Haiku
Human AI teaming Knowledge and Understanding for aviation safety

# D 1.3

# Social, Ethical, Legal, Privacy Issues Identification and Monitoring

**Authors: Elisa Spiller (DBL), Nikolas Giampaolo (DBL)**

## Information Table



This project has received funding from the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

| | |
|---|---|
| Deliverable Number | 1.3 |
| Deliverable Title | Social, Ethical, Legal, Privacy issues identification and monitoring |
| Version | Final version |
| Status | Proposed final |
| Responsible Partner | Deep Blue |
| Contributors | DBL; ECTL; FerroNATS; CHPR; LiU; TAVS; Bordeaux INP; DFKI; ENG; LFV; ENAC; TUI; Suite5; EMBRT; CERTH; LLA |
| Contractual Date of Delivery | 31/12/2022 |
| Actual Date of Delivery | 23/12/2022 |
| Dissemination Level | Public |

## Document History

| Version | Date | Status | Author | Description |
|---|---|---|---|---|
| 0.1 | 05/11/2022 | Outline | Nikolas Giampaolo | |
| 0.2 | 02/12/2022 | First Complete Draft | Nikolas Giampaolo | |
| 0.3 | 16/12/2022 | Proposed final | Simone Pozzi<br>Vanessa Arrigoni | Internal review |
| 1.0 | 21/12/2022 | Final | Brian Hilburn<br>Beatrice Thiebaux | Internal review |

# List of Acronyms

| Acronym | Definition |
|---|---|
| **AI** | Artificial Intelligence |
| **CA** | Consortium Agreement |
| **EC** | European Commission |
| **GDPR** | General Data Protection Regulation |
| **GA** | Grant Agreement |
| **HF** | Human Factors |
| **HLEG** | High-Level Expert Group (on AI) |
| **XAI** | Explainable AI |
| **WP** | Work Package |

## Executive Summary

This deliverable describes the measures taken to ensure that project activities are respectful of human rights, particularly the right to privacy and data protection, and do not generate unethical personal or social effects. The deliverable describes privacy, ethical and other legal concerns, and proposes mitigation measures to address them.

The report is divided into four sections including a short introduction, a description of the main points regarding ethical issues arising from AI, an overview of the ethical issues stemming from the HAIKU research project and the way the consortium plans to mitigate them, involving the processing of personal data and a set of measures that could be implemented to reduce the risk of misuse of the research results. The Appendix contains the AI Ethics issue checklist proposed by the EC (EU GRANTS, 2021).

This report is a living document and will be updated later in the project as new information regarding data collection activities and details regarding specific aspects of Social, Ethical, Legal, Privacy issues become available. The final version will be released at M36, as D1.4

# Table of Contents

# 1. Introduction

This deliverable describes the measures taken to ensure that project activities are respectful of human rights, particularly the right to privacy and data protection, and do not generate unethical personal or social effects. The deliverable describes privacy, ethical and other legal concerns, and proposes mitigation measures to address them.

## 1.1. Deliverable structure

This document is divided into eight sections:

- Section 1 contains a short introduction and document structure.
- Section 2 addresses main points regarding ethical issues arising in the HAIKU project.
- Section 3 provides an overview of the ethical and regulatory framework concerning the protection of personal data.
- Section 4 provides an overview of the ethical and regulatory framework concerning trustworthy AI.
- Sections 5 lists a set of measures and recommendations that could be implemented to reduce the ethical and legal risks related to HAIKU research activity.
- Section 6 presents the measures to prevent potential misuse of research results.
- Section 7 provides the conclusions stemming from the ethical assessment.

# 2. Ethical Considerations

## 2.1. HAIKU Ethical Issues

According to the EC Ethical Self-Assessment Checklist, the ethical issues that arise in the HAIKU project may concern:

- Participation of humans in 'experiments' and data collection activities
- Collection of personal data of front-end operators and staff of the end-user organisations through field observation, interviews - as primary data
- Personal data that come from secondary data—surveys, company video recordings, system logs, internal documents, etc.—provided by end user organisations. (Not foreseen nor planned at this stage; if applicable at a later stage, full compliance to Ethical requirements will be ensured, and the present document updated)
- HAIKU involves a non-EU partner from Brazil and stakeholders providing data from the UK. Non-EU countries will be required to comply with the legal and ethical requirements for participation in EU research and H2020 guidelines
- As in every project that collects and analyses personal data and relevant business data, the potential misuse of the research results are possible
- Development, deployment and/or use of Artificial Intelligence-based systems able to interact, replace or influence human decision-making processes

In light of the above, the following paragraphs will specifically approach each of these issues. Particular attention will be devoted to the new issues related to the use of AI-powered solutions. In this regard, consistently with the approach suggested by the EC and HLEG AI, this document will introduce the AI Ethical Assessment Methodology adopted for the specific purposes of HAIKU.

