



Responsibility and Sustainability In Multimodal Generative AI

Prof. Rita Cucchiara

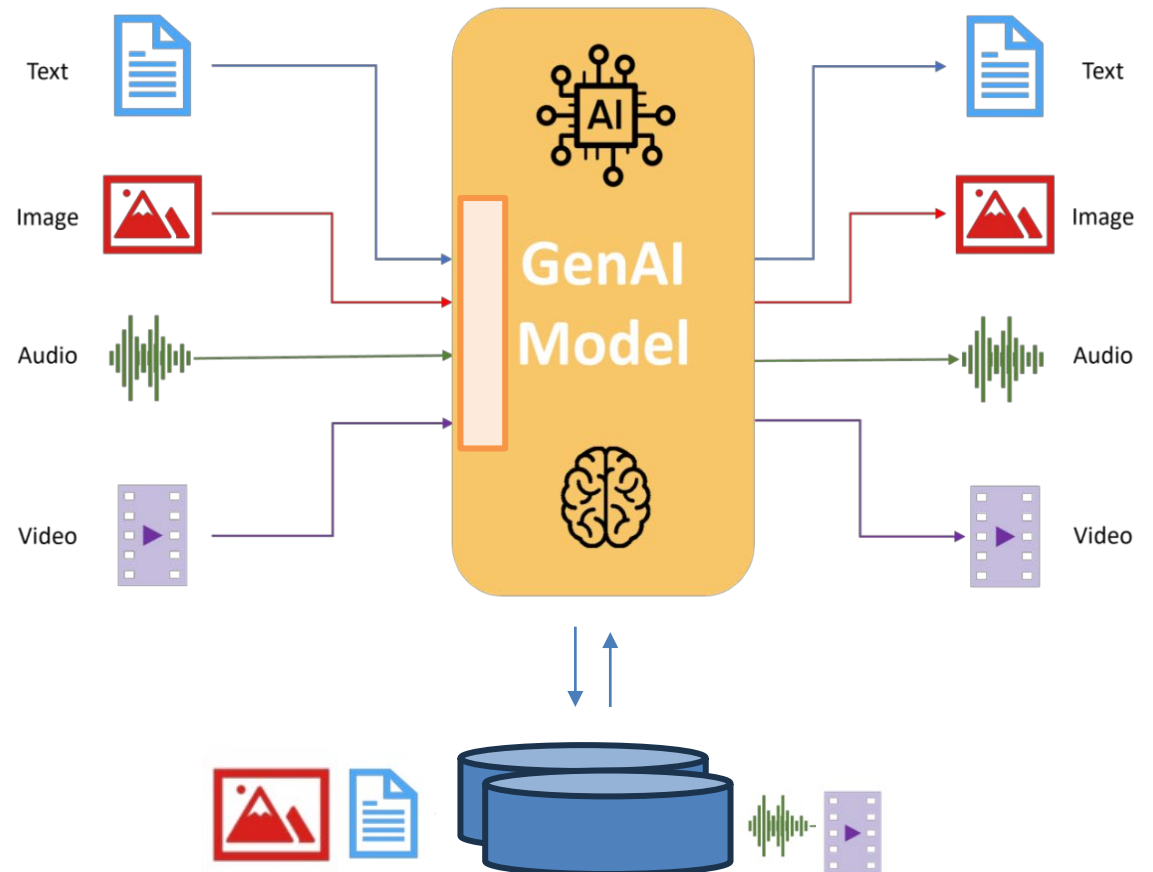
AIRI, Modena ELLIS Unit

Università di Modena e Reggio Emilia, Italia

Multimodal Foundation Models with
Images, Text, Audio, Video...

Worldwide collaborative
Research in Computer Vision, NLP,
Machine Learning

Since 2021,
CLIP (USA), OpenCLIP (EU), EVA-CLIP (CINA)
BLIP (USA), Safe-CLIP (ITA), LION-CLAP (EU)..
a large collaboration with open-source models

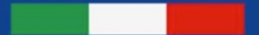


Multimodal Gen AI: some research projects



Piano Nazionale di Ripresa e Resilienza

#NEXTGENERATIONITALIA



ELLIOT (Open Multimedia Foundation Models)

Grecia, Italia (UniTN, UNIMORE, CINECA), Germania, Paesi Bassi, Spagna, Slovenia, Finlandia, Francia,



ELIAS (European Lighthouse of AI for Sustainability)

Italia (UniTN, PoliMI, UNIMORE, FBK, UniMI, IIT, Prometeia), Francia, Paesi Bassi, Germania, Spagna, Slovenia, Romania, ...



