DisAI.eu

D5.4
Ethics, innovation and IPR management plan

Project Title	Improving scientific excellence and creativity in combating disinformation with artificial intelligence and language technologies
Contract Nº.	101079164
Type of Action	HORIZON-CSA
Topic	Disinformation Combating
Project start date	1st Dec 2022
Duration	36 months

Funded by
the European Union

Deliverable title	Ethics, innovation and IPR management plan
Deliverable number	D5.4
Deliverable version	
Contractual date of delivery	30 Nov 2023
Actual date of delivery	30 Nov 2023
Nature of deliverable	Report
Dissemination level	Public
Work Package	WP5
Task(s)	T5.3
Partner responsible	KInIT
Author(s)	Juraj Podroužek (KInIT), Matúš Mesarčík (KInIT), Marián Šimko (KInIT)

Abstract	The deliverable consists of two parts. First, we describe ethics management: rooted in KInIT approach to ethics management, we outline key components guiding our work in the DisAI project. We also discuss the ethics checklist for the research project. Second, we introduce our IPR and innovation management plan.
Keywords	Ethics, Innovation, IPR, Management

Table of Contents

1 Introduction

This document reports planned activities and outputs within ethics, innovation and IPR management. It is closely connected with and complements the Data management plan (Elaborated continuously in deliverables D5.1, D5.2, D5.3).

In the first part, we cover ethics management that is strongly aligned with the institutional approach to ethics management. This includes ethics planning, adhering to values and respecting red lines. It also incorporates an ethical framework being developed in KInIT for both institutional and project level. Elements of the framework involve Ethics checklist, Research data policy, Code of ethics and Ethics committee.

In the second part, we describe our approach to IPR and innovation management. It primarily focuses on assessing all the potential IP matters in the project. It results in the Intellectual property rights management plan including the proposed IP protection for expected project outcomes.

2 Ethics Management

KInIT has an ambition to conduct responsible research of intelligent technologies. We are convinced that every responsible practice should be looking into the future and be sensitive to possible impacts and consequences on people and their wellbeing. Our researchers should proactively consider various alternatives and choose those, which will lead to sustainable development of our society.

Ethics plan

KInIT researchers should conduct research activities with good aims and objectives, follow their own moral compass to stay aligned with our values and universal moral principles like beneficence, non-maleficence or transparency with the emphasis on the fundamental rights of people affected by these research activities. By developing an Ethics plan we aim to provide guidance for KInIT researchers by covering various types of research activities where ethical, legal or societal questions might arise, e.g.:

- Preparing research project proposals
- Preparing industry collaboration proposals
- Writing research papers
- Designing datasets and collection methods

These non-technical questions could be assessed from various perspectives that demand expert consultation or external know-how, namely in these areas of expertise:

- Legal (GDPR and other EU digital regulations)
- Ethics formally given in project binding documents
- Paper submission process/conference papers
- IP management - take into account future exploitation opportunity
- Ethical issues regarding specific cases not covered by other methods

General rules for KInIT researchers that will follow our ethics plan are proposed in Appendix A General rules (Ethics plan). As a support to the ethics plan we prepared guidelines, frameworks and processes for the assessment of ethical, legal and societal issues regarding the research activities in KInIT. Some of them are already completed (e.g. Ethics checklists, Research data policy) and some are in the final phase of their preparation (e.g. Code of ethics, KInIT Ethics committee statute). Every part of the ethics plan is verified with the KInIT Ethics Task Force, which includes members from every research team in KInIT. We hope that these tools will help KInIT researchers to navigate them towards responsible and trustworthy practices.

Values

KInIT research is built upon several key values[1]. Quality as a demand for the best from ourselves and our partners to ensure that we constantly improve and by nurturing our

[1] https://kinit.sk/about-us/

employees to grow and lead by example. Transparency that guides our decisions, how we verify our hypotheses, present our results, but also in the communication and cooperation with other researchers, partners, institutions, and the general public. Honesty in all of our activities with the emphasis on open communication, information sharing and alignment with open-science principles to promote collaborativeness and reproducibility but at the same time by keeping confidentiality of data and information when needed. Diversity as a belief that everyone has the right to participate in the fulfilment of research goals to the best of their abilities and capabilities in diverse and open-minded teams. And curiosity that pushes the boundaries of our knowledge in an environment of creativity and trust, where mistakes are meant to teach us and make us better.

