D3.3 – SSH FRAMEWORK REPORT (V1)

Lead Beneficiary	INESC-ID
Type of Document	[ETHICS]
Dissemination Level	[PUBLIC]
Due Date	31/12/2024
Submission Date	23/12/2024
Main Author(s)	Carla Pacheco (INESC-ID), António Soares (INESC-ID, ISCTE-IUL), Rui Prada (INESC-ID)
Contributors	Weronika Figueiredo (Data Protection Officer at INESC-ID)

PROJECT INFORMATION

Grant Agreement Number	101168355
Acronym	CARMA
Name	Collaborative Autonomous Robots for eMergency Assistance
Topic	HORIZON-CL3-2023-DRS-01-05 - Robotics: Autonomous or semi-autonomous UGV systems to supplement skills for use in hazardous environments
Funding Scheme	HORIZON-RIA - HORIZON Research and Innovation Actions
Start Date	1 September 2024
Duration	36 months
Coordinator	CS GROUP FRANCE
Grant Agreement Number	101168355

DOCUMENT HISTORY

Version	Date	Author/Organisation	Changes
V0.1	1/11/2024	INESC-ID	ToC Initiated
V0.2	9/12/2024	INESC-ID	Initial version
V0.3	19/12/2024	INESC-ID	Revised version
V0.4	20/12/2024	INESC-ID	Final version for submission
V1.0	23/12/2024	Yana Lazarova (CS)	Formatting. Final version for submission

QUALITY REVIEWERS

Name	Organisation
Stephane Cascio	CS Group
Michalis Angelou	CERTH

EXECUTIVE SUMMARY

With the recent integration of collaborative robots across industries, the concerns over their impact in society have emerged drastically. This document aims to provide an overview of the societal, ethical, legal, and privacy issues to consider regarding the activities to be conducted in the course of the CARMA project.

The document describes these concerns, focusing on privacy and data protection as a major legal concern. From an ethical perspective, the document delves into concerns such as safety, privacy, responsibility, trust, false or excessive expectations, labour replacement, fairness and discrimination. From a societal perspective, the document explores concepts such as labour replacement and fairness and non-discrimination.

Furthermore, this document identifies an ethical framework that can be used to oversee the development of the project, in coordination with WP1 (task 1.5). This is the version 1 of this document. Further updates will be made as the project progresses, particularly after identifying the requirements of the end users. to reflect the design choices and the concrete development of the technology and the pilots.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 PURPOSE OF THE DOCUMENT

This deliverable provides an initial overview of the ethical and legal aspects and societal impact of the CARMA project. The research and activities to be carried out during the project intersect with various ethical considerations addressed here. This document outlines the primary international and European legal frameworks relevant to concerns that the CARMA project might raise. Finally, a combined ethical and legal framework is described as a recommendation related to the carrying out of activities. This is the first of two planned versions of this deliverable to be produced throughout the CARMA project.

In the first version, we identify the aspects that can be relevant to the project and establish an initial framework to lay the ground for the work in the project. In the second version we plan to develop a novel framework that specifically addresses the aspects of integration of collaborative robots in first responder teams.

1.2 INTENDED AUDIENCE

This deliverable has public status. That means all the members of the consortium (including the Commission Services) may access its content. Besides, the publication of the content by the members of the project is possible under the rules of the Description of Action and the Grant Agreement of the project. Additionally, it means that the content is freely accessible to the community and aims to reach a diverse audience, including researchers, developers, end-users, policymakers, and the general public.

1.3 STRUCTURE OF THE DOCUMENT AND ROLES

This document begins by setting up the context for the CARMA project, including its description, scope, and the scenarios under consideration. It then explores human rights, with a particular emphasis on the right to privacy, followed by an overview of human rights in the EU. This section highlights the importance of integrating these rights into the development of technology.

Next, the document discusses data protection and privacy in the context of AI systems and introduces key Responsible AI frameworks. The focus then shifts to ethical and societal concerns, identifying potential challenges in human-robot interactions. Finally, the document concludes by emphasizing the need for a framework and principles that were found to be relevant to guide the development and implementation of the CARMA project and recommending a combined one, featuring ethical and legal concerns.

1.4 RELATED DOCUMENTS

Document title	Short description
CARMA DoA	Defines the scope of the work, objectives and overall project timeline
D3.4	SSH Framework Report V2

2 THE CARMA PROJECT: SETTING SCOPE, DESCRIPTION AND CORE SCENARIOS

2.1 CONTEXTUALISING CARMA (DESCRIPTION AND SCOPE)

The CARMA project objective is to develop an innovative, modular and intuitive platform for first-responders of emergency situations, through a user-centred iterative methodology engaging four end-users and several experts, including the social sciences and humanities (SSH) aspects. The platform will integrate a complementary suite of semi-autonomous and autonomous collaborative robots for emergency assistance designed to operate in close collaboration with humans to support and supplement first responders and assist citizens in a wide range of disaster situations.

2.2 CORE SCENARIOS

During the CARMA project, four pilots will be developed and run as use cases. These pilots intend to guide the development, test the feasibility of the project, and determine the effectiveness of its implementation. The four pilots will be run with the collaboration of four end users as follows:

1. MPOL: their scenario depicts a traffic accident in an urban area, involving vehicles transporting dangerous materials, and the pilot will be run in a training site dedicated to the urban operation of the city of Madrid. To built this scenario includes collecting the input data, as well as booking, installing, integrating, and configuring the required equipment and infrastructure, preparing the site as well as MPOL teams and other participants (e.g. citizen representatives) recruited for the pilot.

2. HTRA: their scenario depicts an earthquake in an urban area; the pilot will be executed in a dedicated USAR training centre close to Athens. Developing this includes collecting the input data, as well as booking, installing, integrating, and configuring the required equipment and infrastructure, and preparing the site and the HRTA USAR teams and other participants (citizens, professional firefighters) recruited for the pilot.

3. FMI: their scenario depicts a fire in an underground parking and the pilot will be organized and executed in an underground parking area in Paris. To build this includes collecting the input data, as well as booking, installing, integrating, and configuring the required equipment and infrastructure, and preparing the site and FMI's BSPP fire brigade and other recruited participants (citizens).

4. MARS: their scenario depicts a fire incident in the cargo of a ship and the pilot will be organised and executed in the industrial port of Marseille. To build this includes collecting the input data, as well as booking, installing, integrating, and configuring the required equipment and infrastructure, preparing the site and MARS' crisis managers and BMPM firefighters, along with the other participants (for example, citizens and workers) recruited for the pilots.

