Kluwer Copyright Blog

Robojournalism – A Copyright Study on the Use of Artificial Intelligence in the European News Industry

Alina Trapova (University College London) and Peter Mezei (University of Szeged) · Thursday, March 3rd, 2022

The buzz around AI-generated outputs seems to never stop. While the field is rich on exaggerated claims, there are certain domains that have seen a genuine revolution fueled by AI. One such field is journalism. In the past years, sophisticated AI algorithms have become a meaningful assistant in the European news industry. Going beyond mere computer-assisted reporting (CAR), algorithms are nowadays extensively used in the reporting of sports, weather, and finance. One of the central issues in this respect is the copyright protectability of journalistic outputs generated by or with the help of AI.

Photo by AbsolutVision on Unsplash

In a recent paper (currently under review), we question whether humans can be replaced by AI to generate mechanical/less creative news reports and study whether "robojournalistic" outputs can be granted copyright protection from an EU law perspective. This research project seeks to fill that gap in literature by turning to the application of AI in the specific field of journalism and copyright law. The analysis in the paper combines three perspectives – a law and technology analysis, empirical evidence deriving from the business reality, as well as insights from media and communications studies.

Copyright and journalism – the status quo

The two central interconnected notions of this discussion are originality and authorship. Despite the absence of clear definitions of these two, it has been widely agreed That only output in relation to which it is possible to establish a sufficient level of human contribution (that expresses free and creative choices) would trigger a valid copyright claim.Thus, the notion of the author is an anthropocentric one. Linked to incentives and creativity, this is something that Daniel Gervais calls 'The Human Cause' (see here and here). Originality, on the other hand, following the guidance of the CJEU, has been understood as 'the author's own intellectual creation', where 'free and creative

choices' of the author are reflected in the work. Tying this to journalism, the paper unpacks the extent to which free and creative choices sustain in journalistic output. Most of the times, the creative freedom of the journalists is limited due to the presence of certain constraints such as style, length and structure, which are often imposed by the news editor directly. All these constraints, if very diligently followed, risk restraining excessively the free and creative choices of the human journalist. Thus, it could be argued that some journalistic pieces do not qualify for copyright protection as they would follow too strictly pre-determined rules. Inevitably these lead to one very straightforward and simple conclusion. Not all output, even if directly the product of human hands, benefits from copyright protection under EU copyright law. Free and creative choices and expression of intellectual creation must be present.

Technological perspective

Turning to the technological reality, the journalistic fields that thrive on AI are limited to those areas where there is abundance of numbers and data that need to be reported in a structured and clear manner, namely weather, sports and finance reporting. The main technology used in the field is natural language generation (NLG). While there is no one-size fits all technical model for NLG, consensus exists that in any NLG process process six basic activities need to be performed; these start all the way from input data to a final output text. Even though the order of these may vary, and some of them may be merged together, these stages always come back in one way or another as they represent the stages of any text generation – even when it is 'just' the product of a human author. Reiter and Dale summarise the six activities as follows:

- Content determination – deciding what information should be communicated in the text;
- Discourse planning – the ordering and structuring of the text into a coherent form; for example, ensuring there is a beginning, middle and end;
- Sentence aggregation – the grouping of messages and information into sentences;
- Lexicalisation – deciding which specific words and phrases should express central concepts and relations which appear in the messages;
- Referring expression generation – the selection of specific words or phrases to identify certain information;
- Linguistic realisation – ensuring that the text is grammatically coherent, following rules of syntax, morphology and orthography.

From this technological breakdown, it follows that there are at least two specific stages (discourse planning and lexicalisation), where the choices that are being made could be free and creative enough to trigger a copyright claim. This is however not guaranteed as it may be that the respective editorial policy imposes strict restrictions on creative freedom even during the discourse planning and the lexicalisaliton.

Business perspectives

Next, we carried out a targeted empirical analysis of selected European NLG service providers under various factors. We analysed 10 service providers: *AX Semantics, Retresco, Textomatic* from

Germany; *Syllabs* and *Labsense* from France; *United Robots* from Sweden; *Bakken & Baeck* from Norway; *Arria* and *RADAR* from the United Kingdom; and *Connexun* from Italy.

