Ethics of Artificial Intelligence for Cultural Heritage: Opportunities and Challenges

Simona Tiribelli, Sofia Pansoni, Emanuele Frontoni, and Benedetta Giovanola

***Abstract*—Artificial Intelligence (AI) has witnessed remarkable advancements in recent years and has significantly impacted various domains, including cultural heritage. Indeed, AI technologies offer unprecedented capacities to analyze huge amounts of historical data, enabling researchers and art historians to uncover precious patterns, connections, and insights that might otherwise remain elusive. Also, the efficiency and accuracy of AI techniques play a pivotal role in many cultural heritage-related tasks, such as cataloging and organizing extensive cultural collections, streamlining the management of heritage resources for present and future generations. However, the integration of AI in cultural heritage also brings forth intricate ethical questions. These span over the issues of authenticity, subjectivity, and interpretation biases of an AI-empowered, reproduced, and/or generated artwork up to the legal concerns related to authorship. However, such issues are mostly undefined and unaddressed in the scholarship at the intersection on AI, ethics, and cultural heritage. This paper aims to pave the way to fill such a gap of context-sensitive ethical issues for AI in cultural heritage. To this aim, the paper first analyzes the main opportunities and benefits raised by AI in cultural heritage. Then, matching benchmark, agreed-upon AI ethics principles elaborated in the AI ethics scholarship in the last decade and relevant to cultural heritage, it highlights specific ethical risks that ought to be considered for the development and deployment of trustworthy AI *in* and for cultural heritage. Finally, areas requiring further attention and work, and actors call to intervene, are identified to facilitate next steps for ethics and governance of AI in cultural heritage.**

***Index Terms*—Artificial intelligence (AI), ethics, cultural heritage, cultural computing, digital transformation, generative AI.**

I. Introduction

THANK to the considerable technological progress of the last decades, artificial intelligence (AI) is gradually penetrating various previously unexplored areas, including creative and cultural sectors (hereinafter: cultural heritage). This is due to the unprecedented opportunities it showed to offer in this field such as the preservation and restoration of cultural heritage artifacts, promotion of museum collections through more engaging cultural experiences, user-tailored cultural offerings, and the creation of novel art forms through generative AI, just to mention a few. Indeed, the application of AI systems in the cultural heritage sector presents numerous benefits ranging from the automation of alienating and repetitive processes, the recovery of precious but lost cultural assets, as well as more immersive and fulfilling visitor experiences. For instance, Deep Learning (DL) models can automatically describe works of art, identify the artist, context, and period of the work, and overcome attribution or interpretation disputes [1]. Generative Adversarial Networks (GANs) further improve cultural heritage restoration processes by reconstructing missing parts of frescos or paintings [2]. Extended reality (XAI) techniques such as those for augmented (AR) and virtual reality (VR) help introducing new audiences to art and culture, assisting cultural organizations in fulfilling their educational mandate, and providing a more inclusive and engaging visitor experience.

The diffusion of AI in the cultural heritage field has also been prompted by the drive for digitization supported by two main levers. First, an institutional lever: organizations like the European Commission (EC) and UNESCO stress the importance of digitizing cultural assets through the promotion of programs like Europeana [3]. Indeed, cultural digitization policies seek to democratize culture by promoting accessibility and interoperability, information sharing, and the protection of cultural identity. Second, the wide endorsement of the so-called "adaptation approach" [4] to cultural heritage management and preservation enhances the cultural sector's digital transition.

Indeed, rather than focusing on protecting cultural heritage from the harmful effects of human presence (previous approach, called "mitigation approach"), the adaptation approach contends the preservation of cultural heritage as ensuring cultural continuity and giving heritage new values and meanings that depend on changes in society [4]. Therefore, in this scenario, the digitization of cultural heritage is promoted as an adaptation to the changes in a society that is continuously evolving, responding to the concept of "cultural continuity" [4]. Due to the social relevance of cultural assets and the morally-loaded dimension of cultural heritage, it is also crucial to consider the ethical and social implications

Manuscript received 13 October 2023; revised 18 April 2024 and 8 July 2024; accepted 12 July 2024. Date of current version 16 September 2024. This work has been funded by the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem under Grant ECS00000041 - VITALITY - CUP D83C22000710005. The authors also benefitted from the Jean Monnet Chair (Grant Agreement 101085372) EDIT - Ethics for Inclusive Digital Europe, co-funded by the European Union. *(Simona Tiribelli and Sofia Pansoni are co-first authors.) (Corresponding author: Sofia Pansoni.)*

Simona Tiribelli is with the Department of Political Sciences, Communication and International Relations, University of Macerata, 62100 Macerata, Italy, and also with the Institute for Technology and Global Health, PathCheck Foundation, Cambridge, MA 02139 USA (e-mail: simona.tiribelli@unimc.it).

Sofia Pansoni, Emanuele Frontoni, and Benedetta Giovanola are with the Department of Political Sciences, Communication and International Relations, University of Macerata, 62100 Macerata, Italy (e-mail: s.pansoni@unimc.it; emanuele.frontoni@unimc.it; benedetta.giovanola@unimc.it).

Digital Object Identifier 10.1109/TTS.2024.3432407

that AI technologies may raise when deployed with cultural heritage. Indeed, AI technologies are never neutral: they embed directly or indirectly socio-political, economic, and cultural values and agendas which can improperly affect cultural heritage. Challenges such as incorrect interpretation, historical and cultural biases, and economic and cognitive discrimination can counterbalance benefits provided by AI to one of the finest forms of human expression: art and creativity. Nevertheless, from an ethical and governance standpoint, such risks and issues concerning the specific cultural heritage sector have yet to be explored. This is problematic insofar as to develop digital transformation policies that maximize benefits and minimize risks, it is essential to take a sectoral perspective that complements the generally adopted horizontal view in AI ethics and governance landscape [5], [6]. Although the UNESCO Recommendations on the Ethics of AI [7] and the Ethical Guidelines for a Trustworthy AI [8], proposed by the high-level expert group on AI (HLEG-AI) set up by the EC, can help guide the trustworthy use of these technologies in general, there is still no specific ethical compass for using AI in the cultural heritage sector in an ethical and trustworthy-oriented way.

This paper aims to address such a gap by expanding on the first work of this kind elaborated in the field, entitled "Design of an Ethical Framework for Artificial Intelligence in Cultural Heritage" [9], by further developing context-sensitive ethical considerations, risks, and issues, supported by sector-specific cases and examples, and proposing a call to action to pave the way for the development of effective, sector-specific guidelines for trustworthy use of AI in cultural heritage. In particular, this paper deepens and extends the ethical issues mentioned in the cited paper focused on finetuning AI ethics principles for the cultural heritage sector by mapping onto existing high-level AI ethics frameworks and principles. This works extrapolates and elaborates on opportunities and ethical risks stemming from the use of AI in cultural heritage through the analysis of sector-specific ethical principles developed in such a refined framework. To do so, a cross-search of keywords from the cultural heritage field with those from AI (ML and DL) and AI ethics was conducted to identify scientific scholarship of interest. To identify relevant scholarship, three databases were used: Web of Science (WoS), Scopus, and EBSCO. WoS was chosen for its rigorous journal selection procedure, which is based on publication standards, expert opinions, regular appearances, and the quality of citation data, making it the most reliable database. Scopus was selected because it is the largest database of peer-reviewed literature and EBSCO was included to ensure comprehensive coverage and that no relevant articles were overlooked. In selecting reference articles, we used the following inclusion and exclusion criteria:

1. We included scientific papers focused on the application of AI in the cultural sector, specifically those dealing with generative AI for artwork production, AI for artwork restoration, AI for artwork cataloguing, and AI for creating immersive experiences in the art sector.
2. We included papers addressing AI ethics, as well as those concerning ethics and aesthetics in cultural heritage.
3. We included papers and policy documents that provided ethical recommendations and guidelines on the use of AI.
4. We excluded articles related to AI technologies that do not pertain to cultural heritage.

On the relevant papers collected, we conducted a qualitative content analysis to identify key issues and highlight recurring themes and ethical considerations. In order to reduce subjective bias, content analysis was manually and independently carried out by two authors. The main themes and insights from the fragmented literature were reinterpreted by applying high-level AI ethics principles as analytical lenses. This approach led to the creation of specific, context-sensitive, and clear ethical considerations related to AI in cultural heritage, addressing existing gaps in the literature. These considerations were then discussed in collaborative sessions to reach a consensus on the most critical ethical issues in the field. Three multidisciplinary expert workshops were conducted with co-authors and related partnered stakeholders. A total of 12 experts participated in the working sessions, coming from the fields of AI, AI Ethics, and cultural heritage (including two art curators, two art historians, and two visual artists). The experts included both academics and practitioners in order to ensure that the heterogeneity of professional and academic competencies involved in the field was adequately covered.

