Research and Innovation Action (RIA) H2020 – 957017

SELMA

Stream Learning for Multilingual Knowledge Transfer

D8.1 Ethics Deliverable

Work Package	8
Responsible Partner	Deutsche Welle
Author(s)	Peggy van der Kreeft (DW), Kay Macquarrie (DW), Afonso Mendes (Prib)
Contributors	Joscha Rieber (FhG), João Prieto (Prib), Yannick Estève (LIA), Normunds Gruzitis (IMCS)
Version	3.0
Contractual Date	31 March 2021
Delivery Date	~~29 March 2021,~~ ~~18 November 2022,~~ 31 March 2023
Dissemination Level	Public

Version History

Version	Date	Description
0.1	15.2.2021	Initial Table of Contents (ToC)
0.2	24.2.2021	First draft with template for contribution
0.3	11.3.2021	First full version ready
0.4	16.3.2021	Integrated all contributions from partners
1.0	26.3.2021	Final formatting and layout
1.5	11.11.2022	Added section (5.8) and updated references
2.0	18.11.2022	Resubmission
2.5	27.1.2023	Addressing Ethics Check Report
3.0	31.3.2023	Resubmission II

Executive Summary

The main goal of the ethics report is to give an overview of the SELMA project's ethics, mitigation and awareness strategies and outline possible ethical implications. This document will be updated within the course of the project's developments, as needed. The issues addressed here will be part of the data management, project management and evaluation reports.

SELMA's central concept is to build a deep-learning NLP platform that trains unsupervised language models, using a continuous stream of textual and video data from media sources and make them available in a user/topic-oriented form in over 30 languages.

The knowledge learnt in the form of deep contextual models is transferred to a set of NLP tasks and made available to users through a **Media Monitoring Platform** (Use Case 1) to be able to handle up to ten million story segments per day. The media monitoring platform will be able to transcribe, translate (on demand), aggregate, write abstractive summaries, classify, and extract knowledge in the form of entities and relations and topics and present all this to the user using new visualizations and analytics over the data. The learnt contextual models will also be applied to a **News Production Tool** (Use Case 2), using enriched models for transcription (ASR) and translation (MT), giving journalists in an operational editorial environment a multilingual tool that will be able to learn over time. For testing the NLP components and pipelines of the SELMA platform, **SELMA Basic Testing and Configuration Interface** (Use Case 0) has been additionally introduced. It is used as both an internal testing platform and a public demonstration platform of the SELMA components and pipelines.

Thus, this involves the provision of media monitoring capabilities based on data from publicly available media streams. We use this data to process, track and profile information about people and organizations, which means that the project needs to carefully address ethical issues relating to privacy.

Table of Contents

Table of Tables

1. Introduction

The aim of the SELMA is to address three tasks: ingest large amounts of data and continuously train machine learning models for several natural language tasks; monitor these data streams using such models to improve multilingual Media Monitoring (use case 1); and improve the task of multilingual News Content Production (use case 2), thereby closing the loop between content monitoring and production.

This report presents Deliverable 8.1 Ethics Deliverable and it addresses ethical issues in seven broad categories:

- Protection of personal data
- Copyright protection
- Data Management
- Ethical implications of SELMA technologies
- Social impact of automation
- Sex and gender balance
- Risk assessment of SELMA technologies

In March 2023 the SELMA project was undergoing an Ethics Check by external advisors. The findings of the report (cf. Section 11.2 in the appendix) are addressed throughout this document; two dedicated sections were introduced, see Section 2 Ethics & Data Advisor and Section 9 Risk Assessment.

2.Ethics & Data Advisor

The consortium decided to ask a pair of independent experts advise on the ethical issues relating to the project. We contacted Dr. Els Kindt and Prof. Lorna Woods, and they produced an initial Ethical and Legal check of the project activities. The commission also performed an ethical check. This deliverable was updated to take into account the advice received. Below is the list of advised with answers and actions planned from the consortium:

Table 1 Overview: Advises by The External Ethics Report

#	Advise / Obligation	Further Description	Action / Solution	Status
1	Omit the collection and processing of additional categories of sensitive data for diversity purposes (Monitio use case (UC1))		Collection & processing of sensitive data of Wikidata were eliminated	done
2	Provide lawful basis for processing personal data (e.g. binary gender)	For the (re) processing of e.g., the Wikidata information, in particular for diversity of news source monitoring ('diversity monitoring'), SELMA shall rely for the processing of particular 'sensitive' data on an appropriate basis and exemption of Art. 9.2 GDPR.	Provide a lawful basis to process personal data (e.g. binary gender)	scheduled for May 2023
3	Develop strategies against bias and potential discrimination and freedom of expression		See Tables "Technology Ethics Risk Assessment" and "Platform Ethical Risk Assessment"; cf. D8.1 Chapter 9	done
4	The use of 'local' EU cloud is for this reason highly advised		All SELMA data storage is located within the EU; cf. D6.3	done

5	Assess if SELMA voice technology fells under 'processing of biometric data'	SELMA shall review, e.g. in its DPIA (see below) to what extent biometric data is processed and if so, if the prohibition of art. 9 could apply, safe exceptions.	Provide an assessment in a DPIA	(see #11)
6	Apply the principle of minimizing data	SELMA should have a clear overview of all (public) training and test data sets; this applies also to the 'user feedback' collected by SELMA and shared and used for SELMA improvements.	Cf. D6.3	done / ongoing
7	Apply the principle of "Data protection by design"	These include hiding information, limiting access to authorized (licensed) users only, encryption where appropriate, separating information (e.g., by storing (meta)data in a distributed manner), and aggregating.In terms of the monitoring tools, these could be restricted to those who have need to use the tools rather than making them available to all licensed users; moreover limitations could be placed on the granularity of the searches so as to limit if not exclude the possibility of plurality searches identifying specific individuals.	Cf. D8.1 and D6.3; also this is an ongoing process and will be part of exploitation activities	done / ongoing
8	Apply other GDPR principles such as Transparency, Profiling, Explainability	It would be advised to provide information about the project on the website of the controllers; The transparency and information obligation also applies for UCO (in particular for the public demonstration platform). Profiling: Particular attention is needed to assess the effects of the processing to individual newsmakers, including e.g., journalists, as well as to the individuals, subject of the news	Provide information about personal data processing on the SELMA website and on the public SELMA OSS; Status: scheduled for April 2023 Assess effects to potentially profiling of individuals and the explainability of	scheduled for April 2023

		stories, as to what they become exposed to profiling and/or increased exposure based on reporting news and/or automated Explainability: Automated decision based on 'AI' shall overall also be 'explainable'. To the extent the processing is based on search terms, rather than black box processing, as we discussed, this needs to be described and assessed in the DPIA and explained why this may not lead to increased risks.	automated decision will be part of the DPIA (see below #11)	
9	Clear the roles of partners and the users of the platform	D6.3 states that all partners to the project are joint-controllers. In this case, an appropriate joint-controllership agreement shall be signed amongst partners. Largely, two types of users may be envisaged: are these users merely consulting the platform for information (without storage, editing and other data processing rights, etc) or are they (in addition) allowed to use the platform for organizing their own content, making own selections and compilations of information and the posting of own content. The possibility of restricting the licence has been noted by SELMA but needs to be considered further.	Roles of partner (controller vs processor) has been reviewed and clarified in D6.3. The roles of uses for the SELMA platform and potential licence restrictions are part of the exploitation activities. Status: ongoing	Role of partners has been cleared (done). Roles of uses is ongoing
10	Provide a contact for "Data subjects"	Which responsible entity will be the contact of the data subjects (e.g. the journalists whose news items are processed, users of the platform, …) is also one of the core elements in the joint-controllership between partners to SELMA.	This information will be part of activity #8 and will be made available on both of the public websites	Cf. #8
11	Conduct a DPIA	"(...) because of some 'systematic monitoring ' activities, we deem it important to conduct such DPA. All elements required	A DPIA for SELMA technology will be conducted;	scheduled for June 2023

		are further described in Art. 35 GDPR. This assessment should also include an assessment of fundamental rights and freedoms.		

3. Protection of Personal Data

One aspect of the project on which we will particularly focus is the protection of personal data.