3 HUMAN RIGHTS: THE RIGHT TO PRIVACY

3.1 BACKGROUND INFORMATION: UNIVERSAL DECLARATION OF HUMAN RIGHTS

Article 12 of the UDHR:

No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.[1]

The Universal Declaration of Human Rights (UDHR) is a milestone document in the history of human rights. The UDHR was proclaimed by the United Nations General Assembly in Paris on 10 December 1948 as a common standard of achievements for all peoples and all nations. It sets out, for the first time, fundamental human rights to be universally protected. One of those fundamental rights is the right to privacy. Article 12 of the UDHR to respect for private and family life marked the first time an international instrument laid down an individual's right to protection of their private sphere against intrusion from others, especially from the state.

In the context of the CARMA project, Article 12 is crucial related to protecting one's privacy, which is essential as robots in emergency situations can collect sensitive and personal data (for example, biometric information) to assist victims. Ensuring these systems comply with privacy protection under Article 12 is relevant to prevent misuse or data exposure.

Additionally, and to align with Article 12, robots in emergency situations must operate within a legal framework that clearly defines how to process data and interact with humans.

Regarding honouring reputation and dignity, and in the context of the CARMA project, victims of disasters are already undergoing stressful and vulnerable situations, which means that the use of data or robot-human interaction that includes recording or broadcasting sensitive information can potentially damage their honour or reputation.

3.2 EUROPEAN CONVENTION ON HUMAN RIGHTS

Article 8 of the ECHR:

Right to respect for private and family life:

1. *Everyone has the right to respect for his private and family life, his home and his correspondence.*

2. *There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.*

[1] https://fra.europa.eu/en/publication/2018/handbook-european-data-protection-law-2018-edition

The Council of Europe, established after World War II, aims to promote the rule of law, democracy, human rights, and social development among European states. It adopted the European Convention on Human Rights (ECHR) in 1950, which came into force in 1953. By 2018, the Council had 47 member states, including 27 EU members, though the UK left the EU in 2020. Member states are required to uphold the rights outlined in the Convention in all activities.

The ECHR led to the creation of the European Court of Human Rights (ECtHR) in 1959, based in Strasbourg, France. The Court ensures compliance with the Convention by addressing complaints of rights' violations submitted by individuals, groups, NGOs, or legal entities.

Under Article 8 of the ECHR, the right to privacy can be restricted if three conditions are met : 1) a legal basis for the limitation (legality); 2) pursuit of a legitimate aim (legitimacy), and 3) necessity in a democratic society.

However, the right to privacy is not absolute and may conflict with other rights, such as freedom of expression and access to information. The European Court of Human Rights seeks to balance these competing rights.

3.2.1 HUMAN RIGHTS IN THE EU

Human rights are a fundamental cornerstone of the European Union (EU), embedded in its legal frameworks and institutions. Human rights are rooted in dignity, freedom, democracy, equality, and the rule of law; the EU aims to uphold these rights across its member states and in its interactions with the world. The Charter of Fundamental Rights of the European Union[2], adopted in 2000 and legally binding since the Lisbon Treaty of 2009, consolidates civil, political, economic, and social rights for all EU citizens and residents. These rights are drawn from the constitutional traditions of member states, the European Convention on Human Rights (ECHR), and other international treaties, establishing a robust legal framework to safeguard individual freedoms.

The protection of dignity and personal freedom is central to the EU's human rights agenda. This includes the prohibition of torture, inhuman or degrading treatment, slavery, and forced labour, as well as respect for private and family life. Citizens and residents also enjoy freedom of expression, religion, assembly, and the right to non-discrimination, emphasizing the EU's commitment to pluralism and inclusion. Fundamental economic and social rights, such as fair working conditions, healthcare, education, and social security, are similarly protected, reflecting the EU's recognition of the indivisibility of human rights.

The EU has established institutions to monitor and enforce human rights across its member states. The European Court of Justice (ECJ) ensures compliance with the Charter of Fundamental Rights, interpreting its provisions to address emerging challenges in privacy, data protection, and digital rights. The EU also works in concert with the European Court of Human Rights (ECtHR), which supervises the ECHR, further strengthening the regional human rights architecture. In addition, the European Union Agency for Fundamental Rights (FRA) provides research and advice to policymakers, helping address rights issues ranging from migration to artificial intelligence.

Beyond its borders, the EU promotes human rights as a pillar of its foreign policy. It includes human rights clauses in trade agreements, provides aid to civil society organisations, and imposes sanctions on individuals and regimes responsible for severe rights violations. The EU's Global Human Rights Sanctions Regime, often called the "European Magnitsky Act," facilitates targeted measures against

[2] https://www.echr.coe.int/documents/d/echr/convention_ENG

those implicated in human rights abuses worldwide, demonstrating the EU's global leadership in this domain.

In what concerns the CARMA project, the activities being conducted during the project will oversee the different fundamental rights by the Consortium. This will be done in two ways: first, through compliance related to the European laws that regulate the right to data protection and privacy such as GDPR (article 4)[3] and the European Charter[4] and the second following a combined legal and ethical framework involving the ALTAI (Assessment List for Trustworthy Artificial Intelligence) and the ethical framework from Amigoni & Schiaffonati (2018).

ALTAI framework is a practical tool developed by the High-Level Expert Group on AI (HLEG) to help businesses and organisations evaluating the trustworthiness of their AI systems during the development process.

3.3 AI SYSTEMS: DATA PROTECTION AND PRIVACY

The fundamental right to privacy is protected by international and national law. The elements of privacy are based upon the non-interference principle of Article 8 of the ECHR: 'Everyone has the right to respect for his privacy and family life, his home and his correspondence.' (van Genderen, 2017). A significant element of privacy intrusion today, whether intentional or unintentional, involves the processing of individuals' personal data (ibid).

Artificial intelligence systems rely on the use of vast amounts of data which frequently involves the processing of personal data. AI enables the use of personal data to analyze, predict, and influence human behavior, turning both the data and the results of its processing into highly valuable assets and this occurs in two main ways: first, personal data may be included in datasets used to train machine learning systems, helping to develop their algorithmic models. Second, these trained models can then be applied to personal data to draw inferences or make predictions about specific individuals (Santor and Lagioia, 2020).

Though the processing of data has benefits for society as it enables it to learn more about individuals and their interactions and the acquired social knowledge helps society to progress and develop better governance, it also triggers privacy and data protection concerns. This is because the models emerging from these large sets can make mistakes, discriminate, reproduce human bias or introduce new biases. And even when data processing is fair, it can also raise concerns about surveillance, evaluation, influence and manipulation (Santor and Lagioia, 2020).

Concerns with how data could be used (for example: targeting individuals and personalised treatment of individuals) led to the development of the General Data Protection Regulation (GDPR) in the European Union. The GDPR is the most comprehensive document on personal data with the objective of helping organisations to plan careful interactions with data (Hoofnagle et al., 2019). Data protection then becomes the link between AI systems and legal requirements. Two important definitions under the GDPR are 'personal data'[5] and the activities considered 'processing'[6].