From a procedural standpoint, we proceeded as follows. In the first workshop, the draft considerations were presented to gather stakeholder feedback. This feedback was then utilized by the authors to refine the ethical aspects and considerations emerging form the use AI in cultural heritage. During the second workshop, the reviewed considerations were presented for a second round of feedback and further integration. The final workshop focused on presenting the main results of the paper including the final framework of ethical considerations for AI in cultural heritage to the participants. After presenting the topics the experts were first asked for qualitative feedback through semi-structured questionnaires, which included three main questions: a) Are the presented ethical considerations for AI in cultural heritage adequate and exhaustive according to your background and field of (academic and/or professional) activity? b) If so, what would you add to enrich them based on your expertise? c) If not, what would you modify and why? Subsequently, a collective reflection was conducted on the themes and considerations presented, which arose from content analysis and were re-elaborated through existing high-level AI ethics guidelines. From an ethical standpoint, the methodology to elaborate the specific ethical considerations before and after the collaborative sessions is detailed as follows. Alongside the content analysis of the collected papers, an ethical inquiry was conducted by the authors—comprising two AI ethicists, an expert in cultural heritage, and an expert in AI for cultural heritage. This inquiry involved examining the foremost techniques and applications of AI in cultural heritage, brainstorming on opportunities, as well as current and potential ethical risks. The analysis was framed using the primary AI ethics guidelines to ensure the trustworthy use of AI and the protection of cultural heritage, also taking into account widely discussed moral and societal issues in AI ethics and policy.

These ethical and policy documents along with scholarly articles related to ethical and aesthetic issues, such as cultural heritage's authenticity, interpretation, and dignity, proved to be critically important in incorporating the ethical challenges of the field into general ethical frameworks related to AI ethics. This comprehensive analysis facilitated the design of a call to action, aimed at proposing sector-specific guidelines for the trustworthy use of AI in cultural heritage.

In summary, building on general ethical principles for a trustworthy AI, this paper highlights the ethical challenges that arise from the various applications of AI in the field of cultural heritage and identifies areas where further work is needed and who is called upon to act. The objectives of this work are:

1. Determine the opportunities and benefits brought by the use of AI in the cultural heritage field;
2. Shine a light on the main ethical challenges and harmful consequences that AI can raise in a cultural and artistic context;
3. Identify areas of further work and the key players called into action to ensure ethical use of AI in cultural heritage;
4. Encourage an open discussion and informed debate on the ethical entailment of AI in cultural heritage, laying the foundation for the development of guidelines and measures that promote responsible and informed application of AI in the field, overcoming the generality of national and international standards by offering sector-specific ethical guidance.

The paper is structured as follows: Section II presents the main opportunities and benefits AI brings to cultural heritage regarding conservation, promotion, accessibility, and inclusion, drawing insights from the literature on AI for cultural heritage. Section III explores the ethical risks that AI arises when applied in the cultural context, matching AI and cultural heritage ethics concerns. Particularly, specific controversial scenarios in which the use of AI can undermine the preservation, inclusion, and interpretation of cultural heritage, including the people, groups, and societies at its core, are presented and analyzed. Section IV highlights both the areas requiring further attention toward ethical and trustworthy design and use of AI in the cultural sphere and the actors called into their action. Finally, Section V comments on the results of the discussion proposed in the article and provides final remarks, paving the way for future developments.

II. AI in Cultural Heritage: Opportunities and Benefits

AI presents itself as a powerful and innovative technology whose versatility and applicability in different contexts and sectors allow it to fascinate and conquer the field of cultural heritage. Although AI has only become the focus of public debate in recent years, applying Machine Learning, Deep Learning, and Computer Vision technologies in cultural heritage preservation and promotion has been practiced for nearly a decade. This cutting-edge technology provides three key opportunities, deeply explored in this section: accessibility and inclusion, cultural heritage preservation, and novel forms of creative production.

A. Increased Accessibility and Inclusion

Under their educational mission, museums must ensure the accessibility of their spaces and collections to people with disabilities, such as the physically or the visually impaired. In addition, the United Nations (UN) Convention on the Rights of Persons with Disabilities sets forth the obligation of the adhering countries to "recognize the right of persons with disabilities to take part on an equal basis with others in cultural life" [10, art.30, p.22]. One of the categories of people most disadvantaged in the enjoyment of art is the visually impaired. The main problems of accessibility for visually impaired people belong to two main categories: 1. problems of physical mobility and the consequent inability to move independently in a museum space, and 2. problems related to the lack of non-visual alternatives to artwork [11]. Although it is imperative to make content accessible to all, most museums and galleries' sight-centric design (which privileges sight over other senses) excludes people with visual impairments [12]. Augmented reality (AR) is a great tool to face these problems. Different solutions have been developed to ensure inclusiveness in cultural enjoyment, and many of these technologies can be installed on mobile devices making their use affordable for everyone. Through mobile applications empowered by AR technologies that can provide interactive descriptions of the artwork and 2D visual access, cultural content can be improved, overcoming the obstacles of using tools that are difficult for visually impaired people to use, such as audio guides [13]. In addition, through AR-based mobile applications, visually impaired visitors can independently explore exhibits and places of cultural interest safely and enjoy real-time descriptions. Through technologies such as image sensing and tracking, location, and simultaneous mapping, physical spaces can be augmented both visually as well as through other sensory modalities such as touch and hearing [14]. Auditory virtual tags affixed to objects in museum spaces can be activated based on the proximity between the object and the user's mobile device, providing users with information related to the context and characteristics of the work [15].

As for people with physical disabilities, another way to ensure their accessibility to museum content is by using VR to create a virtual environment and tours. Virtual environments allow navigating within a space and interacting with people in different places [16]. People with disabilities can encounter cultural experiences in VR and AR settings, improving the impression of "being there" or "feeling present" in a given space [16]. Different studies show that virtual visits are an excellent tool for people with disabilities allowing them to access and appreciate works of art [17]. In this context, terrific and yet unexplored opportunities can be offered by the virtual spaces that can provide the possibility of entering and visiting virtual galleries in a realistic and highly immersive way without having to travel miles to reach a physical museum or to enjoy an art exhibition or archaeological site.

Another opportunity presented by AI solutions is the creation of engaging cultural experiences that support audience development and promotional goals by enabling the creation of content and activities that can be customized to different types

of visitors. Immersive technologies can significantly enhance the visitor experience by engaging users in the museum environment. With speech recognition technology and interactive question-and-answer apps combining AR and VR, visitors can interact with virtual characters that can explain and interpret cultural content that emotionally engages them [18]. In this approach, engaging and interactive storytelling can overcome those communication barriers that often prevent art from raising the audience's interest. Using automated planning systems, AI can also create personalized visit itineraries that consider users' preferences. Cultural trails can be customized to suit the preferences and interests of tourists using multimedia tools like a user-friendly app. These systems may increase interest in and demand for museum exhibits and cultural activities. Based on the information entered, the planner develops a tailored path and suggests it to the users, selecting artworks that match their preferences and excluding those that do not [19].

Another benefit AI can bring to the cultural sector, in improving accessibility, is the automatic description and interpretation of artworks. AI algorithms can automatically describe, and catalog works of art. In this way, through high-performance computing (HPC), trained AI models, can significantly improve the information on cultural heritage [20]. To achieve this, tens of thousands of images and text descriptions are used to train a DL model for automatic data labelling. As a result, it is possible to automatically generate descriptions that consider culture, symbolism, and historical context for thousands of cultural heritage archives. The richness of the annotation enables a better browsing experience in collections catalogs and better access to collections. These results can support education, art, and tourism but also scholarly and research activities.

These applications bring significant social benefits: through greater inclusion and accessibility, it is possible to achieve greater democratization of art and culture, leading to a more culturally and artistically informed, sensitive, and aware society.

