SMART 2019/1083
ACTION ON CEF AUTOMATED TRANSLATION
CORE SERVICE PLATFORM
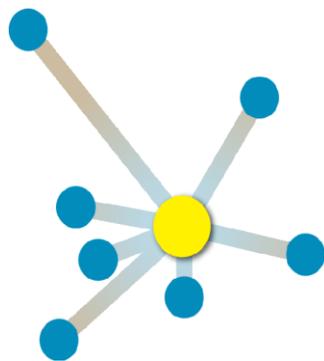
Use Case – Reuse of Emergency Calls embedded in TV Shows

Author(s):	Mickaël Rigault, Hélène Mazo, Valérie Mapelli (ELDA)
Dissemination Level:	Public
Version No.:	V1.0
Date:	2021-11-19
Copyright	© 2021 ELRC

Table of Contents

Executive Summary

In this document, the ELRC legal helpdesk will analyze under which legal conditions audio, video and dialogue subtitles coming from emergency calls embedded in a German TV show can be re-used for developing AI models. This analysis will lead us to review several legal aspects specific to the German legislation including intellectual property and copyright protection, but also to tackle the use and sharing of different types of data (audio, video and transcriptions of the dialogues), and their derivatives, for research and commercial purposes.

1   Introduction

During a research project that aims at building AI models for improving emergency call assistant systems, a researcher found out that some emergency calls made in Germany were included in a TV program produced and showcased by a German broadcaster.

This researcher reached the ELRC legal helpdesk to find out whether these dialogues could be used as publicly available data for research purposes. The researcher also wanted to figure out whether the models trained on crawled data[1] could be made publicly available.

Moreover, although full subtitles were provided during the TV Show, only part of it could be downloaded. The researcher asked whether the missing subtitles were bound to additional restrictions or if the same restrictions applied should the researcher made their own transcriptions of the missing part.

Following a preliminary analysis, we answered to the above-mentioned questions by providing the following legal elements.

2   Copyright exception in German law for research purposes

The German Act on Copyright and Related Rights (Urheberrechtsgesetz – UrhG)[2] provides some exceptions for researchers who use copyrighted content for the purpose of scientific research.

Sect. 60c UhrG provides that 15% of a work may be reproduced, distributed, and made publicly available either for a defined group of people for their own scientific research and for individual third parties to serve the quality check of non-commercial research.

Sect. 60d UhrG also allows to duplicate original material to create a corpus and to make it publicly available for joint research or for evaluation. However, under this provision, the corpus and the original material must be destroyed once the research is finished, but the model that has been trained on the data remains the intellectual property of the researcher.

3   Rights related to TV Shows (Audio & Video Feed)

As general practice, TV shows as such are protected by the copyright reserved to the producer of the TV show and/or the broadcaster. Therefore, to perform any work on such material it is imperative to check the terms of use related to the material and if necessary, negotiate a license with the right holders for reuse of the content.

However, as indicated in section 2 above, copyright exceptions depending on national laws may facilitate the reuse for specific practices. As far as the present request is concerned, the German copyright exception allows for using 15% without further notice to the right holder for scientific research or individual purposes.

---

[1] Thanks to the copyright exception provided for in the German Copyright Act
[2] https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html

4   Status of Emergency Calls

The emergency calls embedded in the TV show constitute a distinct piece of data that we must consider also distinctively in terms of intellectual property rights. One question must be addressed: can those conversations constitute copyright protected creations at all?

We also need to consider the conditions under which the producer could license the reuse of these conversations if they can be considered as copyright protected. Finally, we will examine how to differentiate between actual recordings and reenactments of scripted conversations.

4.1   Copyright protection

In our opinion, these conversations constitute pieces of information that have been produced by public authorities, i.e. the regional or local fire brigades, during the performance of a public service and should therefore be considered distinctively from the TV show. As such they cannot be protected by the producer or broadcaster of the TV show. Instead, they could be reused under the provisions of the Public Sector Information directive[3] that defines the right for open data and re-use of public sector information.

In addition, from the perspective of Intellectual Property protection, we can consider that these conversations, as they occurred in a context of emergency, would not qualify for copyright protection as they lack original and personal character needed for protection under the law.

4.2   Acquisition of a user license from rightful authorities

Although one could call for the PSI regulations, another hypothesis could be that the producer or broadcaster of the TV show has been granted permission to use the recordings of the calls by the authorities that are keeping the archives of these calls.

