eloquence

D 8.2 Ethics Compliance Management Report

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

UKRI UK Research and Innovation

Grant Agreement Number	101135916	**Acronym**	ELOQUENCE
Full Title	Multilingual and Cross-cultural interactions for context-aware, and bias controlled dialogue systems for safety-critical applications		
Project Start Date	01/01/2024	**Duration**	36 months
Type of action	HORIZON Research and Innovation Actions		
Project URL	eloquenceai.eu		
Deliverable	D8.2 – Ethics Compliance Management Report		
Work Package	WP8 – Coordination & Project Management		
Date of Delivery	**Contractual**	30/04/2025	**Actual** 30/04/2025
Type	Report		
Lead Beneficiary	PN		
Author(s)/ Organisation(s)	Anastasiya Kiseleva (AK) and Aayushi Gupta (AG) from Privanova		
Contributor(s)			
Abstract	This Ethics Compliance Management Report outlines the strategic and operational mechanisms adopted by the ELOQUENCE project to ensure robust ethical oversight and compliance throughout its lifecycle. ELOQUENCE aims to develop state-of-the-art multilingual, multimodal, and context-aware conversational AI technologies designed for safety-critical and socially sensitive applications. In response to the unique ethical challenges posed by such technologies—including risks related to human involvement, personal data, and AI accountability—this report consolidates a comprehensive ethics management framework. Key highlights include the establishment and contributions of the Ethics Advisory Board (EAB), the development and implementation of the Ethics Compliance Management Process in ELOQUENCE, and the application of a human-rights-based evaluation methodology for AI systems. Drawing on insights from previous deliverables (notably D6.1 and D9.1), this document serves as both a reference for project partners and external stakeholders. It ensures that the ELOQUENCE work process towards the development and deployment of AI systems remain aligned with European values including fundamental rights.		

Document history

Version	Issue Date	Stage	Description	Contributor
0.1	25/02/2025		Draft synopsis	Aayushi Gupta, PN
0.2	04/04/2025		First complete draft of the deliverable	Anastasiya Kiseleva, PN
0.3	20/04/2025		Peer Review by EUI	Helga MolBaek-Steensig, EUI
0.4	23/04/2025		Peer Review by TID	Jordi Luque Serrano, TID
0.5	28/04/2025		Integration of feedback and final deliverable	Aayushi Gupta, Anastasiya Kiseleva, PN
1.0	29/04/2025		Final Quality check	TID

Dissemination level

PU – Public, fully open, e.g., web	✓
SEN – Sensitive, limited under the conditions of the Grant Agreement	
Classified R-UE/EU-R – EU RESTRICTED under the Commission Decision No2015/444	
Classified C-UE/EU-C – EU CONFIDENTIAL under the Commission Decision No2015/444	
Classified S-UE/EU-S – EU SECRET under the Commission Decision No2015/444	

eloquence

ELOQUENCE Consortium

Participant No.	Participant organisation name	Short name	Country	Role*
1	TELEFONICA INNOVACION DIGITAL SL	TInD	ES	COO
2	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	IT	BEN
3	BARCELONA SUPERCOMPUTING CENTER CENTRO NACIONAL DE SUPERCOMPUTACION	BSC	ES	BEN
4	FONDAZIONE BRUNO KESSLER	FBK	IT	BEN
5	UNIVERZITET U NOVOM SADU FAKULTET TEHNICKIH NAUKA	UNS	RS	BEN
6	EUROPEAN UNIVERSITY INSTITUTE	EUI	IT	BEN
7	VYSOKE UCENI TECHNICKE V BRNE	BUT	CZ	BEN
8	PRIVANOVA SAS	PN	SAS	BEN
9	INOSENS DOO NOVI SAD	INO	RS	BEN
10	TRANSFORMATION LIGHTHOUSE, POSLOVNO SVETOVANJE, D.O.O.	TL	SI	BEN
11	GRANTXPERT CONSULTING LIMITED	GX	CY	BEN
12	OMILIA MONOPROSOPI ETAIREIA PERIORISMENIS EFTHYNIS PAROXIS PLIROFORIKON, TILEPIKOINONIAKON KAI FONITIKON YPIRESION KAI SYSTIMATON	OM	EL	BEN
13	SYNELIXIS LYSEIS PLIROFORIKIS AUTOMATISMOU & TILEPIKOINONION ANONIMI ETAIRIA	SYN	EL	BEN
14	FONDATION DE L'INSTITUT DE RECHERCHE IDIAP	IDIAP	CH	AP
15	BRUNEL UNIVERSITY LONDON	BUL	UK	AP
16	UNIVERSITY OF ESSEX	UESSEX	UK	AP

Role: COO-Coordinator; BEN-Beneficiary; AE-Affiliated Entity; AP-Associated Partner

Table of Contents

List of figures

ABBREVIATIONS AND ACRONYMS	
AI	Artificial Intelligence
GenAI	Generative Artificial Intelligence
G2C	Government to Citizen
B2C	Business to Consumer
C2C	Citizen to Citizen
EAB	Ethics Advisory Board
EU	European Union
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
GDPR	General Data Protection Regulation
CJEU	Court of Justice for the European Union
OECD	Organisation for Economic Co-operation and Development
GPAI	General Purpose AI
FRIA	Fundamental Rights Impact Assessment
XAI	Explainable AI
LLM	Large Language Model
NDA	Non-Disclosure Agreement

Executive Summary

This Ethics Compliance Report sets out a preliminary overview of the aspects relative to ethics in the project. It presents an overall roadmap on how to manage ethical issues that arise during the implementation of the project by developing the Ethics Compliance Management Process in ELOQUENCE. The document, also, outlines the establishment of the Ethics Advisory Board (EAB) and delineates upcoming actions and responsibilities essential for meeting ethics considerations. Through the establishment of an EAB and the enactment of thorough ethical monitoring policies and frameworks, the consortium will be poised to effectively tackle any ethical and legal challenges that may emerge during the project's implementation.

1. Introduction

1.1. Purpose

The purpose of this Ethics Compliance Report is to present an initial framework for managing, monitoring, and addressing ethical considerations throughout the development and implementation of the ELOQUENCE project. As ELOQUENCE seeks to advance state-of-the-art conversational AI technologies in multilingual, safety-critical, and privacy-sensitive applications, it is essential to prioritize ethical integrity in every aspect of the project's lifecycle.

This document outlines key principles, compliance strategies, and operational methodologies that ensure the project aligns with the highest ethical standards set forth by the European Union, particularly in relation to data privacy, user safety, inclusivity, and societal impact. By adopting a proactive approach to ethics management, ELOQUENCE demonstrates a commitment to fostering responsible research and innovation while addressing the complex ethical challenges associated with AI technologies.

1.2. Relevance to other work packages and tasks

Ethics Compliance management is an ongoing and proactive process in the ELOQUENCE project. It began in the earliest stages of the project with outlining a comprehensive plan for fulfilment of ethics post-grant requirements, which was summarised in the Deliverable 9.1. This deliverable provided an action plan to adequately address ethics requirements in the ELOQUENCE project (Ethics Appraisal Procedure) and specified the key areas of consideration: Humans; Personal data; Non-EU Countries; Environment, Health and Safety, and Artificial Intelligence. This D8.2 "Ethics Compliance Report" is the continuation of the ethics compliance management process, following the approach and structure outlined in the Deliverable 9.1.

In addition, this D8.2 "Ethics Compliance Report I" is closely linked to WP6 and the D6.1 "Report on linguistic expression respectful of EU values." As was outlined in Deliverable 9.1, one of the key areas of ethics consideration in ELOQUENCE project is Artificial Intelligence. In turn, the key focus of the D6.1 "Report on linguistic expression respectful of EU values" was an exploration of contemporary law and literature conceptualizations of what AI is compatible with human rights and EU values means and what potential pitfalls are to be avoided. The D6.1 provided a methodology for securing such compatibility through multidisciplinary assessment of alignment of emerging ELOQUENCE AI technology with EU-values. This methodology is integrated into the relevant section of this D8.2 focusing on the ethics compliance management concerning Artificial Intelligence.

1.3. Scope

ELOQUENCE operates at the intersection of advanced conversational AI, human-computer interaction, and multilingual processing. The ethical considerations in this context are multifaceted, encompassing issues related to:

- **Humans**: Ensuring compliance with ethics requirements while engaging external stakeholders and participants in the ELOQUENCE research activities such as implementation of project pilots, dissemination and exploitation activities.
- **Personal data**: Ensuring strong privacy and data protection and compliance with applicable laws, enabling stakeholders to understand the entire lifecycle of personal data on the one hand and the legal challenges on the other.
- **Non-EU countries**: Ensuring that transfers and processing of personal data concerning partners based outside the EU are carried out only for the purposes of the project realisation and complaint with EU law.
- **Environment, Health and Safety:** Ensuring that computational frameworks prioritize energy efficiency and sustainability and that the risks to health and safety are prevented, minimised or mitigated during the development of AI applications and relevant research processes. The latter is closely linked to the next area of ethical consideration: Artificial Intelligence.
- **Artificial Intelligence:** Committing to respecting ethical principles during the development of AI in the ELOQUENCE project and its foreseen deployment.