The protection of personal data within SELMA comprises:

1. the protection of personal data relating to the engagement with stakeholder participants; and
2. the protection of personal data stored in the SELMA platform itself.

Task T6.3 develops and implements a strategy for all data management and protection issues in the project. This plan considers the data collected in SELMA, but also sets down explicit procedures that must be followed by consortium members regarding the protection of identifying data or other type of data to be protected. This is supported by the management processes put into practice in WP7.

3.1 Data related to end users, research participants and other stakeholders

SELMA research itself does not primarily focus on people, but the project will include human participants from various user groups, who will first be involved in the definition of user requirements for the use case studies primarily by DW (Deutsche Welle), as well as Priberam and University of Latvia. End users will also evaluate the systems. The partners will supervise their recruiting and evaluation procedures. Evaluation output/scoring and any other relevant data relating to human participants will be anonymized for protection of the identities of the participants and compliance with EU and international regulations. Still, there is the risk that data and technology (e.g. speaker recognition, named entity linking, topic labelling and topic clustering) is used to profile people. This needs to be considered and proper solutions need to be developed to prevent this kind of profiling especially for non-public individuals and relevantly to journalists.

Personal data is not collected during web crawling from any source, except for mentions to public figures in the text of news articles, these mentions are reported under the freedom of the press principles. News articles available on the web are frequently tagged with their journalist author and this poses a problem about the possible dual-use of using the platform in such a manner that could put those at risk. Addressing these issues is an ongoing discussion together

with the ethical advisors, our current assessment is that we should not allow the possibility of performing analysis using “sensitive” properties regarding people being them public or not.

The consortium discusses data privacy in detail and take all measures necessary to adhere to ethical standards.

3.2 Procedures and criteria to identify/recruit evaluation participants

The project will recruit internal and external participants to contribute to the user requirements. Users will also be recruited through the consortium extended networks as the project progresses, to increase the relevance of the data collected. Recruitments will be conducted in line with the project's ethical guidelines which link directly to the European Commission's ethics self-assessment guidance[1]. In this manner, all participants will have access to detailed information sheets. There will be no recruitment of participants under the age of 18.

Unless attribution is explicitly required, such as utilizing expert opinion, most of the data collected directly, e.g., through interviews and surveys will be made anonymous.

An information sheet will be provided to all participants with details on the nature, purpose, and requirements to participate in the evaluation. The SELMA participant information sheet will be customized for each different type of external participant involvement in the project. This will mainly relate to interviews with key stakeholders during the gathering of end-user requirements and the evaluation of the platform within WP5. Furthermore, if necessary, translations of the participant information sheet will be provided if required to who do not have sufficient understanding of English.

3.3 Protection of data related to the SELMA platform itself

SELMA technologies generally do not focus on acquiring and processing personal data from its users, but obviously some published content may contain some data that can refer to individuals. Also, one Use Case example (Diversity) uses Wikidata labels for content categorisation along diversity characteristics. The aim is to receive an indication of how diverse

[1] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf

a dataset is (e.g., in terms of binary gender, regions and age; no sensitive personal data such as ethnicity, disability and sexual orientation will be analysed). The aim is not to profile individuals, but to get an insight about distributions (e.g., How often were women mentioned in an article, how often were men mentioned?). The focus of this research is to give journalists the tools to assess how and if their reporting covers all societal aspects without discrimination on, gender, ethnicity, cultural backgrounds, etc... These kinds of tools are essential to foster for a more inclusive society and should never be used to track individual journalists, they should be used to get the big picture on how the media deals with minority aspects of society. SELMA will apply methods including regular meetings of the Data Management Committee for the protection of such personal data, in particular regarding the gathering, identification, storage, retention and the destruction of (personal and other) data.

3.4 Data Gathering

Relating to the protection of personal data, the members of the consortium are aware that the project collects data that is considered personal during the data collection phases, specifically multimedia data, data collection from the web. The whole process is described in D6.3. However, only publicly available news and media content from organisations and news publishers will be targeted for data gathering. All efforts will be made to avoid collecting user comments or other user-generated personal data. For the protection of data within the SELMA project the following aspects will be implemented:

- The SELMA Project Management Board, through WP7 **continuously assess** the legal, ethical, and societal impact of the solutions developed within the project and the potential future implementations and deployments based on them.
- The basic approach of SELMA is to **reduce the collection and even initial storage** of personal data to the absolute minimum. The acquired personal data will, under no circumstances, be used for commercial purposes or shared with third parties.
- SELMA **follows the formal procedures** that are explicitly defined within each partner organization and in the Data Management Plan to protect the anonymity of data that is shared among the consortium.
- To ensure an external unbiased view SELMA will regularly evaluate and assess the developments with the external data & ethics advisor.

Only data necessary to the completion of the project will be stored. Social media data will be restricted to news and published media no data from other users which are not publishers or with whom there is no signed agreement is collected. In the case that, for a justifiable reason, the Consortium decides to gather additional data that we find sensitive or that the user consent so requires it, that data will be securely retained, using industry-standard encryption and access control. All data protection documentation is centrally held by the project and is therefore available for audit. More detailed information on D6.3.

All data collection and storage during the life of the project is overseen by the Project Coordinator (PC), a Data Management Committee and an external data & ethics advisor and as such all necessary European legislation and best practice will be adhered to in this area (see chapter 2.5).

3.5 Compliance with national and EU regulations

Underpinning all the procedures above are the regulations set down both nationally and EU-wide regarding the implementation of data protection procedures. Any partner within the project who is collecting personal data must adhere to their country's data protection policy, as well as that of the EU. SELMA will specifically comply with

- the EU Data Protection Directive 95/46/EC[2] and
- the EU's General Data Protection Regulations[3].

Currently, the Data Protection Directive defines personal data as being "any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular in reference to an identification number or to one or more factors specific to their physical, physiological, mental, economic, cultural or social identity".

Processing of personal information is defined as "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure

[2] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046&from=EN

[3] https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction". Specific measures will be analyzed and tailored to the technological and legal framework of each use case.

SELMA activities include processing, tracking, and aggregating information about mentioned people, organizations, locations, events, etc... from texts published in web media press content; also, these activities apply to diversity labels from Wikidata, which are used to classify broadcast content.

When and if additional types of data are required to be ingested by the platform, the project will carefully address ethical issues. It is important that, throughout the project, we also develop an understanding of the impact the technologies we develop may have on people. When SELMA presents such information, the results will be "aggregate views" without exposing any personally identifiable data points. This risk could extend to "regular" broadcast content, where individuals are mentioned. We will make a distinction between data relating to public figures and others. We will provide a clear categorization.

Any personally identifiable information not needed will be destroyed. However, this must also be balanced against the responsibility of the consortium to conduct reproducible research and the project goal of knowledge-base construction. As in the general case awareness is raised of the risk that individuals might be identifiable through the media content.

4. Copyright protection

As the project processes enormous amounts of media, the consortium needs to observe the rules on copyright protection. We assume that most of the videos, images and text that will be used are protected by copyright. This means, in principle, that any copying of these works requires the prior approval by the respective rightsholders. However, as we will be processing these items in the context of scientific research, and more specifically, in the context of text and data mining, we will benefit from certain legal privileges.

Section 60d of the German Act on Copyright and Related Rights, for instance, supports large-scale text and data mining for non-commercial purposes:

> *In order to enable the automatic analysis of large numbers of works (source material) for scientific research, it is permissible*
>
> *1. to reproduce the source material, including automatically and systematically, in order to create, particularly by means of normalisation, structuring and categorisation, a corpus which can be analysed and*
>
> *2. to make the corpus available to the public for a specifically limited circle of persons for their joint scientific research, as well as to individual third persons for the purpose of monitoring the quality of scientific research.*

This means for SELMA, that it would be lawful according to German law to create and process large corpora of texts, images, and videos and to share these within the consortium for scientific and non-commercial purposes. However, at least so far, these corpora cannot be published openly or shared outside the specific research community (meaning the consortium).

When the SELMA project started, other countries, including France, Portugal and Latvia had not implemented a similar copyright limitation for scientific research and Big Data analysis yet. However, this situation will change in the near future as the EU passed a *Directive on Copyright in the Digital Single Market*[4] ("European Copyright Directive") in 2019. At the beginning of

[4] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

the SELMA project, this directive had not been implemented into the laws of the member states. Nevertheless, as the implementation period ended by June 7th, 2021, it is likely that the stipulations of this directive will, in one way or another, be applicable to the SELMA project.