B. Improved Preservation Practices

Natural and human-made challenges such as climate change, natural disasters, conflict, and vandalism have made preserving tangible cultural heritage an increasingly complex challenge to face in the last decades. In this scenario, AI-based technological developments have completely changed how cultural heritage can be preserved and documented, addressing and mitigating damages while improving security measures. As Ocón states "*Digitalization can also provide a sort of memory insurance policy for cultural heritage, in the event of a total or partial destruction, or substantial degradation*" [21]. There are many examples of how, thanks to 3D modeling techniques, destroyed monuments have been revived. AI technologies, for instance, brought back to life Buddhas, which were destroyed in Afghan Bamiyan Valley in 2001, as holographic statues [21]. In a similar vein, the UNESCO-sponsored initiative "Revive the Spirit of Mosul" has been successful in using 3D modeling to restore monuments that were destroyed during the numerous conflicts that decimated the millennium-old city of Mosul [21]. Modern 3D scanning tools and AI algorithms allowed for the mapping of the damage and the planning of the reconstruction of Notre Dame Cathedral in Paris after the catastrophic fire that occurred there in 2019 [21]. In this regard, another great opportunity AI offers to the art field is the capacity to determine the authorship of a painting based on various stylistic characteristics. This makes it possible to solve problems of attribution of artworks avoiding the distribution of paintings with false authorship on the art market. Thanks to DL-based technologies, it is possible to identify a piece's author, context, and period, addressing attribution and interpretation disputes. Art specialists often employ attribution techniques that rely on identifying the unique visual characteristics of paintings, honing their skills by analyzing numerous artworks. In a similar vein, Convolutional Neural Networks (CNNs) are adept at detecting the distinct qualities that capture an artist's unique style. By training CNNs on a large digital collection of art, these networks can discern artist-specific visual traits crucial for reliable attribution [23]. While object recognition might seem straightforward to humans, pinpointing an artist's style and correctly attributing artworks remains a complex task, even for experts. In contrast, deep learning algorithms have shown remarkable accuracy in attributing art quickly and effectively. This technology significantly aids decision-making in museums, auction houses, and among art collectors, providing a robust means to verify the authenticity of artworks in the market.

Another problem that often arises in the restoration of works of art is the complete reconstruction of the artwork when some parts are missing or significantly damaged. In this sense, AI has tremendously aided the study of cultural heritage, notably in automatic object recognition and the translation of ancient languages, where DL algorithms can be used to fill in the gaps in damaged or only partially restored old texts [24]. Moreover, Generative Adversarial Networks (GANs) and DL algorithms broadly have recreated damaged images. These AI models could help conservators and restorers recover frescoes when pieces are damaged or missing [25].

AI technologies can also evaluate and understand vast quantities of geospatial and historical data, providing crucial details about cultural heritage sites and artifacts [26]. This process can offer information for conservation and maintenance efforts aiding decision-making processes in this regard. AI systems can also assist in creating digital cultural heritage models, supporting accurate simulations and reconstructions. For example, in detecting and mapping archaeological sites and other cultural heritage assets, ML techniques are utilized to analyze and interpret the 3D point cloud data produced by LiDAR sensors [27], [28]. Additionally, AI can automate 3D digitization processes, making the procedure time and cost effective, and allowing cultural heritage to be more widely accessible [29]. Indeed, AI-powered virtual reconstruction and restoration enable precise and thorough recreations of historical sites and artifacts, providing visitors and researchers with immersive experiences. AI can contribute to the preservation and transmission of cultural heritage by also offering innovative strategies for protecting our shared history and cultural heritage for future generations [30]. The structural

condition of cultural heritage can also be tracked using AI-based visual inspection systems, in which ML algorithms analyze visual data from sensors and cameras to identify and detect potential structural damage or deterioration [31] quickly and accurately. In this way, AI could track and detect changes to cultural heritage sites and artifacts for damage prevention and in-time intervention. AI-assisted risk and emergency scenario prediction can help to safeguard cultural heritage from upcoming catastrophes like natural disasters [32] or vandalism [33]. AI can also help recover historical property, enabling the identification and restitution of stolen or illegally sold cultural heritage items. In addition, AI can be employed for online crime heritage detection to spot and stop unlawful trading in cultural heritage artifacts [34].

C. Novel Forms of Creative Production

Thanks to generative AI based on GANs, it is now possible to create pieces of painting, music, or literary text using AI. In this way, artists can be assisted in creating new creative languages, leading to a collaborative relationship between AI and human agency and creativity. With the advent of generative AI platforms such as those from OpenAI, Mistral, Google AI, and Anthropic, pretrained algorithms and foundation models that can create creative and artistic content are becoming easily accessible to everyone. With tools such as DALL-E, Claude AI, Suno, Character.ai, Midjourney, or SORA, it is possible to create and conceptualize a work of art through a prompt, which is a text-based input describing the image to be created. While many criticize generative AI tools as automation tool of artistic production, more optimistic perspectives argue that they might create an opportunity rather than a threat [35].

The invention of photography was also criticized at its origin as the end of visual art. However, history has demonstrated that this was not the case: photography has offered the opportunity to move on to new artistic languages that have broken away from realism [35]. In this context, AI can be considered a new technological *medium* that can shape, enrich, and expand the artistic languages we knew to date.

III. AI in Cultural Heritage: Ethical Challenges and Risks

Despite the tremendous opportunities AI offers the cultural sector, it can simultaneously pose several ethical challenges and risks. In this section, we identify at least eight critical ethical risks that should be considered for the responsible and trustworthy use of AI in cultural heritage. These risks were identified by analyzing the applications of AI in cultural heritage through ethical lenses resulting mainly from the reasoned matching of the ICOM Code of Ethics (a reference text that sets standards for the practice of museum professionals) [36], the EU ethical guidelines for trustworthy AI [8], the Recommendation on the Ethics of Artificial Intelligence [7] promoted by UNESCO, and a preliminary sector-specific framework designed for the ethical use of AI in cultural heritage [9].

A. Undermined Authenticity

Various cultural organizations worldwide, including the International Council of Museums (ICOM), the International Federation of Library Associations (IFLA), and the International Council of Archives (ICA), have codes of ethics that aim to guarantee *authenticity*. UNESCO also fully promotes authenticity through the notion of "Outstanding Universal Value", which is the threshold for including a cultural heritage property in the World Heritage List [37]. Despite that, Nakonieczna and Szczepański (2023) defined authenticity as "an ephemeral phenomenon" [37]. They state that "depending on the ideology, the perception of authenticity can be quite diversified, even mutually exclusive" [37]. Regarding the digitization of cultural heritage, authenticity is a hotly contested topic. When referring to "authenticity", the original artwork's substance is emphasized as the sole source of verifiable historical evidence and a connection to the artist [38]. Some contend that interactive 3D reproduction does not accurately depict the original and may even dehumanize cultural heritage [39]. Various authors have long debated and addressed the issue of art authenticity. Indeed, it is frequently believed that a piece of art is unique and capable of inspiring an "aura" that is, a poetic and political gaze in the observer, and that any reproduction will not be able to substitute for the physical experience of the "hic et nunc" (here and now feeling) [40]. While AI provides new tools to make art more engaging through AR and VR, one can contend that since VR is a *medium* to access art, it can weaken or decrease the authenticity of the original artistic space and experience. Replacing physical with virtual contact could lead to a sort of communicative and relational impoverishment that would jeopardize the fulfillment of the educational goals of cultural heritage. Those who claim such risk maintain that museum experiences require personal, social, relational, or strictly physical interaction to increase heritage objects and sites' intrinsic enjoyment and engagement [16]. In this sense it is critical that VR experiences, and thus the AI they depend on for dynamic content generation and optimization, are designed to be highly believable, facilitating interaction and navigability [41]. Some of these issues arise in the case of real-time rendering, a crucial aspect of many VR experiences that impose limitations on the complexity and richness of visual details. This constraint can affect artists who rely on intricate details to convey their artistic vision. A digital sculptor creating highly detailed virtual sculptures may face compromises in visual fidelity due to the real-time rendering demands of VR, potentially impacting the level of detail and nuance in the final piece. In this way, the authenticity of the artwork experience and the message they convey risk being compromised.

The problem of authenticity also affects AI-generated works. If the authenticity of a work of art is based on a poetic relationship between the viewer and the artist, the question naturally arises as to whether this relationship can also be established with a machine-generated piece based on yet-existing artistic training images. Since the authenticity of a work of art is closely linked with its authorship and

provenance, a few key questions arise and will have to be answered though valid arguments in the case of AI-generated art in the near future: Who is the author? Can that work, created by a machine, be said to be authentic?

Another challenge with regard to authenticity is posed by AI automated planning systems that create cultural offerings tailored to consumer tastes and preferences, therefore undermining the educational task of cultural institutions. By offering just what appeals to audiences (popular contents), while potentially omitting less popular and mainstreaming artworks, this technology can diminish the provocative and critical power of the arts, it could also undermine the heterogeneity in creative production leading to more impactful and greater aesthetic standardization.

To prevent this challenge, artists must avoid ceding to content impoverishment actions in the light of a simpler, more engaging, and accessible experience, by working on preserving their intended narrative and meaning while generating viewer engagement. This entails a careful ethical consideration to not let the emphasis on creating an engaging and often standardized or mainstream experience compromise or dilute the unorthodox, unexpected, and sometimes disturbing effect very often invaluably connoting or characterizing artistic expressions.

In summary, preserving cultural heritage authenticity is a significant and necessary ethical problem to address. As Nakonieczna & Szczepański (2023) point out, the concept of "Authenticity of cultural heritage" has evolved over time and now needs a paradigm shift, a redefinition of this concept, taking into account the new technologies that are revolutionizing how the artistic experience is enjoyed: they do not mediate it anymore, while deeply influence and shape it [37]. To the best of our knowledge, no current AI ethics framework considers this requirement to be of utmost significance. Therefore, it is today of paramount importance to define the authenticity of a work of art (what it means and requires to be ensured) and include this core ethical value and principle within a sector-specific ethical framework for ethical AI in cultural heritage.