This report provides guidance on how ethical concerns should be identified, tracked, and mitigated throughout the project's duration, including the establishment of an independent Ethics Advisory Board (EAB) and the development of an Ethics Compliance Log.

1.4. Structure of the Deliverable

The document is structured as follows:

- **Section 2:** Ethics Compliance Procedures and Mechanisms present the methodology for managing and monitoring ethical compliance, including the use of an Ethics Compliance Log to document and address potential risks.
- **Section 3:** Ethical Considerations in ELOQUENCE outlines key ethical issues identified in the project in the areas specified above: Humans; Personal data; Non-EU Countries; Environment, Health and Safety; Artificial Intelligence.
- **Section 4:** Ethics Advisory Board (EAB) describes the structure, roles, and engagement strategy of the EAB, which provides independent oversight and guidance on ethical matters. The section specifies the progress of the EAB activities and planned course of action as the project progresses.
- **Section 5:** Conclusion summarizes the project's commitment to ethical research and innovation.

This report is intended to serve as a living document, evolving as the project progresses to accommodate emerging ethical considerations and best practices. As the project progresses, the updates of the ethics compliance management and the actions taken will be provided in the D8.3 "Ethics Compliance Management Report II" (M34). Through this proactive and structured approach, ELOQUENCE aims to uphold transparency, inclusivity, and societal responsibility in the development and deployment of conversational AI technologies.

2. Ethics Compliance Procedures and Mechanisms

Ethics compliance management is an ongoing process encompassing all the stages of the ELOQUENCE project. The overall Ethics Appraisal Procedure to be followed by the consortium was outlined in the Deliverable 9.1 (Figure 1).

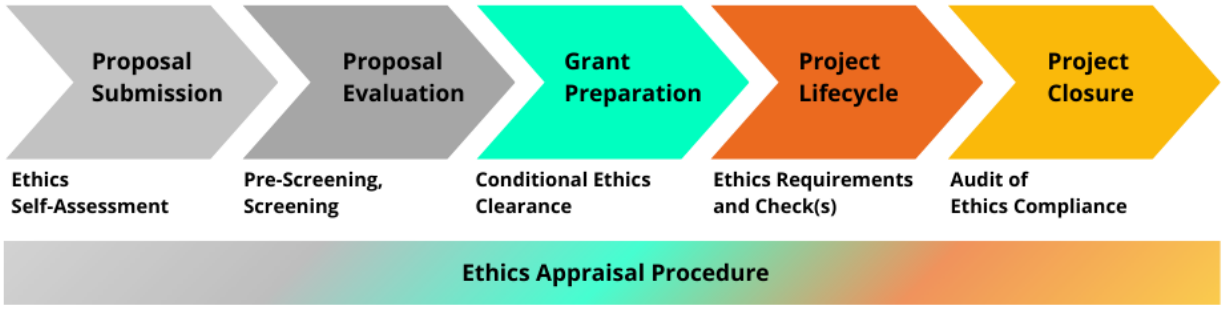

Figure 1. Ethics Appraisal Procedure

As the project is currently in its active phase, this section provides the strategy for ethics compliance management in the fourth stage (Project Lifecycle). It provides a structured and proactive approach (Figure 2) to managing ethical compliance. The methodology includes the following steps:

- Identification of Ethical Issues: Continuous monitoring to identify emerging ethical concerns.
- Risk Assessment: Evaluating the potential impact of identified ethical risks on project activities and outcomes.
- Mitigation Strategies: Developing and implementing measures to address identified risks.
- Documentation: Maintaining detailed records of ethical issues and mitigation actions

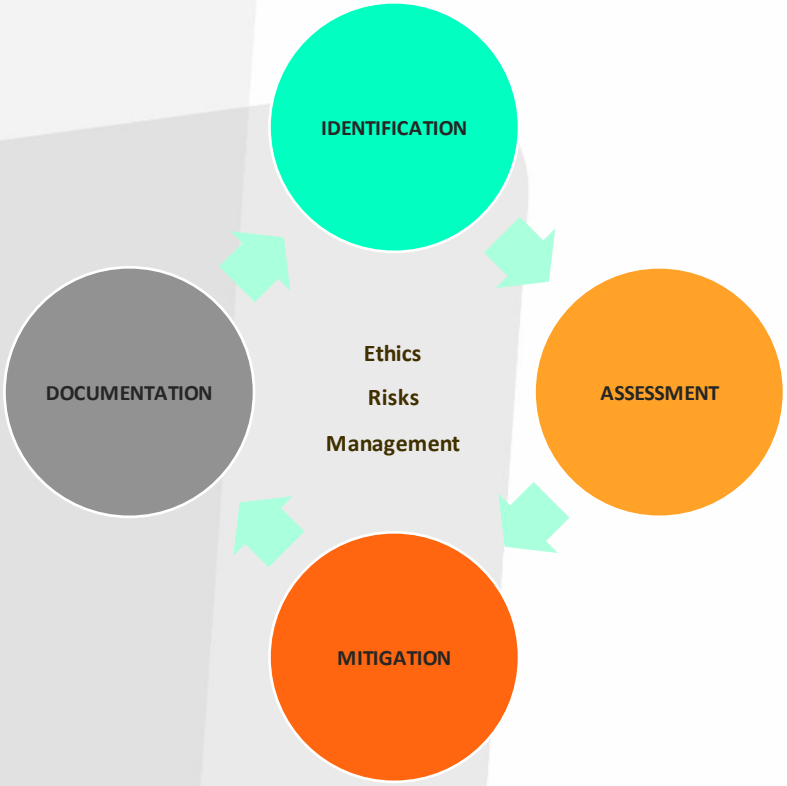

Figure 2. Ethics Compliance Management Process in ELOQUENCE

These steps are performed both at the level of internal control (by project partners) and external control (via the Ethics Advisory Board). While the details of the ethics management procedure are further outlined in the relevant sections (Sections 4 and 0 respectfully), this Section scrutinises the relevant procedures and mechanisms for each of the steps.

2.1. Identification of Ethical Issues

Ethical issue identification in ELOQUENCE is a continuous, iterative process carried out at both strategic and operational levels. This approach acknowledges that ethical concerns are not static or fully predictable at project inception but may evolve dynamically as new technologies, methodologies, and stakeholder interactions emerge.

To ensure that ethics is not treated as an afterthought, ethical reflection is embedded from the earliest design stages of each research activity and pilots. During project meetings, deliverable reviews, and interactions with external stakeholders or research participants, partners are encouraged to raise any newly identified or anticipated ethical concerns. These issues can arise in relation to AI model behaviour, data processing practices, language use in human-AI interaction, or operational decisions that may affect public trust or rights and interested of affected research participants. In addition, meetings of the Ethics Advisory Board and contributions from its members during the meeting and beyond (as reviewers of the deliverables) serve as a crucial mechanism for recognising the potential areas of ethical concerns in the project.

As the ethics management process started at the earliest stages of the project, most of the relevant concerns and issues were outlined in the Deliverable 9.1. The areas of considerations outlined in D9.1 included: ethics procedures to involve humans in the project's activities ensuring their voluntary participation and proper informing; compliance with privacy and data protection requirements; participation of non-EU partners; identification and mitigation of risks to health, safety and environment; trustworthy development and deployment if AI. These areas of considerations are also translated into the structure of this Chapter.In addition, Deliverable D6.1 "Report on linguistic expression respectful of EU values" contributed to understanding of ethical concerns for developing and deployment of AI that respect human rights and EU values. This deliverable serves for re-identification of other issues or re-assessment and mitigation of the previously defined, when applicable.

2.2. Risks Assessment

Once potential ethical issues are identified, a systematic risk assessment process is initiated. This process aims to evaluate the likelihood and potential impact of each identified issue on individuals, communities, regulatory compliance, and the broader societal perception of the ELOQUENCE project. Risk assessment is performed collaboratively between the concerned project partners, with review support from the Ethics Advisory Board (EAB). Each issue is assessed using the following criteria: Likelihood of occurrence (How probable is it that the issue will materialize?) and Severity of impact (What is the extent of the potential harm (e.g., breach of rights, data misuse, social bias)? Risks are then classified into low, medium, or high categories.

2.3. Mitigation Strategies

Following risk assessment, the consortium applies a preventive and responsive approach to ethics risk mitigation. For each identified risk, the following types of mitigation strategies are considered:

- Preventive measures: These are proactive actions aimed at eliminating or reducing the probability of the risk materializing. Examples include data minimization, anonymization, and inclusion of ethical constraints in AI development pipelines.
- Corrective measures: These are responsive actions taken to reduce the impact if a risk occurs. This may involve adjusting consent procedures, revising data processing protocols, or halting specific activities.
- Compensatory actions: In cases where residual risk remains, additional measures such as stakeholder consultation, external expert review, or policy revision may be implemented.