## 2.2. Research activities concerning the use of AI

Ethical and legal issues are always a delicate matter for a study that involves users and especially when collecting personal data and involving the use of AI such as the proposed study. When it comes to technology and data sharing in general, where personal data are collected, there exist obvious public concerns: the difficulty of respecting privacy and confidentiality when third parties have a strong interest in accessing electronically recorded and stored personal health data, or the difficulty in ensuring the security of shared personal monitoring data. Moreover, the use of AI raises additional concerns about human-machine interactions mainly related to the respect of human dignity and autonomy, non-discrimination and liability apportionment in human-AI teaming.

In light of the above, this document primarily relies on the following references and guidelines:

- Ethical Guidelines for Good Research Practice
- Code of Human Research Ethics
- UNESCO Recommendations on the Ethics of Artificial Intelligence
- HLEG AI Ethics Guidelines For Trustworthy AI [10]
- HLEG Assessment List for Trustworthy AI (ALTAI) for self-assessment

- European Commission Note for Identifying serious and complex ethics issues in EU-funded research, published on 5 July 2021
- European Commission Note on Ethics and Data Protection, published on 5 July 2021

The Consortium will conform to the rules and legislation here listed.

The members of the HAIKU project consortium followed the national legislation and the non-EU countries were asked to comply with the legal and ethical requirements for participation in EU research and Horizon Europe guidelines.

## 2.3. Ethical principles for the volunteer participants to HAIKU activities

For each of the research activities, which contain user studies or involve participants in the trials and testing, the necessary ethics approvals (if required) and the free and fully informed consent of the research participants will be obtained. Following the general principle of ethical research, focus groups and structured interviews will be strictly limited to volunteers, mainly employees of partners' organisations.

The HAIKU Consortium will process Personal data on a lawful basis, as provided in Articles 6 and 9 of the GDPR. Notably, the processing of Personal data will be based on **Consent** (for common Personal data) and **Explicit Consent** (for special categories of Personal data).

The procedure for volunteers will strictly adhere to the conditions below:

- Every volunteer has the right to remain anonymous;
- Every volunteer will have his or her data protected as stated in the applicable data protection legislation;
- All data released by the eventual volunteers participating in the project results evaluation brings a duty of confidentiality;
- Every volunteer's free and informed consent is required also while doing interviews, ethnographic observation, non-invasive experimentation, neurophysiological data collection (if used), and accessing personal data records. The purpose of informed consent is to empower individuals to make a voluntary informed decision about whether or not to participate in the research based on knowledge of its goals, procedures and outcomes and therefore no volunteers that are minors and with reduced autonomy or vulnerability will be involved;
- Every volunteer will also have the right to know who will benefit from his or her participation to the experiment and, if they so wish, to receive information of any potential commercial exploitation of the research that involves their participation;
- Participants will be informed about the purpose of the research, the vehicles for the dissemination of the results and people, organisations and stakeholders involved in the research, and also about data treatment and handling;
- Volunteers willing to be informed on the results of the research will be timely updated with the outcomes of the study;
- Volunteers will also be informed of the appropriate insurance cover that is in place, if applicable;
- It will be clearly explained to the volunteers that they can withdraw from the project at any time;
- The data will be collected and processed in a fair and possibly unbiased manner (more details available in D1.2);
- Fair involvement, equal opportunities and equal treatment among people with a different socio-cultural background (e.g. gender, nationality, religion, age, etc.) will be guaranteed.

HAIKU will obtain written consent from all research participants, who will be informed about the research and about the consequences of their participation. They will receive information about the methods, the timeframe and the environment in which their data will be studied. The individuals will have a contact point (a responsible authority within their organisation) in case they need to receive additional information regarding their participation. The participants will be informed about the opt-out procedure.

The information sheets, accompanying the informed consent form, will be written in a language and in terms fully understandable to the participant. The sheets will describe the aims, methods and implications of the research, the nature of the participation and any benefits, risks or discomfort that might be involved. The information sheet will explicitly state that participation is voluntary and that anyone has the right to refuse to participate and to withdraw their participation, samples or data at any time — without any consequences. Only when the research participant has fully understood the information sheet and has given consent without any pressure having been put on them, then consent will be considered valid. For more information about the data collection and the forms utilized in the different experiments please refer to D1.2 “Data Management Plan”.

# 3. Personal Data Protection

## 3.1. Ethical and legal framework

HAIKU research activities comply with the principles of dignity, freedom, equality, solidarity, citizens' rights and justice, as stated in the Charter of Fundamental Rights of the European Union. When performing experiments and collecting data for the HAIKU project, we use particular care for privacy and data protection rights, considering them as fundamental rights to be protected.