We paid attention to seven variables:

- General information of the service (especially the ways these corporations offer their service to their clients, e.g. Software-as-a-Service, Content-as-a-Service);
- The role of humans in the process of content generation (especially whether the service is fully automated or requires substantive human control);
- The number of available languages;
- The number of confirmed clients;
- The sectors that the given corporation is actively present (besides media & publishing);
- The use of service in journalism;
- The availability of the terms of use of the selected corporation's NLG (and if so, what these terms practically include)

Among the many conclusions that can be drawn from this analysis what struck us particularly was that most of the analysed service providers obscure their contractual practices. The publicly available and relevant documents almost unanimously necessitate the client to provide the source data and allow the use of the content without claiming any copyright interest in the input content. Furthermore, although most services advertise the underlying algorithm as fully automated, the final publication of the given content requires some human intervention in the newsrooms. Hence, the copyright protection of the relevant media outputs might effectively arise as a consequence to the potential free and creative choices made at the level of editing, after the NLG process has taken place. Eventually, advertising a service as automated may turn out to be simple window dressing when one studies the reality in the newsroom. The algorithmic creation of contents fits perfectly into the existing copyright business logic and necessitates no extension to any external parties or to the robots themselves.

Media and communications perspective

To comprehensively understand the key implications of robojournalism, copyright lawyers should also take a close look at the topic from the angle of media and communications studies. In particular, we studied the implications of robojournalism for journalists, publishers and readers/consumers. It appears that computing can support journalists to focus on in-depth, investigative activities that give them competitive advantage, rather than taking over their creative role. With respect to publishers, financial considerations seem to play an important and rather frustrating role. Due to the massive amount of resources needed to set up a functioning robojournalism newsroom (including the building of human-robot collaboration in the creative phase), only large media corporations are in the position to afford NLG solutions. It was not possible in our study to measure whether the externalities of robojournalism will be positive or negative for users/consumers. Yet, as a general consequence we can conclude that the massive news production, possible thanks to NLG, might contribute to a substantive devaluation of journalism. This is due to the fact the NLG output is less creative compared to human-authored

narrative pieces, but thanks to the speed with which the technology works readers would find themselves more often than not before an NLG piece. This could eventually impact their perception of journalism as a whole.

Taking all these considerations into account – the long-lasting need for human involvement in news creation, the limited switch to NLG by the bigger media corporations, and the hardly predictable outcomes of robojournalism for users – we argue that there is no convincing evidence in media and communications studies to introduce the copyright protection of automated news for the benefit of A developers.

Conclusions

The three perspectives studied in this paper – technological, business as well as media and communications – demonstrate that there is no clear case for copyright law to be extended to cover output generated by NLG. The current copyright framework is rooted in the presence of a human author and that should remain to be so. The absence of free and creative choices should not be artificially compensated by considerations for potential (if unproven) market failures if copyright protection does not arise for robojournalism output. It can be concluded that robojournalism follows well the negative spaces theory. Being the first one to utilise generative techniques that are trustworthy, transparent, accurate and zeroing discrimination brings enough benefits to companies resorting to NLG techniques, even in the lack of intellectual property, especially copyright protection.

A final draft of the paper can be accessed here. The authors welcome all comments and constructive criticism.

_____

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please subscribe here.

Kluwer IP Law

The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred location. Are you, as an IP professional, ready for the future?

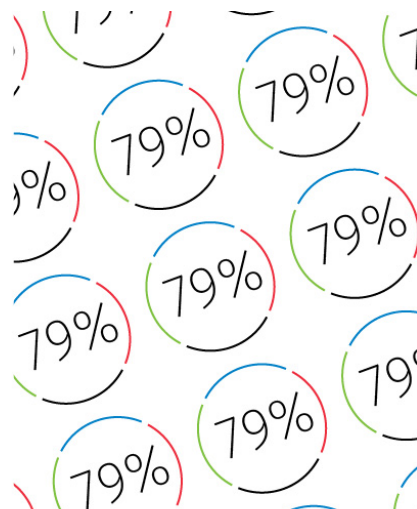Learn how **Kluwer IP Law** can support you.

This entry was posted on Thursday, March 3rd, 2022 at 8:03 am and is filed under Authorship, European Union, Originality, Software
You can follow any responses to this entry through the Comments (RSS) feed. You can leave a response, or trackback from your own site.