All mitigation strategies are developed in coordination with the task leaders, technical developers, and, where appropriate, the EAB. Responsibilities for implementation and follow-up are clearly assigned, and timelines for re-assessment are established. The dynamic nature of the project and AI technologies necessitates regular re-evaluation and adjustment of these strategies.

2.4. Reporting and Documentation

Robust documentation and transparent reporting are essential pillars of ethical compliance in ELOQUENCE. To ensure traceability, accountability, and internal learning, the following practices are implemented:

- Deliverable Review Logs: Ethics feedback and recommendations provided during EAB reviews of deliverables are recorded and tracked, ensuring follow-up action by responsible partners.
- Meeting Minutes and Issue Registers: Project meetings that include ethical discussions generate structured notes, recording concerns raised, decisions made, and items for further evaluation.
- Confidential Records Management: Personal data and sensitive information collected during ethical evaluations (e.g., participant feedback, consent forms) are stored securely in accordance with GDPR and access-controlled within the consortium.
- Integration into Reporting: Ethics-related actions and updates are incorporated into official project reporting cycles, including periodic and final reports to the European Commission.

3. Ethical Considerations in ELOQUENCE

3.1. Humans

As outlined in Deliverable 9.1, "Various activities are carried out as part of the ELOQUENCE project. These activities also involve external stakeholders and research participants. In order to successfully carry out the use cases and the project pilots, it is necessary to collect information resulting from the involvement of external stakeholders and participants. This information may include personal data. External stakeholders and participants refer to individuals who are not members/employees of the project partners and their respective organizations."

The procedures for involvement of external stakeholders and research participants in the ELOQUENCE project, including the activities for the recruitment and participation of individuals in the project's pilots, are scrutinised in Deliverable 9.1. To avoid duplications, this section describes only the updates to the described activities that influence the ethics compliance management.

The risks that can arise out of the involvement of research participants are various: 1. Inadequate informed consent; 2. Privacy breaches and data security; 3. Psychological discomfort or harm; 4. Power imbalances and pressure to participate; 5. Bias in recruitment and participation; 6. Misuse of personal data in dissemination activities. As some of these risks concern privacy and data protection (risk 1; risk 2; risk 6); they will be scrutinized in the next section. This section below represents the ethics management strategy for the remaining potential concerns.

- **Risk: Psychological Discomfort of Harm**
  ***Description*:** Participants of pilots, for example of pilot 2 (mainly in the testing phases) may be exposed to socially sensitive or emotionally charged topics (biases, discrimination, cultural/gender identity), which could cause discomfort.
  Assessment:
    - **Likelihood:** Low to Medium
    - **Impact:** Medium

  Mitigation Strategy:

    - Frame dialogues as hypothetical scenarios rather than personal evaluations
    - Clearly state in the consent form that sensitive topics may arise, and participation is voluntary
    - Provide debriefing sessions after participation to address any concerns
    - Offer access to support resources (e.g., institutional counseling if discomfort arises)

  Reporting and Documentation:

    - Consent forms are documented in Deliverable 9.1 and updated in this deliverable (see section 3.2)
    - Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation."
- **Risk: Power Imbalances and Pressure to Participate**

  ***Description*:** Participants recruited from staff or students (for example, in pilots 1, or pilot 2 or pilot 4) may feel pressured to participate due to hierarchical structures or academic obligations.

  Assessment:
    - **Likelihood:** Medium
    - **Impact:** Medium

  Mitigation Strategy:

    - Explicitly clarify that participation is voluntary and will not influence employment, academic evaluation, or access to resources
    - Avoid recruitment by direct supervisors; designate neutral third parties to handle enrollment and consent
    - Ensure feedback mechanisms where participants can raise concerns or opt-out

  Reporting and Documentation:

- Consent forms are documented in the Deliverable 9.1 and updated in this deliverable (see section 3.2)
- Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation."

- **Risk: Bias in Recruitment and Representation**

  ***Description*:** Limited demographic or backgrounds diversity in recruitment (e.g., students or internal staff only) may lead to biased data and reduced generalizability of AI models.

  Assessment:
  - **Likelihood:** Medium
  - **Impact:** Medium

  Mitigation Strategy:

  - Limitation criteria to participate in pilots are reduced to the maximum possible extent (for example, being over 18 years old (pilot 2)/language knowledge (pilot 2 and lot 4)/medical background (pilot 1)
  - Strive for balanced representation in terms of gender, age, and cultural background within the constraints of voluntary participation

  Acknowledge sample limitations in any reporting or publications and ensure future model training incorporates broader datasets where possible

  Reporting and Documentation:

  - Consent forms are documented in Deliverable 9.1 and updated in this deliverable (see section 3.2)
  - Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation."

3.2. Personal Data

As outlined in deliverable D9.1, "Strong privacy protection is founded by an appropriate governance mechanism within the ELOQUENCE project. This is necessary to ensure the protection of personal data and compliance with relevant data protection laws." The deliverable determined the privacy governance mechanism by specifying how the partners in the ELOQUENCE consortium facilitate privacy-by-design approach, adhere to data protection principles, assess and mitigate risks concerning processing of personal data, ensure data security and provide clear allocation of roles and responsibilities thus facilitating strong accountability in processing of personal data.

The first version of the Data Management Plan (D8.4) was also submitted in M6, thus contributing to the privacy governance mechanism. As it spelled out, it is intended that most project activities should not contain personal data or references to personal data. Therefore, the amount of personal data collected and processed as part of this project is kept to a minimum, as it mostly concerns processing of personal data as part of the project coordination and dissemination, organising stakeholders' meetings or events, and carrying out of pilots. Regarding the last-mentioned activity, the major component of ensuring the data protection rights of research participants is informed consent procedure (as a process accompanying the pilots, now a single time event of signing the document). The consent forms and information sheets were developed since the earliest stages in the project and were submitted as part of deliverable D9.1.

In addition to that, pilot owners (UNS, TID and CNR) constantly monitor if the actual research practices and activities are properly reflected in the relevant documents. In this regard, UNS updated the consent form and information sheet: for example, to respect the preference of participants not to be advertised as part of the project dissemination activities; or to update the description of the data processed during the pilot. The updated versions of the informational sheet and consent form are provided as Annex I of this deliverable.

Below, the risks governance mechanism in data protection area is scrutinised.

- **Risk: Inadequately Informed Consent**

***Description*:** Participants may not fully understand the nature, purpose, or extent of data collection, particularly regarding the use of AI technologies, recording of dialogues, or the integration of data into model development

Assessment:

- **Likelihood:** Medium
- **Impact:** High (may result in violation of autonomy and informed choice; GDPR non-compliance)

Mitigation Strategy:

- Developed tailored Information Sheets and Informed Consent Forms for each pilot
- Ensure prior and voluntary consent with the right to withdraw at any time without penalty

Reporting and Documentation:

- Consent forms and information sheets are documented in the Deliverable 9.1 and updated in this deliverable (see Annex I)
- Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation."

- **Risk: Privacy Breaches and Data Security**

***Description*:** Collection of data across all pilots could lead to privacy breaches if improperly stored or transmitted.

Assessment:

- **Likelihood:** Medium
- **Impact:** High

Mitigation Strategy:

- Implement state-of-the-art encryption and access control measures for all stored data
- Follow data minimization principles: only collect what is necessary
- Apply pseudonymization/anonymization techniques before data sharing or model training

Reporting and Documentation:

- Consent forms and information sheets are documented in Deliverable 9.1 and updated in this deliverable (see Annex I)
- Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation."

- **Risk: Misuse of Personal Data in Dissemination Activities**

***Description*:** Participants in stakeholders' events, workshops, conferences, or webinars may have their personal information shared inappropriately or without consent.

Assessment:

- **Likelihood:** Low
- **Impact:** Medium

Mitigation Strategy:

- Collect only minimal and necessary data for event management
- Inform participants clearly about what personal data is being collected and why
- Use opt-in forms for newsletter sign-ups, publication of names/photos, and sharing of contact info

Reporting and Documentation:

- The description of datasets, purposes of data processing and the relevant procedures are described in D8.4 "Data Management Plan"

3.3. Non-EU Countries

The protection offered by the GDPR should travel with the data. Chapter V (Article 44) of the GDPR (General principle for transfers) states that "Any transfer of personal data which are undergoing processing or are intended for processing after transfer to a third country or to an international organisation shall take place only if, subject to the other provisions of this Regulation, the conditions laid down in this Chapter are complied with by the controller and processor, including for onward transfers of personal data from the third country or an international organisation to another third country or to another international organisation. All provisions in this Chapter shall be applied in order to ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined."