In particular, the HAIKU Consortium ensures ethical and legal compliance with current regulations:

- the EU Charter of Fundamental Rights (art. 7 and 8);
- the European Convention for the Protection of Human Rights and Fundamental Freedoms (art. 8);
- the General Regulation (EU) 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 (GDPR);
- the Directive on Privacy and Electronic Communications (2002/58//EC);
- the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data No.108/1981;
- the Article 29 WP Guidelines on Consent under Regulation 2016/679 (wp259rev.01);
- the EDBO Guidelines 5/2020 on Consent under Regulation 2016/679 (v. 1.1 adopted on 4 May 2020);
- the Article 29 WP Guidelines on Transparency under Regulation 2016/679 (wp260rev.01);
- the Article 29 WP Opinion 05/2014 on Anonymization techniques (wp216);
- the Article 29 WP Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01).

## 3.2. Principles and general rules

For the purposes of this deliverable, from now on the definitions adopted are the following. These latter refer to those contained in Articles 4 and 9 of the GDPR.

- **Personal data** is intended as any information relating to an identified or identifiable natural person (data subject); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.
- **Processing** includes any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.
- **Controller** is intended as the natural or legal person, public authority, agency or another body which, alone or jointly with others, determines the purposes and means of the processing of personal data;

where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law. In the HAIKU project, every partner is a controller and responsible for the data that it collects and processes. All the partners are committed to the respect of the principles and rules listed in the present document and in the following versions of it.

- **Processor** is intended as a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller. On the other hand, **recipient** may be any natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not.

Generally, in light of what provided by the GDPR as well as by FAIR practices [4], personal data will be processed according to the principles outlined by Article 5 of the GDPR, namely:

- **Lawfulness**, **fairness and transparency** in relation to the data subject;
- **Purpose limitation**: Personal data will be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;
- **Data minimization**: Personal data will be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed;
- **Accuracy**: Personal data will be accurate and, where necessary, kept up to date; data were not be modified or falsified.
- **Storage limitation**: Personal data will be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed;
- **Integrity and confidentiality**: Personal data will be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures;
- **Data protection by design and by default**: The controller implemented appropriate technical and organisational measures which are designed to implement data-protection principles in an effective manner and to integrate the necessary safeguards into the processing in order to protect the rights of data subjects. The controller implemented appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed;
- **Accountability**: The controller is responsible for and be able to demonstrate compliance with the above-mentioned principles.

The legal basis for the processing of personal data is **consent of the data subject**. Consent is considered valid if freely given, specific and informed. The data subjects are free to withdraw their consent according Article 6 GDPR. HAIKU research will NOT include children, adults unable to give informed consent, nor vulnerable individuals/groups.

Within the HAIKU research activities, data subjects are able to exercise their rights on the processing of personal data concerning them according to the rules and conditions provided by the General Data Protection Regulation and the related internal policies.

The HAIKU Consortium is committed to ensuring the effective protection of Personal data of all participants to its activities. More detailed information, including those related to the security measures, will be available in D1.2 “Data Management Plan”.

## 3.3. Measures to Prevent Potential Misuse of Research Results

The HAIKU consortium tries to avoid the misuse of the research results. In particular, the HAIKU partners favour a correct and rightful use of the results of the research, as well as of the knowledge of the techniques and theories applied, they avoid any negative repercussion against the public reputation of each partner and of any other involved entity or individual and improper exploitation of research results for obtaining commercial and business advantages, personal profits and any other kind of unlawful application of the research results.

Main mitigation measures to avoid the potential misuse of research results, both from partners of the Consortium and from external malicious actors are:

-Monitoring access to data and researcher actions on data.

-Defining and implementing policy access based on least privilege.

-Protecting data from outsiders.

Multiple security mechanisms and technologies ensure protection from malevolent/criminal/terrorist abuse. For more information please refer to section 5 of D1.2 “Data Management Plan”.

Further, strict procedures are defined for publication of research results, any publication of final or intermediate results must comply with the following statements:

1. Everyone who participated in the collection of data will be acknowledged in the acknowledgements section of the publication, either individually or collectively. Co-authorship of publications will be determined and agreed based on standard academic conventions.
2. Information, when explicitly classified as sensitive, will be presented only if necessary, complying with relevant NDA conditions;
3. Political and socio-economic concerns must be carefully considered in presenting the results as any opinion expressed could affect the public opinion feelings about safety in aviation transport.

# 4. AI Ethical Issues

## 4.1. Ethical and legal framework

HAIKU research activities comply with the principles of dignity, freedom, equality, solidarity, citizens' rights and justice, as stated in the Charter of Fundamental Rights of the European Union. When performing experiments and collecting data for the HAIKU project, we pay particular care to privacy and data protection rights, considering them fundamental rights to be protected.