The European Copyright Directive foresees certain privileges regarding the use of works that are protected by copyright without the rightsholders' approval. Most relevant for the SELMA project is Article 3 that stipulates that:

> "*member states shall provide for an exception to [Copyrights] for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access"*.

This includes the right to store copies of these works, provided that they are stored

> "*with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results*".

However, the directive also points out that these copies may only be retained for as long as is necessary for the purposes of text and data mining (Article 4(2)). Also, these privileges for text and data mining

> "*shall apply on condition that the use of works [...] has not been expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online*".

This means that we can expect privileges for text and data mining activities within SELMA to be justified based on the European Copyright Directive and the respective implementations into the laws of the member states. As member states have a certain leeway as to how they implement the directive, it will be necessary to closely follow the legislative process and make appropriate adjustments throughout the course of the project.

The measures taken within SELMA to cope with the above framework on copyright laws are so regarded in two different views for UC1:

1) For research use within the activities of the project (scalability tests, training of models, platform tests, etc..), we will use all the content that we have available from public media publisher websites.

2) Data that the clients and users can use in their views within UC1 is restricted to the data for which Priberam ensured a license agreement with the media producer or media representative association. Agreements are already in place for Portugal (Visapress), United Kingdom (NLA) and Spain (Cedro).

A discussion, with legal advisors, is still ongoing whether we have lawful basis to show the only the title, a snippet of the article and a link to the original document. Meanwhile (2) is being enforced. Deliverable D6.3 contains additional information on this aspect.

5. Data Management Plan

The consortium published a detailed Data Management Plan in M6 for the SELMA project, which was updated in M18 and will be updated in M36 of the project. The Data Management Plan functions as a central tool for risk mitigation associated with data protection. The Data Management Plan includes the following aspects:

- A clear description of what research and innovation activities project data is used and a description on who is responsible for handling, storing, and destroying the data (data processing)
- A clear description of the purpose of our research and innovation, to make clear that there is a substantial public interest in the work of the project
- A clear description of the safeguards that we will put in place
- Identification of the countries in which data will be processed or reside, together with an understanding of the national privacy and data protection regulations, and engagement with the relevant data protection agencies

The privacy impact assessment includes:

- A description of the information flows in the project (distilled from the Data Management Plan)
- A detailed identification of the privacy and related risks
- Actions taken by SELMA to reduce the identified risks
- Integration of these outcomes into the project plan, in particular the Data Management Plan.

6. Ethical implications of SELMA technologies

In this section possible ethical concerns related to individual components and/or technologies of the SELMA platform are presented; for an initial risk assessment of SELMA technologies and platform see chapter 9.

6.1 Algorithmic transparency

Throughout the project, the consortium members will document and disseminate information about their respective technology components and aim to provide the greatest openness about the purpose, structure and underlaying actions of the sets of steps, models, features, and variables the SELMA platform and the individual components will utilize. In this sense, SELMA aims to make the algorithms and the factors that define them – including the datasets on which the models are trained upon - as transparent as possible. The project strives to use datasets which are diverse in respect to various dimensions such as gender balance and will also include other diversity areas.

6.2 Aggregation of data

The SELMA platform will bring data streams together and create aggregated data. We will further describe the data sets used for the individual use cases and the data sets the platform overall can process. This also goes for the various data sets we will use as ontologies and knowledge bases. We will describe and communicate the quality of the data and its sources, including its accuracy, completeness, uncertainty, as well as its timeliness, magnitude (when training a model), possible bias or other limitations.

6.3 Speaker diarization and speaker recognition

The SELMA platform is mainly used for news and other broadcast media content. In our main use cases, speaker diarization and recognition will primarily be applied to media professionals and persons of public interest who appear in the media items. However, if this technology is open sourced it might also be applied in a wider sense. We will address this issue and bring it to attention of potential users. We will clarify to which extent speaker diarization and the creation of voice databases and identities should be used and restricted in the any use case and we will also assess the feasibility of pseudo anonymization for the open-source use. We

understand that this technology opens up many desirable and versatile use cases, like detailed notes in panel discussions and focus groups. However, we also understand that it bears risks to privacy protection. We will advise future users to implement best practices and high awareness for privacy protection. This technology will be used in use case 2 but only for segmenting the transcriptions, the project will not, at any time, store biometrical data collected with this mechanism.

6.4 Rich automatic speech recognition and machine translation

For the ASR (automatic speech recognition) and ML (machine learning) training, respectively, a huge amount of training data is necessary. We leverage as much publicly available data as possible that is free for use and does not contain sensitive personal content. Additionally, the models are continuously trained on crawled and partner-specific data provided for the project. So, we made sure that the vast amount of needed data is stored responsibly. We aim for an unsupervised stream learning approach that enables us to temporarily store the data just for retraining and model improvements. A balanced dataset regarding speaker gender and age is important for an unbiased model. We take care of that when selecting the training datasets. For the partner-specific data, we will advise the potential partner not to violate given rules when storing and processing their data.

6.5 Expressive and personalized voice synthesis

We have identified three areas that might raise ethical concerns when dealing with artificial voices which all will be further analyzed by the consortium.

First, it is interesting to understand the perception of artificial voices, which categories, attributes, and identities make up for a natural sounding voice, and which attitudes users have related to artificial voices. We will investigate whether people prefer voices that sound extremely natural or whether they should retain attributes that distinguish them as artificial. And if there are particular use cases and formats where users prefer artificial voices over natural ones and vice versa, these aspects are best analyzed through user surveys or panels which we intend to conduct as part of user evaluation.

A second aspect of working with artificial voices are good measures to ensure artificial voices can be recognized as such if needed. This will be mostly on the technical side. During the

project, we will assess the need to make artificial voices clearly identifiable as such and implement possible technical measures to do so.

The third area of consideration regarding the usage, creation and argumentation of artificial voices are measures to understand ownership and copyright. We will address issues such as: What happens if a voice is augmented and changed, and in how far does such change affect copyright?

6.6 Named entity recognition and linking, topic labelling

Named entity recognition, linking and topic labelling are essential to the project. We carefully assess the validity and accuracy, as well as monitor for bias that might arise. Incorrect linking entity might cause misleading information. When a named entity is linked incorrectly, a failed entity disambiguation could be a potential cause. For instance, for the sentence “Paris signed an agreement with the Trump Model Management Agency” the entity Paris should be mapped to the person Paris Hilton, obviously not to the city. Erroneously mapped entities might cause a misunderstanding of a document and knowledge degradation. Even worse, biased data analysis might also cause systematic errors, e.g., a misleading link could harm a person by mapping his or her name to a person who committed a crime or has a different political opinion. So, the consortium aims for highly accurate named entity disambiguation based on state-of-the-art technologies. Comprehensive accuracy evaluations are part of the research done in this domain. As with any system using machine learning technologies or systems that learn statistics from data, error and bias are inherent.

Users of the technologies should be (made) aware of the nature of algorithms and the potential ethical problems and threats that incur from their use.

6.7 Abstractive summarization

Summarization is a complex and exciting part of the project. The nature of this technology and its application is to filter the relevant part and enable the users to engage with a vast quantity of content pieces. To ensure best outcomes, we will provide transparent documentation and evaluate if and how the perception of content changes when summarized with the tool. During evaluation we will not only investigate how well the summarization works, but also how and in which context SELMA summarization tools will be used by the media professionals, and what effect the summarization technologies have on information gathering. Abstractive

summarization rewrites the original text in a compressed way, this rewriting might lead to an incorrect layout of the original facts. One of our major efforts during the research is to minimize this problem but, as always, users should be aware and warned about the possible problem. We will also look for biases that might arise from the automatic distillation of large-scale news content and for ways to address this aspect of implementation of machine learning assisted summarization.