According to the ethics and data protection report published by the Commission, there are two different approaches that can be followed in case data has to be transferred outside EU. The first approach applies to few non-EU countries that have received an 'adequate determination' from the EU indicating that their data protection framework offers a level of protection equivalent to the one provided under EU law. The transfer of personal data from non-EU countries is subject to strict data protection requirements under Chapter V of the GDPR. According to the report, the conditions that need to be applied for data transfers to non-EU countries to be lawful are listed below:

- the explicit consent of the data subject (which requires them to be informed in advance of any such transfers);
- an 'adequacy determination' by the European Commission in respect of the country in question;
- a data-transfer agreement containing EC standard contractual clauses giving effect to EU data protection law; or
- binding corporate rules covering both sender and recipient and approved by a national supervisory authority.

ELOQUENCE consortium is composed of project partners from 10 countries including Serbia, the United Kingdom and Switzerland outside the EU. Two project partners are based in Serbia. Serbia is an associated country in the sense of Horizon Europe Regulation 2021/695. Consortium partners registered in the UK and Switzerland benefit from EC's adequacy decisions on transfers of personal data. Necessary transfers of personal data to and from partners based outside the EU will be only for the purpose of the project realization. Appropriate safeguards required by relevant requirements of EU law (particularly those laid down by Chapter V and Art. 32 of the GDPR) are to be ensured for the purpose of lawful data transfer.

- **Risk: Lower standards to personal data protection in countries outside the EU**
  ***Description*:** Countries outside the EU (such as Serbia) can have different data protection laws that do not provide the level of personal data protection as guaranteed under the GDPR
  Assessment:
  - **Likelihood:** Medium
  - **Impact:** Medium

  Mitigation Strategy:

  - Fulfillment of one of the conditions for data transfers outside the EU (explicit consent/data-transfer agreement)
  - Serbia is an associated country in the sense of Horizon Europe Regulation 2021/695, it is bound to respect the relevant framework in research activities
  - UK and Switzerland benefit from EC's adequacy decisions on transfers of personal data
- **Reporting and Documentation:**
  - All taken measures are documented in the relevant safeguards (explicit consent/data-transfer agreement) and in the Data Management Plan (D8.4).

3.4. Environment, Health and Safety

As outlined in Deliverable 9.1, by relying on the relevant guidelines and practices, the project consortium has assessed potential risks and harm to health, safety, and the environment that could arise from the implementation of the project activities. The assessments have concluded that there are no risks and harms on health and safety, whereas potential risks and harms may affect the environment during the development of the AI.

- **Risk: Significant energy consumption associated with training and running AI models**
  ***Description*:** The computational demands of AI algorithms, especially deep learning models, require large amounts of electricity. This reliance on energy-intensive computation can contribute to an increase in carbon emissions and exacerbate the environmental impact of the energy sector.
  Assessment:
    - **Likelihood:** Low (at the scale of the research project activities)
    - **Impact:** High

  Mitigation Strategy:

    - Innovations in AI design, such as more energy-efficient processes, and advances in software optimization will help to reduce the energy footprint of AI systems
    - The relevant project partners will prioritize sustainable practices in AI development, including responsible sourcing of materials, efficient recycling of electronic components and the adoption of ethical guidelines that consider environmental impact alongside technological advancement
    - The relevant project partners (UESSEX, BUT, TID) have been using BSC cloud infrastructure and EuroHPC benchmark calls (https://access.eurohpc-ju.europa.eu/calls/41) to ensure access and use of energy-efficient computing resources. For example, a BSC most advanced supercomputer (Marenostrum5) is a highly energy efficient supercomputer, fully powered with green energy, and utilizing heat reuse technology.
- **Reporting and Documentation:**
    - Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation."

3.5. Artificial Intelligence

Proper management of the ethics concerns that can potentially arise out of the development and deployment of AI is in the core of the ELOQUENCE project, as it seeks to advance state-of-the-art conversational AI technologies. These technologies, while becoming more and more sophisticated, efficient and promising progress in different domains, also raise serious ethics concerns, in particular with respect of fundamental rights of individuals. The scope of fundamental rights that can be potentially affected by AI, including GenAI applications, was analysed and determined in deliverable D6.1 "Report on linguistic expression respectful of EU values." These rights include: right to privacy and personal data; right to non-discrimination; freedom of information and expression; democracy and human agency.

The EU takes a proactive approach in mitigating those concerns by developing ethics guidelines and most recently, adopting the relevant law – the EU AI Act. In 2019, the High-Level Expert Group on AI set up by the European Commission developed the Ethics Guidelines for Trustworthy AI.[1] These Guidelines established the following key principles (requirements) of Trustworthy AI: accountability; human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity and non-discrimination; societal and environmental well-being. These principles were laid down as the foundation for the AI Act.

[1] European Commission: Directorate-General for Communications Networks, Content and Technology High-Level Expert Group on AI set up by the European Commission, *Ethics guidelines for trustworthy AI*, Publications Office, 2019, **https://data.europa.eu/doi/10.2759/346720**

An overview of existing soft law and emerging hard law governing AI in Europe and globally, including the architecture, main requirements of the AI Act, and its Fundamental Rights Impact Assessment framework were provided in deliverable D6.1 "Report on linguistic expression respectful of EU values." In addition, the deliverable developed the methodology for human rights and EU-values assessment of GenAI to be applied in the ELOQUENCE project – the tool of evaluating emerging ELOQUENCE outcomes by interdisciplinary panels of five: (1) the team that has developed the product as the 'submitting partner' who is capable of answering important questions related to the technical documentation, the training of the AI, the data and the design choices behind the object being assessed; and (2) a chairperson from EUI, Work Package leader of WP6. Three other experts are appointed to assess the object from the 'outside' so to speak, their fields of expertise will vary depending on the nature of the object being assessed. The panel will always aim to have a balance of professionals with technical expertise and expertise in the legal, cultural, ethical or political field. Usually, one of the three comes from another ELOQUENCE partner and two from the ELOQUENCE Community of Experts. The EUI team developed the questionnaire, addressing a range of concerns, most importantly for Ethics compliance:

- **Definitions:** What is assessed? (the assessment template the submitting partner is therefore tasked with the important duty to clearly describe the necessary background information about the object of assessment (including by attaching relevant documentation) and describe what the envisioned use of the product will be);
- **Transparency**: The submitting partner should include information on what has been done to facilitate Explainable Artificial Intelligence (XAI), while other assessors are tasked with determining to what extent it is clear for users that they are interacting with an AI, how it works and what kinds of information it gathers about them, including whether they are being profiled/categorised in any way;
- **Robustness, Reliability and trustworthiness:** These elements go to the very core of what the purpose of the object is and whether it is able to reliably fulfil it. An important factor here is the question of hallucinations which plagued early GenAI in particular. The question relates to what extent the object fulfils its function and produce useful and truthful content. Assessors here are invited to consider in particular the intended function of the object;
- **Bias and Discrimination:** The question of bias in the assessment concerns bias in a cultural or legal sense, rather than in the technical sense (which is the topic of Q5). The submitting partner is invited to disclose any known biases in the training data and what measures were taken to identify and mitigate them. The other assessors are invited to ascertain whether the product produces biased content on account of for example sex, gender, sexual orientation, gender identity, ethnicity, national origin, language, disability or political opinion either due to biases in the data or due to overcompensations during mitigation attempts. Both the submitting partner and the other assessors are also invited to investigate how the object responds to situations where the real world is biased, whether it includes any additional context or warnings for users that the results it produces are biased because its training data contains certain biases. Such biases will often take the form of information left out or perspectives failed to be included;
  This question of discrimination relates to the prevention of to the production of directly discriminatory and offensive content including hate speech. The submitting partner is invited to disclose results from any adversarial testing conducted in-house and to disclose what has been done to prevent the AI from producing hateful and discriminatory content, and how any eventual negative consequences of mitigating measures have been addressed. The other assessors are invited to assess how the object addresses situations where content is offensive out-group but not in-group, and how it addresses situations where it is requested to reproduce or summarise factually, ethically or legally problematic content. Assessors are also invited to engage in adversarial testing to see if the object be convinced to produce illegal or discriminatory content;
- **Multi-linguality and Cross-cultural knowledge:** This question addresses an important feature of the ELOQUENCE project and other projects launched under the same call – the attention to low-resource languages and cross-cultural knowledge. The submitting partner is invited to disclose what languages the AI currently supports and which it is envisioned to incorporate next, and the other assessors are invited to test the AI in as many languages as possible – particularly low resource languages;
  Privacy and data-protection: The submitting partner is invited to provide information on what kind of data the application gathers and how it gathers it – such as whether it outright requests data from. Other

assessors are invited to test how the system responds to attempts to retrieve data from other users (by for instance pretending to be two different members of a household if testing a smart home system, or two different callers to a call centre etc) and to consider what the risks of the current data management plan would be in case of data leaks, mission creep or other use contrary to the stated function of the object.

The questionnaire with all the specified considerations and specific descriptions and relevant questions to be addressed is included in D6.1 "Report on linguistic expression respectful of EU values" while results of completed assessments are being analysed and will be included in deliverable D6.2 "Emerging ELOQUENCE technology-approved by the ELOQUENCE Community."