In particular, the HAIKU Consortium ensures ethical and legal compliance with current regulations and guidelines:

- the EU Charter of Fundamental Rights (spec. art. 1, 7, 8 and 21);
- the European Convention for the Protection of Human Rights and Fundamental Freedoms (art. 8 and 14);
- HLEG AI Ethics Guidelines For Trustworthy AI
- HLEG Assessment List for Trustworthy AI (ALTAI) for self-assessment
- Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts (COM/2021/206 final)
- Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) (COM(2022) 496 final)

## 4.2. Definition

For the purposes of this deliverable, from now on the definitions adopted are the following.

- **Artificial intelligence (AI)** generally refers to "to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions"[10].

However, "as a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)"[10].

In light of what is required by HAIKU research activities, the general definition adopted from the High-level expert group on artificial intelligence will be tailored to the different necessities of each use case taking into account that HAIKU integrates different AI applications (e.g. Machine Learning, digital assistant, etc.)

## 4.3. Principles and General Rules

Generally, in light of what provided by the HLEG AI Ethics Guidelines For Trustworthy AI [10], within the HAIKU research project AI relies on the following principles:

1. **Respect for Human Agency**: To allow individuals to make their own choices and act in accordance with them, respect is required. Three additional specific principles that define fundamental human rights include respect for human agency: dignity, freedom, and autonomy;
2. **Privacy and Data governance**: Privacy and data security are fundamental human rights that must be upheld at all times;
3. **Fairness**: Equal rights and opportunities should be afforded to everyone, and unfair advantages and disadvantages should be avoided;
4. **Individual, Social and Environmental Well-being**: AI systems should contribute to, and not harm, individual, social and environmental wellbeing;
5. **Transparency**: Stakeholders should be able to comprehend AI programs' objectives, inputs, and operations;
6. **Accountability and Oversight**: humans should be able to understand, supervise and control the design and operation of AI based systems.

In light of the above, the development, deployment and use of AI is compliant with the following requirements:

1. **Human agency and oversight:** A machine cannot be in full control: entailing the right of end users not to be subject to a decision based solely on automated processing. As a result, human oversight should always be present. Humans should always have the ability to change a system's decision in the end. AI developers ought to take into consideration the kind of technical measures that ought to be put into place to guarantee human oversight when designing an AI product or service. To ensure human control, they should, for instance, provide a stop button or a procedure for aborting an operation.
2. **Technical robustness and safety:** Algorithms for trustworthy AI must be secure, dependable, and robust enough to deal with errors or inconsistencies throughout the entire AI system's lifecycle. When developing algorithms, all vulnerabilities should be taken into account in practice. In order to comprehend and reduce the dangers of hacking and cyberattacks, this necessitates testing AI systems. In the event that someone uses the AI system they are developing for harmful purposes, AI developers should establish procedures that are able to assess the safety risks involved.
3. **Privacy and data protection:** As a matter of principle, all AI stakeholders in the EU must adhere to the General Data Protection Regulation (GDPR), which places a high value on data privacy and security. Data owned by citizens should not be used to harm or discriminate against them, and citizens should have complete control over their own data. Data anonymization and encryption should be used by AI developers to accomplish this.
4. **Transparency:** To ensure that AI is impartial, transparency is essential. A number of measures to ensure transparency in the AI industry are outlined in the AI guidelines. For instance, the data sets and procedures utilized in the construction of AI systems ought to be traceable. In addition, AI systems ought to be easy to identify, and humans ought to be aware that they are interacting with an AI system. In addition, AI systems and human decisions are subject to the explainability principle, which states that humans should be able to understand and follow them.
5. **Diversity, non-discrimination and fairness:** When designing AI products and services, the guidelines place a strong emphasis on avoiding unfair bias. In practice, AI developers should ensure that their algorithms are not biased in design (for example, by using an insufficient data set). Stakeholders who might be impacted in some way by AI systems ought to be consulted and a part of their design and

implementation. The entire range of human abilities, skills, and requirements should be taken into account when designing AI systems, and accessibility for people with disabilities should be ensured.

6. **Societal and environmental wellbeing:** Positive social change and AI system sustainability and environmental responsibility should be promoted by using AI systems. Measures that ensure the environmental friendliness of AI systems should be encouraged (such as choosing a method that uses less energy) and the social impacts of these systems on people's physical and mental health should be monitored and taken into account. Also, it is important to look at how AI systems affect democracy and society, including how they affect elections.
7. **Accountability:** It is necessary to establish mechanisms that guarantee accountability for AI systems and the outcomes they produce. AI systems whose use violates fundamental rights should be subject to both internal and external independent audits. Decisions regarding the trade-off—that is, the decision to choose to fulfil one ethical requirement over another—should be continuously evaluated in situations where the implementation of the key ethical requirements creates conflicts between them. Redress mechanisms that are accessible should be implemented.

In particular, considering the ethical principles and the fundamental rights involved in the development, deployment and use of the AI-powered solutions within HAIKU research, the Consortium opts for a proactive approach. This choice is consistent with the three-pronged approach suggested by the EC and HLEG AI, and relying on the so-called Ethics by Design.