B. Inaccurate Interpretation of the Artwork

The issue of inaccurate automatic interpretation of the artwork fits into broader ethical framework principles, particularly in the issue related to the accountability of AI systems. According to UNESCO's Recommendation on AI ethics [7], AI systems ought to incorporate methods for oversight, impact assessment, verification, and due diligence to maintain AI accountability. For this reason, a preliminary sector-specific Ethical Framework entails reliability as a fundamental ethical principle for using AI in the cultural heritage context [9]. The latter framework stresses how the ethical requirement of Human Agency and Oversight (see HLEG-AI 201) becomes crucial to address accountability of AI in cultural heritage, which always requires the cultural heritage expert in the loop. In particular, [9] argues how human agency can tackle one of the main obstacles to using AI for the automatic description and cataloging of works: the quality and quantity of data required and available to train AI models [20]. Training an automatic captioning model requires a dataset with corresponding images and descriptions. Cultural heritage metadata usually does not refer to the visual content of the artwork that, by contrast, is crucial to formulate a complete analysis of the work. In addition, a highly symbolic component characterizes a work of art: the same element may have different meanings depending on the artistic movement to which it belongs or the historical period of reference. For this reason, the dataset's annotation phase is of paramount importance. Thus, if the descriptions used to train the model are incomplete, there is a risk of transmitting an inaccurate description to the end user. Consequently, omitting critical artwork features might change its social, cultural, and artistic meaning. An example of a risk that may occur is that an AI algorithm may misinterpret the symbolism in a painting, associating an ordinary object with a particular meaning that the artist does not intend—for instance, interpreting a butterfly as a symbol of freedom when the artist intended it to represent fragility. Another problem may arise in the automatic interpretation of conceptual art, which often relies on ideas and contexts rather than traditional artistic elements. AI may struggle to interpret and appreciate an artwork's conceptual nature, resulting in a superficial analysis that misses the deeper meaning. For this reason, to be reliable, the development of automatic captioning systems for describing works of art requires extensive annotation of datasets related to archives and catalogs, which can be done only and exclusively by *domain experts* such as art historians, archaeologists, linguists, and philologists. Here the role of human agency and oversight as a requirement becomes clearer. It is worth specifying that the choice of such experts is also important. The lack of representation and cultural diversity (also in terms of artistic influences, backgrounds, ethnicity, age, and gender) might lead to another kind of risk and ethical challenges, belonging to the category of biased interpretation.

C. Biased Interpretation

Diversity and nondiscrimination are two fundamental ethical requirements outlined in the UNESCO [7] recommendations for using AI and the EC Guidelines for Trustworthy AI [8]. These high-level requirements arise from the general AI ethics principle of justice and fairness, present in almost every framework developed to date for AI [6]. These requirements must be considered when applying AI in cultural heritage, for example, to automatically interpreted works, in order to avoid the risk of culturally limited or biased interpretations. Although experts often try to define objective criteria for reading works of art, interpretations often vary according to the cultural and epistemic background of the art critics.

One might question the importance of interpretation. Indeed, according to Susan Sontag "Interpretation is not an absolute value [· · ·] Interpretation must itself be evaluated, within a historical view of human consciousness. In some cultural contexts, interpretation is a liberating act [· · ·]. In other cultural contexts, it is reactionary, impertinent, cowardly, stifling" [42, p. 4]. In this definition of interpretation, highly subjective aspects indeed emerge; aspects that vary according to diverse personal encounters people can have with art. For this reason,

in criticism, as in art, there are many currents of thought, each with its own particularities and different readings.

When the interpretation of a work is made through AI models, the choice inevitably falls on the most dominant and accessible via data reading of the work (*informativeness bias*). In this way, only the most popular interpretation is transmitted and communicated, often leading to a cultural-historical bias and the marginalization of less dominant, but no less significant, interpretations and hence understanding [43]. This happen especially in interpreting the cultural heritage of minorities such as indigenous communities or communities with a colonial heritage, which tend to be vulnerable to be improperly influenced and read by ethnocentric, Western-centric lenses [44]. In the attempts to assimilate indigenous knowledge and spirituality into a "Western" worldview interpreting cultural heritage metadata in such a way risk undermining local cultures' veracity [45]. Particularly, this risk mainly occurs when AI models are trained on datasets that reflect biases prevalent in historical art collections or among critics. This challenge can lead to a skewed understanding of art history, reinforcing stereotypes, as well as overlooking contributions from marginalized groups. Also, there is a risk of underrepresentation of diverse perspectives: If the training data lacks diversity in terms of artists' backgrounds, styles, or cultural influences, the AI model may struggle to interpret artworks that deviate from the dominant narratives present in the training data. High-level ethical principles and requirements of diversity, fairness, and nondiscrimination are violated in these situations [7], [8]. The automatic interpretation of works thus runs the risk of failing to reflect minorities while flattening multiple readings of a work. Therefore, the inability of AI to reflect all the different and myriad interpretations of a work become a relevant challenge to consider.

In this context, as some authors have started to observe, it is critically important to deeply expand and revise (decolonize) core AI ethics principles by considering also other cultures besides the Western to develop truly representative ethical systems of values to embed in AI [45]. The ethical frameworks characterized by principles and values that guide AI ethics run the risk of being biased by the vision of Western moral philosophy; that is, they risk failing to consider ethical perspectives belonging to other, more marginalized, and vulnerable cultures (e.g., Ubuntu ethics in Sub-Saharan Africa), which deserve equal attention for a globally sensitive AI ethics discourse [46]. Similarly, the application of AI in cultural heritage faces the risk of becoming a channel for perpetuating patterns of colonialism. For this reason, the ethics of AI for cultural heritage must move beyond the widely acknowledged risks such as privacy and autonomy issues to consider AI's role through cultural heritage in nurturing forms of colonialism [47]. In this scenario, it is possible to identify two major ethical issues: The first concerns risk of cultural identity erasure of minorities resulting from the failure to consider cultural heritage's multiple cultural and semantic facets. Data abstraction processes, i.e., the ability to automatically provide a semantic characterization of data [48], can cause this risk, implementing data simplification procedures and subsequent generalization of details. The second issue entails that a globally sensitive and inclusive ethics of AI for cultural heritage requires the decolonization of cultural consumption. That is, it is necessary to conceive cultural heritage as something created by and for the target communities, and consequently, it is only through their reading, customs, and traditions that it must be interpreted and enjoyed. On the contrary, if this effort is missing, the application of AI will lead to a cultural heritage's innovative fruition. Still, only in terms of the medium and not in terms of content: in fact, it will only replicate the patterns of cultural reproduction, appropriation, and exploitation already in place.

D. Digitization as an Excuse to Destruction

Another principle common to the EC guidelines and UNESCO recommendations is sustainability of AI [7], [8], intended as societal and environmental wellbeing. That is, there is a need for AI to consider the protection of the environment, the natural ecosystem, and biodiversity and contribute to more just and fair societies [9]. For this reason, AI should be used in a way that succeeds in reducing its energy consumption, which is often directly proportional to its performance levels. Also, it should be used to support innovative reconstruction and restoration systems. This principle aims to limit the risk associated with digitization and subsequent virtual reproduction of artworks: the possibility that digital representation will supersede the significance of the physical one, justifying the destruction of certain physical cultural heritage sites and artifacts that are of historical and cultural significance to local communities while not on the World Heritage List –therefore not protected by relevant legislation. For example, accurate 3D modeling and reproductions that allow digital access to cultural heritage can justify interventions to demolish it to construct new structures for purely economic and commercial purposes. In the face of this risk, Ocón [21] examines the challenges that Southeast Asian cities faced in balancing the preservation of unprotected cultural heritage with urban development needs. In particular, the author highlights how digital archiving, 3D modeling, and VR have been used to document and virtually preserve precarious sites such as The Bukit Brown Cemetery in Singapore and the Pekeliling Flats and A&W Petaling Jaya in Malaysia before they were altered or demolished.

While these technologies have been crucial for conserving these sites, it is important to also consider the risks of relying solely on digital preservation. Using digital methods as justification for the physical destruction of cultural heritage could result in the loss of tangible elements that often hold significant identity value for the reference group.

In addition, the application of AI in cultural heritage entails significant human and environmental costs that frequently conflict with cultural heritage's social and ecological sustainability goals [45]. For example, the development of AI systems requires the occupancy of space (e.g., data centers), technological infrastructure, and environmental resources, which can adversely affect areas that are crucial for the preservation and well-being of specific historical, natural and cultural sites, potentially degrading these environments and diminishing their historical and cultural value.

