GA – Project number:	**101158046**
Project Acronym:	AUTOMATA
Project Title:	AUTOmated enriched digitisation of Archaeological liThics and cerAmics
call/topic:	HORIZON-CL2-2023-HERITAGE-ECCCH-01-02
type of action	HORIZON RIA
Principal Investigatot	Prof Gabriele Gattiglia, UNIPI
Tel:	+39 050 2215228
E-mail:	gabriele.gattiglia@unipi.it

This project has received funding from the European Union's HORIZON RIA research and innovation programme under grant agreement N. 101158046

D 2.4 Ethical guidelines for trustworthy AI

Version: 1.0

Revision: first release

Work Package: 2 - Needs analysis, methodologies, specification, design
Lead Author (Org): Veronica Neri (UNIPI), Silvia Dadà (UNIPI)
Contributing Author(s) (Org): Nevio Dubbini (MIN), Gabriele Gattiglia (UNIPI), Martina Naso (UNIPI)
Due Date: M8
Date: 30/04/2025

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	**Public**	**X**

C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Description
0.1	17.03.2025	Gabriele Gattiglia	Document creation
0.2	13.04.2025	Veronica Neri, Silvia Dadà	Content added
0.3	15.04.2025	Veronica Neri, Silvia Dadà	Content added
0.4	16.04.2025	Gabriele Gattiglia	Content revision
0.5	22.04.2025	Gabriele Gattiglia	Revision
0.6	23.04.2025	Nevio Dubbini, Martina Naso	Content Revision
0.7	29.04.2025	Gabriele Gattiglia	Content added
0.8	30.04.2025	Veronica Neri, Silvia Dadà	Content added
0.9	30.04.2025	Gabriele Gattiglia	Final revision

Disclaimer

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Abbreviations

WP: Work package

M: Month

UNIPI: Università di Pisa

UBM: Université Bordeaux Montaigne

UoY: University of York

INRAP Institut National de Recherches Archéologiques

AMZ: Arheoloski Muzej u Zagrebu

QB: QBrobotics Srl

HUJ: The Hebrew University of Jerusalem

MIN: Miningful srls

KCL: King's College London

IIT: Fondazione Istituto Italiano di Tecnologia

UB: Universitat de Barcelona

CL: Culture Lab

Executive summary

This deliverable outlines the ethical framework underpinning the use of Artificial Intelligence (AI) within the AUTOMATA project. AI and Machine Learning (ML) technologies are central to AUTOMATA's approach to digitising ceramic and lithic artefacts, particularly by enabling real-time calibration, data acquisition, and enhancement of 3D model construction. While these technologies promise to increase efficiency, accuracy, and automation, they also introduce specific ethical challenges that must be carefully addressed to ensure a responsible deployment.

The document first examines the relevant ethical and legal frameworks. It identifies several critical ethical issues associated with the use of AI in digitisation, including the risk of diminishing human oversight, the propagation of biases through training data, over-reliance on automation, and the potential neglect of physical artefacts in favour of digital surrogates. Furthermore, concerns related to data privacy, equitable access to digitised resources, and the impact of automation on employment are discussed. Legally, the AI systems utilised in AUTOMATA are categorised as low- or minimal-risk under the EU AI Act (2024), which encourages voluntary adherence to ethical codes of conduct despite the absence of mandatory regulatory obligations.

Building upon international guidelines—particularly the Ethics Guidelines for Trustworthy AI (EU High-Level Expert Group on AI, 2019) and UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021)—AUTOMATA's ethical framework is structured around several key principles: digital preservation, accessibility, accountability, autonomy, explainability and transparency, human oversight, right to privacy and data protection, technical robustness, and sustainability. These principles are tailored to the specific context of archaeological digitisation and aim to ensure that the AI systems contribute positively to the field while safeguarding fundamental rights and societal trust.

The framework promotes digital preservation practices that ensure long-term access to high-quality 3D data, enhancing rather than replacing the material conservation of artefacts. Accessibility is addressed by committing to open, scalable, and inclusive dissemination of digitised heritage, with particular attention to ensuring equitable access across diverse communities, including individuals with disabilities. Accountability measures are embedded through transparent decision-making processes, auditability, and continuous human oversight, maintaining expert involvement throughout the digitisation workflow.

AUTOMATA also prioritises explainability by designing AI systems whose operations can be clearly understood by users, and transparency through the open documentation of workflows, algorithms, and data sources. Data protection and privacy considerations are integral to the system's architecture, ensuring compliance with relevant regulations and safeguarding sensitive information. Furthermore, technical robustness and resilience are promoted to guarantee reliable and secure operation, while sustainability measures aim to minimise the environmental footprint of the AI components across their lifecycle.