Ethics by Design can be used to make sure that the ethical requirements of an AI system or technique are properly addressed during its development. The purpose of this strategy is to make it as clear as possible what is required, to describe the specific tasks that must be carried out, and to assist developers in considering ethical considerations when creating an AI system [2].

In some AI projects, the relevant ethical issues may be discovered during the development phase, while in others, they may be discovered after the system has been deployed. By addressing ethical issues during the development phase rather than attempting to fix them later in the process, Ethics by Design aims to prevent ethical issues from occurring in the first place. By proactively utilizing the principles as system requirements, this is accomplished. A five-layer model describes Ethics by Design [2].

The aim of Ethics by Design is to make people think about and address potential ethics concerns, while they are developing a system [2].

The main ethical requirements for AI and robotics systems above can be summarised as:

- The dignity, freedom, and autonomy of human beings must not be compromised by AI systems.
- The right to personal data protection and privacy must not be violated by AI systems. They must use data that is necessary, representative, and free of bias.
- AI systems must be developed without discrimination and with an inclusive, fair agenda.
- It is necessary to take precautions to ensure that AI systems do not inflict harm on individuals, societies, or the environment, rely on harmful technologies, persuade others to take harmful actions, or lend themselves to function creeps.
- AI systems ought to be as open as possible to their users and stakeholders.
- To address noncompliance and ensure compliance with these principles, human oversight and accountability are required.

# 5. HAIKU Ethical Issues and Mitigation Measures

Considering the specific research activities proposed by the HAIKU project, the use of AI is aimed at the development of digital assistants. In particular, the different solutions presented in the use cases involve technologies with different level of automation and autonomy. Their analytical, decisional and operational scope of action may significantly vary according to the specific needs of the contexts of use.

The general concerns emerged from EC Ethical Self-Assessment Checklist about the development, deployment and/or use of Artificial Intelligence-based systems able to interact, replace or influence human decision-making processes here should be furtherly detailed. More specifically these are the main AI ethical issues related to the HAIKU project:

- Human-AI relationship
- Explainability
- Safety
- Accountability
- Biases

These issues will be specifically addressed in the following pages, explaining the mitigation measures and intended requirements the Consortium aims at implementing in HAIKU research activities. Eventually, the document reports the list of measures adopted to mitigate emerging ethical risks over the project

## 5.1 Human- AI relationship

The scope of HAIKU is to develop a human-AI relationship for different AI aviation applications that preserves and enhances **Human Autonomy**. In particular, HAIKU aims to answer the following research question: How to support the human with AI, without the human feeling under-valued or losing motivation and 'investment' in safe operations, or becoming 'out-of-the-loop', losing situation awareness and the ability to take back control when needed?

The answer relates to deciding where the AI can best help without detrimental impact on human performance. HAIKU is particularly focused on this aspect, which is also in line with leading European thinking on the way we would all like AI to evolve, as stated by Humane AI (https://www.humane-ai.eu/). The core of the humane AI vision is as follows: "The development of robust, trustworthy AI systems capable of understanding humans, adapting to complex real-world environments, and appropriately interacting in complex social settings. The overall vision is to facilitate AI systems that enhance human capabilities and empower individuals and society as a whole while respecting human autonomy and self-determination."

The HAIKU project is implemented via six aviation Use Cases, capturing a range of aviation segments, time horizons (from tactical to strategic), task complexity, and human impact. Each Use Case requires a tailored AI concept - with specific purposes, expected AI benefits for operations and for humans, underlying values and design principles. The Use Cases proceed as parallel strands but adopt a harmonized approach in terms of methods, validation exercises, and expected outcomes. This ensures cross-fertilisation among the Use Cases, whenever lessons learnt, and findings allow. The rationale for this approach is that AI-powered assistants can support humans in a variety of tasks. They can support the execution of simple tasks - be they for perception, decision-making or execution, using the established cognitive framework - or they can partner with humans to monitor and manage overall system status. In some Use Cases, AI will simply speed up human performance or substitute for the humans, while in other instances the use of AI will transform the current

tasks and deeply change the nature of work. Effective human-AI partnerships should be designed differently in all of these different cases, depending on the task complexity and on the expected benefits of "work as done with AI."