6.8 Cross-Analysis and Filtering of Data

The cross-analysis of vast amounts of data opens up novel ways to detecting content and conduct filtering over various topics and entities. It can be used to help people (find areas with unbalanced media representation), but at the same time it can also be used against them if this technology is used in a harmful way (e.g., by identifying persons who express something which is against a political view, for example in non-democratic countries). In SELMA these technologies are primarily applied in UC1 Media Monitoring within the Advanced Use Case and here especially in the Diversity Use Case Application enabling users to detect various diversity categories for public figures which are on Wikidata including gender and age (Wikidata stores even more diversity categories including ethnicity and disability, but the use of this data would not be lawful under GDPR rules). The aim is to get indications on how diverse a dataset is both in terms of quantitative measurements. In order to mitigate potential risks, the SELMA project will closely and regularly evaluate the technological developments and will establish a risk analysis with a focus on ethical and privacy concerns towards a potential use of the technology in the Use Cases and the commercial platform. The consortium will continue to debate and consult external advisers on the matter whenever the questions arisen by the technology so require. The final product legal framework should address these ethical considerations, binding the users with the principles against discrimination on gender, religion or sexual orientation.

6.9 SELMA platform

We understand that open sourcing tools might both increase the probability of socially beneficial uses of the technology, but also use of the technology for unintended use.

Open-source use

By definition, open-source licensing cannot limit any type of usage even if it is unethical. The provider of open-source software can also not be liable for any unethical use of their product, we strongly believe that the benefits of open-source release outweigh the risks. And while the current definition of an open-source license excludes any limitations of usage, we are aware that there are voices and initiatives in the open-source community campaigning for a change in the definition of one that focuses on ethical considerations like the concept of the Hippocratic license by Coraline Ada Ehmke[5]. A Hippocratic Source license would specifically prohibit the use of software to violate universal standards of human rights and embodying the principles of ethical software. We will closely monitor any developments that might occur in this area.

Commercial use

The main ethical implications from the commercial use of the platform again appear to arise from the possible use of personal data to target individuals. As mentioned above, this is something we have addressed in our policies relating to the use of personal data throughout the project. Commercial use is foreseen in terms of proprietary software to be implemented in products such as Monitio and plain X.

[5] https://github.com/EthicalSource/hippocratic-license

7. Social impact of automation on jobs and employment

As the project deals with the automation of workflows, increasing productivity and reducing the time needed to complete editorial tasks performed, it is essential to understand the potential impact of this on editorial jobs. There is a widespread concern that novel technologies, especially AI applications and automation, may replace or kill some existing positions. On the other hand, the same technology may redefine and enrich job positions in a positive way.

The project provides support for low-resource languages and allows wider coverage of such regions. It also helps to automate monotonous tasks, enabling media professionals to focus on more creative and skilled aspects of the journalistic work.

The goal is for this technology to reduce the time needed for laborious editorial tasks, and lead to better performance and job satisfaction.

Additionally, we note that the combined effect of the economic crisis with the media crises has resulted in newspapers being closed and newsroom depletion due to job cuts, hence innovation is welcomed, since it will help journalists to deal with the increased workload. Innovation will also help improve quality, which will result in better audiences and products, which will also help keep media jobs or even create new ones.

Overall, the view prevails that early adoption of innovative language technologies, involving media professionals in such tool development, and focusing on a human-centric approach for the workflow creates a positive effect on the role of media professionals. It opens up more opportunities for optimized implementation of HLT applications, providing innovative solutions to media companies and ensuring that they can expand to new markets and target new audiences fast and stay competitive.

8.Sex and Gender Balance

The project ensures sex and gender balance in different ways:

It takes gender differences into account in market trend analysis, and in developing user scenarios, ensures gender balance during testing and user evaluation, selecting people, setting up questionnaires, etc. It also ensures that workshops, conferences, evaluation sessions, and hack events avoid any gender bias and actively address specific groups to enable diversity.

Furthermore, the tools in the platform make it possible to better understand gender representation in the media. For this matter, we are working on a specific use-case application making diversity aspects in media more visible.

9. Risk assessment

The following tables address the internal consortium assessment of the risks posed by each of the developed technologies, possible dual-use, the data involved for training, the measures minimize bias and the risks involved on using the technologies.

9.1 Technology Ethics Risk Assessment

The risk analysis is made based on the following six principles (1): Protecting human autonomy, Promoting human well-being and safety and the public interest; Ensuring transparency, explainability and intelligibility; Fostering responsibility and accountability; Ensuring inclusiveness and equity; Promoting AI that is responsive and sustainable.

Table 2 Technology Ethics Risk Assessment

Technology	Benefit	Potential dual-use	Datasets used for training	Measures to minimize bias (if appropriate)	Risk: hallucinations, explainability
Clustering	Enabling comprehensive coverage of a specific story	none so far	Public dataset Published by	So far bias problems where not detected for the task nor in the datasets	The model does not output why the clustering decisions are being taken. It is not a major problem in our context since it has to be validated by a human. This should be taken into account if such a model is used

Technology	Benefit	Potential dual-use	Datasets used for training	Measures to minimize bias (if appropriate)	Risk: hallucinations, explainability
			Rupnik[6]		in a pipeline with subsequent automatic decisions. The final user should be alerted to the kind of automatic processing being done, and the possible errors associated with the model, such as aggregating unrelated stories
Summarization	Synthetising the of key ideas of a text	None so far, in the models developed within the consortium. Similar technologies could, in principle, be used to generate false or misleading information	public: XSum, CNNDailyMail, and others DW news	Ensure that training data is diverse, ensuring that the base Language Models have a good language coverage	Summarizations can be wrong or misleading, the models can hallucinate and produce misleading and wrong information. Minimizing these is one of the objectives of the current work. The final user should be alerted of these kinds of behaviours

[6] Rupnik, J., Muhic, A., Leban, G., Skraba, P., Fortuna, B., Grobelnik, M.: News across languages - cross-lingual document similarity and event tracking. In: Journal of Artifcial Intelligence Research, Special Track on Cross-language Algorithms and Applications. (2015)

Technology	Benefit	Potential dual-use	Datasets used for training	Measures to minimize bias (if appropriate)	Risk: hallucinations, explainability
Multilingual summarization	Enabling short synthesis of key ideas of a translated text. Enables general understanding of text in multiple languages	As above	DW data, and other public datasets	Ensure a good coverage of languages, and the use of base Language Models with appropriate Language cover	The same risks as for the monolingual case. Plus the risks associated with the transfer to the target language
Topic Detection	Understanding large collections of text data, by assigning “tags” or categories	Aggregation against topics IPTC topics together with other variables, poses risks depending on the intended usage	Closed LUSA dataset and the licensed Finish dataset	Until now no biases have been detected in the datasets, the addition of the Finish dataset ensures a more diverse labelling. The output models should be monitored and if bias problems are to be detected the appropriate debiasing measures should be taken in the datasets	Documents can be classified with wrong topics. The final user should be alerted
NER	Finding specified entities in a text (e.g. person, location)	Detection of mentions to people in text can be used to monitor specific people	NER-annotated datasets created within the project	Ensure a good language coverage, diverse data. The datasets used in SELMA are news datasets. Use in other domains requires creation of test sets	Mention span can be wrong, undetected or classification erroneous
Entity linking	Linking entities with its corresponding description in a knowledge base	Linking mentions unique identifiers in text can be used to monitor specific people.	public datasets, Aida-Yago CONNL, TAC, AQUAINT, ACE2004, ClueWEB, sVoxel, Wikipedia, Wikidata	Actuality is the main concern, new entities keep appearing, entities context drift in time	Entities might be ambiguous and incorrectly labelled

Technology	Benefit	Potential dual-use	Datasets used for training	Measures to minimize bias (if appropriate)	Risk: hallucinations, explainability
Speech recognition	Recognition of spoken language into text	Detect content in speech / video in large datasets with harmful / criminal intentions	DW data for self-supervised training. For supervised training: CommonVoice, TEDLIUM-3, mtedx, ESTER+EPAC+REPERE+QUAERO (French broadcast news)	Ensure that training data is diverse in speakers, dialects, languages, gender etc. Also see our research paper about gender bias (Interspeech 2022: https://arxiv.org/pdf/2204.01397.pdf)	Speech recognition might be wrong, especially for named entities. Recent deep neural models suffer the same hallucination problems as the summarization models. Our work also focusses in minimizing those
Speech translation	Translation of speech in one language typically to text in another	As above, with even less barriers	DW data for self-supervised training and data distributed in the framework of the IWSLT evaluation campaigns from 2019 until 2023: mainly MuST-C corpus, HowTo corpus, and data collected in Tamasheq language from broadcast news radio station and translated into French under the supervision of LIA, in association to ELRA, the European Language Resource Agency. Also use of the mtedx data	As above	Speech translation might be wrong, especially when translating from low-resourced languages
Speech Synthesis	Convert text into speech	Usurpation / audio deepfake	DW data (journalist voices + scripts)	Consent of informed professional adults	Unintelligibility, sound artifacts, lack of naturalness, repetitions

