Ref. Ares(2024)2338058 - 28/03/2024

Funded by the European Union

HORIZON EUROPE PROGRAMME: HORIZON-CL4-2022-DIGITAL-EMERGING-02

SoliDAIR

Solid, rapid and efficient adoption of Data, AI & Robotics applications in production

Deliverable D1.2: Risk & Ethics management plan

Primary Author(s)	Andreas Frommknecht | FHG
Deliverable Type	Report
Dissemination Level	Public
Due Date (Annex I)	31.03.2024 (Month 6)
Pages	14
Document Version	Final
GA Number	101120276
Project Coordinator	Andreas Frommknecht | Fraunhofer IPA (FHG) (andreas.frommknecht@ipa.fraunhofer.de)

Contributors	
Name	**Organisation**
Andreas Frommknecht	Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.

Formal Reviewers	
Name	**Organisation**
Anesa Begovic	I2m Unternehmensentwicklung GmbH
Christoph Mitteregger	AUTFORCE Automations GmbH

Version Log			
Rev #	**Date**	**Author**	**Description**
0.1	23.01.2024	Andreas Frommknecht (FHG)	Draft
0.2	08.03.2024	Lena Loercher (FHG)	Draft
1.0	21.03.2024	Anesa Begovic (I2M)	Quality review
2.0	22.03.2024	Alexander Katzbeck (AUT)	Quality review
3.0	27.03.2024	Andreas Frommknecht (FHG)	Formatting check
4.0	28.03.2024	Andreas Frommknecht (FHG)	Coordinator review and approval, deliverable ready for submission

Project Abstract

SoliDAIR aims to accelerate the uptake of Artificial Intelligence (AI) and Robotics in European manufacturing, using Data as an enabler. It will co-develop and demonstrate tailored solutions to digitalise and automate visual inspection and physical testing, enable predictive quality control and process optimisation. The SoliDAIR project tackles the problem of AI & Robotics systems not being extensively used in the production industry, because it is not clear whether they are safe and when or why they will fail, by researching, developing and testing methods that are as solid and trustworthy as possible to be adopted by the European industry, while being cost-efficient to develop and replicate.

New methods and tools will be developed by research and technology providers, which leverage the current state of the art in visual AI, AI for process data, and smart & collaborative Robotics. The developed technologies will be applied and demonstrated in 4 industry use cases (UC Brose, UC CIE, UC BOS, UC AUT) to prove their functionality and applicability in real production environments. The objective is to improve production processes through digitalised and automated quality control for high volume, high rate and flexible manufacturing. The developed methods shall be efficiently and easily adaptable and replicable, so they can be easily applied to new use cases outside the consortium.

Table of Contents

Public Summary

This deliverable, “Risk & Ethics management plan” provides a comprehensive description of the management strategy for risk, ethical and legal issues during the project. The corresponding plan include risks identification, risk analysis, contingency plans of technical and organisational risks and an ethical risk management plan.

1 Introduction

1.1 Rational of this deliverable

This deliverable, “Risk & Ethics management plan” provides a comprehensive description of the management strategy for risk, ethical and legal issues during the SoliDAIR project. The corresponding plan includes risks identification, risk analysis, contingency plans of technical and organizational risks. The ethical risk management plan includes ethical dimension of the objectives, methodology and likely impact, human interaction with AI and the environmental impact of the SoliDAIR project. This leads to compliance with ethical principles and relevant legislations.

So, it addresses the goal of work package WP1 “Project Management & Coordination” in the direction of legal and ethical aspects and supplements in this way Deliverable D1.1 Project Handbook. With its basic guidelines it has a strong impact on WP2 “Methodological framework & Generic modules”, WP3 “Use case implementation and deployment” and WP4 “Use case demonstration, evaluation, and optimization”.

Attainment of the objectives and explanation of deviations

The objectives of this deliverable are achieved without any deviations.

2 Technical and organizational risks

This section presents the specific aspects of the Risk Management Plan (RMP) for SoliDAIR project. The RMP is a process including the steps below:

- Risk identification
- Risk analysis
- Risk mitigation strategies
- Risk monitoring
- Documentation and reporting
- Review and update

2.1 Risk identification

Risk identification is the first step in the risk management process which aims to identify and document the potential risks that may impact the successful execution of the project. The risk identification phase plays a crucial role in setting the stage for effective risk management. By proactively identifying risks, the project team can better understand the potential challenges and uncertainties they may face, allowing them to develop appropriate strategies to mitigate or respond to those risks.