E. Socio-Economic and Cognitive Accessibility

Building on the core AI ethics principle of justice and fairness, the sector-specific framework developed for using AI in the cultural heritage context recognizes accessibility as a fundamental principle [9]. Indeed, it is essential to ensure equal possibilities of access to and enjoyment of these technologies independently of different kinds of institutions and audiences. As echoed in the UNESCO recommendations, spreading awareness and literacy of these technologies to domain experts and target audiences is crucial to making these technologies accessible to all [7].

To further address these challenges, it is imperative to also consider the provision of adequate economic and educational resources. Developing AI models to promote and sustain cultural heritage requires significant economic investment and mainly specialized human resources [5]. As a result, there is a risk of disparities between larger and wealthier institutions that can afford to make significant investments and smaller institutions with niche productions that cannot bear the high costs. Similarly, because of the high investment required to develop such technologies, there is a risk that cultural heritage of economically weaker countries will be excluded from the digitization process [49]. The risk of unfairness and discrimination related to the need for significant economic resources to develop AI systems, occurs when only wealthier and more developed countries can afford AI technologies to preserve and restore their cultural heritage. In contrast, economically disadvantaged countries are likely to be left behind and find the process of reconstruction, preservation and accessibility improvement challenging to implement, leading to a distinction between "A-list cultural heritages" and "B-list cultural heritages"

Another problem is the cognitive accessibility of digital content. Digitization allows better access to cultural heritage, even for people with physical disabilities. For people with cognitive disabilities, there is a risk that these processes will marginalize them even more [50]. Indeed, it is essential to ensure that the digitization of cultural heritage does not result as an end in itself, but that digitalized artworks, services, and experiences are also cognitively accessible and enjoyable by the target audience. For this reason, once access to digital collections has been enabled or immersive experiences have been developed through AR and VR, it is also essential to ensure that these products and content are consciously consumed. To this end, it is essential to design these products considering the nature and source of potential users' cognitive vulnerabilities to develop interfaces that compensate for, rather than exacerbate, such imbalances or diversities (e.g., neurodiversity).

For instance, the target audience may lack the necessary training to utilize AI technologies appropriately, which creates an additional barrier. This risk is especially significant when enhancing the cultural heritage experience by making it more personalized and inclusive. Without a proper understanding of how to interact with AI-driven tools, users may feel frustrated, incapable, or confused, and even excluded by design. It may also ruin their overall experience and potentially discourage them from further exploring AI-enhanced cultural content. Additionally, AI applications often require understanding complex interfaces or interpreting AI-generated recommendations. Untrained users may find these aspects challenging, leading to misuse or underutilize AI tools. Furthermore, without proper or personalized training, people with disabilities or older adults may find it particularly difficult to use AI technologies, limiting their access to art and cultural education. These are just a few examples of risks that can occur. Therefore, as outlined in the UNESCO recommendations, the development of AI systems must be accompanied by processes of education and literacy about how these technologies work by improving their accessibility.

F. Universal Access and Top-Down Approach

One of the sector-specific principles developed for AI in cultural heritage [9] is meaningful participation. This prescribes that the use of AI technologies should be implemented in an inclusive manner through a bottom-up approach involving different stakeholders and competencies [9].

Many contend that cultural heritage digitization is frequently adopted top-down and as a means rather than an aim, omitting several vital aspects of the community's needs. As Taylor and Gibson (2016) argue "*certain digitization activities can subtly reinforce non-democratic structures*" [51]. For example, this might be the case of digitizing the cultural heritage of indigenous communities. This would ask for both hearing and centering their voices and considering the consequences that such operation (i.e., the resulting free access to that heritage) would bring to the target community. The target communities conceive some cultural sites as sacred or symbolic, and their access is allowed only to people of a specific gender or age. For this reason, there is a risk that the digital reproduction of a cultural heritage site and its resulting free access will not align with the worldview of the target communities. Neglecting communities' needs and values reinforces different forms of discrimination [39]. Consequently, these risks are at odds with the commitment of cultural institutions to present a variety of perspectives and to promote cultural diversity and mutual dialogue between cultures [39], as well as with the ethical principles of fairness, equity and nondiscrimination promoted by the general EU and UNESCO recommendations [7], [8]. Moreover, the involvement of communities' members might help overcome the risk of excluding them from benefiting from AI's economic and cultural opportunity in cultural heritage without violating its sacred and identity value.

For some communities, cultural heritage has sacred value, and using disruptive technologies such as AI to enjoy this heritage could be a form of heresy. To respond to this risk, it becomes necessary to introduce an additional domain-specific ethical principle: dignity, i.e., the need to ensure the respectful preservation of culturally sensitive material, such as material of sacred, identity significance, and its intrinsically human nature [9]. For this reason, it is desirable to engage in an open dialogue with the communities that live the cultural heritage to represent it in a manner consistent with their customs and traditions.

G. Responsibility Attribution for Actions and Decisions Based on AI Outcomes

According to the EU guidelines, explainability is another essential principle and condition for trustworthy AI [8]. The need for explainable AI is associated with the risk of a black box effect. Indeed, the better AI performs, the more difficult it is to explain the solutions provided by such models. However, since cultural heritage is a common good and an expression of the values of the relevant community, the solutions and decisions on cultural heritage obtained by AI models should be explainable to defend AI-based results with a rational explanation that non-technical users, including audiences and communities, can understand. For example, AI can significantly support restoration practices, reconstructing missing parts of frescoes or old damaged buildings. However, if one cannot explain why the AI arrived at that solution, it is difficult to assess if that solution is flawed or trustworthy. Most importantly, if that decision proves to be incorrect, it would become hard to determine who is responsible for the resulting consequences and changes on the cultural heritage. The complexity arises from the involvement of multiple actors in AI systems, including system designers, software producers, and relevant data providers. Moreover, assigning responsibility is further complicated by AI systems' ability to autonomously modify themselves (self-adapt) based on the processing of new data or their continuous interaction with their environment. This lack of accountability risks damaging invaluable cultural assets and undermines public trust in AI applications within this sensitive field.

For this reason, explainability is a key pro-ethical condition for assessing and distributing responsibility. Indeed, one of the most significant ethical challenges is the need for an understanding of who is responsible for decisions made based on the results processed by AI systems. Assigning responsibility for decisions based on results derived by AI systems is a highly debated topic in several areas of AI application, such as healthcare and justice. However, the use of AI in the cultural sector still needs to be better considered. Regardless of the extent to which AI is used in cultural heritage preservation practice, it is essential to clarify who bears responsibility for decisions based on algorithms' results, predominantly when those decisions cause harm or lead to adverse or incorrect outcomes. It is essential to define whether the algorithms can be considered, to some extent, agents or whether the responsibility for their decisions lies only with individuals such as developers, computer scientists, or end users such as art historians and archaeologists [52]. Given the risks listed above and knowing that AI systems, as well as human intelligence can fail, it is vital to recognize those responsibilities when such threats occur working on this issue ex ante. Specifically, determining who is responsible for situations such as the misreading or misinterpretation of a work of art, the failure of a sensor to monitor the deterioration of cultural heritage resulting in its damage, or the incorrect cataloging of a work of art, is crucial for a reliable and trustworthy use of AI in cultural heritage. In addition, decisions based on AI algorithms can also result in economic injuries to buyers and sellers in the art market, such as in the case of misattribution of the work. The issue of allocating responsibility is also strictly linked to the new challenge in AI-generated contexts of attributing authorship for art images and products especially in cases of copyright infringements. For example, the question of who can be held responsible for copyright infringement of a work used by the AI system as a training image to create the AI-generated creative content must be clarified by design. We expand this challenge in the next section.

H. A New Concept of Authorship

Further concerns to consider are today widely related to the so-called rise of Generative AI which makes it possible to create realistic art products such as poetic texts, melodies, or paintings [2]. As previously outlined Generative AI produces creative content using training data from existing art media [35]. Despite the opportunities Generative AI can bring to art and creativity we outlined above, it also poses significant sector-specific issues related to the determination of authorship such as the risk of copyright infringement and the challenge of the creative industry to ensure fairness in the labor market [53]. Since generative AI produces creative results by training on existing images often copyrighted and part of cultural heritage, copyright infringement can frequently occur. In particular, the final outputs could be considered derivative works for which the original artists could claim authorship, at least jointly [53]. If, on the one hand, there is a risk of alleged copyright infringement of the training images, there is also the challenge of determining authorship of the new creative work created by Generative AI. Therefore, the question arises about who has the right of authorship to a creative output achieved by generative AI. Indeed, text-to-image systems underlie the creation of artistic images using generative AI. Tools such as Midjourney, DALL-E, and SORA work effectively when the prompts used are specific and well-written [54], [55]. This has given rise to a new research area called "prompt engineering," where it is essential to determine the extent to which the right of authorship rests with the prompt engineer, the AI system, or its developers. The importance of establishing the authorship of the work lies in the primary purpose of copyright law, which is to promote creative production as a social practice to support the dialogue between the viewer and the artist, who has a political role in society [56]. From this aspect arises the question of authenticity. Suppose only art that serves as a medium for poetic dialogue between artist and viewer, that is, between two human beings, is understood as authentic. In that case, understanding the author of an AI-generated work with which the viewer enters a dialogue is of utmost importance to assess AI-generated artworks' authenticity. Consequently, resolving the work's authorship dilemma is fundamental to defining its authenticity.