Technology	Benefit	Potential dual-use	Datasets used for training	Measures to minimize bias (if appropriate)	Risk: hallucinations, explainability
Speaker Diarization	Segmenting individual speakers within an audio stream for efficient indexing, transcription, and analysis	Unauthorized surveillance, eavesdropping, and invasion of privacy by tracking and analyzing audio data without consent	VoxCeleb open-source dataset consisting of audio recordings with labeled speakers	Ensure that training data is diverse and representative, including speakers from various demographics, languages, dialects, and accents	Segmentation errors may lead to incorrect attributions, impacting the reliability and trustworthiness of audio data analysis. However, it is not a big concern in our UC2 nor used in UC1. Moreover, inaccurate or biased diarization models could result in unfair treatment of speakers from specific demographics or with particular accents
Story Segmentation	Dividing input streams into meaningful units which are valuable for information retrieval, content recommendation, and summarization	Facilitate the distribution of harmful or illegal content by automating the process of breaking down such content into smaller, more easily shareable segments	DW data; Datasets could include news articles or any other public collection of text, audio, or video data with natural breaks and transitions	As above	When analysing long segments of audio, it is important to divide the speech into semantically connected pieces. The use case is News bulletins where different stories are reported. The models do not hallucinate and explainability is not an issue

Technology	Benefit	Potential dual-use	Datasets used for training	Measures to minimize bias (if appropriate)	Risk: hallucinations, explainability
ASR & Speaker Identification	ASR transcribes spoken language into text, while speaker identification distinguishes between individual speakers. This makes audiovisual content more accessible and searchable. Speaker clustering also improves the readability and organization of transcriptions	Voice data is protected and not shared without permission	Widely-used open-source datasets like Multi-lingual LibriSpeech and CommonVoice apart from DW data	We do not expose speaker-oriented models. Clustering models do not save fingerprints. Speakers are not automatically associated with persons	Errors in speaker identification may lead to incorrect attributions, impacting the reliability and trustworthiness of audio data analysis and ASR transcriptions

9.1 Platform Ethical Risk Assessment

Table 3 Platform Ethical Risk Assessment

Technology	Benefit	Risk / dual-use	Solution
UC1 - Monitoring	Search & analysis in vast news content datasets, allows journalist to better fact check their stories, further investigate. Allows decision make to make better informed decisions, etc.	A platform like this can, when used by to ingest personal data as social media or other sensitive data, can be used for tracking citizens and journalists.	The platform will not be shared as open-source. No social-media is ingested into the platform. All data in the platform, except for press data will not contain references to individuals. Authors of articles will not be available for aggregation.
	Aggregation of content on topics, organizations, people, events, media publisher, location, date of publishing and profession. These aggregations are essential for the users of the platform to produce analysis on specific topics of interest. Aggregations are allowed on public figures reported by journalists using their right to inform, allowing for a better scrutiny and fostering bigger transparency.	Users are able to conduct searches on individuals, e.g. journalists / authors of articles.	Mitigation strategies on the possible aggregations regarding people are being studied. Authors of new articles will not be available as aggregations in order to protect them. Ethical and legal advice on these matters and on how to cope with the relevant issues are still in progress.
	Automatically aggregate news about a specific story and enable to follow that story over time.	No risks have been found.	
	Produce graphs connecting events, topics, people etc.	A useful tool to understand the underlying connection	As above.

Technology	Benefit	Risk / dual-use	Solution
		between entities and explore the content. The same kinds of problems as for the aggregation when personal data is involved.	
	Aggregations on specific characteristics of people extracted from Wikidata properties. Like ethnic group, religion, gender, profession etc. would enable the study on how journalists report about minorities. Raising awareness on the topic would lead for a more inclusive news reporting thus leading for a more inclusive society.	There is the risk that if proper measures on the granularity of the aggregations are not in place, this would allow the targeting of categories which fall under the definitions of personal sensitive data in GDPR.	Sensitive properties from Wikipedia will not be collected, see the diversity use case.
UC2 - Production	The tool helps to create transcripts, subtitles, translations and voice-overs in many languages.	Data processing through Commercial NLP or AI models reduces control over the data. Is a voice an authentic voice or being generated?	Only work with trusted engines and have commercial contracts. User Data should not be sent to jurisdictions for which GDPR compliance is not assured. Users from different jurisdictions have different restrictions. GDPR compliance must be enforced for the platform, see D6.3 for the measures taken. If you use synthetic voice, you indicate that it is AI generated. All AI and ML methods are explained to the users in accessible documents.
UC0 - SELMA Open-Source	UC0 is based on UC2, but with open-source tools and models. No data is stored for this use-case. It is used for debugging the components of the platform.	Models can be used for nefarious reasons.	Open-source models should state their behaviour in the various situations.

10.Conclusion

The aim of this report is to introduce the key ethical questions that SELMA must address.

Three major areas are identified which have a direct relation to the work in SELMA:

1. **Protection of Personal Data / Privacy:** Privacy is a major ethical issue arising from SELMA, especially relating to the work we do on press media analysis. Privacy is designed into the SELMA data management infrastructure which is defined in D6.1 Initial Data Management Plan (M6, M18) and will be refined in one following deliverable (M36).
2. **Ethical implications of SELMA technologies and platform:** SELMA technologies have a direct impact on workflows and on new ways of automation. Next to management and impact reports, data protections and ethical conflicts will also be an important part of technical deliverables.
3. **Bias in data:** SELMA deals with very large content streams from various sources. The bias in data plays a vital role in the selection of the data streams and will be documented in deliverables. Making aware that data is biased will play a key role in our blog communication.

These ethical challenges are closely monitored and discussed during the project and are an integral part of the data management, project management and evaluation reports.

11.Appendix

11.1 Ethics

Correction: Concerning "4. Personal Data" both fields should state "Yes" (Yes, our research involve personal data collection and/or processing. And: Yes, our research involve further processing of previously collected personal data (secondary use)).

Note: In the "Ethics Summary Report" from April, 22nd 2020 it was correctly marked, stating: Yes, this research involves personal data collection and / or processing.

4 - Ethics

1. HUMAN EMBRYOS/FOETUSES		Page
Does your research involve Human Embryonic Stem Cells (hESCs)?	○ Yes ◉ No	
Does your research involve the use of human embryos?	○ Yes ◉ No	
Does your research involve the use of human foetal tissues / cells?	○ Yes ◉ No	
2. HUMANS		**Page**
Does your research involve human participants?	○ Yes ◉ No	
Does your research involve physical interventions on the study participants?	○ Yes ◉ No	
3. HUMAN CELLS / TISSUES		**Page**
Does your research involve human cells or tissues (other than from Human Embryos/ Foetuses, i.e. section 1)?	○ Yes ◉ No	
4. PERSONAL DATA		**Page**
Does your research involve personal data collection and/or processing?	○ Yes ◉ No	
Does your research involve further processing of previously collected personal data (secondary use)?	○ Yes ◉ No	
5. ANIMALS		**Page**
Does your research involve animals?	○ Yes ◉ No	
6. THIRD COUNTRIES		**Page**
In case non-EU countries are involved, do the research related activities undertaken in these countries raise potential ethics issues?	○ Yes ◉ No	
Do you plan to use local resources (e.g. animal and/or human tissue samples, genetic material, live animals, human remains, materials of historical value, endangered fauna or flora samples, etc.)?	○ Yes ◉ No	
Do you plan to import any material - including personal data - from non-EU countries into the EU?	○ Yes ◉ No	
Do you plan to export any material - including personal data - from the EU to non-EU countries?	○ Yes ◉ No	
In case your research involves low and/or lower middle income countries, are any benefits-sharing actions planned?	○ Yes ◉ No	
Could the situation in the country put the individuals taking part in the research at risk?	○ Yes ◉ No	