Red lines

Although we understand that all research should be conducted as a free and open human endeavour, we are aware of red lines and prohibited practices that exist also in the field of intelligent technologies. We respect such red lines and avoid domains that pose harm to human life or health or directly manipulate human beings and violate basic human rights. Therefore, we refuse collaboration in research activities that support such practices. In general, we declare that we will not knowingly conduct research that directly or indirectly leads to unfair harms on others and we will not partake in activities that may lead to such harms, such as:

- Support of undemocratic processes and regimes
- Bribery and frauds in all forms of corrupt behaviour
- Direct military operations
- Irresponsible plundering of natural resources
- Exploitation of children vulnerability and marginalised groups
- Gambling and narcotics
- All activities endangering the fundamental democratic values

Research data policy

KInIT acknowledges that research in intelligent technologies carries various legal and ethical risks regarding the data collection, processing and sharing. With the aim to mitigate these risks, KInIT is implementing its own Research Data Policy. Research Data Policy specifies processing of personal data for research purposes in KInIT. It defines scope of research activities, research purposes, compliance with specific obligations and specific rules for most common research activities like research of systemic risks or training of AI models. As for the annotation process there is also the process that will help to support the wellbeing of annotators and research participants.

Code of ethics

The code of ethics will serve to conceptualise and understand our view on responsible research in intelligent technologies and to introduce the set of principles that should serve as the moral compass for our employees. We are working on a second draft of KInIT code of ethics that will be aligned with other international codes such as The European Code of Conduct for Research Integrity[2] and should be understood as a way of setting out our moral views that we propose to follow. It will also guide the cooperation processes with our partners to maintain trustworthy partnerships (which are elaborated in Trustworthy partnerships checklist). It will define our values, red lines and state eight principles aligned with these values:

- Principle 1: We maintain excellent quality in the process and outputs of our research
- Principle 2: We strive to achieve right competencies to strengthen our excellency and integrity
- Principle 3: We come to our decisions in a transparent and responsible manner
- Principle 4: We support honest and fair methods that respect people and their rights
- Principle 5: We value multidisciplinarity with respect to various stakeholders
- Principle 6: We create an inclusive and respectful work environment
- Principle 7: We cultivate curiosity and freedom of seeking new challenges
- Principle 8: We support the solutions with positive impacts on people and society

Ethics committee

In many cases where serious doubts arise that can not be overcome by researchers themselves or where its decision is formally needed, the ethics committee should be contacted. We plan to establish the KInIT ethics committee in the following months. Committee members will be divided into working groups depending on the perspective that needs to be assessed. KInIT researchers should contact the working group that relates to the type of issue they need to cope with:

- Legal working group - 2 members (GDPR representative, IT lawyer)
- Ethics working group - 2 members (IT ethicists)
- Projects working group - 1 member (projects office representative)
- CEO/CRO or other board representatives - 2 members

The members of the ethics committee will be nominated as domain experts (legal and ethics working group) or as KInIT representatives (projects working group and board representatives) and will work in a close collaboration with the Ethics Task Force.

Ethics checklist

Ethics checklist is designed as a self-assessment tool to help ensure that our research project meets ethical standards and respects the rights and well-being of any participants involved. Ethics checklist consists of more than 30 questions in three categories – General questions related to research activity, Data and Algorithms.

[2] https://allea.org/code-of-conduct/

Researchers must provide as much detail as possible and ensure that all information is accurate and up-to-date.

In the DisAI project we have assessed two research activities – claim matching research and gender stereotypes in machine translation and language modelling. Regarding the claim matching we research how to automatize claim retrieval – the task of retrieving an appropriate fact check based on an input claim – by using AI methods considering both multilingual (multiple languages) and multimodal (including visual information) variants of this task. Regarding the research of gender stereotypes we collect stereotypical statements and observe how AI models behave when they are exposed to them to measure degrees of stereotypical reasoning.