By implementing this ethical framework, AUTOMATA ensures that the use of AI technologies not only advances the project's scientific and technical goals but also adheres to a model of responsible innovation. This approach strengthens societal trust in AI applications within cultural heritage and contributes to the broader objective of creating human-centric, trustworthy AI systems.

1 Introduction

As reported in D2.3 (paragraph 6.4), AI and ML will significantly contribute to the digitisation of ceramic and lithic artefacts in the AUTOMATA system by enabling real-time processing, particularly for calibration, ensuring accurate data acquisition, and identifying regions of interest on artefacts for sensor analysis. These technologies will enhance efficiency, accuracy, and automation in the digitisation process, with a key challenge being the minimisation of overall digitisation time.

AI systems enhance digitisation by supporting automatic calibration based on initial scans and accurate data acquisition without human intervention. They identify optimal viewpoints for artefact inspection using real-time sensor data and computer vision, detecting key features and recommending the best angles for data collection. Additionally, ML models filter and prioritise relevant information, distinguishing meaningful data from noise. Real-time analyses verify data accuracy, identifying issues like acquisition errors or sensor failures, ensuring high-quality, interpretable data.

With respect to the AI-based data enhancement of 3D model quality (D2.3, paragraph 6.4.1), AI can assist in the 3D model construction process by improving data quality at different stages. ML algorithms can refine scanning outputs through automated adjustments, enhancing the raw data before further processing. Subsequently, AI supports stages such as segmentation, point-cloud cleaning, alignment, registration, and mesh optimisation. Although many of these stages can be automated, manual intervention will still be required.

The integration of AI in the system during the digitisation includes several key enhancements. For instance, AI can automatically identify and segment artefacts, distinguishing them from the background and removing unnecessary data. It also cleans point-clouds by identifying and eliminating outliers. Additionally, AI improves alignment algorithms by recognising recurring parts of objects across multiple scans. For mesh quality assurance, AI mends holes and reduces point density in areas with low geometry. Where real-time processing is necessary, AUTOMATA aims to incorporate these AI steps into the live acquisition pipeline to streamline the digitisation process and reduce post-processing efforts.

The AUTOMATA system will have a multi-tiered data management infrastructure for secure storage, efficient access, and long-term sharing of high-volume digitisation results. Integration with the cloud supports automation, scalability, and FAIR data principles. During development and acquisition, project partners manage and back up data locally to prevent loss. Data are temporarily stored and indexed through the RIS3D platform (D2.3, paragraph 6.6.1), which integrates 3D geometry, spatial metadata and analytical results. Validated data is archived with the Archaeology Data Service (ADS) for long-term preservation and accessibility, using DOI and Dublin Core metadata. ADS complies with CoreTrustSeal certification and supports all file formats utilised by AUTOMATA. Typically, licensing follows the Creative Commons Attribution (CC BY 4.0) standard. This integration guarantees that AUTOMATA outputs are interoperable, secure, and accessible to archaeologists, heritage professionals, and researchers.

Considering the unique characteristics of these systems and their intended purpose, several potential ethical issues arise (Gattiglia, 2025). To effectively manage them, it is essential to consider the legal and ethical framework surrounding the use of AI. The deployment of AI in digitisation must be guided by ethical principles to ensure that the technology is used responsibly. This includes considering the potential long-term impacts on society and the environment.

2 Ethical and legal frameworks

2.1 Ethical issues

One of the strengths of AUTOMATA is the automation of some steps in the digitisation process, which will reduce the digitisation time (< 5 min) (D2.3, paragraph 6.5). However, automation of calibration, data acquisition, and data processing may pose some ethical issues:

- The reduction of human intervention may decrease human involvement and undermine the integrity of ethical decision-making in the field (Dennis, 2020).
- AI algorithms can inadvertently perpetuate biases present in the training data. This could lead to erroneous outcomes, particularly in the identification and prioritisation of data, which may affect the integrity of the digitisation process.
- While AI enhances efficiency and accuracy, there is a risk of over-reliance on automated systems.
- Digitisation is undoubtedly a valuable tool for preserving artefacts; however, it poses the risk of relying solely on digital preservation, potentially leading to the neglect or destruction of the physical artefact (Tiribelli et al., 2024).
- Interaction with robotic systems raises concerns about potential harm and job security.
- Automation of procedures, such as digitisation, can lead to the reduction of work for many people. On the one hand, however, such digitisation would probably NEVER have been achieved without automated systems; on the other hand, the control and verification of what the system accomplishes make the presence of the human being necessary.