Finally, linked to this aspect, generative AI poses a significant challenge to the future of the cultural heritage labor market [35]. Indeed, suppose authorship of AI-generated works rests with a computer. In that case, this will bring the risk of technological substitution not only for repetitive jobs

that follow explainable processes but also work requiring more creativity or domain expertise. For this reason, these systems have begun to worry professional groups such as composers, designers, and writers who, until a few years ago, saw their jobs at low risk of technological substitution because creativity is a process that is traditionally human, to date not easily codifiable in concrete rules [35].

Furthermore, those in the industry who cannot use these tools due to a lack of technical skills will inevitably be unable to benefit from their opportunities. In this scenario, generative AI challenges the role of humans as creative agents in society [57], risking delegating to algorithms even the most human of capabilities, namely creativity and invention. Consequently, human creativity, imagination, and even divergence from AI logic might be channeled and discouraged.

IV. Call to Action

While many ethical issues related to the use of AI in the cultural heritage field fit within broader AI ethics frameworks, challenges emerge in the cultural heritage field that do not precisely fit within existing ethical frameworks. Aspects such as authenticity, dignity, shared responsibility, and meaningful participation are unique in their kind and specific application of AI in the cultural context. Hence, our call to action is to consider the risks and opportunities we discussed in order to manifest the urgency of a sector-specific, actionable ethical guidelines for using AI in cultural heritage.

Given the great amount of possibilities AI offers to cultural sectors, it is necessary to promote its conscious development while considering the risks and ethical challenges that its application may entail in a field, namely that of cultural heritage, which until recently was the exclusive domain of human beings.

Despite the significant efforts of actors such as UNESCO with its Recommendations on the Ethics of AI [7], these guidelines provide only the high-level groundwork for determining the requirements that AI technologies should meet to ensure their trustworthy use in the field of cultural heritage. Considering the specific ethical risks and sensitive issues identified above for AI in cultural heritage, these guidelines need to be further developed in a context-sensitive way to be truly effective and actionable in this field. This section will contribute to pave the way in this direction by highlighting the areas that need research and work to create a sector-specific ethical and policy compass and regulation. For each area of work, we define also the involved stakeholders. It is worth outlining that a few of the pathways we propose are more actionable in the near future (e.g., establishing criteria for authenticity through consensus votis among experts and stakeholders), while others require further time and effort due to their complex nature (e.g., to define the authorship of AI-generated works of art as it required knowledge of data input and training of the systems that are often opaque).

First, the conditions under which AI technologies do not compromise the authenticity of the artistic experience must be identified. Furthermore, it is crucial to reformulate the concept of artistic authenticity and the criteria for determining what is authentic in light of the incursion of AI in cultural heritage. Moreover, it is essential to define the criteria under which digitization makes the artistic experience more accessible and comprehensive while retaining the centrality of the physical artwork. In other words, it is crucial to understand in which conditions digital artistic experience can undermine cultural heritage's authenticity, leading to substituting the physical one. To work on this aspect, the contribution of experts in the field, such as art historians, cultural mediators, and aesthetics, who can define the boundaries within which the artistic experience retains its authenticity, is essential.

Second, it is necessary to develop greater literacy regarding AI and digitization processes [7] so that it does not remain the exclusive preserve of a highly educated and elitist audience. For this reason, it is essential to introduce digital and computer science skills at all levels of education, from elementary school to higher levels. Redesigning educational systems with interdisciplinary capabilities offers the opportunity to combine artistic skills with cutting-edge technological solutions. To ensure the preservation of cultural heritage and encourage innovation and creativity, a collaborative approach between AI and professionals in the field is necessary, rather than viewing them as mutually exclusive. That would be possible through education, the acquisition of innovative skills, and greater awareness of how AI tools work. Only in this way AI can become a beneficial tool rather than a harmful substitute. Better education and a greater understanding of how AI works will help develop systems that foster creative human agency rather than replace or undermine it. Working on these goals necessitates the involvement of technologists and technical experts who define the minimum requirements for the informed use of these technologies to spread. Their involvement needs to arise from national and international institutional actors (e.g., UNESCO or Committee on Culture and Education (CULT)), defining new education policies based on an understanding of how technologies work, which would ensure a more informed and conscious use of them. Furthermore, AI systems must ensure justice as fairness and nondiscrimination when applied to cultural heritage. For this reason, in a living heritage context, the needs and cultural values of the local community must be respected and considered. Therefore, it is necessary to develop workflows to implement digitization projects with a bottom-up approach, starting from the needs of the population stewarding cultural heritage. In other words, defining criteria for evaluating digitization actions based on a community-based approach is essential. In line with the recommendations of UNESCO, a multi-stakeholder approach must be adopted, to ensure the centrality of the reference group [7]. In particular, the cultural sector's public, private, and nonprofit stakeholders are called upon. This line of effort is undoubtedly the most complex, requiring more significant commitment. The needs of the local community change depending on the area in which it operates, so the definition of general reference standards requires a noteworthy collective and collaborative effort. Moreover, needs and preferences also change over time, hence, ongoing monitoring and revision may also be needed.

Finally, further work needs to be done on the legal side. According to EC Ethical Guidelines for a trustworthy AI, to be

deemed trustworthy, AI systems should meet the requirement of accountability [8]. This means that, on the one hand, clear regulations must be developed for algorithms' accountability. On the other hand, appropriate redress and compensation mechanisms must be defined in case of harm or dangerous outcomes, for example, when a cultural heritage site is damaged, mining the history and memory of a community, when a misattribution of a work of art causes economic harm to the owner or buyer, or in case of incorrect interpretation or cataloging. In addition, an international copyright law update that protects artists' rights and their works from copyright infringement by generative AI is particularly urgent. Furthermore, it is necessary to clearly define whether AI-generated creative works can be considered artistic expressions and, if so, who owns their authorship. In this case, the scientific community and governments are called upon to reflect and develop new regulations and a unified perspective to provide legal clarity on this issue.

V. Conclusion

This article attempted to shed light on the potential of AI for the promotion and preservation of cultural heritage considering the ethical implications that these applications may have, such as the loss of authenticity of the artwork, the attribution of responsibility for decisions based on the results of algorithms, the risk of discrimination against minorities, the implications related to universal access to cultural heritage and, finally, the issue of authorship and the role of the artist in our society. AI has proven to be able to enter every aspect of our existence, and it is doing so in an increasingly rapid and high-performance manner, making it difficult for institutions to respond promptly to the diversity of challenges that arise with this technology. The importance of this work lies in its call for institutions, the scientific community, and experts in the field to action for establishing specific criteria and standards that go beyond the generalist approach of the guidelines that have existed to date and develop clear and specific industry guidelines for harnessing AI's benefits and preventing its risks for cultural heritage, thereby promoting an informed use of these technologies without jeopardizing the positive social impact of art and culture. Aware that so much work needs to be done in this direction, we hope our study will pave the way for further multidisciplinary and multi-stakeholder efforts, involving professionals from different sectors – from art historians to archaeologists, from computer scientists to jurists – in the field of ethical AI for cultural heritage, as far as necessary to improve a relational, mutually flourishing approach between AI agency and human agency and creativity.

Acknowledgment

All the authors have contributed to editing this paper. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

References

[1] A. Felicetti, M. Paolanti, P. Zingaretti, R. Pierdicca, and E. S. Malinverni, "Mo.Se.: Mosaic image segmentation based on deep cascading learning," *Virtual Archaeol. Rev.*, vol. 12, no. 24, pp. 25–38, Jan. 2021, doi: 10.4995/var.2021.14179.

[2] S. Shahriar, "GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network," *Displays*, vol. 73, Jul. 2022, Art. no. 102237, doi: 10.1016/j.displa.2022.102237.

[3] W. Tabone, *Technology for Cultural Heritage, Tesserae Spring Edition*, Heritage Malta, Kalkara, Malta, pp. 46–54, 2019.

[4] G. Boccardi, "From mitigation to adaptation: A new heritage paradigm for the anthropocene," in *Perceptions of Sustainability in Heritage Studies*, Berlin, Germany: De Gruyter, 2015, pp. 87–98.