One of the most pressing ethical issues we have identified were the possible negative impacts on moral integrity, physical and mental well-being of researchers who might be exposed to various disinformation narratives, including narratives that attack minorities, human rights etc. With the aim to prioritize the well-being of researchers and annotators when engaging with distressing or disturbing content we proposed guidelines for them consisting of simple steps that will help them not only to identify the signs and symptoms indicating the erosion of well-being but also to establish clear mechanisms to address and minimize these issues.

Although all the collected posts are already present in a publicly available space on social networks and the publicly available fact-checks, we have also considered possible impacts on data privacy and human dignity as some personal data might be part of the social media posts (e.g., names mentioned in their content) and can be possibly misused by other stakeholders in the future. We decided not to release them freely to protect the privacy of social media users. However, as all the data will also become part of the Monant data platform collection, there needs to be appropriate measures taken to maintain data rights of the authors of the social media posts in the Monant platform and its datasets.

One of the issues that emerged during the assessment of gender stereotypes research was the uncertainty in key concepts such as stereotype, fairness, bias, or discrimination that might lead to inaccurate evaluation of our research. Therefore, we followed the outcomes from the workshop where we discussed various concepts and definitions of fairness with focus on concepts of distributive and group fairness. The concept of stereotype and the list of stereotypes were derived from discussions and qualitative workshops with experts on gender bias and approved by some of these experts. Still it holds that the definitions and list reflect the perception of researchers, data creators or bias experts and they should be always reconsidered when starting new research activity.

Also the researchers should be aware of the diversity of the research team and various forms of evaluation bias that might be present in their research, e.g. when considering non-binary language variations or gender diversity of data creators and annotators. These limitations should be properly communicated in all publicly available outcomes of their research activities.

3 IPR and Innovation Management plan

Introduction

The aim of the DisAI project is to enhance the scientific excellence of KInIT and the consortium partners in trustworthy AI and multimodal natural language processing and multilingual language technologies to combat disinformation. It is increasingly important to develop tools and methods for disinformation combating also for low resource languages including Slovak language. The ambition of KInIT is to strengthen AI to combat disinformation, an area that especially in Slovak digital space contributes to the development of AI technologies and tools for low resource languages and growing awareness of online disinformation. Together with leading partners German Research Center for Artificial Intelligence, Centre for Research and Technology Hellas and Copenhagen University (Partners), a joint research project on claim matching as well as networking and mobility activities will be organised. Importantly, transfer of know-how on innovation and creativity for facilitating industry-academia collaboration and boosting the skills of the research managers and administrative staff will be part of the capacity building programme.

The goal of the task is to outline intellectual property (IP) related strategies to ensure that relevant knowledge is made accessible to those who need it and will benefit from it. The DisAI consortium members are aware that appropriate intellectual property rights (IPR) regulations and their management are critical for the project results and, therefore, general IPR rules will apply to the project. Bearing in mind the collaborative nature of the project, IP will be managed on an equitable basis by the partners who generate the knowledge and know-how. This knowledge will be protected in the form of intellectual property rights including copyright, database rights or protection of know-how and trade secrets.

Legal framework

The DisAI IPR Management plan within the consortium and third parties complies with the rules defined in the Grant Agreement and conditions agreed in the Consortium Agreement including the general rules and recommendations defined for projects of the European Union’s current Framework Programme for Research and Innovation “Horizon Europe” (2021-2027).

Specific rules regarding intellectual property rights, results and background are set out in Annex 5 of the Grant Agreement.

Further delineations of intellectual property rights are also presented in the Article 8 and Article 9 of the Consortium Agreement. The DisAI Consortium Agreement particularly covers:

- Ownership and joint ownership of Results
- Transfer of Results

- Obligations regarding dissemination Results
- Handling of Background IP
- Handling of Access Rights
- Specific Provisions for Access Rights to Software.

Definitions

According to the Grant Agreement (Article 16.1) *Background* is defined as "any data, know-how or information — whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights.