Concerning data processing and integration (D2.3, paragraph 6.6):

- Managing large volumes of data, especially those stored in the cloud, raises concerns about protecting sensitive data and preventing unauthorised access.
- The storage and management of large volumes of data raise concerns about the environmental impact and sustainability of AI technologies.
- Inequalities in access to information could occur. In this regard, it is necessary to question the assumption of universal availability of digitised materials (Manžuch, 2017).

2.2 Legal framework

2.2.1 AI Act (2024)

The primary legal framework is the EU AI Act (2024), which has as its main goal to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the EU Charter of Fundamental Rights (Art.1).

AI Act classifies AI systems based on their intended purpose using a risk-based approach. This classification distinguishes between AI systems that pose unacceptable, high, limited, and low or minimal risk.

Fig. 1. Risk Pyramyd (EU AI Act 2024).

The AI systems that will be used in AUTOMATA are classified as low- or minimal-risk, meaning they do not pose significant risks. These systems fall into a residual category, encompassing all AI systems not included in higher risk levels (Casonato and Olivato, 2024). While low- or minimal-risk AI systems are not subject to specific regulatory obligations, they are encouraged to follow voluntary codes of conduct.

2.3 Ethical guidelines

2.3.1 EU HLEG- AI (2019) and UNESCO (2021)

In accordance with the regulatory framework, low- or minimal-risk systems are encouraged to adhere to codes of conduct or guidelines voluntarily. AUTOMATA refers to the main guidelines and codes of conduct in the international landscape (Jobin et al., 2019; Floridi & Cowls, 2019; Correa et al., 2023).

Although there is still no specific ethical compass for the use of AI in the cultural heritage sector (Tiribelli et al., 2024), the UNESCO Recommendations on AI Ethics (2021) and the Ethical Guidelines for Trustworthy AI, proposed by the High Level Expert Group on AI (HLEG-AI 2019) established by the European Commission, may help guide the reliable use of these technologies (Giannini & Makri, 2023). These guidelines aim to ensure AI technologies contribute positively to society while mitigating potential risks and ethical concerns. Starting from a fundamental rights-based approach, both these documents identify the ethical principles and related values that must be respected in the development, deployment, and use of AI systems.

It is also useful to briefly recall the UNESCO Charter for the Preservation of Digital Heritage (2003) with respect to the concepts of “digital preservation” and “digital continuity.” This document aims to outline principles for safeguarding digital resources and ensuring the preservation of digital heritage. Digital preservation encompasses the processes designed to ensure the ongoing accessibility of digital materials and to represent the original content to users accurately. Continuity of digital heritage is about maintenance throughout the life cycle of digital information, from creation to access. The Charter promotes the long-term preservation of digital heritage through the design of reliable systems and procedures that produce authentic and stable digital objects (UNESCO, 2003, art.5).

According to the EU Ethical Guidelines for Trustworthy AI, a trustworthy AI system must be lawful, adhering to all applicable laws; ethical, aligning with relevant ethical values and principles; and robust, meeting socio-technical requirements to prevent unintended harm.

The proposed principles are:

1. respect for human autonomy;
2. prevention of harm;
3. fairness;
4. explainability.

Following these principles, AI systems should meet seven requirements, which can be implemented using both technical and non-technical methods:

1. human intervention 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.

These principles and requirements aim to ensure that AI systems are developed and used in a trustworthy and ethical manner.

UNESCO (2021) proposes the following principles:

1. proportionality and do no harm;
2. safety and security;
3. fairness and non-discrimination;
4. sustainability;
5. right to privacy and data protection;
6. human oversight and determination;
7. transparency and explainability;
8. responsibility and accountability

2.3.2 AUTOMATA Ethical Framework

Considering the possible ethical issues previously listed (2.1) and the legal-ethical framework (2.2 and 2.3), AUTOMATA will pay special attention to the principles described in the following.

1. **(Digital) Preservation**: AUTOMATA should ensure continuity of access to digital materials during the entire life cycle of digital information, from creation to access. Digital preservation should be promoted by ensuring accurate data acquisition and improving the quality of 3D models. The use of robust data management infrastructures and adherence to FAIR data principles ensure long-term preservation and accessibility of digital heritage. Digital preservation will not replace, but rather enhance, the material preservation of artefacts.
2. **Accessibility**: AI technologies employed in AUTOMATA will improve accessibility by streamlining the digitisation process, making high-quality digital artefacts more readily available to researchers,

heritage professionals, and the public. Integration with cloud infrastructure will support scalable and efficient access to digitised data. Accessibility must also be ensured regardless of age, gender, ability or personal characteristics. Of particular importance is the accessibility of this technology for people with disabilities.