The deployment and use of AI systems may implicate fundamental rights and their underlying values. In order to ensure an adequate Human-AI interaction, the consortium abides to the following principles, following from the High-Level Expert Group on AI[10]:

- **Respect for human dignity.** Human dignity encompasses the idea that every human being possesses an "intrinsic worth", which should never be diminished, compromised or repressed by others – nor by new technologies like AI systems. In this context, HAIKU aims to ensure that all people are treated with respect due to them as moral subjects. AI systems are therefor designed to augment, complement and empower human cognitive, social and cultural skills. HAIKU aims to do this by following human-centric design principles securing human oversight over work processes in AI systems.
- **Freedom of the individual.** Human beings should remain free to make life decisions for themselves. In an AI context, freedom of the individual for instance requires mitigation of (in)direct illegitimate coercion, threats to mental autonomy and mental health, unjustified surveillance, deception and unfair manipulation. HAIKU commits to enabling individuals to wield control over their lives, including (among other rights) protection of the freedom to conduct a business, freedom of expression, the right to private life and privacy, and freedom of assembly and association.
- **Respect for democracy, justice and the rule of law.** AI systems should serve to maintain and foster democratic processes and respect the plurality of values and life choices of individuals. In this respect, HAIKU ensures that AI systems do not undermine democratic processes, human deliberation and mandatory laws and regulation.
- **Equality, non-discrimination and solidarity - including the rights of persons at risk of exclusion.** Equal respect for the moral worth and dignity of all human beings must be ensured. HAIKU commits to the aim that the system should not generate biased outputs. In order to mitigate this ethical issue, the consortium respects potentially vulnerable persons and groups, such as workers, women, persons with disabilities, ethnic minorities, children, consumers or others at risk of exclusion.
- **Prevention of harm** AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity. HAIKU commits to the fact that the AI systems and the environments in which they operate must be safe and secure, technically robust, and not open to malicious use.

## 5.2 Explainable AI

HAIKU aims to maintain the human in the loop in AI learning for each of the aviation applications. Human-in-the-loop learning can lead to more explainable AI (XAI), as the user will have more visibility on the underlying mechanisms, ideally steering a continuous learning process.

In order to ensure an adequate Human-AI interaction, the consortium abides by the following principles:

- **Explicability** Explicability is crucial for building and maintaining users' trust in AI systems. HAIKU ensures that AI processes aretransparent, the purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested.
- **Traceability.** The HAIKU consortium ensures that the data sets and the processes that yield the AI system's decision, including those of data gathering and data labelling as well as the algorithms used, is documented to the best possible standard to allow for traceability and an increase in transparency.

This enables identification of the reasons why an AI-decision was erroneous which, in turn, could help prevent future mistakes. Traceability facilitates auditability as well as explainability.

- **Explainability** concerns the ability to explain both the technical processes of an AI system and the related human decisions (e.g. application areas of a system). The HAIKU consortium ensures that the decisions made by an AI system can be understood and traced by human beings.
- **Communication** AI systems should not represent themselves as humans to users; humans have the right to be informed that they are interacting with an AI system. The HAIKU consortium makes sure that AI systems are identifiable as such. In addition, the HAIKU consortium provides the option to decide against this interaction in favour of human interaction. Beyond this, the AI system's capabilities and limitations are communicated to end-users in a manner appropriate to the use case at hand.

## 5.3 Safety and Security of AI

AI is a game-changer, and it cannot be assumed that existing safety, security and Human Factors methods will be sufficient to give assurance that a new system is ready for use in live aviation operations.

The HAIKU consortium abides by the following principles:

- **Technical robustness** which requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm.
- **Resilience to attack and security**. AI systems, like all software systems, should be protected against vulnerabilities that can allow them to be exploited by adversaries, e.g. hacking.
- **Fallback plan and general safety**. AI systems should have safeguards that enable a fallback plan in case of problems. This can mean that AI systems switch from a statistical to rule-based procedure, or that they ask for a human operator before continuing their action.
- **Accuracy**. Accuracy pertains to an AI system's ability to make correct judgements, for example to correctly classify information into the proper categories, or its ability to make correct predictions, recommendations, or decisions based on data or models.
- **Reliability and Reproducibility.** It is critical that the results of AI systems are reproducible, as well as reliable. A reliable AI system is one that works properly with a range of inputs and in a range of situations.

To ensure the development of safe and secure AI assistants, HAIKU takes the following approach. The approach begins by establishing the state of the art in areas such as safety and security in Air Traffic Management/Air Navigation Services. The state of the art in Human Factors (e.g. the SESAR Human Performance approach, as well as EASA Certification Specification CS 25 1302), as well as relevant Cybersecurity standards, are reviewed. Contemporary liability and legal standpoints are also assessed.

*Figure 2. Approach for Safety, Security and Human Factors assurance*

The second stage considers new safety hazards, human performance issues, and cyber threats that could arise with Human-AI Systems (HAIS). This second stage identifies potential hazard scenarios that could lead to accidents, and these inform the derivation of prototype liability principles.

The third stage is to apply these upgraded or new tools to the HAIKU use cases. Since the use cases are at varying TRLs, and using varying 'levels' of AI, this is an advantage in terms of seeing how well new approaches work at different TRLs, and how scalable they are for different levels of HAIS. The task performs the typical steps of safety, security and Human Factors analysis, according to current best practices and standards, including EUROCONTROL SAM, European Risk Classification Scheme (ERCS), ER-023 (AIR 7209), JARUS SORA, SESAR SecRAM, and Human Performance Assessment Methodology.