The ethics issues arising out of the development and deployment of AI in the ELOQUENCE projects were extensively discussed in the meeting of the Ethics Advisory Board (see more details in Chapter 4). The outcomes of these discussions are summarised below, following the methodology described in Chapter 2.

- **Risk**: **Bias in Data and Model Outputs**

***Description*:** Bias in Data and Model Outputs can be caused by skewed demographic or linguistic representation in training data; Reinforcement of cultural, social, or gender stereotypes during fine-tuning or bias amplification in synthetic data or during adaptation to specific domains.

Assessment:

- **Likelihood:** Medium
- **Impact:** Medium (at the scale of research pilots)

Mitigation Strategy:

- Data Auditing: Conduct systematic audits on all training datasets (open-source and private) to identify imbalances across demographic groups, languages, and discourse types.
- Bias Detection in Pre-processing: Integrate automated tools to flag discriminatory patterns before data enters training pipelines.
- Fairness-Aware Training: Apply fine-tuning methods that actively reduce bias (e.g., adversarial training, fairness constraints).
- Diverse Data Sampling: Ensure balanced representation of underrepresented voices and conversational patterns in both real and synthetic data.
- Human-in-the-Loop Reviews: Involve diverse evaluators to validate fairness in output across scenarios.

- **Reporting and Documentation:**
  - Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" D5.4 "Pilots Evaluation," governance and management of data in the project are documented in Data Management Plans (D8.4, D8.5 and D8.6). EAB meetings where the issues and mitigation strategies were discussed are described in Chapter 4 of this deliverable.

- **Risk**: **Privacy and Data Protection Violations in the AI context**

***Description*:** This risk might be caused by the use of industrial/private data with sensitive or identifiable information/lack of clarity on anonymization processes/data leakage through model memorization.

Assessment:

- **Likelihood:** Medium
- **Impact:** Medium (at the scale of research pilots)

Mitigation Strategy:

- GDPR-Compliant Consent: Ensure clear, granular consent procedures are used when private data is involved.
- Anonymization and Minimization: Remove or mask identifiers before training. Use differential privacy where applicable.

- Federated Learning: For use cases involving sensitive environments (e.g., home-based or healthcare interactions), apply decentralized model training.
- Data Access Control: Limit access to raw data; enforce secure storage and audit trails.
- Data Lifecycle Policies: Define timelines for data retention and secure deletion protocols.

Reporting and Documentation:

- Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation," EAB meetings where the issues and mitigation strategies were discussed are described in Chapter 4 of this deliverable.

- **Risk**: **Lack of Transparency and Explainability**
  - ***Description*:** This risk may arise from several factors. Fine-tuned models can become overly complex, making decision logic difficult to interpret. Synthetic data may reduce the traceability of model behavior. In addition, communication with users about AI involvement may be insufficient. T**Likelihood:** Medium
  - **Impact:** Medium (at the scale of research pilots)

Mitigation Strategy:

- Model Interpretability Tools: Integrate explainability layers (e.g., SHAP, LIME, attention visualization) into model evaluation.
- Traceable Training Pipelines: Maintain full documentation of data sources, pre-processing steps, model changes, and hyperparameters.
- User Disclaimers: Clearly inform users when interacting with AI (especially in sensitive domains like healthcare/emergency response).
- Transparent Datasets: Where possible, publish metadata and ethical justifications for dataset use.

Reporting and Documentation:

- Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation," EAB meetings where the issues and mitigation strategies were discussed are described in Chapter 4 of this deliverable.

- **Risk**: **Misuse or Misunderstanding of AI Decisions**

***Description*:** This risk might be raised if/when the users believing AI judgments are authoritative or human-equivalent or within the use of AI in high-stakes domains without proper safeguards.

Assessment:

- **Likelihood:** Medium
- **Impact:** Medium (at the scale of research pilots)

Mitigation Strategy:

- Disclaimers in Deployment: Label all AI-driven recommendations as supportive, not definitive, especially in medical/emergency contexts.
- Professional Oversight: Embed human review and override mechanisms in critical systems.
- Scenario Testing: Simulate high-risk use cases to test AI behavior under stress or ambiguity.

Reporting and Documentation:

- Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation," EAB meetings where the issues and mitigation strategies were discussed are described in Chapter 4 of this deliverable.

- **Risk**: **Reinforcement of Ethical Violations in Deployment**

***Description*:** this issue might be caused by undetected biases or errors in live AI deployments/scaling of problematic responses or behaviors over time.

Assessment:

- **Likelihood:** Medium
- **Impact:** Medium (at the scale of research pilots)

Mitigation Strategy:

- Post-Deployment Monitoring: Use continuous monitoring frameworks to audit real-world AI interactions for ethical violations.
- Feedback Loops: Implement user reporting features and integrate responses into model re-training cycles.
- Ethics Dashboards: Provide real-time oversight tools for stakeholders to monitor model decisions, bias alerts, and transparency indicators.
- Regular Fairness Audits: Schedule internal audits during each project phase and report findings to the EAB and the European Commission.

Reporting and Documentation:

- Planning and carrying out of pilots are documented within WP5, in Deliverables D5.1 "Pilot Requirements & Usability Evaluation"; D5.3 "Pilots Integration" and D5.4 "Pilots Evaluation," EAB meetings where the issues and mitigation strategies were discussed are described in Chapter 4 of this deliverable.

4. Ethics Advisory Board

4.1. Structure and Composition

The Grant Agreement details in framework of T8.4 that:

> *"[The task will] establish an independent Ethics Advisory Board (EthAB) for the project by bringing on board external experts with relevant background in EU Ethics Appraisal Scheme, Privacy and Data Protection and other areas relevant to the ELOQUENCE research activities. The independent board members will be invited to review any of the deliverables of interest to them during each year's reporting period and to offer their views on ethical and social issues concerning the project and/or to which they offer solutions ."*

The members of the EthAB have leading and unique expertise in areas that are critical to the project. The EthAB is independent and has no formal responsibility towards the project. The EthAB is regularly informed about the status of the project and the current perceived ethical and data protection issues and is invited to comment on them. In particular, the independent board members are invited to review the deliverables of interest and to express their opinions on ethical and social issues related to the project and/or propose solutions. In this way, the EthAB, with the support of the PC, ensures that all relevant activities, including the processing of personal data and the development of trustworthy AI, are continuously compliant with applicable regulations.

The external members of the EthAB have been selected on the basis of their scientific achievements, expertise, and network, their role and expertise in national and international policy-making, and their role as and connection to relevant stakeholders.

To date, the following members have accepted to be a part of the project's EthAB:

- **Adriana Minovic** is a seasoned lawyer specializing in diverse regulatory and compliance issues across various industries, including ICT, Telecom, Media, R&D, i-Gaming, Aviation, and Life Science. With expertise in Data Protection, Intellectual Property, Competition & State Aid Law, AML, and Anti-Corruption, Adriana has built a reputation for her comprehensive understanding of legal complexities in innovative and data-driven sectors. Her key competencies include GDPR compliance, information and cyber security policy development, regulatory impact assessment, and public policy advocacy. With a focus on strategic litigation, regulatory strategy, and ethics evaluations, Adriana Minovic demonstrates a commitment to ensuring legal compliance and ethical practices within complex regulatory environments.
- **Ashwinee Kumar (LL.M)** is a dedicated legal and ethical expert specializing in cybersecurity, data protection, and IT law. With extensive experience at Vrije Universiteit Brussel, he has contributed as a Legal and Ethical Manager on high-impact EU-funded projects like INTREPID, LOCARD, and FASTER, addressing forensic evidence collection, emergency response, and drone threat mitigation. His expertise extends to GDPR compliance, transnational IT contracts, and e-commerce law, having previously collaborated on the ARC initiative to raise awareness about data protection among SMEs. Ashwinee holds a Ph.D. in Law from Vrije Universiteit Brussel and an LL.M. in Intellectual Property & IT Law from the University of Göttingen. His diverse academic background, including a B.Tech in Electrical Engineering, enables him to bridge the gap between law and emerging technologies. Passionate about ethical AI, cybersecurity, and digital rights, he continues to shape policies and frameworks that drive technological innovation while ensuring legal compliance.
- **Hanne Elsen** is a seasoned legal professional with a wealth of experience spanning diverse roles. As a DPO at Universiteit Gent, she ensures regulatory compliance and data security. Simultaneously, she serves as an independent expert for the European Commission, contributing her legal acumen. With a background in sales, Hanne brings a unique perspective to her legal career, having honed her negotiation and communication skills. Her experience also includes roles in mediation and research coordination, where she showcased her aptitude for problem-solving and policy analysis.
- **Amal Marc**, with nearly two decades of expertise in new technologies, digital regulation, and human rights/ethics, brings extensive experience to her roles. She is currently the Founder and Independent Advisor at Sublimis, focusing on digital compliance, cyber risks, and eco-responsible business. Her previous

roles encompassed domains such as legal, cybersecurity, and compliance, including engagements in European R&D programs. Amal's diverse background spans legal, cybersecurity, and compliance fields, driving impactful projects aligned with the global climate agenda and promoting responsible business practices.