7. ENVIRONMENT & HEALTH and SAFETY		Page
Does your research involve the use of elements that may cause harm to the environment, to animals or plants?	○ Yes ◉ No	
Does your research deal with endangered fauna and/or flora and/or protected areas?	○ Yes ◉ No	
Does your research involve the use of elements that may cause harm to humans, including research staff?	○ Yes ◉ No	
8. DUAL USE		Page
Does your research involve dual-use items in the sense of Regulation 428/2009, or other items for which an authorisation is required?	○ Yes ◉ No	
9. EXCLUSIVE FOCUS ON CIVIL APPLICATIONS		Page
Could your research raise concerns regarding the exclusive focus on civil applications?	○ Yes ◉ No	
10. MISUSE		Page
Does your research have the potential for misuse of research results?	○ Yes ◉ No	
11. OTHER ETHICS ISSUES		Page
Are there any other ethics issues that should be taken into consideration? Please specify	○ Yes ◉ No	

I confirm that I have taken into account all ethics issues described above and that, if any ethics issues apply, I will complete the ethics self-assessment and attach the required documents. ☒

11.2 Ethics and Personal Data Report

2.0

Memo

From : E.J. Kindt and L. Woods

Re : Input Ethical and Legal Check of SELMA – Stream Learning for Multilingual Knowledge Transfer

Date : 21 March 2023

1. Introduction

Below you find a first assessment of some legal and ethical aspects as raised in and required by the Ethics Check Report of November – December 2022 of the SELMA project.

The assessment is based upon 1) a review of the D6.1 Initial Data Management Plan, the D6.3 Interim Data Management Plan and the D.8.1 Ethics Deliverable; 2) a presentation of the SELMA project by the coordinator Kay Macquarrie and Peggy van der Kreeft on 27 February 2023, followed by 3) a review and analysis of the presented slides and the Ethics Check Report of 2022.

This report provides an assessment and input from mainly an AI and a data protection regulation perspective. It does not cover IP issues. As to data protection, only some key points are briefly discussed, without aiming to provide a full data protection analysis.

2. Summary of the relevant facts and use cases

The SELMA project ('SELMA') is building further on news gathering and production platforms developed before, including in the SUMMA project. The core of the SELMA development (UCO SELMA OSS) is a natural language processing platform, on which various applications are built, including Monitio, an advanced content analysis and Media Monitoring Platform (Use case 1 or UC1). The second SELMA application is content production, referred to as a News Production Tool/Plain X (Use case 2 or UC2), which would also be available when in operational use to external journalists for news item creations. Both use cases are presented and made available for internal testing and demonstration purposes. Additionally, there is the SELMA Basic Testing and Configuration Interface (Use case 0 or UC0) which also is publicly available under: https://selma-project.github.io/ .

The technologies used include deep-learning multilingual natural language processing (NLP), models and analytics and rely on unsupervised learning and unstructured data. The unstructured data used consist of a continuous stream of text, audio and video data from world wide media sources. The content (re)used, produced and processed on SELMA is limited to content obtained from professional media providers, such as journalistic content, excluding user generated social media content produced by the public at large. [1] The output will be user/topic oriented.

[1] See also D 6.3, pp. 28-29.

1

2.0

3. A continuous evolving legal landscape : the upcoming AI Act

3.1. AI Act Proposal

The European Commission proposed in 2021 a law on artificial intelligence (AI), which will categorize AI applications into three categories of risks. The proposal would classify systems in 'prohibited practices', 'high risk systems', and 'general purpose AI systems'. Other regulators have also issued guidelines as to the use of AI. The SELMA platform uses AI[2] and is very likely to be subject to this upcoming legislation. We do not further analyse these guidelines and the AI Act proposal as to its consequences for the SELMA platform in this memo.

3.2. The need for addressing algorithmic bias in machine and deep learning

Bias in algorithms has come at the forefront of attention in the field of AI. Such bias is detrimental if it affects fundamental rights, such as when it leads to unjust discrimination, or freedom of expression, and has many other negative effects.[3] One of the main issues is that when using AI in combination with deep learning, the bias, if any, is fed back in the whole learning mechanism ('feedback loops').[4] Another concern is the diversity in languages which may be not well matching with using existing NLP technology[5] and with any machine learning at all.[6]

Bias is also a threat for the SELMA development and the AI used.[7] While SELMA may have as objective unbiased and diverse news compilations, not only the origin, but also the search and selection terms and the AI algorithm will be of key importance. This may require attention in different areas.

>>>>> For example, when searching for news on particular (political) movements or organisations, searching with tags such as the name of the group and/or a particular label, (e.g. 'terrorist' group), is likely to find and present news items from particular parts of the world only, seeing these organisations as linked to terrorism (e.g., the West), while this may not be considered as such in specific countries or other parts of the world (e.g., the East) and SELMA hence risking of incorporating 'biased' or one-sided news items and opinions/views, excluding reports about the same organization not seeing them as a terrorist organization.[8]

---

[2] AI is defined in the initial proposal as 'artificial intelligence systems' (AI systems)' which means 'software that is developed with one or more of the techniques and approaches listed' in the Annex I and 'can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with' (Art. 3(1) AI Act Proposal. This definition has been further modified and specified during the further negotiations.

[3] See FRA, *Bias in algorithms – Artificial intelligence and discrimination*, December 2022, 106 p. (FRA, Bias, 2022).

[4] Such feedback loops hence are critical for prediction algorithms.

[5] FRA, Bias, 2022, p. 78.

[6] S. Wachter, B. Mittelstadt and C. Russell, 'Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI', *West Virginia Law Review*, vol 123, No 3.

[7] See for an analysis of automated detecting hate-speech and the role of terms used : Ibid. p. 62.

[8] See on the same topic, also D 8.1, Ethics report, p. 18.

2.0

Various *mitigation strategies* have been discussed in various fora. One includes working and reviewing the *classifiers used by/for the model*. But since text searches are important for SELMA, control over and review of the *search terms*, and used language models, as good as possible, amongst other means, may remain important ways to detect any possible risk for bias, as well as review of the outcomes. Changing the selection/approach to underlying training data/channels is another common approach.

>>>>> Much of the mitigation hence requires specific *attention and human intervention*[9] in the controlling mechanisms. Fairness as such cannot be automated.[10]

Overall, a due analysis of bias risks, and this from various viewpoints (e.g., discrimination, freedom of expression, ….), an analysis of the AI model, a strategy, a detection and correction mechanisms, training of people (and of the algorithms) and testing are all essential components of *addressing (automated) bias in an effective manner*.

It is evident that SELMA needs to develop strategies, which need to be tested and evaluated throughout the project and later on.

4. EU Data Protection regulation : Applicable to personal data activities in the EU

4.1. General

Briefly summarized, the EU personal data protection rules and obligations apply if the data processing activities concern (1) personal data and (2) the processing takes place in the context of activities of an establishment in the EU.[11] Both criteria are fulfilled. Indeed, SELMA is using texts, spoken or written, originating from (professional) news contributors (i.e., journalists, podcast makers, etc) for detecting and combining content and news trends, in various languages. For this purpose(s), information relating to identified individuals is automatically processed, by SELMA and its partners established in the EU for its research activities, and hence the General Data Protection Regulation *will apply* (from a territorial and material scope perspective). At the same time, it remains important to qualify which information on the platform would be personal data, e.g., would the data identifiers/internal data format[12] contain information relating to individuals/sources and hence possibly qualify as personal data ? We understand that the entity identifiers (see D. 6.3, sect. 4.4) would refer to individuals. A second question is the extent to which any exceptions apply, notably those relating to research (in relation to the development of the platform by the SELMA team) and the journalism exception (in relation to the UC1 and UC2 as deployed). Although the obligations of

[9] FRA, Bias, 2022, p. 3; see also G. Sartor and A. Loreggia, *The impact of algorithms for online content filtering or moderation,* Study for EP, 2020, p. 23.
[10] See also A. Balayn and S. Gürses, *Beyond Debiasing : Regulating AI and its inequalities*, EDRI, 155 p.; S. Wachter, e.a., 'Why Fairness cannot be automated: Bridging the gap between EU Non-discriminatory law and AI', in CLSR, 105567.
[11] See Art. 3 GDPR.
[12] See D. 6.3 p. 23.

users as controllers would not be the responsibility of the SELMA team, unless in case of joint-controllership as SELMA designs the tool, it would seem good practice to consider how the tools will be used by others. The difficulty in assessing the extent to which the journalism exception applies relates to the fact that under Art 85 GDPR (which is similar to the analogous provision in the Data Protection Directive), the detail of the exception lies to a large extent with the Member States opening up the possibility of different standards[13] applying to journalistic data controllers depending on the Member State in which they, respectively, are established.