The final stage finalises the case-based approach and associated guidance, and suggest how regulation could accommodate this guidance, based on the principles of performance-based regulation, first adopted in 1998 by ICAO through Assembly Resolution A32-14 and on the principles of risk-based regulation.

| **Safety** | **Security** | **Human Factors** | **Legal** |
|---|---|---|---|
| Description of system functions and the relationships between these functions | Scoping and Assets Identification | Understand the concept | Understand the concept |
| Hazard identification | Identification of vulnerabilities & threats<br>Impact Assessment | Understand the Human Performance implications | Identification of liability issues |
| Risk assessment and definition of safety objectives | Risk evaluation | Conduct HP activities | Analysis of liability allocation and acceptability of liability risks for all stakeholders |
| Define and validate Safety Mitigations and requirements | Risk treatment<br>Security Controls<br>Security Requirements | Improve and validate the concept and the HP requirements | Definition of mitigations for liability risks, related design recommendations |
| Collate findings and conclude on the transition to the next TRL | | | |

## 5.4 Accountability

To ensure the accountability of the AI systems developed in the project, the HAIKU consortium abides to the following principles:

- **Auditability.** Auditability entails the enablement of the assessment of algorithms, data and design processes. Evaluation by internal and external auditors, and the availability of such evaluation reports, can contribute to the trustworthiness of the technology.
- **Minimisation and reporting of negative impacts.** Identifying, assessing, documenting and minimising the potential negative impacts of AI systems is crucial for those (in)directly affected.

The consortium approach to legal liability considers the allocation and mitigation of possible liability risks from the earliest stages of the design process. In this regard, the consortium will apply DBL's "liability by

design" approach. The approach is entailed in a specific methodology called Legal Case. The Legal Case is a methodology with an associated tool intended to support the integration of automated technologies into complex organisations. Its purpose is to address liability issues arising from the interaction between humans and automated tools, ensuring that these issues are clearly identified and dealt with at the right stage in the design, development, and deployment process. It offers a structured approach and process for the identification, analysis and mitigation of liability attribution issues related to the introduction of new operational concepts and tools in complex environments. Originally designed for the Air Traffic Management domain, its usage can be easily extended to other domains. The Legal Case process consists of the following four steps:

● Understand context and concept. Collecting information on the operational concept and the context of its deployment, as well as the legal and regulatory aspects. This step includes the identification of the level of automation of the concerned system, its impact on roles (see Figure 14), tasks and responsibility and a set of use cases considered relevant for the following legal analysis.

● Identify liability issues. Identifying the possible liabilities, and determining the associated liability risks.

● Address the liability allocation. Analysing the acceptability of liability risks for all stakeholders, proposing mitigations that may improve liability allocation, and making design recommendations accordingly.

● Collecting findings and Systemic Analysis. Results of the study, highlighting the liability issues associated with the object of study and the ways to deal with legal risks, as well as making further recommendations. An overview of the Legal Case process is depicted below, also showing where the Human Factors and Safety analysis feed the analysis.

## 5.5 AI Biases

To ensure that the AI systems developed in the project do not present unwanted and unjustified biases, the HAIKU consortium abides by the following principles:

- **Avoidance of unfair bias.** Data sets used by AI systems (both for training and operation) may suffer from the inclusion of inadvertent historic bias, incompleteness and bad governance models. The continuation of such biases could lead to unintended (in)direct prejudice and discrimination against certain groups or people, potentially exacerbating prejudice and marginalisation. The HAIKU consortium aims to remove identifiable and discriminatory bias in the collection phase where possible. Bi Gender aspects are taken into account when developing HAIKU Use Cases, Digital Assistants, and validation activities. Gender has been shown to be relevant in the domain of digital assistance and AI[7]. Gender inequalities might indeed be reinforced due to the application of AI [5]. For instance, Bansal, Kockelman, and Singh showed that men are more willing to accept Digital Assistants in their vehicles than women [1]. Furthermore, the consortium puts in place oversight processes to analyse and address the system's purpose, constraints, requirements and decisions in a clear and transparent manner.
- **Accessibility and universal design**. Systems should be user-centric and designed in a way that allows all people to use AI products or services, regardless of their age, gender, abilities or characteristics. The HAIKU consortium considers Universal Design principles addressing the widest possible range of users, following relevant accessibility standards. This enables equitable access and active participation of all people in existing and emerging computer-mediated human activities and with regard to assistive technologies.
- **Stakeholder Participation.** In order to develop AI systems that are trustworthy, it is advisable to consult stakeholders who may directly or indirectly be affected by the system throughout its life

cycle. In this regard, the HAIKU consortium solicits regular feedback and set up longer term mechanisms for stakeholder participation throughout the whole process of implementing AI systems.