The EthAB was formally established following the invitation of the PC to the selected experts. The experts were proposed by the PN. After accepting their role on the panel, all EthAB members signed the Non-Disclosure Agreement (NDA).

Advisors are set to play a crucial role in ELOQUENCE, their responsibilities will cover the oversight and contextualization of the ongoing research endeavours from the lenses of needed ethical measures. The external members of the EthAB are set to bring new perspectives and offer a comprehensive understanding of the project's compliance with best ethical practices. The EthAB is particularly expected to address specific ethical challenges anticipated by the consortium. The members should be, as per the GA, knowledgeable about EU Ethics Appraisal Scheme as well as in matters of Privacy and Data Protection. Their ability to assess, and mitigate ethical risks, ensures a robust accompaniment for the ELOQUENCE project. Diversity is also a cornerstone in the selection of advisors. They are expected to bring an array of viewpoints from various relevant fields and geographical distribution. The advisors represent a diversity of sectors, they thus represent the law sector, business, as well as academia.

4.2. Roles and Responsibilities

The primary roles and responsibilities of the EAB include:

- Ethical Oversight: Monitoring the project's activities to ensure compliance with ethical principles and legal requirements.
- Risk Assessment: Identifying potential ethical risks and proposing mitigation strategies.
- Advisory Function: Providing guidance on ethical challenges, including data protection, bias mitigation, and user safety.
- Review of Deliverables: Evaluating key project outputs to ensure they meet ethical standards.

The EAB will serve as a crucial resource for the consortium, helping to navigate complex ethical issues and maintain accountability.

4.3. Engagement Strategy

The engagement strategy for the EAB includes:

- Regular Meetings: To discuss ethical concerns, review project progress, and provide recommendations.
- Ad Hoc Consultations: Additional meetings or consultations as needed to address urgent ethical issues.
- Documentation: Detailed minutes of EAB meetings, including action points and recommendations, will be maintained and shared with relevant stakeholders.
- Collaboration Tools: Use of secure communication platforms to facilitate ongoing dialogue and collaboration among EAB members.

Given the nature of the ELOQUENCE project, the EAB members commit to a NDA. This non-disclosure agreement has been formally signed by each external EAB member. Consequently, any information shared with the EAB is held under a stringent obligation of confidentiality in strict adherence to the provisions outlined in the NDA attached to this deliverable.

4.4. Reporting and Feedback Mechanism

The EAB provides regular reports to the project's management team, summarizing key findings, recommendations, and any outstanding ethical issues. These reports will be integrated into the project's overall governance framework to ensure timely and effective decision-making.

4.5. Ethics Advisory Board Meetings

4.5.1. First Ethics Advisory Board Meeting

The first meeting of the Ethics Advisory Board for the ELOQUENCE project took place on 26th April, 2024. The meeting commenced with introductions of the Ethics Advisory Board members, each bringing significant expertise in law, ethics, and AI governance. This introduction reinforced the collective commitment to guiding the project towards the highest ethical standards while ensuring compliance with European and international regulations.

Key Topics Discussed

The ELOQUENCE project team provided a comprehensive overview of the project, highlighting its ambitious objectives in developing multilingual AI-driven conversational systems. The discussion acknowledged the ethical complexities and challenges posed by large language models (LLMs), including:

- Ensuring alignment with human values and avoiding biases.
- Privacy and data protection concerns, particularly in sensitive interactions.
- Transparency and accountability in AI decision-making.

Nine Work Packages were outlined in detail, each demonstrating the project's commitment to ethical compliance with European values, the AI Act, and GDPR. The team emphasized that pilot studies would serve as key validation mechanisms, allowing for an evaluation of how AI technologies impact real-world scenarios and ensuring that user concerns are adequately addressed.

During the meeting, several ethical concerns were discussed, including:

1. Call Prioritization and Fairness: How AI-driven decision-making in emergency response scenarios should prioritize calls without introducing biases. The ethical responsibility of AI in ensuring fair and unbiased outcomes for all users, regardless of demographic or linguistic background.

2. Medical Assessment Compliance and Risk Management: The importance of adhering to medical and healthcare standards when integrating AI for triaging or assisting in emergency medical calls. Ensuring that AI systems are not replacing professional medical judgment but rather serving as a support tool.

3. User Disclaimers and Transparency: Clear communication of AI involvement in conversations to avoid misleading users. Ensuring that users are fully aware of data usage policies, particularly in sensitive or high-stakes interactions.

The first EthAB meeting successfully established a strong ethical foundation for the ELOQUENCE project. It reinforced the need for ongoing ethical scrutiny and compliance monitoring throughout the project lifecycle, active user involvement in the development and testing phases to ensure ethical acceptability and clear regulatory alignment with EU and international standards.

4.5.2. Second EthAB Meeting

The second meeting of the Ethics Advisory Board for the ELOQUENCE project was held on 27th November 2024. The session focused on reviewing the progress and preliminary findings from key project tasks, particularly those related to data collection, pre-processing, and fine-tuning of Large Language Models (LLMs). Discussions centred on bias in training data, potential ethical risks, and strategies for mitigating these risks to ensure fairness, privacy, and compliance with European AI standards.

Key Topics Discussed

1. Findings from Task T1.1 – Data Analysis and Pre-processing

The project team presented initial findings from T1.1, which focuses on the collection and processing of open-source and industrial-based (private) data for training ELOQUENCE's AI-driven conversational models. The discussion covered:

- Challenges in handling semi-structured, unstructured, and multi-turn dialogues, particularly in multilingual settings.
- Differences between open-source and private datasets, including:
    - Variability in data quality, representativeness, and completeness.
    - Potential biases in datasets, which may disproportionately represent certain user groups or conversation styles.
- Pre-processing techniques used to clean, filter, and structure raw data to ensure model reliability while preserving diversity.

Ethical Concerns Identified:

- Bias in source data: Many datasets contain inherent biases reflecting social, cultural, and linguistic disparities, which may reinforce stereotypes when integrated into AI models.
- Privacy implications: The use of private datasets raised concerns regarding data protection, anonymization, and compliance with GDPR.
- Transparency in data handling: The need for clear documentation on how data is collected, filtered, and used in training.

EthAB Recommendations:

- Enhance dataset diversity by ensuring balanced representation across different demographics, languages, and conversational styles.
- Implement bias detection mechanisms in the pre-processing stage to flag potentially discriminatory patterns in the data.
- Adopt privacy-preserving techniques such as differential privacy and federated learning when handling sensitive data.
- Ensure transparency in data selection through public documentation and ethical review of private datasets used in training.

2. Insights from Deliverable D2.1 – Fine-Tuning and Simulation for LLMs

The team provided an update on D2.1, which outlines methodologies for fine-tuning LLMs and generating augmented conversational training data. Key points included:

- Comparison with State-of-the-Art (SOTA) models, identifying areas where ELOQUENCE's approach excels or needs refinement.
- Challenges in adapting existing fine-tuning methodologies to the project's unique use cases.

- Integration of synthetic data augmentation to improve performance in low-resource languages and specialized domains.
- Implementation of a shared codebase for fine-tuning LLMs on selected datasets, ensuring traceability and reproducibility of results.

Ethical Concerns Identified:

- Bias amplification during fine-tuning: Even when diverse training data is used, fine-tuning can unintentionally magnify biases present in pre-trained models.
- Synthetic data risks: While augmentation techniques can improve model robustness, they must not introduce artificial biases or deviate from real-world linguistic structures.
- Lack of interpretability: Fine-tuned models often become more complex, making it difficult to explain how decisions are made, impacting trust and accountability.

EthAB Recommendations:

1. Use fairness-aware fine-tuning techniques that actively counteract biases identified in pre-trained models.
2. Validate synthetic data augmentation using human evaluation and fairness metrics to ensure realistic, unbiased content generation.
3. Establish explainability standards by integrating model interpretability tools that allow users to understand how AI-generated responses are formed.
4. Regularly benchmark fine-tuned models against ethical AI guidelines and industry best practices to ensure alignment with European AI Act principles.

3. Bias in AI and Ethical Risk Mitigation Strategies

A key part of the meeting focused on the ethical risks associated with bias in AI models, particularly in relation to ELOQUENCE's use cases. The board engaged in an in-depth discussion on:

- Sources of bias in conversational AI, including:
    - Training data bias (demographic imbalance, overrepresentation of dominant language styles).
    - Algorithmic bias (fine-tuning decisions reinforcing discriminatory patterns).
    - User interaction bias (misinterpretation of user intent due to cultural differences).
- Ethical risk scenarios, such as:
    - AI-generated responses perpetuating stereotypes or offensive language.
    - Unintended discrimination in automated call prioritization systems.
    - Misinformation risks in AI-generated explanations.