If the producer received a license to reuse the emergency calls in the TV program, then it may be possible that the researcher could obtain a license to reuse the calls for research.

However, the researcher reported that the handling of the calls being decentralized in Germany, it would be difficult to enter into an agreement with each competent authority.

4.3   Examination of the originality of the recordings

At a first sight, one may think that real emergency calls were recorded. However, another possibility could be that the producer of the TV Show had only access to a scripted version of the emergency conversations and decided to have actors reenact the conversations in a recording ultimately integrated into the show. If this holds true, then the rights to the performance of the voice actors would probably be passed to the producer.

In this case, verifications should be run to find out whether the production properly disclaims that the conversations have been reenacted at the beginning or the end of the show.

5   Creation of partial transcriptions of the show

Much like what we analyzed for the TV shows in section 3, creating transcriptions from the show is covered by the right to create derivatives that belongs to the producer of the TV show.

---

[3] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L1024&from=EN

Thus, it would not be possible to create a transcript of the whole show without obtaining related rights from the producer of the TV show.

Nevertheless, copyright exceptions could also apply in that regard, if only audio excerpts (up to 15% in Germany law) were exploited instead of the whole show. Consequently, it would be possible to extend the protection granted to researchers to derivatives created from the protected excerpts.

6 Use of download manager systems to retrieve audio and transcriptions

The researcher also reported that Youtube-dl [4] was used to download the audio and transcriptions from the show. Regarding this practice, we should pay attention to Sect. 95a of the UhrG that provides that effective technological measures implemented to protect a copyrighted work may not be circumvented.

When reviewing the ARD Library FAQ[5], we found out that if downloading videos is allowed, the feature does only allow to save the content within the dedicated app or in the device internal memory. Therefore, we can assert that there are technical measures that have been implemented to restrict the reproduction of the TV show.

On that account, we can conclude that using Youtube-dl to download the programs from the Library constitutes a circumvention to the technical measures set by the ARD.

7 Redistribution of the data and models trained on data

7.1 Redistribution of data

The copyright exception of article 60c UhrG provides that work reproduced for the purposes of non-commercial research can only be made publicly accessible to (1) a defined group of people for their own scientific research and (2) to individual third parties to evaluate the quality of scientific research.

Following this provision, it is impossible to make data extracted under the provision of this exception available to any entity for commercial purposes. We would like to emphasize that the notion of "commercial purpose" encompasses the monetary exchanges and the transfer to commercial entities.

However, redistribution for further scientific research is completely feasible and its terms should be aligned with the restrictions of the copyright exception.

7.2 Redistribution of NLP models

The second item of interest we would like to bring up is the NLP models intended to be built up by the researcher. The researcher informed us that those models would be trained on the basis of data collected only within the provisions of the copyright exception for scientific research.

---

[4] https://ytdl-org.github.io/youtube-dl/
[5] https://hilfe.ardmediathek.de/faq/Geraete-%26-Funktionen/Kann-ich-Sendungen-offline-speichern-und-ansehen%3F/

In our analysis, language models are considered as computer programs that are copyrighted objects under the provision of Sect. 69a et seq. of the German Act on Copyright and Related Rights (Urheberrechtsgesetz – UrhG).

The untrained version of such model would therefore be the intellectual property of its creator. So would also be the trained version as it would only be an enhanced version of the original program that has been modified with new parameters by its programmers and trained with data that has been collected in compliance with exceptions to copyright.

8   Conclusion

To summarize, we can say that the following statements shall be followed when dealing with copyright protected works and public data collected from German TV shows with embedded emergency calls:

- German law allows to collect up to 15% copyright protected material for research.
- Data produced by public authorities can be separated from the copyright protected material in application of the PSI Directive.
- Inclusion of public domain data into copyright protected works does not prevent from collecting this data for research purposes.
- It is allowed to create derivative works from data collected in compliance with copyright exceptions.
- It is not allowed to circumvent technological measures set by copyright owners to prevent crawling of websites.
- Redistribution of data is only allowed (1) for specific persons in the framework of their own research and (2) for third parties for research evaluation purposes.
- Intellectual property of untrained Machine Learning models belong to their creators; trained models are also protected when they are trained with data exempted from copyright protection.