According to the Grant Agreement (Article 16.2) *Results* are defined as "any tangible or intangible effect of the action, such as data, know-how or information, whatever its form or nature, whether or not it can be protected, as well as any rights attached to it, including intellectual property rights."

In the context of the Consortium Agreement and Grant Agreement *Access rights* means rights to use results or background.

Results

According to the Annex 5 of the Grant Agreement, Partners which have received funding under the grant must adequately protect their results — for an appropriate period and with appropriate territorial coverage — if protection is possible and justified, taking into account all relevant considerations, including the prospects for commercial exploitation, the legitimate interests of the other Partners and any other legitimate interests.

Transfer of results

According to the Article 8.3.1 of the Consortium Agreement each Partner may transfer ownership of its own Results, including its share in jointly owned Results, following the procedures of the Grant Agreement Article 16.4 and its Annex 5, Section Transfer and licensing of results, sub-section "Transfer of ownership".

Ownership

According to the Annex 5 of the Grant Agreement, results are owned by the Partners that generate them. Article 8 of the Consortium Agreement confirms that results are owned by the Partner that generates them.

The potential joint ownership shall be agreed in writing between Partners, if they have jointly generated results and it is not possible to establish the respective contribution of each beneficiary, or separate them for the purpose of applying for, obtaining or maintaining their protection.

According to the Article 8.2 of the Consortium Agreement, unless otherwise agreed:

- each of the joint owners shall be entitled to use their jointly owned Results for non-commercial research (including but not limited to research contracts in national and European funded projects with third parties provided it does not lead to any commercial or monetary benefit to third parties involved in such cooperative research project), and teaching activities on a royalty-free basis, and without requiring the prior consent of the other joint owner(s).
- each of the joint owners shall be entitled to otherwise exploit the jointly owned Results and to grant non-exclusive licences to third parties (without any right to sub-license), if the other joint owners are given: (a) at least 45 calendar days advance notice; and (b) fair and reasonable compensation.

Dissemination of results

Dissemination is governed by the Article 8.4 of the Consortium Agreement. During the project and for a period of 1 year after the end of the project, the dissemination of own (including jointly owned) Results by one or several Partners including but not restricted to publications and presentations, shall be governed by the procedure of Article 17.4 of the Grant Agreement and its Annex 5, Section Dissemination, subject to the following provisions.

Prior notice of any planned publication shall be given to the other Partners at least 10 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement by written notice to the coordinator and to the Partners proposing the dissemination within 4 calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is permitted. Further provisions specifying the process are presented in the Consortium Agreement.

Partners agreed to comply with Open science policy as foreseen by Annex 5, Article 17 of the Grant Agreement. Following the guidelines of the EU for Open Access to scientific knowledge produced within the European funded projects, DisAI is committed to utilise measures for allowing open access to outputs wherever the joint commercial and industrial exploitation allows doing so. Public project deliverables will be uploaded on the project website to be open and freely accessible. All project scientific publications will be published as open access under suitable licences; DisAI partners will also make use of Open Research Europe to increase the visibility of results and to enhance in-project and post-project awareness.

Our overall aim is to make the DisAI project part of the dynamic open science community, promoting the adoption and support of FAIR principles, i.e., Findable, Accessible, Interoperable, Reusable (with exceptions notably for commercial purposes) in the fields of industry-academia cooperation, i.e., promoting a set of guiding principles that will lead to a higher degree of reusability and sustainability of data and digital artefacts in the domain.

Background

We summarise information about background based on the DisAI Consortium agreement, Attachment 1 and provide it in the Appendix B Background of the project.

Intellectual property rights management plan

Intellectual Property management is the process of effectively capturing, sharing and organising IP. In DisAI, it aims to provide the best use of the IP generated in different work packages within the consortium.

Main IP challenges during the project implementation is to analyze the existing and potential knowledge creation and management tools, and to discuss possible IP protection (e.g. patents, copyright or know-how). Proper handling of IP rights during the project implementation is important as the management and ownership of the results is a basis for possible future exploitation. After the project is finished,, challenges of the IPR management are definition of joint exploitation possibilities,identification of possible IP ownership arrangements and related responsibilities, and discussion of potential agreements options on use of IP resulting from the project.