[3] https://gdpr-info.eu/art-4-gdpr/
[4] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:en:PDF
[5] https://gdpr-info.eu/art-4-gdpr/
[6] https://gdpr-info.eu/art-5-gdpr/

3.3.1 Preliminary consideration – entities controlling the processing of personal data

It is crucial to define what entities – in the context of the project (and, potentially, at the implementation stage) – will be considered as data processing controllers. There are several options to be considered: joint controllership of the consortium members or individual controllership in the scope of the activities carried out.

3.3.2 Data Protection Impact Assessment

The GDPR defines in article 35.1. that, "where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data".

As for the research project, the GDPR states in paragraph 4 of article 35 that: "The supervisory authority shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment pursuant to paragraph 1". In this context, as it is expected to carry out processing activities regarding special categories of personal data (as defined in article 9.1. of GDPR) for the purpose of scientific research, to the extent personal data are processed (as opposed to synthetic data), a data protection impact assessment is required (according to line 6) of the Regulation n. º 798/2018 referred to above.

As for the application of the technology being developed, it is expected that a data protection impact will be required as well, based on article 35.3 (a) of GDPR which affirms that "A data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person".

3.3.3 Privacy by Design and by Default

Data protection by Design and by Default (DPbDD)[7] is a requirement imposed by article 25 of the GDPR that defines the rule of data protection being designed into the processing of personal data and as a default setting and applicable throughout the processing lifecycle. The core of the provision is to ensure appropriate and effective data protection both by design and by default, imposing appropriate measures and safeguards in the processing to ensure that the data protection principles and the rights and freedoms of data subjects are effective.

In line with Article 25.1, the controller shall implement appropriate technical and organisational measures which are designed to implement the data protection principles and to integrate the necessary safeguards into the processing in order to meet the requirements and protect the rights and freedoms of data subjects. Both appropriate measures and necessary safeguards are meant to serve the same purpose of protecting the rights of data subjects and ensuring that the protection of their personal data is built into the processing. Technical and organizational measures and necessary

[7] https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf

safeguards can be understood in a broad sense as any method or means that a controller may employ in the processing. Being appropriate means that the measures and necessary safeguards should be suited to achieve the intended purpose, i.e. they must implement the data protection principles effectively. The requirement to appropriateness is thus closely related to the requirement of effectiveness.

In the context of the CARMA project, it is particularly important to ensure the compliance with the data minimization principle as well as to guarantee that the architecture of the entire solution – from the platform to the UGVs – respects the DPbDD provisions.[8]

3.3.4 VULNERABILITY OF DATA SUBJECTS AND THE LAWFUL BASIS

The issue of the vulnerability of data subjects is particularly relevant when considering the application of the technology being developed. In disaster situations, individuals are vulnerable – not necessarily *per se* but as a result of exceptional circumstances. They are under stress, often traumatized, with reduced mental capacity and, sometimes, physically injured and/or disabled. Some of these individuals can possess added layers of vulnerability – i.e. resulting from their age, language skills (consider minorities not necessarily being able to communicate in the language of rescuers) etc.

Specifically, in the context of personal data processing, the vulnerability of data subjects can have a significant impact on the lawful basis of processing. In disaster situations, typically the protection of vital interests will apply according to article 6.1 d) of GDPR.

However, regarding special categories of personal data (such as health data, i.e. vital signs), the protection of vital interests of data subject or of another natural person can only be relied on when the data subject is physically or legally incapable of giving consent – article 9.2.c). This legal aspect should be considered at the design stage of the technology.

3.3.5 AUTOMATED DECISION-MAKING

When developing the technology, the impact of automated decision-making – expected in the context of victim detection – should be considered. It is also important to differentiate between a tool that serves to support the first responder in his decision-making process and a tool that actually makes decisions that can have an impact on individuals' rights.

Automated decision-making may be unfair and create discrimination, for example by giving priority to certain types of victims. Transparent and clear criteria are fundamental – transparency of processing is a basic requirement of the GDPR. Under Article 12.1 the controller must provide data subjects with concise, transparent, intelligible and easily accessible information about the processing of their personal data. The quality of data used in algorithms supporting the decision-making processes is also of utmost importance. In this context, please consider the recommendations of WP29 on Automated decision-making[9]: "If the data used in an automated decision-making or profiling process is inaccurate, any resultant decision or profile will be flawed. Decisions may be made on the basis of outdated data or the incorrect interpretation of external data."

[8] For more information on DPbDD refer to Guidelines 4/2019 on Article 25 Data Protection by Design and by Default issued by the European Data Protection Board.

[9] Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679" issued by WP29, adopted on 3 October 2017 (as last revised and adopted on 6 February 2018).

Taking into account the circumstances of processing, the right of individuals to oppose to automated processing seems to be less relevant.

3.3.6 When does GDPR apply?

The GDPR Article 4 defines “personal data” 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”. It also defines “processing” as “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”.

GDPR applies when “personal data” is “processed”, which means that all activities - collecting, storing, disclosing and erasing - are considered as data processing (Hoofnagle et al., 2019). The GDPR defines three essential actors in addition to "personal data" and "processing: "data subjects," "controllers," and "processors”. Controllers and processors have responsibility on how and why data is processed.

1. 'Data subjects' are individuals whose personal data will be processed.
2. 'Controllers' are those individuals who determine the purposes and the means of processing personal data (organisations, for example)
3. 'Processors' relate to the entities that will operationalize personal data on behalf of controllers—in such cases, there is a clear hierarchy. For example, suppose Company Y gathers and analyzes survey data on Company X's customers, as Company X instructed. In that case, Company X is the controller, and Company Y is the data processor. Suppose two organisations collaborate and need to determine the objective (why) and how the processing of personal data will be. In that case, they would be seen as joint controllers and share the regulatory burden and liability for errors.

Regarding the CARMA project, we aim to identify specific data protection concerns by discussing some practical examples. For example, we organized a co-design session with the project's end users to delve into the different emergency scenarios—car accident, fire in an underground car parking, earthquake, and fire in a harbour—and capture details of their work processes.

From running these simulations, we expect to determine specific data concerns and identify the best approaches based on the multiple steps involved in an emergency situation. The pilots that are going to be carried out afterwards will help to refine end-users requirements as well.

3.3.7 Designing AI systems: framework

Ethical Frameworks for AI: an Overview

The fast-paced development of Artificial Intelligence systems has raised concerns about the potential harm it can cause if not developed and deployed without addressing certain ethical guidelines to ensure it aligns with societal values.

Several organisations, regions and countries across the world have established ethical guidelines to address concerns such as fairness, accountability, transparency, and the potential societal impact of AI. Some of the most well-known are the frameworks created by the 1) European Union (EU), the 2) Organisation for Economic Co-operation and Development (OECD), and the 3) United Nations (UN). These frameworks present similar points, considering transparency, fairness, accountability, and alignment with human values.