The HAIKU project aims to deliver solutions that are not gender-specific, however gender differences are considered in enrolling experimental subjects and controlling for potential gender variations with the appropriate sample size. The project takes concrete actions to meet all the three objectives on gender equality:

- Fostering gender balance in Horizon Europe research teams;
- Ensuring gender balance in decision-making, to reach the Commission's target of 40% of the under-represented sex in panels;
- Integrating gender/sex analysis in R&I.

## 5.6 Measures to Prevent Potential Ethical Risks

Specifically, regarding the misuses of AI results, the HAIKU consortium will take the following actions:

- Establishment of the ethical advisory committee
- External ethical advisors will be asked to provide an ethical assessment of each use case (UC) at the end of M12.
- Training and seminars on the ethical issues arose from the ethical assessment will be provided to all the UC leaders and technical providers in order to assist them with the monitoring of the ethical issues.
- In case unforeseen ethical issues emerge over the research activities, the Consortium will ask the Ethical advisors committee for an expert opinion
- DBL will ultimately carry on the monitoring of ethical issues at M22 and M34 and synthesize the results in D1.4.
- The mapping of these issues is done on the basis of the ALTAI methodology defined by the High-level expert group on artificial intelligence.
- The ALTAI definition is used throughout the development of the project to assess potential emerging issues.

Table 1. Preliminary risk assessment of the ethical requirements based on severity, likelihood and impact (where rates are intended as follows: 1 low, 2 medium, 3 high risk).

| Principle | Requirements | Risk |
|---|---|---|
| Human- AI relationship | Respect for human dignity | 2 |
| | Freedom of the individual | 1 |
| | Respect for democracy, justice and the rule of law | 1 |
| | Equality, non-discrimination and solidarity - including the rights of persons at risk of exclusion | 2 |
| | Prevention of harm | 3 |
| Explainable AI | Explicability | 2 |
| | Traceability. | 1 |
| | Explainability | 3 |

| | | |
|---|---|---|
| | Communication | 2 |
| Safety and Security of AI | Technical robustness | 2 |
| | Resilience to attack and security | 3 |
| | Fallback plan and general safety | 1 |
| | Accuracy | 2 |
| | Reliability and Reproducibility | 3 |
| Accountability | Auditability | 3 |
| | Minimisation and reporting of negative impacts. | 2 |
| AI Biases | Avoidance of unfair bias. | 3 |
| | Accessibility and universal design. | 2 |

| | | |
|---|---|---|
| | Stakeholder Participation. | 1 |

# 6. Conclusions

This deliverable describes the measures taken to ensure that project activities are respectful of human rights, particularly the right to privacy and data protection, and do not generate ethically unwanted personal or social effects.

The report presents the overview of the main points regarding ethical issues arising from AI, an overview of the ethical issues stemming from the HAIKU research project and the way the consortium plans to mitigate them, involving the processing of personal data and a set of measures that could be implemented to reduce the risk of misuse of the research results. Furthermore, the deliverable assessed the possible issues of the AI systems relating to human autonomy, explainability, safety and security, accountability and biases.

This report is a work in progress, and it will be updated later on in the project as additional data collection-related information becomes available.

# 7. References

[1] Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. Transportation Research Part C: Emerging Technologies, 67, 1-14.

[2] Dainow, B., & Brey, P. (2021). Ethics By Design and Ethics of Use Approaches for Artificial Intelligence. Brussels: European Commission.

[3] EU Grants (2021). How to complete your ethics self-assessment. *Resource document. European Commission. https://ec. europa. eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/how-to-complete-your-ethics-self-assessment_en. pdf. Accessed*, *14*.

[4] European Commission (2017). H2020 Programme-Guidelines on FAIR Data Management in Horizon 2020.

[5] Friedman, B., & Hendry, D. G. (2019). Value sensitive design: Shaping technology with moral imagination. Mit Press.

[6] Leavy, S., O'Sullivan, B., & Siapera, E. (2020). Data, power and bias in artificial intelligence. arXiv preprint arXiv:2008.07341.

[7] Madiega, T. (2019). EU guidelines on ethics in artificial intelligence: Context and implementation.

[8] Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. Journal of applied social psychology, 27(10), 864-876.

[9] Stahl, B. C. (2021). Ethical issues of ai. In Artificial Intelligence for a Better Future (pp. 35-53). Springer, Cham.

[10] Timan, T., & Mann, Z. (2021). Data Protection in the Era of Artificial Intelligence: Trends, Existing Solutions and Recommendations for Privacy-Preserving Technologies. In The Elements of Big Data Value (pp. 153-175). Springer, Cham.

[11] European Commission (2018). High-level expert group on artificial intelligence. In Ethics Guidelines for Trustworthy AI; European Commission: Brussels, Belgium. Available online: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0 434-11ea-8c1f-01aa75ed71a1 (accessed on 11 December 2022).