3. **Accountability**: The use of AI in digitisation requires clear accountability mechanisms to ensure that any errors or biases in the AI processes are identified and addressed. Auditability and traceability of (the working of) AI should be ensured. This includes maintaining transparency in AI decision-making and ensuring that human oversight is in place.
4. **Autonomy**: AUTOMATA employs artificial intelligence systems in digitisation by automating tasks such as calibration, data acquisition and 3D model building, reducing the need for human intervention and enabling more efficient workflows. The distribution of functions between humans and AI systems should follow anthropocentric design principles and leave ample opportunities for human choice.
5. **Explainability and Transparency**: AUTOMATA should make digitisation workflows open and understandable. This includes documenting AI algorithms, data sources, and decision-making criteria. AI systems used in digitisation should be designed to provide clear explanations of their processes and decisions. This is important for ensuring that users can understand and trust the outcomes produced by AI.
6. **Human Oversight**: Despite the automation of many stages, human oversight remains crucial to ensure the accuracy and quality of digitised data. Expert-in-the-loop intervention is necessary for tasks that require specialised judgment and to address any issues that AI systems may not handle adequately. Appropriate training of supervisory personnel should therefore be provided if necessary.
7. **Right to Privacy and Data Protection**: The digitisation process must comply with data protection regulations to safeguard sensitive and personal information. This includes implementing measures to protect data privacy and ensuring that data management practices are secure.
8. **Technical Robustness**: AI systems must be technically robust to ensure reliable and accurate digitisation. This includes developing algorithms that can handle various challenges in the digitisation process, such as identifying and eliminating outliers, improving alignment, and ensuring mesh quality. Measures or systems should be in place to ensure the integrity and resilience of the AI system against possible attacks. The behaviour of the system in unexpected situations and environments should be considered.
9. **Sustainability**: Considering the long-term impacts on society and the environment, AUTOMATA should include measures to reduce the life cycle environmental impact of the AI systems by reducing energy expenditure for each action performed, and, consequently, paying attention to global warming.

These principles are not more important than others, but are the most exposed to risk according to the previous ethical assessment (2.1).

Considering this ethical framework, AUTOMATA will adhere to the recommendations outlined in pertinent documents throughout the entire lifecycle of its AI systems. These ethical guidelines will be integrated into the design, factored into system selection, and assessed beforehand. Any violations will be promptly addressed and rectified.

Although it is no substitute for accountability, obtaining certification of the reliability of systems and the ethicality of their use is desirable (reference is made to GoodAILab of University of Pisa, Scuola Normale

Superiore and CNR, a research and service centre for the validation of reliable, transparent, robust, safe and ethical artificial intelligence systems based on the criteria just mentioned).

These principles and values guide the responsible use of AI in digitisation, ensuring that the technology is used ethically and effectively to enhance the preservation and accessibility of digital heritage.

3 Conclusions

In conclusion, the integration of Artificial Intelligence (AI) and Machine Learning (ML) technologies in the digitisation process offers significant advancements in efficiency, accuracy, and automation. AI plays a crucial role in real-time data processing, calibration, and the identification of areas of interest on artefacts, leading to improved sensor analysis and data acquisition. The automation of tasks such as calibration and 3D model construction enhances the quality of digital preservation, while machine learning models streamline the filtering and prioritisation of relevant data, ensuring the accuracy of the acquired information. However, despite the promising benefits, key challenges remain.

The ethical considerations surrounding the use of AI in digitisation are of paramount importance. As highlighted, automation can reduce human intervention, but this may raise concerns regarding the loss of human oversight and decision-making. The risk of algorithmic biases influencing data prioritisation and object segmentation must be carefully managed. Additionally, there is a concern that digitisation might prioritise digital preservation over the physical conservation of artefacts, potentially neglecting the tangible heritage itself.

The legal context, particularly the EU AI Act (2024), classifies the AI systems employed in AUTOMATA as low-risk, thus exempting them from stringent regulatory obligations, but encouraging adherence to voluntary codes of conduct. This framework underpins the ethical use of AI by focusing on principles that safeguard public trust and ensure compliance with legal standards. The integration of AI in digitisation must therefore align with both ethical guidelines and legal requirements, ensuring long-term digital preservation while respecting the rights of individuals and communities. Ethical guidelines, such as those proposed by UNESCO (2021) and the Ethical Guidelines for Trustworthy AI (EU HLEG-AI, 2019), provide a framework for ensuring that AI technologies contribute positively while addressing potential harms.