Correspondingly, the following critical risks have been identified on the organisational and technical level. Some of them are only relevant for certain use cases, if so, this is explicitly mentioned:

- R1: Overall delays and deviations from the time plan
- R2: Method development or validation falls under confidentiality
- R3: The framework might be limited to the SoliDAIR use cases and not applicable to other EU industries
- R4: Explainability is not achievable for certain AI-algorithms

- R5: Insufficient real data is available or can't be generated in the needed amount
- R6: Transparency of AI solution is not achieved in the defined range (UC BROSE)
- R7: Aimed improvement in data efficiency is not achieved (UC BROSE)
- R8: Lack of acceptance by end-users, although AI solution might fulfil all technical requirements (UC BROSE)
- R9: System performance lacking or not all NOK defects identified (UC CIE)
- R10: AVI cameras do not properly 'read' all machined surfaces, holes and threads with sufficiently high resolution (UC CIE)
- R11: Quality control robotic system is not compatible with production speed (UC CIE)
- R12: Prediction is still inaccurate despite trying all available AI-algorithms (UC BOS)
- R13: Simulation-based synthetic data generation doesn't provide sufficient NOK data, in time for AI training (UC BOS)
- R14: AI-based end-of-line prediction model doesn't meet the runtime requirements of the production line (BOS UC)
- R15: Invalid input from the humans due to different interpretation and background (UC AUT)
- R16: Incorporation of AI features over the progress in manufacturing steps does not yield accurate results (UC AUT)
- R17: Recommender system suggestions are not accepted by workers and thereby not executed (AUT UC)
- R18: Exploitation plan targets are not clear, measurable or achievable in timeframe

2.2 Risk analysis

After identification of the risks, each risk is analysed and evaluated in order to prioritize its significance, and likelihood of occurrence. Risk analysis involves assessing the potential consequences of each risk and determining the probability of its realization. The assessment of the risk helps in understanding the impact of the risk on the project and aids in making informed decisions on which risks need immediate attention. This risk analysis and evaluation involves considering the interdependencies with other WPs and potential cascading effects.

These risks are linked to the different work packages as shown in the following table. Their likelihood and severity are also listed.

Risk Number	WPs	Likelihood	Severity
R1	WP1	Medium	High
R2	WP2	Low	Medium
R3	WP2	Medium	Low
R4	WP2	Medium	Low
R5	WP2, WP3, WP4	High	Low
R6	WP3, WP4	Medium	Medium
R7	WP3, WP4	Medium	Medium

R8	WP3, WP4	Medium	Low
R9	WP3, WP4	Medium	Medium
R10	WP3	Medium	Medium
R11	WP3, WP4	Medium	Medium
R12	WP3, WP4	Low	Medium
R13	WP3, WP4	Medium	Medium
R14	WP3, WP4	Medium	Medium
R15	WP3, WP4	Medium	Medium
R16	WP3, WP4	Medium	Medium
R17	WP3, WP4	Medium	High
R18	WP5	Medium	High

2.3 Contingency plan

Once the risks are evaluated and their potential impact is understood, the next step is to develop appropriate risk mitigation strategies. These strategies aim to minimize the probability and/or the impact of the identified risks. Depending on the nature of the risks, mitigation approaches can vary, including risk avoidance, risk transfer, risk reduction, and risk acceptance. By proactively addressing risks with suitable mitigation strategies, the project team can enhance the chances of project success and reduce potential negative outcomes.

For each risk mitigation measures have been outlined as described in the following contingency plan:

- R1: The coordinator and the WP leaders will monitor the tasks' progress periodically (3-6 months) and ensure the project is on track. Delays will be detected early and will be addressed with the consortium partners (see deliverable D1.1 Project Handbook).
- R2: Developed methods become abstracted and are not limited on special data. Because of the broad consortium a validation can always take place with another industry partner, even if a data exchange is not possible.
- R3: The method development for all functions in WP2 will first be done on generic data, which is open source available. This leads to methods that are applicable in a wide range. A peer-review of the framework will assess wide applicability. If needed, higher generalisation will take place.
- R4: If the explainability approach fails for certain AI-algorithms the transparency will be achieved by clearly outlining the training process and a detailed evaluation process of these algorithms.
- R5: Data efficiency methods like synthetic data generation and/or anomaly detection will be used to compensate lack of real data. If this still is not sufficient, public available data sets will be used for development. The transfer possibility on the real situation will always be assured.
- R6: The AI solution will be split up in submodules and it will be focused on the most critical submodules for transparency and trustworthiness efforts.