DisAI IPR management aims to effectively capture IP generated in different work packages and the relevant ownership, and to use the best efforts to exploit the results adequately protected where justified.

IPR management plan follow these steps:

- Results identification
- Results description/recording with identification of ownership,
- Establishing the relevant intellectual property rights protection measures
- Planning and performing the dissemination and exploitation,
- Evaluation

The Background of the project has been identified in the Consortium Agreement and is also part of this deliverable.

Results expected in the project may be derived from deliverables as presented in the Annex 1 of the Grant Agreement. The expected results and the proposed following IPR management of results are:

No	Result	Public or non-public	Proposed IP protection
D1.1, D1.2	Scientific strategy report	Non-public	Copyright and Non-Disclosure agreements
D1.3	Scientific webinar series report	Public	Copyright
D1.4	Shared task report	Public	Copyright
D1.5	Replication challenge report	Public	Copyright

D1.6	Summer school report	Public	Copyright
D2.1	AI methods for claim matching	Public	Copyright
D2.2	Infrastructure documentation	Non-public	Copyright and Non-Disclosure agreements
D2.3	Collected labelled dataset	Public	Copyright and Database Rights
D2.4, D2.5	Scientific staff mobilities scheme	Non-public	Copyright and Non-Disclosure agreements
D3.1	KInIT research management/ administration assessment report	Non-public	Copyright and Non-Disclosure agreements
D3.2	Training report on opportunity seeking and scientific proposal writing	Public	Copyright
D3.3	Training report on innovation, technology transfer and networking	Public	Copyright
D3.4	Training report on research management support and operation	Public	Copyright
D4.1, D4.2	Plan for dissemination and exploitation including communication	Public	Copyright
D4.3	Visuals and branding materials	Public	Copyright and Design rights
D4.4, D4.5	Report on dissemination, communication and synergy activities	Public	Copyright
D4.6	Report on CLARIN integration and other liaison activities	Public	Copyright
D5.1, D5.2, D5.3	Data management plan	Public	Copyright
D5.4	Ethics, innovation and IPR management plan	Public	Copyright
D5.5	Progress report	Non-public	Copyright and Non-Disclosure agreements

For each project Result, key elements will be identified, like whether the ownership is of two or all three partners, the name of the owners, the country of establishment of the owners and whether the results will be disseminated and/or exploited. Each Result will be assessed if it is likely to be commercially used, valuable for increasing the scientific knowledge, or if it is not for commercialization nor for public dissemination but can be used for internal improvements and research.

Role of the Scientific Coordination Committee

Management of IPR is part of the tasks of the Scientific Coordination Committee (SCC) established based on Article 6.3.1 of the Consortium Agreement. According to the Article 6.3.1.2, SCC shall take decisions regarding the intellectual property rights.

SCC decides how dissemination and exploitation strategies will be applied together with continual monitoring and evaluation of the IPR.

Innovation Management

The goal of innovation management is establishing a systematic approach to foster innovation in the project. The primary goal is to identify and implement activities that contribute to the structured and consistent enhancement of the project's innovation capacity and performance.

This endeavour is centred around tho major initiatives:

- New ideas generation and management
- Development of new projects
- Protection and exploitation of project’s results

Protection and exploitation of project’s results are involved as a part of the innovation management agenda. Besides this deliverable, these are addressed also in the deliverable D4.2 Dissemination and Exploitation Including Communication 2.

