusion/abstract, finding examples (including e.g. linguistic structures)
Translation	Translation of the entire manuscript, translation of source texts/quotations
Compression/expansion	Using AI to shorten or expand existing human-generated input to meet word/page limit
Generation from prompt	Generation of an entire article from a prompt with a human assuming editorial control over the output.

Table 2: AI use categories to enrich the CRediT taxonomy

This classification contains two essential modifications compared to the USCO and STM classifications.

First, the authors of this article believe that “editing”, as a relatively simple and mechanical task in which LLMs clearly outperform humans should be distinguished from other assistive uses, in which the gap in performance between humans and AI is much more debatable.

Second, in the authors’ view translation deserves to be distinguished from both the “assistive uses” and from “expressive inputs”, and constitute a class of its own. The use of AI to take down the language barrier should be viewed differently from other uses of AI, as it assists rather than substitutes the human thinking process. At the same time, translation seems too substantial to fall within the “assistive use” category.

The proposed classification is also meaningful from the point of view of copyright law: “generation from prompt” would most likely exclude the output from copyright protection, whereas all the other uses should have little impact on the output’s copyright status

As a final step, the specific tool used in the process and its version should also be disclosed in the metadata. In short, the revised *Author* metadata can be represented as in Table 3:

Author	<i>Person</i>
CRediT role	Writing – original draft OR Writing – review & editing
AI use category	None Assistive use Editing Generation of minor elements Translation Compression/expansion Generation from prompt
AI tool	<i>Tool name and version</i>

Table 3: Author metadata disclosing the use of AI

6. Conclusion and Next Steps

The article proposes a clear and structured approach to disclosing the use of AI in author metadata, based on the well-established CRediT taxonomy. It addresses one of the most significant challenges faced by researchers, publishers, and legal scholars in the age of AI. It supports transparency and trust, reduces the risk of misattribution (thereby safeguarding the author’s moral right to claim authorship), and helps to align research and publishing practices with the copyright system. It preserves the centrality of

human authorship while acknowledging the growing role of AI in the writing process.

The authors further contend that the proposed approach can be extended to other roles within the CRediT taxonomy that may involve AI assistance.

The authors intend to engage with the National Information Standards Organization (NISO) and its CRediT Standing Committee to advocate for the incorporation of this approach into a future version of the taxonomy. They will report on the outcome at a next conference.

7. Bibliographical References

Aad, G. et al. (2015), [Combined Measurement of the Higgs Boson Mass in pp Collisions at \$\sqrt{s} = 7\$ and 8 TeV with the ATLAS and CMS Experiments](#), *Physical Review Letters*, 114, 191803

DCMI (2020). DCMI Metadata Terms. Available at:

<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> (last visit: 20.02.2026).

Hosseini, M., Kerridge, S., Allen, L., Kiermer, V. and Holmes, K. (2026). CRediT Roles and Example Research Tasks That Could be Attributed to Them. Available at: <https://doi.org/10.5281/zenodo.18421448> (last visit: 20.02.2026).

Hosseini, M., S. Kerridge, L. Allen, V.K. Kiermer, and K. Holmes (2026). Enhancing Understanding and Adoption of the Contributor Roles Taxonomy (CRediT). *Learned Publishing* 39, no. 2: e2048. <https://doi.org/10.1002/leap.2048>.

Kamocki, P. and A. Witt (2022). [Ethical Issues in Language Resources and Language Technology – Tentative Categorisation](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 559–563, Marseille, France. European Language Resources Association.

Kamocki, P. and A. Witt (2024). [Ethical Issues in Language Resources and Language Technology – New Challenges, New Perspectives](#). In *Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies @ LREC-COLING 2024*, pages 19–23, Torino, Italia. ELRA and ICCL.

STM (International Association of Scientific, Technical & Medical Publishers) (2025). Recommendations for a Classification of AI Use in Academic Manuscript Preparation. Available at: https://s3.eu-west-2.amazonaws.com/stm.offloadmedia/wp-content/uploads/2025/04/23020709/STM_AI_Classification_Recs_19_Sept2025-1.pdf (last visit: 20.02.2026)

USCO (United States Copyright Office) (2025). *Copyright and Artificial Intelligence, Part 2:*

Copyrightability. January 2025. Available at: <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf> (last visit: 20.02.2026).

Wiley (2025). *Explanations 2025: The Evolution of AI in Research*. Available at: <https://www.wiley.com/en-de/about-us/ai-resources/ai-study/> (last visit: 26.03.2026)