1. *European Union (EU) Ethical Framework for AI*[10]

The EU has been laying the ground for addressing ethical concerns through its Ethics Guidelines for Trustworthy AI published in 2019 by the European Commission's High-Level Expert Group on AI (HLEG). HLEG's framework builds on three verticals: 1) AI must be lawful, adhering to existing laws; 2) it has to be ethical, respecting principles such as fairness and human dignity; and 3) robust, ensuring technical reliability and safety. The document outlines seven key requirements for trustworthy AI: transparency, accountability, data governance, and societal and environmental well-being. The EU further supports these principles with the AI Act, which proposes risk-based regulation to safeguard the risks that AI systems can pose.

The core principles of the EU guidelines is that the Commission needs to develop a human-centric approach to AI that is respectful of European values and principles. This is to emphasize placing human values at the core of the development, deployment, use, and monitoring of AI systems. This approach ensures respect for fundamental rights, as stated in the Treaties of the European Union and the Charter of Fundamental Rights of the European Union.

The guidelines include:
1. Human Agency and Oversight
2. Technical Robustness and Safety
3. Privacy and Data Governance
4. Transparency
5. Diversity, Non-discrimination and Fairness
6. Societal and Environmental Well-being
7. Accountability

An Assessment List for Trustworthy AI (ALTAI) was proposed providing an initial approach for the evaluation of Trustworthy AI systems. ALTAI is intended for flexible use which means that organisations can extract relevant elements to the particular AI system they are building. They can also add elements if required, taking into consideration the industry where they operate in. ALTAI framework intends to help organisations identify the risks an AI system can potentially generate and how to minimize those risks while maximising the benefit of AI systems.

In more detail, each guideline aims to (based on Radclyffe, Ribeiro, Wortham, 2023):

a) Regarding **human agency and oversight**, the objective is for organisations to identify how maturely they have considered the role of a human-in-command, human-in-the-loop, or human-on-the-loop in relation to the AI system.

Sample questions that can be asked:

[10] https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf

1. Is the AI System designed to interact, guide or take decisions by human end-users that affect humans or society?

2. Did you ensure a "Stop" button or procedure to safely abort an operation when needed?

3. Did you take any specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?

b) In the context of **technical robustness and safety**, organisations should adopt a proactive approach to managing risks associated with AI systems. This means anticipating potential challenges and integrating measures to minimize these risks during the design phase. Addressing safety and risk is crucial, as these concerns are among the most significant for AI systems.

Sample questions are:

1. Did you red-team and/or pen-test the system?

2. Did you define risks, risk metrics, and risk levels of the AI system in each specific use-case?

3. Did you put in place a well-defined process to monitor if the AI system is meeting the intended goals?

c) Regarding **privacy and data governance**, AI systems require large amounts of data to function. If the system is making recommendations about a person's preferences or predicting their behavior, it may need to use personal or identifiable information. This means that designers must ensure privacy is built into the system from the start and maintain strong data management practices to protect user information.

Sample questions are:

1. Did you establish mechanisms that allow flagging issues related to privacy concerning the AI system?

2. Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?

3. Did you align the AI system with relevant standards or widely adopted protocols for (daily) data management and governance?

d) When addressing **transparency**, organisations can break the guideline into three key areas:

1. Data and Model Provenance: Ensuring the origins and sources of data and models in the AI system are well-documented and controlled.
2. Explainability: Determining how well the AI system can clarify its decision-making process.
3. User Communication: Ensuring clear and accessible disclosure to users about the AI system's existence and how it operates.

Sample questions are:

1. Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)?

2. Did you explain the decision(s) of the AI system to the users?

3. In cases of interactive AI systems (e.g., chatbots, robo-lawyers), do you communicate to users that they are interacting with an AI system instead of a human?

e) When addressing diversity, non-discrimination, and fairness in AI, a significant risk lies in data collected from an unequal society and processed by non-diverse teams. Without careful attention, AI systems could amplify these inequalities. Therefore, it is essential to prioritize diversity and inclusion throughout an AI system's entire lifecycle to mitigate potential harm and ensure fairness.

Sample questions:

1. Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?

2. Did you take the impact of the AI system on the potential end-users and/ or subjects into account?

3. Did you consider diversity and repetitiveness of end-users and/ or subjects in the data?

f) The key consideration regarding **societal and environmental well-being** is the risk that improper use of technology could disrupt the underlying structure of society. It is crucial to strike a balance between addressing the needs of current generations and preserving resources and opportunities for future generations.

Sample questions are:

1. Are there potential negative impacts of the AI system on the environment?

2. Does the AI system impact human work and work arrangements?

3. Could the AI system have a negative impact on society at large or democracy?

g) The primary concern regarding **accountability** is the potential for harm if an AI system behaves unpredictably or operates without apparent oversight. This issue is compounded by the complexity of integrating AI systems with other AI or non-AI systems, making it difficult to determine who controls the system or addresses any resulting harm.

Sample questions are:

1. Did you establish a process to discuss and continually monitor and assess the AI system's adherence to this Assessment List for Trustworthy AI?

2. Did you ensure that the AI system can be audited by independent third parties?

3. Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential gray areas?

2. *OECD Principles on AI*[11]

The OECD's Principles on AI, adopted in 2019, represent a collaborative effort involving 42 member and non-member countries. The OECD AI Principles are the first intergovernmental standard on AI. They promote innovative, trustworthy AI that respects human rights and democratic values.

Adopted in 2019 and updated in 2024, they comprise five values-based principles such as:

- Inclusive growth, sustainable development and well-being.
- Human rights and democratic values, including fairness and privacy.
- Transparency and explainability.
- Robustness, security and safety.
- Accountability.

The OCDE document defines an AI system as a "machine-based system that, with explicit or implicit objectives, processes input to produce outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." It notes that AI systems differ in autonomy and adaptiveness after deployment.

It describes the AI cycle as a process consisting of several phases:

1. Design, Data, and Models: A context-dependent sequence that includes planning and design, data collection and processing, and model development.
2. Verification and Validation: Ensuring the system meets its intended objectives and operates as expected.
3. Deployment: Implementing the AI system for use in its intended environment.
4. Operation and Monitoring: Overseeing the system's performance, addressing issues, and ensuring it meets its objectives.

These phases often occur iteratively rather than sequentially. Additionally, the decision to retire an AI system can arise during the operation and monitoring phase based on performance, changing needs, or other factors.

4. ***United Nations (UN) Initiatives for Ethical AI***[12]

The UN's approach to ethical AI is based on its commitment to human rights and sustainable development. The UNESCO Recommendation on the Ethics of AI, adopted in 2021, intends to foster ethical, transparent, and accountable AI globally. The document outlines human rights protection, environmental sustainability, cultural diversity, and gender equality. UNESCO emphasizes education, capacity-building, and promoting international cooperation to address AI access and governance disparities. The UN also leverages its Sustainable Development Goals (SDGs) to align AI ethics with broader global priorities, such as eradicating poverty and reducing inequalities.

