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)	Olga Kisselmann (DW), Peggy van der Kreeft (DW), Kay Macquarrie (DW)
Contributors	Joscha Rieber (FhG), João Prieto (Prib), Yannick Estève (LIA), Normunds Gruzitis (IMCS)
Version	2.0
Contractual Date	31 March 2021
Delivery Date	~~29 March 2021,~~ 18 November 2022
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

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.

Thus, this involves the provision of media monitoring capabilities based on data from publicly available media streams. We will 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

1.Introduction

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

- Protection of personal data
- Copyright protection
- Ethical implications of SELMA technologies
- General ethical concerns related to open-source release of novel analytics technologies
- The social impact of automation
- Sex and gender balance

We will describe in general terms the consortium's mitigation strategies for central aspects that will be covered in this project.

2. 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 relating to the SELMA platform itself.

Task T6.3 will develop and implement a strategy for all data management and protection issues in the project. This plan will consider the data collected in SELMA, but will also set 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 will be supported by the management processes put into practice in WP7.

2.1 Data relating 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 labeling 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.

Potential collection of personal data and its processing during web crawling, analysis of social media and multimodal data will also be considered. The consortium will discuss data privacy in detail and take all measures necessary to adhere to ethical standards. Details on the recruitment of participants and personal data protection are detailed below.

2.2 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 their 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 by biaswho do not have sufficient understanding of English.

2.3 Protection of data relating to the SELMA platform itself

SELMA technologies generally do not focus on acquiring and processing personal data, 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 a dataset is (e.g., in terms of gender, ethnicity and disability). The aim is not to profile individuals, but to get an inisght about distributions (e.g., How often was an article published by a man, how often by a woman). SELMA will apply methods including regular of the Data Management Committee

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

for the protection of such personal data, in particular regarding the gathering, identification, storage, retention and the destruction of (personal and other) data.

2.4 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 and social media. However, only publicly available news and published media content 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 will **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 will be 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 will **follow the formal procedures** that are explicitly defined within each partner organization to protect the anonymity of data that is shared among the consortium.

Only data necessary to the completion of the project will be stored. Social media data will be restricted to news and published media and its numerical reach data. 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 will be centrally held by the project and will therefore be available for audit.

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

2.5 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 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 profiling information about people and organizations from published media content, and published media content published on social media; also, these activities apply to diversity labels from Wikidata, which are used to classify broadcast content. The ethical issues regarding this use of social media are the same as the general case, bringing no additional privacy issues.

If for any unforeseen reason the project needs to get additional data from social media, then the project will carefully address ethical issues relating to privacy issues relating to this. It is

[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

important that, throughout the project, we also develop an understanding of the impact the technologies we develop may have on people. We take precautions to protect regular social media users, by for example not including their comments and any other user-generated content into the analysis. We restrict it to certain personal indicators, such as numbers of likes, dislikes, and shares. 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, especially regarding social media. 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-based construction.

The consortium will not deal with images and videos contained in social media posts unless they came from media publishers. As in the general case awareness is raised of the risk that individuals might be identifiable through the media content.

3.Copyright protection

As we intend to process enormous amounts of media and other kind of content items, the consortium will furthermore need 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 the SELMA project, this directive had not been implemented into the laws of the member states. Nevertheless, as the implementation period will end 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

[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

implement the directive, it will be necessary to closely follow the legislative process and make appropriate adjustments throughout the course of the project.

4. 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 initial Data Management Plan includes the following aspects:

- A clear description of what research and innovation activities project data is used and a description 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.

5. Ethical implications of SELMA technologies

In this section possible ethical concerns related to individual components and/or technologies of the SELMA platform are presented.

5.1 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.

5.2 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.

5.3 Speaker diarization and speaker recognition

The SELMA platform will mainly be 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.

5.4 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 will leverage as much publicly available data as possible that is free for use and does not contain sensitive personal content. Additionally, the models will continuously be trained on crawled and partner-specific data provided for the project. So, we need to make sure that the vast amount of needed data is stored responsibly. We will 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 will 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.

5.5 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?

5.6 Named entity recognition and linking, topic labeling

Named entity recognition, linking and topic labeling are essential to the project. We will 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 will aim for highly accurate named entity disambiguation based on state-of-the-art technologies. Comprehensive accuracy evaluations will be 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 aware of the nature of algorithms and the potential ethical problems and threats that incur from their use.

5.7 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.

5.8 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, sexual orientation and religion. The aim is to get indications on how diverse a dataset is both in terms of quantitative and qualitative measurements. The application is foreseen to use NLP tasks (Named entity, speaker diarization, speaker recognition, topic labeling and topic clustering) and is based on established diversity categories such as gender, ethnicity and disability (cf. D1.1. Use Case Description, section “Diversity” and in the requirements list “Diversity Detection” (96-102)).

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.

5.9 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

6. 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 will provide support for low-resource languages and will allow 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.

7.Sex and Gender Balance

The project will ensure sex and gender balance in different ways:

It will take gender differences into account in market trend analysis, and in developing user scenarios, ensure 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 will 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.

8. 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 social media analysis. Privacy will be designed into the SELMA data management infrastructure which will be defined in D6.1 Initial Data Management Plan (M6) and refined in two following deliverables (M18, M36).
2. **Ethical implications of SELMA technologies and platform:** SELMA technologies have a direct impact on workflows and new ways of automation. Next to management and impact reports, ethical issues 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 will play 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 will be closely monitored and discussed during the project and will be an integral part of the data management, project management and evaluation reports.

9. Appendix

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. ☒