Appendix A: General rules (Ethics plan)

1. Check whether the research activity is aligned with the KInIT Code of ethics
   - if the research activity is in conflict with principles stated in Code, or you are afraid of potential harms of research on society, or you think the activity could lie behind the red lines, contact KInIT ethics committee (*Ethics working group*)
2. Read the Checklist for ethical partnerships to identify whether the potential business partner (PBP) is suitable for us or such partnerships could pose too high (reputational, moral, but also business) risks
   - If PBP is not recommended or PBP needs further deliberation ask and wait for further instructions from KInIT ethics committee (*Projects working group*)
3. Check the alignment of planned activities with Research Data Policy
   - If there are any conflicts or open questions that you can not resolve, contact KInIT ethics committee (*Legal working group*)
4. Check the possibility of keeping some intellectual property rights in KInIT possession or look for the potential of the re-usability of project outcomes in other research activities.
   - If you are not sure about IP rights and potential of re-usability of project outcomes, discuss them with KInIT ethics committee (*Legal working group*)
5. Fill the Ethics checklist regarding ethical and legal issues in the project. Check whether there are some ethical or legal risks related to data and AI
   - If you're not sure how to complete the form and which questions apply to your project, contact KInIT ethics committee (*Ethics working group*)
6. Ensure that you are acquainted with relevant document that should guide your actions (Code of research ethics, Checklist for partnerships, Research data policy) and that you have completed ethics checklist
   - If you need formal approval of proposal from KInIT ethics committee or you answered "Yes" for some questions in Ethics checklist, contact KInIT ethics committee and ask for the approval
7. Ethics committee will review the Ethics checklist and provide an formal opinion
   - If ethics committee will not identify any considerable ethical and legal issues it will approve research activity
   - In case of minor issues the ethics committee will give you a recommendation how to improve or cope with the issues in your research activity
   - If ethics committee identify the presence of considerable ethical and legal issues based on the information received during the conduct of the research project it can recommend to undergo full ethics-based assessment facilitated by Ethics team KInIT

Appendix B: Background of the Project

Background	Owner	Specific restrictions and/or conditions for implementation[3]	Specific restrictions and/or conditions for Exploitation[4]
Monant, Research platform for monitoring and detection of false information	KInIT	Access rights to Monant data and services through an API will be provided as necessary for the implementation of the action.	To be negotiated and based on a separate license agreement.
Image Verification Assistant: this is a software component implementing a number of image forensics algorithms and exposes them through a REST API and a web-based user interface.	CERTH	The service is offered for free to DisAI partners upon request but rate limited. Higher than normal limits may be made available to partners during the project as and when necessary.	Paid license is available for commercial use.
Context Aggregation and Analysis: this is a software component implementing metadata collection and analysis for images from Facebook and Twitter and videos from YouTube, Facebook and Twitter making the results available through a REST API and a web-based user interface.	CERTH	The service is offered for free to DisAI partners upon request but rate limited. Higher than normal limits may be made available to partners during the project as and when necessary.	Paid license is available for commercial use.
Near-duplicate multimedia retrieval: this is a software component that enables client applications to build and update large indexes of image and video content, which they can search by example. It is available through a REST	CERTH	The service is offered for free to DisAI partners upon request but rate limited. Higher than normal limits may be made available to partners during the project as and when necessary.	Paid license is available for commercial use.

[3] Article 16.4 Grant Agreement and its Annex 5, Section "Access rights to results and background", sub-section "Access rights to background and results for implementing the Action"
[4] Article 16.4 Grant Agreement and its Annex 5, Section "Access rights to results and background", sub-section "Access rights for exploiting the results"

API and a web-based user interface.			
Deepfake detection: this is a software component that detects deepfakes in images and videos. The results are available through a REST API and a web-based user interface.	CERTH	The service is offered for free to DisAI partners upon request but rate limited. Higher than normal limits may be made available to partners during the project as and when necessary.	Paid license is available for commercial use.
Visual Location Estimation: this is a software component that provides an estimate of the depicted location (if any) in an image based on the visual analysis of its content exposes it through a REST API and a web-based user interface.	CERTH	The tool is offered for free to DisAI partners upon request but rate limited. Higher than normal limits may be made available to partners during the project as and when necessary.	Paid license is available for commercial use.
MeVer NetworkX: This is a web-based tool that enables end users to browse/explore and gain insights from large social networks of Twitter, Facebook and Telegram.	CERTH	The tool is offered for free to DisAI partners upon request but rate limited. Higher than normal limits may be made available to partners during the project as and when necessary.	Paid license is available for commercial use.