There are also other frameworks which have emerged from organizational contexts such as Microsoft Responsible AI Principles[13] and Google AI Principles[14] among others. The biggest difference between these frameworks and the EU, OECD and UN is that these last three are rooted in the commitment to

[11] https://oecd.ai/en/ai-principles
[12] https://oecd.ai/en/ai-principles
[13] https://www.microsoft.com/en-us/ai/principles-and-approach
[14] https://ai.google/responsibility/principles/

human rights. Microsoft and Google are big technology organisations located in the Uniyed States where products are developed taking a more market-oriented approach to AI systems development.

From all the frameworks that have been examined, the EU Ethical guideline appears to be a more comprehensive framework and therefore the one to be used in the CARMA project.

Besides these frameworks, some efforts are going on in Europe to work towards the standardization of AI through the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC)[15]. The objective is to standardize AI rules and guidelines that are proliferating and making the development of AI systems confusing. Discussions are ongoing and nothing formal has emerged yet, but perhaps it would be valuable to follow these discussions.

Finally, the General Purpose AI Code of Practice[16] is taking shape at the EU level. Drawn from the AI Act[17], which is called "The Code", this first draft of the Code addresses critical considerations for providers of general-purpose AI models and for providers of general-purpose AI models with systemic risk through four objectives and dedicated working groups:

1. Transparency and copyright-related rules
2. Risk identification and assessment for systemic risk
3. Technical risk mitigation for systemic risk
4. Governance risk mitigation for systemic risk

[15] https://www.cencenelec.eu/european-standardization/cen-and-cenelec/
[16] The document is available here: file:///Users/alex/Downloads/First_Draft_GeneralPurpose_AI_Code_of_Practice__sFt8VTOzxWsmJhIfdMVYPLM42C0_109946.pdf
[17] https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

4 ETHICAL AND SOCIETAL CONCERNS

4.1 INTRODUCTION

The development, implementation, and integration of collaborative robots in first responder teams raises several ethical concerns, namely safety, privacy, responsibility, trust, false or excessive expectations, labour replacement, fairness and discrimination (Battistuzzi et al., 2021).

Further, holding workshops for the design and testing of damage scenarios with the co-participation of developers, associated partners, and end users is a way of contextualising ethics in the development of collaborative robots in first responder teams (Daun et al., 2024).

In this section, we address these ethical and social issues, which significantly influence this technology's social acceptance and future development.

4.1.1 SAFETY

The development of robotic technology for emergency response operations aims to provide important improvements for first responders, for instance, including the introduction of autonomous assistance functions into firefighting operations to enhance safety (Schneider & Wildermuth, 2017). However, there is a need for objective assessment tools to systematically address and assess all safety-related aspects (Berx et al., 2023).

Considering ethics, to technically ensure the safety of human team members and victims in an emergency, the use of soft actuators (e.g., flexible joints or soft arms) and the admittance (the system's ability to accept effort inputs and produce flow outputs) and impedance (accepting flow inputs and producing effort outputs) control, can increase safety and reduce the risk of harm during close human-robot collaboration (Hua et al., 2023).

Developing and testing damage scenarios helps to find solutions for minimising risks and uncertainties in the use of robots, and previous research has shown that the robot operator's ability to assess the limits and reliability of AI-assistance functions is essential for success (Daun et al., 2024). Robots with different characteristics and capabilities working in a team with humans can facilitate first responders' decision-making by providing information in an emergency situation (Allison et al., 2024).

However, the interaction of first responders with robots in high-stress environments is a source of concern, as there is evidence that people may have difficulty understanding how to interact with collaborative robots (Harriott et al., 2012).

Furthermore, the amount of data provided to humans must be considered, as too much information can lead to overload, and too little data can lead to presumptuous autonomy – in both cases, the team's results tend to be less secure (Allison et al., 2024).

4.1.2 PRIVACY AND DATA PROTECTION

There are concerns about privacy due to the development and implementation of robots in collaboration with first responders, considering the collection of sensitive data from people involved in the emergency occurrence. The issue of privacy during the development and implementation of

emergency assistance robots must be addressed by all interested parties, namely policy makers, manufacturers, users and other stakeholders.

Robots for emergency assistance will be equipped with sensors, 3D cameras, dynamic control devices, among other technologies based on AI. This equipment aims to exchange data between members of the rescue team and other possible stakeholders related to the emergency occurrence. This data is crucial for recognizing the environment, moving robots, and assisting victims. However, the collection and storage of information relating to victims and others involved in the incident raises concerns about whether this personal data will be adequately protected. There is a risk of damage to the private lives of the people involved, especially in the case of improper use by unauthorized third parties.

The use of robots involves the extensive collection of personal data and therefore implies concerns about privacy and the potential leakage of personal information, even more so in the case of sensitive rescue operations (Battistuzzi et al., 2021). This data can be about both the first response teams and the victims and other people in the disaster area, and all people involved have the right to privacy and control over their personal data (Boada et al., 2021).

Due to the critical nature of emergency operations, the benefits of data collection far outweigh the possibility of harm, but the information collected by robots must be restricted to rescue organisations and used only for the purpose of rescue (Battistuzzi et al., 2021; Harbers et al., 2017).

Furthermore, the deployment of robots raises cybersecurity issues, and it is essential to ensure that robots do not become targets of cyberattacks (Wang et al., 2021). For instance, terrorists who are involved in a disaster could hack the robotic systems and subvert them to cause even greater damage. Taking this into consideration, a robust safeguard system is needed to protect them from unauthorized access.

4.1.3 Responsibility and Accountability

A significant ethical issue is determining who is responsible for the actions of autonomous or semi-autonomous robots (Boada et al., 2021). This includes addressing moral responsibility for potential harm caused by robots, so that, for example, both designers and human operators are held accountable (Battistuzzi et al., 2021). It is necessary to define who should be held responsible for the malfunction of a robot and its consequences and for any damage caused to the physical and psychological integrity of people (e.g., mechanical accidents or damage caused by violation of autonomy or privacy).

Furthermore, the decision-making process of autonomous robots should be aligned with the human values involved (machine ethics), considering any ethical dilemmas that may arise (e.g., security versus privacy) (Boada et al., 2021).

Rescue robots can have different degrees of autonomy depending on the emergency situation. As an assistance robot without autonomy, it is (tele)operated by a first responder.

However, this operator can ask the robot to autonomously perform a task following a pre-defined procedure, or even, at a higher degree of autonomy, the operator can ask the robot to choose between multiple tasks.

A possible conflict situation may occur when a rescue robot has to choose to respond to the command of one operator over another, for instance, in situations in which an operator needs to override the command of another operator. This possible conflict imposes the need to establish rules and

procedures (doctrines) that consider the rescue robot in its ontology and establish clear parameters in terms of machine ethics.