[5] K. Izsak, A. Terrier, S. Kreutzer, T. Strähle, C. Roche, and M. Moretto (Publ. Office Eur. Union, Luxembourg City, Luxembourg). *Opportunities and Challenges of Artificial Intelligence Technologies for the Cultural and Creative Sectors*, 2022, doi: 10.2759/144212.

[6] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Mach. Intell.*, vol. 1, no. 9, pp. 389–399, 2019.

[7] "Recommendation on the ethics of artificial intelligence." UNESCO. 2021. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137

[8] (Publ. Office Eur. Union, Luxembourg City, Luxembourg). *Ethics Guidelines for Trustworthy AI*, 2019, doi: 10.2759/346720.

[9] S. Pansoni, S. Tiribelli, M. Paolanti, E. Frontoni, and B. Giovanola, "Design of an ethical framework for artificial intelligence in cultural heritage," in *Proc. IEEE Int. Symp. Ethics Eng., Sci. Technol. (ETHICS)*, West Lafayette, IN, USA, 2023, pp. 1–5, doi: 10.1109/ETHICS57328.2023.10155020.

[10] D. MacKay, "The United Nations Convention on the rights of persons with disabilities," *Syracuse J. Int'l L. & Com.*, vol. 34, p. 323, 2006.

[11] S. Asakawa, J. Guerreiro, D. Ahmetovic, K. M. Kitani, and C. Asakawa, "The present and future of museum accessibility for people with visual impairments," in *Proc. 20th Int. ACM SIGACCESS Conf. Comput. Accessibil.*, 2018, pp. 382–384, doi: 10.1145/3234695.3240997.

[12] F. Candlin, "Blindness, art and exclusion in museums and galleries," *Int. J. Art Design Educ.*, vol. 22, no. 1, pp. 100–110, Mar. 2003, doi: 10.1111/1468-5949.00343.

[13] D. Ahmetovic, C. Bernareggi, K. Keller, and S. Mascetti, "MusA: artwork accessibility through augmented reality for people with low vision," in *Proc. 18th Int. Web All Conf.*, 2021, pp. 1–9, doi: 10.1145/3430263.3452441.

[14] K. Sheehy, H. Garcia Carrizosa, J. Rix, J. Seale, and S. Hayhoe, "Inclusive museums and augmented reality: Affordances, participation, ethics, and fun," *Int. J. Inclusive Museum*, vol. 12, no. 4, pp. 67–85, Jan. 2019, doi: 10.18848/1835-2014/CGP/v12i04/67-85.

[15] N. Cherukuru, R. Manuel, A. J. Lauer, T. Scheitlin, and B. Bhagchandani, "Using augmented reality (AR) to create immersive and accessible museums for people with vision-impairments," in *Proc. MuseWeb*, 2021, pp. 1–23.

[16] N. Dìaz-Rodrìguez and G. Pisoni, "Accessible cultural heritage through explainable artificial intelligence," in *Proc. Adjunct Publ. 28th ACM Conf. User Model., Adapt. Pers.*, 2020, pp. 317–324, doi: 10.1145/3386392.3399276.

[17] G. Kostoska, A. P. Vermeeren, J. Kort, and C. Gullstrom, "Video-mediated participation in virtual museum tours for older adults," in *Proc. 10th Int. Conf. Design Emot.*, 2016, pp. 1–10.

[18] K. Shitut, J. Geigel, J. Decker, G. Jacobs, and A. Doherty, "Triggering the past: Cultural heritage interpretation using augmented and virtual reality at a living history museum," in *Proc. GHC*, 2021, pp. 61–70, doi: 10.2312/gch.20211406.

[19] S. Agostinelli, F. Battaglini, T. Catarci, F. Dal Falco, and A. Marrella, "Generating personalized narrative experiences in interactive storytelling through automated planning," in *Proc. 13th Biannu. Conf. Ital. Conf. (SIGCHI CHITALY'19)*, 2019, pp. 23–25.

[20] A. Reshetnikov, "What AI can bring to glam? Experience of 'saint George on a bike' project," in *Proc. Manuscript Papers Prepared Workshop*, 2022, pp. 1–12.

[21] D. Ocón, "Digitalising endangered cultural heritage in southeast asian cities: Preserving or replacing?" *Int. J. Heritage Stud.*, vol. 27, no. 10, pp. 975–990, Feb. 2021, doi: 10.1080/13527258.2021.1883711.

[22] K. Jacquot and R. Saleri, "Gathering, integration, and interpretation of heterogeneous data for the virtual reconstruction of the Notre Dame de Paris roof structure," *J. Cult. Herit.*, vol. 65, pp. 232–240, Jan. 2024.

[23] N. Van Noord, E. Hendriks, and E. Postma, "Toward discovery of the artist's style: Learning to recognize artists by their artworks," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 46–54, Jul. 2015, doi: 10.1109/MSP.2015.2406955.

[24] R. Rizk, D. Rizk, F. Rizk, and A. Kumar, "A hybrid capsule network-based deep learning framework for deciphering ancient scripts with scarce annotations: A case study on phoenician epigraphy," in *Proc. IEEE Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Lansing, MI, USA, 2021, pp. 617–620, doi: 10.1109/MWSCAS47672.2021.9531798.

[25] J. Cao, Z. Zhang, A. Zhao, H. Cui, and Q. Zhang, "Ancient mural restoration based on a modified generative adversarial network," *Herit. Sci.*, vol. 8, no. 1, pp. 1–14, Jan. 2020, doi: 10.1186/s40494-020-0355-x.

[26] R. Pierdicca and M. Paolanti, "Geoai: A review of artificial intelligence approaches for the interpretation of complex geomatics data," *Geosci. Instrum., Methods Data Syst.*, vol. 11, no. 1, pp. 195–218, Jun. 2022, doi: 10.5194/gi-11-195-2022.

[27] Ø. D. Trier, J. Reksten, and K. Løseth, "Automated mapping of cultural heritage in Norway from airborne LiDAR data using faster R-CNN," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 95, Mar. 2021, Art. no. 102241, doi: 10.1016/j.jag.2020.102241.

[28] A. Argyrou and A. Agapiou, "A review of artificial intelligence and remote sensing for archaeological research," *Remote Sens.*, vol. 14, no. 23, p. 6000, Nov. 2022, doi: 10.3390/rs14236000.

[29] L. Espina-Romero and J. Guerrero-Alcedo, "Fields touched by digitalization: Analysis of scientific activity in scopus," *Sustainability*, vol. 14, no. 21, Nov. 2022, Art. no. 14425, doi: 10.3390/su142114425.

[30] X. Wang, "Artificial intelligence in the protection and inheritance of cultural landscape heritage in traditional village," *Sci. Program.*, vol. 2022, pp. 1–11, Jan. 2022, doi: 10.1155/2022/9117981.

[31] M. Mishra, T. Barman, and G. V. Ramana, "Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage," *J. Civil Struct. Health Monit.*, vol. 14, pp. 103–120, Jan. 2024, doi: 10.1007/s13349-022-00643-8.

[32] F. Granata and F. Di Nunno, "Artificial intelligence models for prediction of the tide level in Venice," *Stochastic Environ. Res. Risk Assess.*, vol. 35, no. 12, pp. 2537–2548, Apr.2021, doi: 10.1007/s00477-021-02018-9.

[33] G. Fangi, W. Wahbeh, E. S. Malinverni, F. Di Stefano, and R. Pierdicca, "Archaeological syrian heritage memory safeguard by low cost geomatics techniques," in *Proc. IMEKO Int. Conf. Metrol. Archaeol. Cult. Herit.*, Lecce, Italy, 2017, pp. 23–25.

[34] D. Abate et al., "Significance. Stop illicit heritage trafficking with artificial intelligence," *Int. Archives Photogram. Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 729–736, May 2022, doi: 10.5194/isprs-archives-XLIII-B2-2022-729-2022.

[35] Z. Epstein et al., "Art and the science of generative AI," *Science*, vol. 380, no. 6650, pp. 1110–1111, Jun. 2023, doi: 10.1126/science.adh4451.

[36] "ICOM Code of Ethics for Museums—As approved by the 21st general assembly of ICOM in Seoul, Republic of South Korea, 8 October 2004," *Int. J. Cult. Property*, vol. 13, no. 4, pp. 393–408, 2006, doi: 10.1017/S094073910606022X.

[37] E. Nakonieczna and J. Szczepański, "Authenticity of cultural heritage vis-à-vis heritage reproducibility and intangibility: From conservation philosophy to practice," *Int. J. Cult. Policy*, vol. 30, no. 2, pp. 220–237, 2023.

[38] U. S. Malik, L. N. Tissen, and A. P. Vermeeren, "3D reproductions of cultural heritage artefacts: Evaluation of significance and experience," *Stud. Digit. Herit.*, vol. 5, no. 1, pp. 1–29, Jun. 2021, doi: 10.14434/sdh.v5i1.32323.