The ethical framework of the AUTOMATA project is grounded in a comprehensive approach that balances technological advancement with responsibility, ensuring AI systems are deployed in a manner that respects fundamental ethical principles. First, the principle of **preservation** is emphasised, as AI helps improve the quality of 3D models and ensures the long-term accessibility of digital heritage through robust data management infrastructure, following FAIR (Findable, Accessible, Interoperable, and Reusable) data principles. This ensures the continuity of digital heritage, which is critical for both current and future research.

Accessibility is another core tenet, with AI technologies designed to streamline digitisation processes, making digital artefacts readily available to researchers, professionals, and the public. The project also prioritises equitable access, ensuring that digitisation technologies are accessible to diverse populations, including individuals with disabilities.

The principle of **accountability** is also central to the framework, with clear mechanisms in place to ensure that errors, biases, or failures in AI processes are identified and addressed. Transparency in AI decision-making is crucial to maintaining trust in the system, and human oversight remains integral to ensuring the accuracy and reliability of digitised data. Furthermore, **autonomy** is maintained by allowing human oversight in critical decision-making stages, ensuring that AI systems support, rather than replace, human judgment.

The framework also emphasises **explainability and transparency**, requiring that AI workflows be documented and the decision-making criteria be made clear to users. This openness helps ensure trust in AI systems and their outputs. **Human oversight** is crucial throughout the digitisation process, as manual intervention may still be needed to address challenges that AI systems cannot handle autonomously.

The framework also ensures compliance with **data protection regulations** to safeguard sensitive information and emphasises the **technical robustness** of AI systems to ensure they function reliably and securely under various conditions. Lastly, AUTOMATA should implement **sustainability**-focused strategies to minimise the environmental footprint throughout the AI system's life cycle.

In order to prevent ethical validation from becoming a matter of ticking boxes, a continuous process of identifying and implementing requirements, evaluating solutions and improving results throughout the life cycle of the AI system, and involving stakeholders in that process must be put in place.

By adhering to these ethical principles, the AUTOMATA project ensures that AI systems contribute positively to cultural heritage preservation while mitigating potential risks and promoting societal trust.

References

AI Act (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

Casonato, C., Olivato, G. (2024). AI Regulation in Europe: Exploring the Artificial Intelligence Act. In: Fabris, A., Belardinelli, S. (eds) Digital Environments and Human Relations. Human Perspectives in Health Sciences and Technology, vol 150. Springer, Cham. https://doi.org/10.1007/978-3-031-76961-0_5

Dennis, L.M. (2020) 'Digital Archaeological Ethics: Successes and Failures in Disciplinary Attention', *Journal of Computer Applications in Archaeology*, 3(1), p. 210–218. https://doi.org/10.5334/jcaa.24

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1, 2-15. https://doi.org/10.1162/99608f92.8cd550d1

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., ... & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, *4*(10). https://doi.org/ 10.1016/j.patter.2023.100857

Gattiglia, G. (2025). Managing Artificial Intelligence in Archeology. An overview, *Journal of Cultural Heritage,* 71, 225-233, https://doi.org/10.1016/j.culher.2024.11.020

Giannini, E., Makri, E. (2023). Cultural Heritage Protection and Artificial Intelligence; The Future of Our Historical Past. In: Moropoulou, A., Georgopoulos, A., Ioannides, M., Doulamis, A., Lampropoulos, K., Ronchi, A. (eds) *Transdisciplinary Multispectral Modeling and Cooperation for the Preservation of Cultural Heritage*. TMM_CH 2023. Communications in Computer and Information Science, vol 1889. Springer, Cham. https://doi.org/10.1007/978-3-031-42300-0_32

HLEG-AI (2019). *Ethics Guidelines for Trustworthy AI*, 2019, doi: 10.2759/346720.

Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat Mach Intell* **1**, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2

Manžuch, Z. (2017). Ethical Issues In Digitization Of Cultural Heritage, *Journal of Contemporary Archival Studies*, 4, 4. DOI: 10.2788/8472

UNESCO (2003). Charter for the Preservation of the Digital Heritage. Available: https://unesdoc.unesco.org/ark:/48223/pf0000179529

UNESCO (2021). Recommendation on the ethics of artificial intelligence. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137