Autonomous robots must have a minimum of interactive skills to deal with ethically sensitive situations, but a robot is a machine and has no conscience, so it cannot assume any kind of responsibility (Torras, 2024). Consciousness refers, for example, to self-consciousness and moral conscience, and can be conceptualized ethically as a self-reflection on the nature of one's own acts, whether obligatory or prohibited, right or wrong. If the agent recognizes an act of one's own as a moral violation, an internal sanction eventually arises, a bad self-consciousness in the form of feelings of remorse, guilt or shame (Beauchamp & Childress, 2019).

Psychologically, self-consciousness originates from the pleasant or unpleasant feelings generated by homeostasis, that is, psychophysiological processes for individual preservation. A robot would need the perception of its own body and feelings, with humanlike homeostatic characteristics, to emulate a conscience (Devillers, 2021; Man & Damasio, 2019), which is outside the scope of this project.

For these reasons, protocols based on transparency and traceability are necessary to assign responsibility for the actions of robots (Boada et al., 2021). For example, accountability for a robot that can learn from its own experience can be shared among its developers, manufacturers, owners, and all those with whom the robot has interacted (Torras, 2024).

4.1.4 TRUST

Developing and maintaining the trust of the first responders with collaborative robots requires the reliable performance of the robots and implies the ethical concern of eliminating or minimising false or excessive expectations about the capabilities of the robots (Battistuzzi et al., 2021).

4.1.5 COOPERATIVE DESIGN

Robots with different characteristics and capabilities working in a team with humans can facilitate first responders' decision-making by providing information in an emergency situation (Allison et al., 2024).

However, the interaction of first responders with robots in high-stress environments is a source of concern, as there is evidence that people may have difficulty understanding how to interact with collaborative robots (Harriott et al., 2012).

Furthermore, the amount of data provided to humans must be considered, as too much information can lead to overload, and too little data can lead to presumptuous autonomy – in both cases, the team's results tend to be less secure (Allison et al., 2024).

The easy integration of collaborative robots and their developers with first responders is an important factor for successful emergency response. There is evidence that an appropriate team structure and easy-to-use and stable tools are crucial for the successful deployment of collaborative robots with first responders (Steinbauer et al., 2014). Therefore, co-design bringing together robot developers and first responders is a key factor in designing collaborative robots that offer a user-friendly interface and are well integrated with the human team of which they will be part.

Cooperation between first responders and developers is necessary to analyse needs in different emergency scenarios, design the corresponding integrated robotic system, and experiment with human-robot teamwork in realistic missions (Kruijff-Korbayová et al., 2021).

Furthermore, holding workshops for the design and testing of damage scenarios with the co-participation of developers, associated partners, and end users, is a way of contextualising ethics in the development of first responder robots (Daun et al., 2024).

Examples of important requirements previously identified by workshops are adaptation of robot operational capabilities, robust network connectivity (Kruijff-Korbayová et al., 2021) and assistance functions, and including users in training workshops (Daun et al., 2024).

Additionally, the manipulation tasks of all robot joints require precise control by the user, through an interface with a high-bandwidth and low-latency connection to the robot (Daun et al., 2024).

4.1.6 Societal Concerns

Some ethical issues are also societal concerns, namely the fear of job loss (labour replacement) due to increased automation in disaster response (De Graaf & Allouch, 2016; Wang et al., 2021), and issues related to fairness and non-discrimination that robotic systems may create due to the lack of information about users, e.g., vulnerable minorities (Torras, 2024), or even in the relationship between first responders among themselves (Battistuzzi et al., 2021).

1. ***Labour Replacement*.** The implementation of robots has raised the spectre of job losses in the most diverse sectors, leading to a general concern about how the job market could be readjusted to deal with the effects of AI and automation (Pavlidou et al., 2011). The worry of possible job displacement caused by the introduction of robots can have an emotional impact on first responders and implies an ethical consideration (Battistuzzi et al., 2021).

2. ***Fairness and Non-Discrimination*.** Ensuring that the deployment of robots does not lead to unfair treatment or discrimination among first responders is an ethical issue, which includes considering how robots might affect the dynamics and hierarchy within teams (Battistuzzi et al., 2021). Robots can negatively affect social equality in terms of access and quality of assistance, e.g., if a robot does not have information about certain users in its database, it will not be able to provide similar assistance to all people (Boada et al., 2021). Vulnerable minorities who do not have enough data stored may not benefit equally from the service of robots (Torras, 2024).

   These ethical and societal concerns highlight the need for a proactive approach in the development and deployment of robots in first responder teams. Stakeholder involvement in the process must be ensured, and the use of robots must be guided by an appropriate ethical and societal framework (Battistuzzi et al., 2021; Harbers et al., 2017; Kruijff-Korbayová et al., 2021).

4.2 Ethical concerns for search and rescue robots

New technologies such as robotics can be conceptualized as social experiments, considering their experimental nature and the inherent uncertainty that this entails (Amigoni & Schiaffonati, 2018). Indeed, unexpected occurrences may arise in operations with robots in real contexts (Steinbauer et al., 2014). In this sense, the evaluation of the social impact of the use of robots in specific contexts can be conceptualized as an exploratory experiment, diverging from the traditional notion of controlled experience.

Search and rescue robots are an opportunity to test socio-ethical approaches to the development of robots and their interaction with humans and environments (Amigoni & Schiaffonati, 2018).

There is a need for normative ethical guidance regarding first responder robots and their respective AI systems (Daun et al., 2024). Considering the explorative character of the new technology as a social experiment, Amigoni and Schiaffonati (2018) applied the ethical framework proposed by van de Poel (2016) for experimental robotics to the field of robots for search and rescue. This ethical framework with 16 conditions is listed in Table 1. There are three groups/principles: *avoiding harm* (nonmaleficence, 1–7), *seeking to do good* (beneficence, 8 and 9), and *respect for autonomy and justice* (10–16).

Table 1 : Ethical Framework for Search and Rescue Robots

Number	Condition	Principle
1	Absence of other reasonable means for gaining knowledge about risks and benefits	Non-maleficiance
2	Monitoring of data and risks while addressing privacy concerns	
3	Possibility and willingness to adapt or stop the experiement	
4	Containment of risks as far as reasonably possible	
5	Conciously scaling up to avoid large-scale harm and to improve learning	
6	Flexible setup of the experiment and avoidance of lock-in of the technology	
7	Avoid experiments that undermine resilience	
8	Reasonable to expect social benefits from the experiment	Beneficience
9	Clear distribution of responsabilities for setting up, carrying out, monitoring, evaluating, adapting, and stopping the experiments	
10	Experimental subjects are informed	Respect for autonomy and justice
11	The experiment is approved by democratically legitimized bodies	
12	Experimental subjects can influence the setting up, carrying out, monitoring, evaluating, adapting and stopping of the experiment	
13	Experimental subjects can withdrawn from the experiment	
14	Vulnerable experimental subjects are either not subject to the experiment or are additionally protected or particulary profit from the experimental technology (or a combination)	
15	A fair distribution of potential hazards and benefits	
16	Reversibility of harm or, if impossible, compensation of harm	

Source: adapted from Amigoni & Schiaffonati, 2018

Van De Poel (2016) proposed the following specifications for the principles present in this ethical framework:

Non-maleficence: duty not to cause harm, minimize risks and take precautions against possible risks or harm.