EthAB Recommendations:

1. Develop frameworks that can monitor AI-generated responses for potential ethical violations.
2. Implement human-in-the-loop verification, especially in high-risk applications like emergency response and healthcare AI interactions.
3. Conduct fairness audits at different stages of the project to identify and address hidden biases.

Conclusion and Next Steps

The second EthAB meeting provided valuable insights into the ethical challenges of training AI conversational models, particularly concerning data biases, fine-tuning risks, and privacy considerations. The discussions reinforced the importance of:

- Proactive bias mitigation strategies at every stage of AI development.
- Robust ethical review mechanisms to ensure transparency, fairness, and compliance with European AI regulations.

The next EthAB meeting will be organized in May 2025.

5. Conclusion

Ethical compliance is a fundamental pillar of the ELOQUENCE project, underpinning its commitment to developing AI that is trustworthy, inclusive, and socially responsible. Throughout the project, the consortium has demonstrated a proactive, transparent, and accountable approach to managing ethical risks—particularly those associated with AI development and deployment. The ethics compliance process is deeply integrated into project governance, with structured documentation, partner engagement, and expert oversight through the Ethics Advisory Board. Specific methodologies—such as the developed framework for the Ethics Compliance Management Process in ELOQUENCE and multidisciplinary AI value assessments —enable ELOQUENCE to respond dynamically to the evolving ethical landscape of AI research. As the project progresses, these systems will be further reinforced and evaluated through follow-up deliverables (e.g., D8.3), ensuring a continuous and responsive ethics management cycle.

Annex I: Updated Consent Form and Information Sheet by UNS

Informed Consent Form

We, University of Novi Sad, hereby ask for your consent to the collection, storage, processing and other use of audio recordings of dialogue conversations containing your voice, created in the context of the realization of Pilot 4 within the ELOQUENCE Project, as well as the textual corpus of their loose transcriptions. The aforementioned activities are aimed at the execution of the Project and its deliverables.

You have the right to obtain information about your data, access to it, restriction of its processing, as well as to advance a dispute to the Supervisory Authority (Privacy Authority) in compliance with Chapter III of GDPR. Full information regarding the processing of data is provided before the participation at the Project.

Full information regarding the processing of the data collected through your contribution, provided before the participation in the Project, is attached to this form as Information Sheet (Privacy Policy).

I, the undersigned, _____, born on _____ / _____ / _____
(Participant Surname and Name)

by completing this form and with my signature at the bottom, within the meaning and for the purposes of the EU Regulation No. 2016/679 – General Data Protection Regulation (GDPR) and aware of the general objectives and characteristics of the project, as well as of the conditions and commitments I assume as a participant:

☐ freely consent to the participation in the Project;

☐ have read and fully understood the Information Sheet (Privacy Policy);

☐ am aware of being able to withdraw at any time, without having to give justifications or incurring in costs or penalties;

☐ consent to the collection, storage, processing and other use of audio recordings of dialogue conversations in which I participated;

☐ consent to the production of transcriptions of these conversations as well as their revision;

☐ consent to the sharing of obtained transcriptions to other researchers or institutions inside or outside the EEA;

☐ I know who to contact if I have any question about the Project and my privacy;

☐ I have been given a copy of this informed participation and consent form;

☐ I have had all my questions answered to my satisfaction;

I have granted the above authorisations of my own free will and they are also valid as a disclaimer. In the event of revocation of these authorisations, I understand that the audio recordings containing my voice will be restricted from any subsequent processing or research, while their loose textual transcriptions, already produced prior to the revocation, can still be used without limitation.

Date: _____

Information Sheet

1. Scope of this policy

This information sheet (hereafter "Privacy Policy") describes how the data created through your contribution is collected, stored, processed and otherwise used in the context of the project **"Multilingual and Cross-cultural**

interactions for context-aware, and bias-controlled dialogue systems for safety-critical applications (ELOQUENCE)“, funded under the H2020 research programme, contract no. 101135916 (hereafter the "Project").

This information sheet concerns the data included in an audio corpus of human spoken dialogues as well as the textual corpus of corresponding transcriptions, created within the set of activities related to Pilot 4, owned by the Faculty of Technical Sciences, University of Novi Sad, one of the Project partners. The two corpora are intended for training an AI system supporting similar human-machine dialogues intended for collecting information from human callers in a call centre offering medical advice to young parents. The owner of the developed corpora will be the Faculty of Technical Sciences, University of Novi Sad.

This Privacy Policy includes a description of your data protection rights. In this Privacy Policy:

- "We" or "us" refer to the Faculty of Technical Sciences, University of Novi Sad, who will store, process and share the collected data, as will be described below. We can be contacted by sending an email to pdpeloquence@uns.ac.rs.
- "Data Protection Legislation" means the Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (the "GDPR"), as well as any legislation and/or regulation implemented or created pursuant to the GDPR and the e-Privacy legislation, or which amends, replaces, re-enacts or consolidates any of them, and all other national applicable laws relating to processing of personal data and privacy that may exist under applicable law.
- The terms "third party", "personal data", "processing", “Supervisory Authority”, shall have the meanings set out in the applicable Data Protection Legislation.

2. What personal data is processed?

In the context of the Project, data created through your contribution will be collected, stored, processed and otherwise used by the Faculty of Technical Sciences, University of Novi Sad.

No data directly related to you or any other participant (such as personal information, medical data or behavioral data) will be collected at any point. However, we will collect, store, process and may otherwise use the recordings of your voice, as will be described in more details below.

The data collected and/or developed will include audio data and textual data, and will be obtained through recording sessions (audio data), followed by semi-automatic transcription and subsequent revision (textual data). Each recording session will involve two participants, and you will participate either as a **medical expert** (e.g. nurse) or a volunteer hereafter referred to as a **caller**. The two participants will enact a conversation such as those that may actually occur in a call center where medical experts offer advice to callers about problems with their babies (including newborns). The issue that will be discussed in a particular conversation (e.g. the specific problem) and certain directions of that conversation will be previously arranged between the participants but the conversation will not follow a predefined script. The callers (predominantly young parents themselves) will discuss imaginary issues with babies based on similar events from their own experience. The conversations will be carried out over mobile telephones, and only the audio recording (hereafter referred to as **recording**) of each conversation will be captured. As the primary purpose of the recordings is to obtain textual transcriptions that are optimized for being used as LLM training material for a medical chatbot, in the transcription process the number of filler words and hesitations will be reduced, grammar and sentence structure will be made more formal, overlapping speech will be identified and converted into a more formal structured representation (by hand). This will results in a set of documents loosely based on the transcriptions of existing dialogue recordings, and these documents will hereafter be referred to as **loose transcriptions**. If you participate as a medical expert, you may also be engaged in the revision of these transcriptions in order to maximize their verisimilitude.

The legal basis for us to collect, store, process and otherwise use your data is your consent, which you may withdraw at any time you choose and at your own initiative by contacting us at pdpeloquence@uns.ac.rs. The withdrawal of

your consent will not affect the lawfulness of the collection and processing of your data based on your consent up until the moment where you withdraw your consent.

3. How long is your personal data stored?

The recordings of dialogues in which you participated will be stored on secure servers of the Faculty of Technical Sciences for an unlimited time and protected by passwords known by authorized team members. Appropriate technical and organisational measures will be implemented in order to ensure an appropriate level of security of the recordings. The recordings will be stored under filenames not revealing the identity of either participant in the recording session, and will not be accompanied by any metadata related to them. The recordings will not be made publicly available to any party inside or outside the European Economic Area (EEA), but may be used by the Faculty of Technical Sciences, Novi Sad, for other research, including, but not limited to transcription and usage in language model training and evaluation.

The (revised) loose transcriptions will be stored on the servers of the Faculty of Technical Sciences for an unlimited time, and may be shared with third parties, as explained in Section 4 in more detail.

As a participant, you have the right to withdraw from this activity at any time without giving a reason. In this case, the audio recordings to which you have contributed will be restricted from further processing, but since they do not contain your personal data (such as personal information, medical data or behavioral data), no deletion of data will occur. Furthermore, as the textual corpus of loose transcriptions contains no data directly or indirectly related to you, we retain the right to use without limitation any loose textual transcriptions already produced prior to the revocation.

4. How is your personal data shared with third parties?

No personal data related to you, disclosable to other consortium entities or any third-party will be collected from you at any point. The audio corpus, containing recordings of your voice, will not be shared with any other consortium entity or third party during or after the Project lifetime.

The textual corpus including loose transcriptions will at some point be made publicly available and thereafter will be accessible and usable by anyone (e.g. other researchers or institutions inside or outside the EEA) for research purposes, including, but not limited to transcription and usage in language model training and evaluation. The textual corpus, however, includes no data directly or indirectly related to the participant in the conversation.

The collected data may be shared with government authorities and/or law enforcement officials if mandated by law or if required for the legal protection of the data owner's legitimate interests in compliance with applicable laws.

What are your rights?

You are entitled to request information regarding the recordings of your voice, including, but not limited to specific purposes, storage, sharing and security, and you are also entitled to receive a copy of such data in a machine readable format.