- R7: If the results of the data efficiency module should not be sufficient, additional human capacity will be used for training-data generation.
- R8: Additional trainings and workshops will take place, to achieve the necessary conviction of the end-users.
- R9: Split the output of the production cell in two QC systems running in parallel so the required speed is achieved. Decrease camera field of view or increase resolution.
- R10: Replace cameras with higher resolution, or more specialised to the task.
- R11: Improve quality control cycle time, through optimized path, additional cobot/robot or AVI cameras.
- R12: If proved insufficient, a strategy for adding sensors will be explored, and a new dataset will be captured from the production line.
- R13: BOS will evaluate this early in the project, and if needed, will intentionally produce additional NOK parts to complement and balance the dataset.
- R14: Carry out pruning methods to reduce the computational footprint of the trained models in production.
- R15: To correct invalid input of expert data, the data will be periodically validated (e.g. once a week) by another team or person which (upon approval) increase the weighting to 100%.
- R16: Determination of most relevant sensor placing along the manufacturing process and most suitable sensor type for the individual positions will enable to generate more specific data as input for the AI solution and thus increase the prediction quality of the AI findings.
- R17: A real-time capable recommender system will be developed which suggests non-standard correction measures that are executed in the remaining manufacturing steps of the individual gearbox concerned (result for worker).
- R18: Targets will be refined right after project start; exploitation plan balances both incremental improvements and breakthrough R&D steps harmonized among the partners.

2.4 Risk monitoring, documentation, reporting, review and update

As described in D1.1 “project handbook”, the project coordinator (FHG), supported by the project management partner (I2M), will be responsible for the risk monitoring and documentation. The WP leaders have the responsibility to manage the risks within their own WP and inform the project Coordinator and SB regarding the risk status for reporting, reviewing and updating the risks and ethics management plan, whenever necessary.

2.5 Further Risks

Further very use case specific risks are outlined in Deliverable 3.1 “Requirements and technical specifications for each use case”.

3 Ethics risk management

The aim of the ethic risk management is to minimize almost any ethical risk, which can occur during the project. To achieve this different ethical and legal aspects are considered as outlined below.

3.1 Ethical dimension of the objectives, methodology and likely impact

The consortium of SoliDAIR is fully committed to handle any ethical issues related to Artificial Intelligence that may arise during the course of the project with great care and has undergone the Ethics Self-Assessment for AI and defined an Ethics Risks assessment process (part of Task 1.4), provided in the PART B Annex. Task 1.4 supports the project over the complete runtime of the project. It ensures that the AI systems developed and deployed will adhere to the maximum extent possible to the ethical principles and legal requirements.

3.2 Human interaction with AI

The SoliDAIR project's main focus is on developing AI-enabled manufacturing control and inspection systems for production support and optimisation. Therefore, only manufacturing-related data is used.

When human operator input is integrated into the AI, as complementary input data or as feedback to its inferences and decisions, they are largely related to the product or production line, and at no point related to the human himself/herself. As such, personal information has no relevance for the project scope. Nevertheless, interaction of the developed AI systems with humans is necessary and encouraged. The systems need to adhere to the principles of transparency, explainability and trust (Trustworthy AI strategies). As these are key topics of the SoliDAIR project, they will be thoroughly addressed in every technical aspect of the project. For achieving sufficient transparency, the human operators will always be informed or aware that they are interacting with a "smart" AI-enabled system, and will be trained to interact adequately, therefore there is no risk of "unawareness".

In all the SoliDAIR systems developed (use cases), the AI is, in principle, there to support and help humans in decision making (e.g. quality inspection). To reassure acceptance of the AI-Systems user opinion will be considered during the whole project. The content of questions for feedback acquisition will be as bias-free as possible and the users will be selected in a representative and discrimination-free way.

The AI always supports and automates manual tasks, but never controls human decision making. Human intervention/override functions are planned within the AI systems. The developed AI systems will never manipulate humans. In fact, the aim is to follow "human in the loop" approaches and ensure human supervision of the AI.

3.3 Environmental impact

The AI systems are meant to improve production performance, leading to sustainability improvements, not the other way around. So, no negative effects are foreseen, instead positive ones, especially the CO_2 reduction (see PART B Annex), will be achieved.