4.2 Transfer of personal data

The GDPR also contains specific rules for personal data transfers outside the EU. If sent to countries where there is no 'adequate level of protection', specific measures need to be taken to ensure essential equivalent protection.[14] The data management shall take into account the flows of data outside the EU, both during the testing, development and demonstration phases as during the production phase, and foresee appropriate safeguards in case data are sent to third countries not guaranteeing an 'adequate level of protection', including legal instruments as to protect any personal data processed by SELMA and transferred outside the EU, including for on-ward transfers. This includes transfers to 'processors', and use of the cloud (see D.6.3, p. 27). For any cloud services, use of 'local' EU cloud is for this reason highly advised. 'Incoming' personal data (sent into the EU, e.g., from Brazil to Germany) would in principle not be subject to the transfer rules briefly summarized above.

5. Monitio: The use of 'sensitive' criteria for filtering and grouping content

Another parameter and legal criterion is the restriction as relating to the use of 'sensitive' data as set forth in the General Data Protection Regulation. This kind of special categories of data is explicitly listed in Art. 9.1 GDPR. It states that the processing of these data is prohibited, unless an exception (of the list of Art. 9.2 GDPR) applies and there is a legal basis for doing so.[15]

We understand that SELMA could review, collect and use such additional data, e.g., relating to the journalist or person being reported about, which would fall in the list of 'sensitive data', such as e.g., ethnic origin, but also gender[16], and which data was collected from open sources such as Wikidata (though we note that not all diversity monitoring searches would implicate personal data). Hence, for the (re) processing of e.g., the Wikidata information, in particular for diversity of news source

[13] Article 29 Working Party, *Data Protection Law and the Media*, Recommendation 1/97, 25 February 1997, pp 6-7, available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/1997/wp1_en.pdf
[14] See Chapter V GDPR.
[15] The list of legal grounds is mentioned in Art. 6 GDPR. See also below.
[16] While gender is strictly speaking not listed as 'sensitive data' in Art. 9 GDPR, reference to male/female could remain sensitive, as (i) it could lead to inferred information about e.g. sexual orientation and (ii) also be used for discrimination, which is the reason of the list of the special categories of data, needing special protection.

monitoring ('diversity monitoring'[17]), SELMA shall rely for the processing of particular 'sensitive' data *on an appropriate basis and exemption of Art. 9.2 GDPR*.

In our view, (subject to the possible availability of journalism exceptions) only 'necessity for reasons of substantial public interest' (Art. 9.2.g GDPR) could possibly apply for the processing of particular 'sensitive data' of journalists or persons being reported about, provided SELMA can motivate this (1) on 'proportionate' (national or EU) legislation in relation to the aim pursued and (2) SELMA takes suitable and specific measures to protect the fundamental rights of the individuals whose data are processed ('data subjects'). SELMA shall hence *identify any such relevant legislation* and assess whether the other conditions (proportionality of such legislation and safeguards) are met.

Members States (MS) have the possibility to deviate of some of the principles and obligations of the GDPR for *processing for journalistic purposes* (see Art. 85 GDPR), in the objective for reconciling the right to data protection and the right to information and freedom of expression. In such case, MS need to notify such to the EU Commission. We noted that Germany has notified various deviations based on Art. 85.3 GDPR .

>>>>> SELMA shall hence verify if any of such national deviations could apply to their processing of 'sensitive' information for (i) research and development and (ii) for later use as a news content provider platform. In addition, it seems to us that also for (all) the other EU countries where the news platform and its content would be made available, it would be necessary, absent EU legislation allowing the use of such 'sensitive' data, such as gender for (news) diversity purposes, to check for the appropriate exemption/legal basis for the specific category of data processed to meet diversity goals. This hence needs to be further reviewed. Nonetheless it seems unlikely that when this tool is used - even by journalistic bodies - for management purposes (monitoring of diversity) that would constitute a journalistic purpose and therefore any relaxation of data protection rules for journalists would not apply here.

>>>>> In any case, where by the processing of other 'sensitive' information, such as race or ethnic origin, religion, medical condition,[18] the proportionality of the necessity to process such information to reach the objective of diversity, is not present, we *advise to omit the collection and processing of such additional categories of sensitive data for diversity purposes (*this was also discussed during our online meeting of 28 February 2023) as the approach that SELMA would take during development and later use).

Absent a legal basis and an appropriate exemption, the processing of the special categories of data relating to an individual is forbidden (Art. 9.1 GDPR).

>>>>> The above, however, may not apply in case of ranking and selection of news articles based *on the topic regarded as sensitive* and while this topic could not be related to specific individuals/journalists. However, this may be a difficult distinction to make, as topics and journalists may be closely connected.

[17] About diversity monitoring in the employment context, see, e.g., Claeys & Engels, Which data can companies collect as part of a diversity policy ?, 4.2.2021, available at https://www.claeysengels.be/en-gb/news-events/which-data-can-companies-collect-part-diversity-policy

[18] See for the categories, D 6.3, p. 25. See also the Ethics Check Report, 2022,

2.0

6. The use of Voice and Speech technologies for content production

SELMA also processes voice and speech content with require special attention and adequate safeguards.

6.1 Speech and Text

As most data processed in speech and text is usually so-called unstructured data, applying the GDPR principles becomes challenging.

> >>>>> For example, the notion of 'data subject' in speech and text is already ambiguous, as it may refer (1) to the individual who produced the data ('identified contributors'), (2) the individual(s) mentioned in the text ('individuals mentioned') and (3) other individuals whose identity can be inferred ('inferred identities'). Furthermore, attributes of data subjects, especially identified contributors, are embedded in the data and could be extracted.[19]

Any information relating to these data subjects requires adherence to the data protection regulations. This may raise particular challenges depending on the 'category' of data subject mentioned above. While the processing of text and speech of identified contributors may be rather straightforward, individuals mentioned in text could raise questions as of when this information is personal data.[20]

Another difficulty is that attributes used to link individuals in machine learning systems may not be necessarily explainable.[21]

6.2 Voice and other biometric information

As to the use and processing of voice and voice snippets, other challenges arise. As speech, if spoken live, contains information about characteristics of an individual by the voice, and if such information is processed, e.g. for the transformation of speech to text, SELMA shall assess to what this would fall under 'processing of biometric data' and which specific restrictions apply. The definition of biometric data in the GDPR is however rather specific. It requires the processing of personal data resulting from '*specific technical processing* relating to the physical, physiological or behavioural characteristics of a person' allowing or confirming the unique identification of a person.[22] Only if SELMA's processing of voice would qualify as the above, the processing of the information would be processing of biometric

[19] See Privacy in Speech and Language Technology, S. Fischer-Hübner, e.a. (eds), Dagstuhl Reports, 2023, Vol. 12, Issue 8, pp. 60-102 ('Speech and Language Technology, Dagstuhl Report 2023'), p. 16.
[20] See e.g., General Court, *OC v European Commission*, T-384/20, C-479/22 P, 4 May 2022, ECLI:EU:T:2022:273, available in French.
[21] *Ibid.*
[22] See Art. 4 (14) GDPR.

information. Furthermore, (only) if the biometric data are processed for 'uniquely identifying', this would be forbidden, unless one of the exemptions would apply.[23]

>>>>> SELMA shall review, eg in its DPIA (see below) to what extent biometric data is processed and if so, if the prohibition of art. 9 could apply, safe exceptions.

>>>>> Furthermore, specific risks of processing of such information, such as facial images and voice, shall be considered (e.g., scraping, misuse for deepfakes, etc) and limited.

7. Research and Development in SELMA : data minimization requirement

SELMA should have a clear overview of all (public) training and test data sets used or created during research and development[24] and assess the implications of data protection requirements (purposes specification (e.g., if it will be shared with the wider research community), legal basis (see also below), transparency, accuracy, …).