You have the right to request that the recordings of your voice should be suspended from being processed in case they are unlawfully or unnecessarily processed.

To exercise any of these rights, you can get in touch with us using the details set out in Section 1. If you have unresolved concerns, you have the right to lodge a complaint with an EU data protection authority where you live, work or where you believe a breach may have occurred.

In the event personal information is compromised as a result of a security breach and where the breach is likely to result in a high risk to your rights and freedoms, we will make the necessary notifications, as required under the Data Protection Legislation.

We see no potential risks for you associated with your participation in this activity, such as discomfort or inconvenience. Your participation in this activity may contribute to the development of a system for AI support to more efficient call centres for young parents seeking advice related to their babies.

Questions, comments, remarks, requests or complaints regarding this Privacy Policy are welcome and should be addressed to: pdpeloquence@uns.ac.rs.

Annex II: Ethics Advisory Board Report on the ELOQUENCE Project

Developed by EthAB Chairperson: Ms. Adriana Minovic

Introduction

As Chairperson of the Ethics Advisory Board for the ELOQUENCE project, I have attended three ethics advisory board meetings, including the initial introductory meeting. This report is based on my initial assessment of the project, drawing from the information presented in these meetings. My analysis is conducted with an unbiased perspective, focusing solely on the ethical considerations relevant to the project's goals, methodologies, and implementation strategies.

The ELOQUENCE project aims to develop AI-based language technologies that enhance natural language understanding, human-machine interaction, and multilingual communication. While the project aligns with European values and regulations, several ethical concerns must be proactively managed to ensure compliance with the highest standards of fairness, transparency, and accountability. This report outlines key ethical issues and provides recommendations to address them effectively.

My initial observations highlight critical ethical considerations, including data privacy and protection, AI bias and discrimination, explainability and transparency, security risks, and ethical commercialization. These areas require careful attention to ensure that ELOQUENCE not only meets regulatory standards but also upholds ethical integrity in its development and deployment.

Potential Ethical Issues

Data Privacy and Protection

ELOQUENCE involves processing significant amounts of online and voice data, raising concerns about data protection, user consent, and security. While the project commits to GDPR compliance, ensuring strict adherence to these principles, especially regarding informed consent and data minimization, remains a key challenge.

Special attention must be given to how data is collected, stored, and processed to prevent unauthorized access or misuse. Proper encryption protocols, secure data storage solutions, and strict access control mechanisms should be in place to protect user data. Additionally, users must have a clear and accessible way to provide or withdraw consent, with full transparency about how their data will be used. Independent audits and privacy impact assessments should be conducted regularly to ensure ongoing compliance and address emerging risks.

AI Bias and Discrimination

One of the primary concerns in AI-driven language models is the potential for bias, which can result in unfair treatment of certain user groups. Bias in training data can reinforce stereotypes and discrimination, disproportionately affecting marginalized communities.

While ELOQUENCE aims to implement bias-mitigation strategies, such efforts must go beyond initial assessments. A diverse dataset that reflects the full spectrum of cultural, linguistic, and socio-economic backgrounds must be continuously updated. Bias detection algorithms should be employed to identify and correct any emerging biases in real time. Furthermore, third-party reviews, including input from ethicists and affected communities, should be incorporated to ensure a more inclusive and equitable AI system.

Explainability and Transparency

AI models often operate as "black boxes," making it difficult to understand how decisions are made. The explainability of AI-generated responses is crucial for user trust, particularly in safety-critical applications such as emergency response services and legal or healthcare consultations.

To address this, ELOQUENCE must prioritize developing user-friendly explanation tools that provide clear, comprehensible rationales behind AI-generated decisions. Where AI is used in critical situations, mechanisms should be in place to allow human oversight and intervention. Standardized reporting practices should be introduced to document AI decision-making processes, enabling both end-users and regulators to audit and evaluate the AI system's fairness and reliability.

Misuse and Security Risks

AI-driven virtual assistants pose significant security risks, including adversarial attacks, misinformation, and manipulation. Cybercriminals may attempt to exploit vulnerabilities in the system to spread harmful or misleading content. Additionally, AI-powered dialogue systems can be susceptible to manipulation by users, potentially leading to biased or unethical responses.

To mitigate these risks, the ELOQUENCE project should implement rigorous security measures, including robust authentication methods, anomaly detection systems, and continuous monitoring of AI outputs. Ethical hacking and red-team testing should be conducted regularly to identify vulnerabilities before they can be exploited. Moreover, AI-generated content should be fact-checked against verified knowledge bases to prevent the spread of misinformation.

Ethical Use of AI for Commercial Purposes

The potential secondary use of AI-generated personalized content for advertising or other commercial purposes raises serious ethical concerns. Users must have control over their data and the extent to which it can be repurposed. Without proper governance, AI-driven personalization could lead to manipulation, exploitation, or violations of user autonomy.

To address this, ELOQUENCE must develop clear policies outlining how user data will be used beyond the primary research purpose. Any commercialization of AI-driven insights must require explicit user consent, and alternative business models that do not rely on excessive data monetization should be explored. Additionally, partnerships with external entities must be carefully vetted to ensure compliance with ethical AI standards and prevent unethical commercialization practices.

Best Practices and Recommendations

To ensure the ethical integrity of the ELOQUENCE project, a series of best practices and recommendations must be implemented. These recommendations focus on mitigating ethical risks while fostering trust, transparency, and accountability in AI-driven language technologies.

Strengthening Data Governance

- Conduct a comprehensive **Data Protection Impact Assessment (DPIA)** to evaluate risks associated with data collection, storage, and processing.
- Adopt a **Privacy-by-Design** approach, ensuring that all AI systems incorporate privacy safeguards from inception.

- Implement **strict access controls and encryption protocols** to prevent unauthorized data access and mitigate security threats.
- Develop transparent **user consent mechanisms**, ensuring that users clearly understand how their data is collected and used, with easily accessible opt-in and opt-out options.
- Regularly audit and update **data anonymization techniques** to ensure compliance with GDPR and other data protection regulations.

Enhancing Fairness and Bias Mitigation

- Establish a **bias-detection and correction framework**, using diverse datasets that accurately represent various demographic, cultural, and linguistic groups.
- Utilize **fairness-aware machine learning techniques** to actively detect and mitigate biases in training data and AI models.
- Implement a continuous **monitoring and feedback loop**, allowing researchers and external auditors to assess AI fairness and recommend improvements.
- Conduct **third-party ethical reviews**, including consultations with underrepresented communities to ensure inclusivity and fairness.

Improving AI Transparency and Explainability

- Develop **explainable AI models**, ensuring that users understand the logic behind AI-driven decisions.
- Provide **user-friendly dashboards** that allow stakeholders to inspect and validate AI-generated outputs.
- Implement **human-in-the-loop oversight**, particularly in safety-critical applications, where AI outputs must be subject to human verification.
- Regularly publish **transparency reports**, detailing model performance, bias mitigation efforts, and user feedback.

Addressing Security and Misuse Risks

- Employ **adversarial testing and ethical hacking** to identify vulnerabilities in the AI system before deployment.
- Establish **real-time monitoring** mechanisms to detect and prevent potential adversarial attacks or manipulation of AI-generated content.
- Develop a **misinformation detection system**, ensuring that AI-generated outputs are validated against credible sources before dissemination.
- Create clear **ethical use policies**, outlining acceptable and prohibited uses of the AI system to prevent exploitation or misuse.

Ensuring Responsible Commercialization and Secondary Data Use

- Define clear **guidelines on secondary data use**, ensuring that AI-generated data is not repurposed without explicit user consent.
- Explore **alternative business models** that prioritize ethical considerations over excessive data monetization.
- Establish **partnership review protocols**, ensuring that external collaborations align with ethical AI standards.
- Introduce **consumer education initiatives**, empowering users with knowledge about how AI systems operate and how their data is used.

Conclusion

The ELOQUENCE project demonstrates a strong commitment to ethical AI development. However, proactive measures must be taken to address potential risks related to data privacy, bias, transparency, security, and commercialization. By implementing robust governance frameworks, bias mitigation strategies, and transparency measures, the project can set a benchmark for responsible AI development in Europe.

As Chairperson of the Ethics Advisory Board, I strongly recommend continuous ethics oversight, stakeholder engagement, and alignment with emerging EU regulations (e.g., AI Act) to ensure ELOQUENCE remains at the forefront of ethical and trustworthy AI innovation.

eloquence

Telefónica

ICAR Istituto di Calcolo e Reti ad Alte Prestazioni

idiap RESEARCH INSTITUTE

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

FBK Fondazione Bruno Kessler FUTURE BUILT ON KNOWLEDGE

Brunel University London

University of Essex

EUI

BRNO UNIVERSITY OF TECHNOLOGY

Privanova

INOSENS

TRANSFORMATION LIGHTHOUSE

GrantXpert

Omilia Conversational Intelligence

SYNELIXIS

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

UKRI UK Research and Innovation