Beneficence: duty to do good, remove existing harm and prevent harm or risk that do not originate in the experiment; cause more good than harm, creating or increasing benefits.

Respect for autonomy: duty to protect and guarantee autonomy, the ability of individual or group autonomous choice.

Justice: duties related to distributive justice, protection of vulnerable groups, and prevention of exploitation, in addition to fair protocols in procedural justice.

5 CONCLUSIONS

This document provides an initial analysis of the ethical, legal, and societal implications associated with the CARMA project. It is the first of two planned reports to be delivered throughout the project's timeline.

The rapid adoption of emerging technologies, particularly artificial intelligence (AI), has sparked growing concerns about their implications for human rights, data protection, and citizen privacy. While AI systems hold great potential to benefit society as a whole, they also pose risks, such as compromising personal privacy and damaging individual reputations if not carefully managed.

The objective and scope of the CARMA project are to support citizens throughout emergency crises by using robotic technologies. As AI systems play a central role in this project, it is critical to identify and address associated concerns.

This deliverable provided an overview of the key challenges from legal, ethical and societal perspectives, offering a framework to guide and oversee its development processes. The proposed framework addresses critical legal, ethical and societal issues to ensure responsible and effective implementation.This framework's guiding principles are based on the ALTAI framework discussed above.

However, several other principles can be identified during the project, particularly during the pilot phases, which will be reflected in the coming second version.

Examples of ethical actions that should be developed in the CARMA project, considering the ethical framework of Amigoni and Schiaffonati (2018), are combined with the ALTAI framework, as we can see in Table 2. The items 1 to 7 concern non maleficence, 8 and 9 concern beneficence, and 10 to 16 concern respect for autonomy and justice.

Table 2 : Framework recommendation based on ALTAI combined with the ethical framework of Amigoni and Schiaffonati (2018)

Guideline principle	Description	Key action/recommendation
Human agency and oversight	To identify how maturely they have considered the role of a human-in-command, human-in-the-loop, or human-on-the-loop in relation to the AI system	12) Victims must be able to influence the development of the CARMA project experiments to the extent possible. This may be difficult to establish, considering that victims are generally not able to play a role in defining the actions of the robotic system. 13) Victims must have the option to withdraw from the CARMA project experiment, if they have the desire and preserved autonomy to do so. For example, a victim who does not wish to be transported by a search and rescue robot should have the possibility to

Guideline principle	Description	Key action/recommendation
		request that they not be transported in this way. The numbering of these items refers to Table 1, principle of respect for autonomy and justice (Amigoni & Schiaffonati, 2018).
Technical robustness and safety	To manage risks associated with AI systems, anticipating potential challenges and integrating measures to minimize these risks during the design phase. Safety and risk as these concerns are among the most significant for AI systems	1) Testing in controlled environments to verify that the robots are functioning correctly, in order to eliminate or minimize the risk of harm to people. This action is foreseen in the CARMA project co-developed with the first responder teams. 2) The robotic system must be monitored to prevent physical harm to all people involved in the emergency, as well as preserving the privacy of all involved. 3) The robotic system must be interruptible or adaptable to avoid harm to people. Items 2 and 3 should be discussed between the CARMA project developers and the first responders during the co-design process. 4) All possible measures must be implemented to eliminate or minimize the risk of damage, e.g., the use of soft materials in areas where the robot is most likely to come into contact with people. The numbering of these items refer to Table 1, principle of nonmaleficence (Amigoni & Schiaffonati, 2018).
Privacy and data governance	AI systems need large amounts of data to be able to work and may need to use personal or identifiable information. Privacy must be built into the system from the start and maintain strong data	This item should be discussed between CARMA project developers and first responders during the co-design process.

Guideline principle	Description	Key action/recommendation
	management practices to protect user information	
Transparency	1) Data and Model Provenance: Ensuring the origins and sources of data and models in the AI system are well-documented and controlled. 2) Explainability: Determining how well the AI system can clarify its decision-making process 3) User Communication: Ensuring clear and accessible disclosure to users about the AI system's existence and how it operates	10) Experimental situations in which human subjects are involved require the informed consent of these same subjects. However, by default, emergency situations make it impossible to collect informed consent from victims. One possible alternative is for robots to alert victims to their presence through sounds and lights, in order to make victims aware of a possible interaction with them. 11) The CARMA project experiments must obtain approval from democratically legitimized bodies. The numbering of these items refers to Table 1, principle of respect for autonomy and justice (Amigoni & Schiaffonati, 2018).
Diversity, non-discrimination, and fairness	Risk lies in data collected from an unequal society and processed by non-diverse teams. It is essential to prioritize diversity and inclusion throughout an AI system's entire lifecycle to mitigate potential harm and ensure fairness	15) The distribution of potential risks and benefits must be fair and equitable among different victims in an emergency, e.g., decisions about the prioritization of care for a victim by first responders. The numbering of this item refers to Table 1, principle of respect for autonomy and justice (Amigoni & Schiaffonati, 2018).
Societal and environmental well-being	This is related to the incorrect use of technology that can disrupt the underlying structure of society. Balance between addressing the needs of current generations and preserving resources and opportunities for future generations is required	5) Collaborative robots should be implemented first in emergency scenarios in small areas without casualties, gradually increasing the area and the number of casualties in different environments. The intention is to gain increasing knowledge about the performance of robotic systems in emergency scenarios, thus reducing the eventual risk of large-scale damage. 6) Avoid fixating on a specific technological option, i.e. do not

Guideline principle	Description	Key action/recommendation
		ignore possible new robot and system options to improve results in the case of replacing robotic systems already implemented. 7) Resilience in dealing with unexpected risks in an emergency is higher for first responders than for victims, considering their training to work in dangerous conditions. Therefore, risk minimization must assume these individual differences in order to prevent resilience from being undermined. 8) The implementation of the CARMA project must offer real benefits and advantages to the community, which outweigh the possible risks associated with the robotic system, e.g., the increased speed of mapping the disaster environment caused by the robots will supposedly reflect in the speed at which first responders assist potential victims. 14) Victims are the target subjects of the CARMA project experiments, considering that the objective of rescue and rescue robot technology is precisely to benefit victims, who by default are vulnerable. 16) If damage is unavoidable, it must be compensated in some way. The numbering of these items refer to Table 1, 5-7 principle of nonmaleficence, 8 principle of beneficence, 14 and 16 principle of respect for autonomy and justice (Amigoni & Schiaffonati, 2018).
Accountability	Accountability is associated with an AI system behaving unpredictably or operates without apparent oversight	9) There should be transparency in the assignment of responsibilities to the members of the first responders' team and

Guideline principle	Description	Key action/recommendation
	And the risk increases when integrating AI systems with other AI or non-AI systems, making it difficult to determine who controls the system or addresses any resulting harm	other people involved in the configuration, execution, monitoring, evaluation, adaptation and interruption of the experiments. One of the purposes of the CARMA emergency simulations is to clearly define which team members have which responsibilities, in a co-design with the first responders. The numbering of this item refers to Table 1, principle of beneficence (Amigoni & Schiaffonati, 2018).