[39] Z. Manzuch, "Ethical issues in digitization of cultural heritage," *J. Contemp. Archival Stud.*, vol. 4, no. 2, p. 4, Jan. 2014, doi: 10.2788/8472.

[40] D. Goldblatt, B. L. Brown, and S. Patridge, "The work of art in the age of mechanical reproduction," in *Aesthetics: A Reader in Philosophy of the Arts*, 4th ed., England, U.K.: Routledge, 2017, pp. 66–79.

[41] E. Ch'Ng, Y. Li, S. Cai, and F.-T. Leow, "The effects of VR environments on the acceptance, experience, and expectations of cultural heritage learning," *J. Comput. Cult. Herit.*, vol. 13, no. 1, pp. 1–21, Feb. 2020, doi: 10.1145/3352933.

[42] S. Sontag, "Against interpretation," in *Against Interpretation and Other Essays*. New York, NY, USA: Macmillan, 2001.

[43] A. Gupta and N. Kapoor, "Comprehensiveness of archives: A modern ai-enabled approach to build comprehensive shared cultural heritage," 2020, *arXiv.2008.04541*.

[44] A. Maas, F. Rottensteiner, and C. Heipke, "Classification under label noise based on outdated maps," *Annal. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 4, no. 1, pp. 215–222, May 2017, doi: 10.5194/isprs-annals-IV-1-W1-215-2017.

[45] H. Whaanga et al., "He Matapihi Mā Mua, Mō Muri: The ethics, processes, and procedures associated with the digitization of indigenous knowledge—The Pei Jones collection," *Catalog. Classif. Quart.*, vol. 53, nos. 5-6, pp. 520–547, Jan. 2015, doi: 10.1080/01639374.2015.1009670.

[46] S. Mhlambi and S. Tiribelli, "Decolonizing ai ethics: Relational autonomy as a means to counter ai harms," *Topoi*, vol. 42, pp. 1–14, Jul. 2023, doi: 10.1007/s11245-022-09874-2.

[47] L. Squires (Univ. Washington Inf. School, Seattle, WA, USA). *Beyond Bias: Ethics & AI in Cultural Heritage*, 2021. Accessed: Sep. 16, 2023. [Online]. Available: https://ischool.uw.edu/capstone/projects/2021/beyond-bias-ethics-ai-cultural-heritage

[48] G. Cima, M. Console, M. Lenzerini, and A. Poggi, "A review of data abstraction," *Front. Artif. Intell.*, vol. 6, pp. 1–14, Jun. 2023, doi: 10.3389/frai.2023.1085754.

[49] A. Tzouganatou, "Openness and privacy in born-digital archives: Reflecting the role of AI development," *AI Soc.*, vol. 37, no. 3, pp. 991–999, 2022.

[50] M. Gea, X. Alaman, P. Rodriguez, and V. Martinez, "Towards smart & inclusive society: Building 3D immersive museum by children with cognitive disabilities," in *Proc. EDULEARN16*, 2016, pp. 5260–5268, doi: 10.21125/edulearn.2016.2240.

[51] J. Taylor and L. K. Gibson, "Digitisation, digital interaction and social media: Embedded barriers to democratic heritage," *Int. J. Herit. Stud.*, vol. 23, no. 5, pp. 408–420, 2017.

[52] J. Huggett, "Algorithmic agency and autonomy in archaeological practice," *Open Archaeol.*, vol. 7, no. 1, pp. 417–434, Jun. 2021, doi: 10.1515/opar-2020-0136.

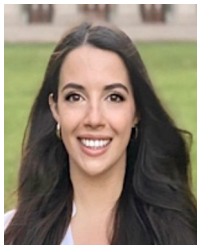
[53] P. Mezei, "From leonardo to the next rembrandt–the need for AI pessimism in the age of algorithms," *UFITA*, vol.84, no. 2, pp. 390–429. Jul. 2020, doi: 10.5771/2568-9185-2020-2-390.

[54] S. Flynn, "Wipo conversation on intellectual property (IP) and artificial intelligence (AI)," Am. Univ. Washington College Law, Washington, DC, USA, Working Paper, 2020.

[55] J. Oppenlaender, "A taxonomy of prompt modifiers for text-to-image generation," *Behav. Inf. Technol.*, 2023, to be published.

[56] C. J. Craig. "The AI-copyright challenge: Tech-neutrality, authorship, and the public interest." SSRN.com. Jan. 2022. [Online]. Available: http://dx.doi.org/10.2139/ssrn.4014811

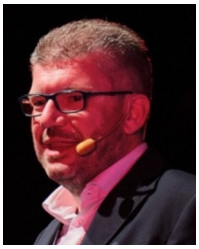
[57] A. Lavazza and M. Farina, "Infosphere, datafication, and decision-making processes in the AI era," *Topoi*, vol. 42, pp. 843–856, Apr. 2023, doi: 10.1007/s11245-023-09919-0.



Simona Tiribelli received the European Ph.D. degree (cum laude) in global studies. justice, rights, politics from the University of Macerata in 2021, with a dissertation in ethics of artificial intelligence.

She is an Assistant Professor of Ethics with the University of Macerata, Italy, where she teaches Ethics of Artificial Intelligence (AI), Communication Ethics, and Global Justice and Technology. She is a Key Person of the Jean Monnet Chair in Ethics for Inclusive Digital Europe, funded by the European Commission. She is also the Director of AI Ethics with the Institute for Technology and Global Health, the PathCheck Foundation's Research and Innovation Center founded at the Massachusetts Institute of Technology, USA. Previously, she held positions as a Visiting Scholar of AI Ethics with the NYU Center for Bioethics, New York University in 2023, and as a FULBRIGHT-Awarded Fellow with the Massachusetts Institute of Technology Media Lab in 2020. She is an expert in the field of applied ethics with a particular focus on digital ethics and ethics of AI. She is also a Co-Founder of GAIA srl (AI Ethics and Governance), innovative start-up and university spin off focused on the ethical-legal assessment of AI solutions and AI Ethics advisor for companies in Europe and U.S. On these topics, she authored two books and a significant number of articles in leading scientific journals (AI and Society, Science and Engineering Ethics, and Ethics and Information Technology), and she delivered more than 60 talks in academic institutions, such as Harvard University, Tufts University, Toronto University, and many more in Europe, Canada, and U.S.

Sofia Pansoni is currently pursuing the Ph.D. degree with the Department of Political and Communication Science and International relationship, University of Macerata. She graduated in Economics and Management for Arts, Culture, Media, and Entertainment with a focus on art markets, heritage, and culture. She has always been interested in multidisciplinary solutions that combine management with technological innovation for the promotion and economic sustainability of arts and culture. During her professional experience, she gave the opportunity to advocate for the digitization and promotion of cultural heritage, collaborating with cultural and research institutions. During these years, she has worked in the field of digital marketing, analyzing data, researching, and applying digital tools to improve the promotion and communication of the creative and non-profit industries.



Emanuele Frontoni is a Full Professor of Computer Science with the University of Macerata and the Co-Director of the VRAI Vision Robotics and Artificial Intelligence Lab. His research interests include computer vision and artificial intelligence with applications in robotics, video analysis, human behavior analysis, extended reality, and digital humanities, with a specific focus on AI for cultural heritage. He has published extensively on topics, such as augmented reality, point cloud semantic segmentation, and the use of AI for the preservation and interpretation of cultural assets. His work bridges advanced AI techniques with practical applications in the cultural sector. He is also involved in several national and international technology transfer projects in the fields of AI, deep learning, data interoperability, cloud-based technologies, and big multimedia data analysis, extended reality, and digital humanities.

Dr. Frontoni is a member of the European Association for Artificial Intelligence, the European AI Alliance, and the International Association for Pattern Recognition.

Benedetta Giovanola holds the Jean Monnet Chair Ethics for inclusive digital Europe co-funded by the European Union, and is a Full Professor of Moral Philosophy with the University of Macerata, Italy, where she has also been the deputy rector and a vice-rector for international relations. She is a regular Visiting Professor with Tufts University and previously held visiting positions with UIBE, Beijing, and KU Leuven, and was a Huygens Fellow with Radboud University Nijmegen. She is the Vice-President of the Italian Association of Moral Philosophy and a Past Member of the Executive Board of the Italian Association of Analytic Philosophy. She is an Expert of Digital Ethics and Ethics of AI, as well as of the integration of EU values in digital technologies. On these topics she is involved in a number or research and technology transfer projects and serves as an ethics advisor for many international projects. She is also the President and a Co-Founder of GAIA srl, an innovative start up and university spin off focused on AI ethics and governance, and providing training and assessment of AI systems ethical impact and legal compliance. She has published seven books and more that 150 papers.