An Ethical Framework for Trustworthy Neural Rendering applied in Cultural Heritage and Creative Industries

L. Stacchio[1], **E. Balloni[2]**, L. Gorgoglione[2], R. Pierdicca[2], A. Mancini[2],
E. Frontoni[1], B. Giovanola[1], S. Tiribelli[1], M. Paolanti[1], P. Zingaretti[2]
[1]University of Macerata, [2]Università Politecnica delle Marche

CVPR SEATTLE, WA JUNE 17-21, 2024

I Introduction

Motivations

- Artificial Intelligence (AI) advancements have impacted various sectors, including **Cultural Heritage (CH)** and **Creative Industries (CI)**
- **Neural Rendering (NR)** techniques, such **as Neural Radiance Fields (NeRFs)** and **3D Gaussian Splatting (3DGS),** improve digitization of 3D objects from 2D images
- **Ethical concerns** remain largely **unexplored**, particularly around the authenticity, intellectual property, and the data-driven decision implications of NR-generated content.

Key Objectives

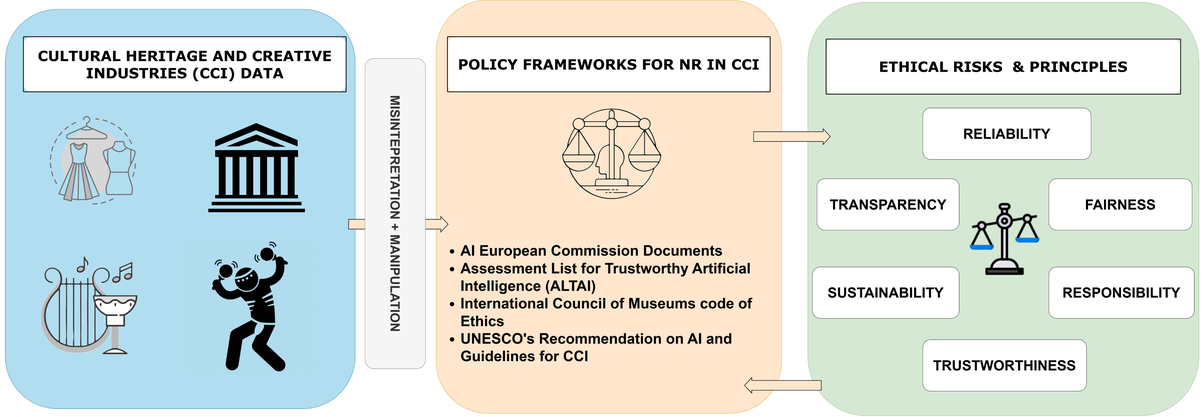
- Identify **ethical pitfalls of NR** paradigms in CCI
- Design an **ethical framework** based on global guidelines
- Provide multidisciplinary **guidelines** for developing NR solutions

Acknowledgements

CTE Square Pesaro
CUP D74J22000930008
FSC MISE 2014-2020

II Methodology

Framework

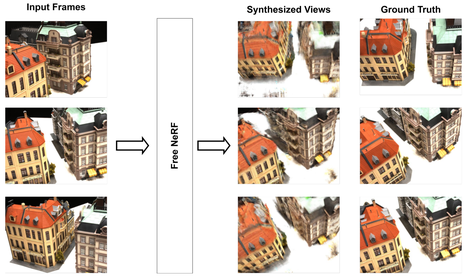


Ethical Principles in NR

- **Transparency and Explainability:** Clear communication of processes, algorithms, and outcomes
- **Reliability:** Accuracy and consistency of NR reconstructions
- **Trustworthiness:** Stability and generalization in various environments
- **Sustainability:** Minimizing environmental impact
- **Fairness:** Ensuring unbiased and fair results
- **Responsibility:** Ethical data ownership and authenticity

III Results

Ethical Principle	Challenges	Technical Risks	Possible Solutions
Transparency and Explainability	Understanding complex AI models and validate data collection process	• Lack of interpretability • Missing description of data collection steps • Lack of controllability for erroneous reconstructions	• Provide clear and detailed **documentation** of data collection processes • Use visual and interpretative methods to make model decisions **understandable** • Maintain open **communication** about processes and outcomes to stakeholders.
Reliability	Ensuring accuracy of reconstructions	• Inconsistent outputs due to few or one-shot • Hard camera estimation due to data scarcity • Novel view synthesis and geometrical outputs with low veridicity • Bias of pre-trained NR methods	• Establish rigorous **testing protocols** to ensure accuracy and consistency • **Collect comprehensive data**, ideally 50-150 images per object • Conduct **bias analysis** on training data and pre-trained models.
Trustworthiness	Demonstrating stability and generalization in different (social) environments	• Lack of visual generalization • Inconsistent Geometrical Representation • Missing social considerations into the system's functionality	• Develop **frameworks** to demonstrate model **reliability** and **stability** • Integrate **social considerations** and **collaborate** with domain experts.
Sustainability	Minimizing environmental impact	• High computational demand • Energy cost to create and maintain a capture setting	• Optimize model architectures and training processes to **reduce energy consumption** • Use **renewable energy** and **energy-efficient hardware** • Implement protocols to minimize the number of cameras and **optimize GPU usage**.
Fairness	Unbiased and fair results	• Biased NR prior knowledge • Artifacts caused by NR paradigms which exploit regularization, synthetic generation or ignore high-frequency details	• Use auxiliary networks to **detect and correct biases** • Ensure **datasets** are diverse in **culture, periods, and regions** • **Enhance** model architectures and optimization strategies
Responsibility	Ethical data ownership an authenticity	• Misuse of generated data • Accountability for unfaithful generation • Intellectual property	• Follow **ethical data ownership** and secure handling protocol • Implement **validation protocols** to ensure **data authenticity** • Adhere to **legal and intellectual property** frameworks and protect data integrity.



Reliability on few-shot models: FreeNeRF network trained on 3 images from the DTU dataset with the same setting provided by the original authors and 3 synthesized novel views compared against their ground truths