>>>>> A table with mentioning of the resources and compliance would be a useful way for a quick overview, e.g. in the record to be kept.

To the extent SELMA will process personal data, it shall adhere to data minimization, in particular during the research and development phase, but also thereafter. This means that only the data that is *necessary* for the research, for gathering the content and for any training of the filters/algorithms, shall be made collected, processed and made further available. This shall require an *assessment* for each of the collection and processing activities of particular data, such as to whether names, countries, etc are required. This approach seems to be followed as set out in D.6.3, p. 9.

Additional precautions need to be taken for any other user participant data involved in requirement setting and evaluation as to process a minimum of information as necessary (anonymization, yet pseudonymization if the research goals cannot be reached).[25] This applies also to the 'user feedback' collected by SELMA and shared and used for SELMA improvements.[26] Specific guidance as to anonymization techniques, including generalization and randomization, were provided by the group of data protection authorities in 2014.[27]

[23] Art. 9 GDPR.
[24] See also D 8.1, Ethics report, p. 17 and D6.3.
[25] See Art. 89 GDPR.
[26] See D. 6.3, p. 21.
[27] See Art. 29 Working Party, Opinion 5/2014 on Anonymisation techniques, 2014, 37 p.

2.0

8. Other 'privacy-by-design' strategies

In addition to data minimization, other strategies to increase privacy and data protection where required should be considered. 'Data protection by design' is also an obligation.[28] These include *hiding* information, limiting access to authorized (licensed) users only, encryption where appropriate, *separating* information (e.g., by storing (meta)data in a distributed manner), and *aggregating.*[29] A description of these technologies are part of the DPIA (see below). They should be designed and decided before the start of the processing and regularly evaluated during the use of the platforms and use cases.

>>>>> SELMA should consider these strategies, which are mainly technical, carefully, in view of the risks for data subjects involved (e.g., journalists in war territories, reporting on delicate issues), taking into account the technologies that exist.

>>>>> Special attention needs to be given to tools and technology for excluding, filtering and controlling the news input from published news sources only, excluding social media user-generated content. In this context, the use of Twitter and YouTube feeds shall also be carefully assessed and addressed.

>>>> In terms of the monitoring tools, these could be restricted to those who have need to use the tools rather than making them available to all licensed users and logs kept of the searches using that tool (also of relevance for accountability and transparency); moreover limitations could be placed on the granularity of the searches so as to limit if not exclude the possibility of plurality searches identifying specific individuals.

These technical measures also come on top of the organizational measures as imposed by data protection legislation, such as *informing* and providing *transparency*, allowing data subjects to *control*, applying and *enforcing* data protection and *demonstrating* compliance.

9. Other Important GDPR principles and obligations: a selection

Other important elements for the (1) research and the development and (2) the later use of the platforms are briefly mentioned below.

[28] Art. Art. 25 GDPR.
[29] See J.-H. Hoepman,'Privacy design strategies' in *IFIP International Information Security Conference,* pp. 446-459, Springer, 2014, pp. 446-459.

8

2.0

9.1 Transparency

Information about the data processing in SELMA on the websites. One of the core principles of data protection is *transparency* towards the individuals whose personal data are collected.[30] This includes information relating to the content information about professionals whose bibliographical data is collected and processed, including for the use case UC1. In the latter, information from other sources, from Wikidata is used as well.

In order to adhere to this principle, and subject to national law variants[31], the information obligation is of key important (Art. 13 and 14 GDPR). In case the information is not directly collected from these individuals, the info should be provided within a reasonable period after having obtained the data and at least within one month (Art. 14.3.(a) GDPR) . In the case at hand, SELMA could also state that the data is disclosed, in which case the timing of the information would be the first disclosure. As in the present case, it would be impossible or involve a disproportionate effort, SELMA could rely on an exception for this information to be provided to the individuals.[32] In case of research and development, and provided suitable safeguards (such as for data minimization) are taken, this exception could be invoked. At the same time, and conforming guidelines about the information and transparency principle, it would be advised *to provide information about the project on the website of the controllers* (e.g., coordinator of the project, Deutsche Welle , …) that this project and research is ongoing and content from various sources, including some personal information of producers of such content, is processed. This would be applicable for (1) R&D phase and (2) operational use.

The transparency and information obligation also applies for UCO (in particular for the public demonstration platform).

9.2 Profiling and automated individual decision-making

Particular attention is needed to assess the effects of the processing to individual news makers, including e.g., journalists, as well as to the individuals, subject of the news stories, as to what they become exposed to profiling and/or increased exposure based on reporting news and/or automated decision-making.[33] The central storage of information about these individuals also entails risks.

---

[30] See, for those core principle, Art. 5 GDPR. See also Art. 12 GDPR.

[31] See Art. 85 GDPR. This article allows for notification by members states of exemptions or derogations from various principles and obligations under GDPR for purposes of reconciling freedom of expressing, including by processing for journalistic purposes, with data protection. These notifications of all the members states, including of e.g., Germany, can be consulted on the EU Commission's website at https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu/eu-member-states-notifications-european-commission-under-gdpr_en

[32] See Art. 14.5 GDPR.

[33] Important is also to note that no special categories of data may be used.

2.0

Automated decision based on 'AI' shall overall also be 'explainable'. To the extent the processing is based on search terms, rather than black box processing[34], as we discussed, this needs to be described and assessed in the DPIA and explained why this may not lead to increased risks.

10. Roles of SELMA partners and of the users of the SELMA platform

10.1 The SELMA partners during the development

D6.3 states that all partners to the project are joint-controllers. In this case, *an appropriate joint-controllership agreement* shall be signed amongst partners.

10.2 Operational phase: the users of the SELMA platform

Depending on possible varying roles of users of the SELMA platform, SELMA shall reflect on these and determine any varying authorizations model(s) for these various users, including to protect the information. Largely, two types of users may be envisaged: are these users *merely consulting* the platform for information (without storage, editing and other data processing rights, etc) or are they (in addition) allowed to use the platform for organizing their own content, making own selections and compilations of information and the posting of own content.

>>>>> In the latter case, SELMA should evaluate, determine and make clear if these users would also become responsible, including for data protection purposes, and hence become (co)controller. This responsibility for any personal data processing activities should also be reflected in the 'terms of use' or licence for the platform. The possibility of restricting the licence has been noted by SELMA but needs to be considered further. Transparency and information about the collection of user data shall also be provided. This applies also to e.g. to the 'user feedback' collected by SELMA and used for SELMA improvements.

11. Legal basis

From data protection regulation purposes, each processing of personal data *requires a legal basis* ('lawfulness'). These are set out in Art. 6 GDPR. SELMA needs to specify, including in its data records, for each (group of the) processing activities for specific purposes, which legal basis is appropriate. In

[34] E.g., when the deep learning processing would be based on e.g., (machine learned) inferred attributes (from text or speech).

10

case SELMA would rely on legitimate interests, this shall always be balanced with the fundamental rights and freedoms of data subjects involved.

>>>>> For example, for the (re) processing of the Wikidata information, such as gender, in particular for diversity monitoring, SELMA shall rely on a (new) appropriate legal basis, in addition to purpose specification, especially during operational use, but also during the research and development.

>>>>> For special categories of data (so-called 'sensitive' data), including eg gender, the processing shall also rely on one of the (ten) exceptions set out in Art. 9 GDPR (see above).

In addition to the need of an appropriate legal basis, and qualification for an exemption if 'sensitive' data are processed, *all the other principles shall be respected*, including data minimization (art. 5.1.c GDPR) and accuracy (art. 5.1.d GDPR).

12. Rights of data subjects

The GDPR confers specific rights to the data subjects when their personal data is processed. These rights are set out in Art. 12-22 GDPR. Which responsible entity will be the contact of the data subjects (e.g. the journalists whose news items are processed, users of the platform, …) is also one of the core elements in the joint-controllership between partners to SELMA.

13. DPIA

A data protection impact assessment is required if e.g. 'new technologies' are used, taking into account the nature, scope, context and purposes of the processing, likely to result in a high risk. Advice of the DPO is needed if such is appointed. Because of the above, but also because of some 'systematic monitoring ' activities, we deem it important to conduct such DPA. All elements required are further described in Art. 35 GDPR. This assessment should also include an assessment of fundamental rights and freedoms.