3.4 Compliance with ethical principles and relevant legislations

3.4.1 Compliance with ethical principles

Even though the ethical risks identified in the SoliDAIR project are limited, the project consortium is committed to ensuring that the AI systems developed and in use cases deployed will adhere to the maximum extent possible to the ethical principles and legal requirements. For that reason, the following processes during the project are and will be established:

- An initial Trustworthy AI assessment was performed at the beginning of the project to support the design of the AI systems (use cases).
- The development of the use cases will be monitored for any unforeseen risks throughout the project duration, at least once per year.
- A second Trustworthy AI ethic assessment will be done for all use cases on M22 prior to deployment and demonstration (Milestone 3).
- Human Machine Interfaces (HMI) will be developed based on principles of transparency, explainability and trust (Trustworthy AI strategies).
- Humans and AI will cooperate in the best way, ensuring supervision, validations and acceptance of AI predictions, inferences and decisions (human in the loop).
- Final validation of the AI systems will ensure adherence to the above, and overall correct performance in line with specifications and expectations.

3.4.2 Compliance legal requirements

The consortium will follow all relevant European regulations and scientific standards to perform ethical research. Especially the EU AI Act will be considered and taken into account to ensure the legal compliance of every technical development during SoliDAIR in the future. The following lists some of the basic regulations and guidelines taken into account:

- EU AI Act: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
- EU Guidelines on ethics in artificial intelligence: https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2019)640163
- EU Code of Research Integrity: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/european-code-of-conduct-for-research-integrity_horizon_en.pdf
- VDE SPEC 90012 V1.0 (en): VCIO based description of systems for AI trustworthiness characterization https://www.vde.com/resource/blob/2177870/a24b13db01773747e6b7bba4ce20ea60/vde-spec-90012-v1-0--en--data.pdf
- Charter of Fundamental Rights of the European Union (2000/C 364/01): https://www.europarl.europa.eu/charter/pdf/text_en.pdf

For every AI development and deployment work performed in the SoliDAIR project the legal requirements of the EU and the involved national countries will be obeyed. All activities carried out within SoliDAIR comply with the applicable international, EU and national laws, regardless of whether these activities are carried out in the EU or in non-EU countries such as Turkey. If any doubt or conflict in legal requirements arise, the consortium commits to uphold the EU and the stricter of the two legal frameworks, and ensures that all activities are permitted in EU Member States. Any non-EU-project partner obliges itself to all European rules considering the shared data.

4 Link to other work packages

The guidelines described in this deliverable have a strong impact on WP2 “Methodological framework & Generic modules”, WP3 “Use case implementation and deployment” and WP4 “Use case demonstration, evaluation, and optimisation”.

5 References

EU AI Act:
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

EU Guidelines on ethics in artificial intelligence:
https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2019)640163

EU Code of Research Integrity:
https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/european-code-of-conduct-for-research-integrity_horizon_en.pdf

VDE SPEC 90012 V1.0 (en): VCIO based description of systems for AI trustworthiness characterization
https://www.vde.com/resource/blob/2177870/a24b13db01773747e6b7bba4ce20ea60/vde-spec-90012-v1-0--en--data.pdf

Charter of Fundamental Rights of the European Union (2000/C 364/01):
https://www.europarl.europa.eu/charter/pdf/text_en.pdf

6 Acknowledgements and disclaimer

The authors would like to thank the partners in the project for their valuable comments on previous drafts and for performing the review.

#	Partner	Partner full name
1	FHG	FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV
2	BRO-B	BROSE FAHRZEUGTEILE SE & CO. KOMMANDITGESELLSCHAFT
3	CIE	FUNDACION CIE I+D+I
4	BOS	BOSCH SANAYI VE TICARET AS
5	AUT	AUTFORCE AUTOMATIONS-GMBH
6	SISW	AUTFORCE AUTOMATIONS-GMBH
7	UGS	UG SYSTEMS GMBH & CO. KG
8	THL	TWI ELLAS ASTIKI MI KERDOSKOPIKI ETAIREIA
9	VIF	VIRTUAL VEHICLE RESEARCH GMBH
10	I2M	I2M UNTERNEHMENSENTWICKLUNG GMBH

LEGAL DISCLAIMER

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document shall be liable or responsible, in negligence or otherwise, for any loss, damage or expense whatever sustained by any person as a result of the use, in any manner or form, of any knowledge, information or data contained in this document, or due to any inaccuracy, omission or error therein contained.

Funded by the European Union

Abbreviations and Definitions

Term	Definition
AVI	Automated Visual Inspection
QC	Quality Control
SB	Steering Board
WP	Work Package