Note: Please refer to pages 14–16 for examples of specific questions that can be asked when applying this framework

Finally, we want to emphasize the critical role of INESC-ID's Data Protection Officer and the Ethical Committee throughout the project. Their support and guidance are integral to ensuring compliance with data protection and ethical standards.

6 REFERENCES

Allison, M., Farmer, M., & Song, Z. (2024). Towards Distributed Learning to Support Situational Awareness for Robotic Team Augmented Humanitarian Disaster Response. *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, 0370–0374. https://doi.org/10.1109/CCWC60891.2024.10427713

Amigoni, F., & Schiaffonati, V. (2018). Ethics for Robots as Experimental Technologies: Pairing Anticipation with Exploration to Evaluate the Social Impact of Robotics. *IEEE Robotics & Automation Magazine*, *25*(1), 30–36. https://doi.org/10.1109/MRA.2017.2781543

Battistuzzi, L., Recchiuto, C. T., & Sgorbissa, A. (2021). Ethical concerns in rescue robotics: A scoping review. *Ethics and Information Technology*, *23*(4), 863–875. https://doi.org/10.1007/s10676-021-09603-0

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics*. (8_{th} ed.). Oxford University Press.

Berx, N., Brescia, A., Aqamarina, R., Curcio, E. M., Pintelon, L., & Carbone, G. (2023). Stakeholders' perspectives on safety-related human–robot collaborative scenarios. *International Journal of Advanced Robotic Systems*, *20*(5), 17298806231200095. https://doi.org/10.1177/17298806231200095

Boada, J. P., Maestre, B. R., & Genís, C. T. (2021). The ethical issues of social assistive robotics: A critical literature review. *Technology in Society*, *67*, 101726. https://doi.org/10.1016/j.techsoc.2021.101726

Daun, K., Bark. F., Tateo, D., Peters, J., Heinlein, J., Wendt, J., Heidemann, N., Kruijff-Korbayová, I., Kohlbrecher, S., Friedrich, J., Martin, D., Schmidt, M.W., Hillerbrand, R., von Stryk, O. (2024). A Holistic Concept on AI Assistance for Robot-Supported Reconnaissance and Mitigation of Acute Radiation Hazard Situations. *2024 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, n.a.

De Graaf, M. M. A., & Ben Allouch, S. (2016). Anticipating our future robot society: The evaluation of future robot applications from a user's perspective. *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 755–762. https://doi.org/10.1109/ROMAN.2016.7745204

Devillers, L. (2021). Human–robot interactions and affective computing: The ethical implications. In J. von Braun, M. S. Archer, G. M. Reichberg, & M. Sánchez Sorondo (Eds.), *Robotics, AI, and Humanity* (205–211). Springer International Publishing. https://doi.org/10.1007/978-3-030-54173-6_17

Harbers, M., De Greeff, J., Kruijff-Korbayová, I., Neerincx, M. A., & Hindriks, K. V. (2017). Exploring the Ethical Landscape of Robot-Assisted Search and Rescue. In M. I. Aldinhas Ferreira, J. Silva Sequeira, M. O. Tokhi, E. E. Kadar, & G. S. Virk (Eds.), *A World with Robots* (Vol. 84, pp. 93–107). Springer International Publishing. https://doi.org/10.1007/978-3-319-46667-5_7

Harriott, C. E., Buford, G. L., Zhang, T., & Adams, J. A. (2012). Human-human vs. Human-robot teamed investigation. *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, 405–406. https://doi.org/10.1145/2157689.2157820

Hoofnagle, C. J., Van Der Sloot, B., & Borgesius, F. Z. (2019). The European Union general data protection regulation: what it is and what it means. *Information & Communications Technology Law*, *28*(1), 65-98.

Hua, M. T., Langås, E. F., Zafar, M. H., & Sanfilippo, F. (2023). From Rigid to Hybrid/Soft Robots: Exploration of Ethical and Philosophical Aspects in Shifting from Caged Robots to Human-Robot Teaming. *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1794–1799. https://doi.org/10.1109/SSCI52147.2023.10372032

Kruijff-Korbayova, I., Grafe, R., Heidemann, N., Berrang, A., Hussung, C., Willms, C., Fettke, P., Beul, M., Quenzel, J., Schleich, D., Behnke, S., Tiemann, J., Guldenring, J., Patchou, M., Arendt, C., Wietfeld, C., Daun, K., Schnaubelt, M., Von Stryk, O., … Surmann, H. (2021). German Rescue Robotics Center (DRZ): A Holistic Approach for Robotic Systems Assisting in Emergency Response. *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 138–145. https://doi.org/10.1109/SSRR53300.2021.9597869

Man, K., & Damasio, A. (2019). Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence, 1*, 446–452. https://doi.org/10.1038/s42256-019-0103-7.

Pavlidou, N., Tsaliki, P. V., & Vardalachakis, I. N. (2011). Technical change, unemployment and labor skills. *International Journal of Social Economics*, *38*(7), 595–606. https://doi.org/10.1108/03068291111139230

Santor, G., & Lagioia, F. (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence.

Schneider, F. E., & Wildermuth, D. (2017). Using robots for firefighters and first responders: Scenario specification and exemplary system description. *2017 18th International Carpathian Control Conference (ICCC)*, 216–221. https://doi.org/10.1109/CarpathianCC.2017.7970400

Torras, C. (2024). Ethics of Social Robotics: Individual and Societal Concerns and Opportunities. *Annual Review of Control, Robotics, and Autonomous Systems*, *7*(1), 1–18. https://doi.org/10.1146/annurev-control-062023-082238

Van De Poel, I. (2016). An Ethical Framework for Evaluating Experimental Technology. *Science and Engineering Ethics, 22*(3), 667–686. https://doi.org/10.1007/s11948-015-9724-3

Wang, R., Nakhimovich, D., Roberts, F. S., & Bekris, K. E. (2021). Chapter 5 Robotics as an Enabler of Resiliency to Disasters: Promises and Pitfalls. In F. S. Roberts & I. A. Sheremet (Eds.), *Resilience in the Digital Age* (Vol. 12660, pp. 75–101). Springer International Publishing. https://doi.org/10.1007/978-3-030-70370-7_5