ELSA (European Lighthouse on Secure and Safe AI)

Germania, Finlandia, Svezia, Spagna, Italia (UNIMORE, IIT, CINI, Leonardo), Spagna



IT4LIA (Italy for Artificial Intelligence)

CINECA, UNIMORE, UniTO, UniBO, Sapienza, FBK, AI4I, ICSC, ART-ER, ACN, IFAB, SFC/Confindustria, ...



MINERVA

MINERVA (AI and HPC)

Italia (CINECA, UNIMORE, iGenius), Francia, Germania, Spagna



FAIR (Future Artificial Intelligence Research)

CNR, FBK, IIT, INFN, PoliMi, PoliTo, SISSA, Bocconi Sapienza, UniBA, UniNA, Sapienza, UniTN, UniPI, UniCT, UniBO, UniCal, UNIMORE



ITSERR (Italian Strengthening of the ESFRI RI Resilience)

CNR, UNIMORE, UniOR, UniPA, UniTO



Fit4MedRob (Fit for Medical Robotics)

CNR, UniPV, UniNA, Sant'Anna, UniFI, UniGE, UNIMORE, UniPI, UniSI, Campus Bio-Medico, IIT



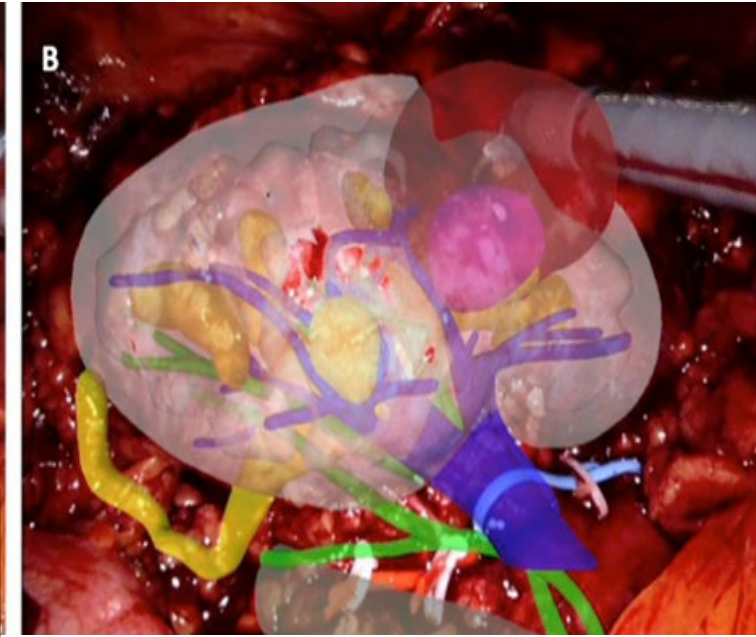
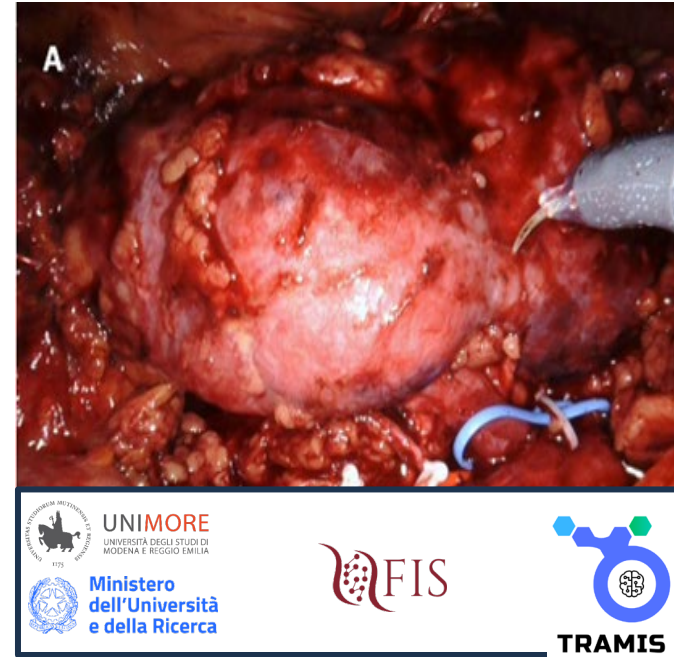
TRAMIS

Trustworthy Robotic Assistant for Improved Minimally Invasive Surgery UNIMORE, FIS 2Italia

Multimodal GenAI: from theory to applications

Multimodal GenAI : a direction for future Research and applications

- Textual Description of Medical Images
- Surgery robot support
- Visual/audio human-robot Interaction
- Sustainable Image editing for Tourism
- Multimodal Document summarizing
- Responsible Image Generation in Tourism
- E-Commerce, Automotive...



Multimodal GenAI: from theory to applications

Multimodal GenAI : a direction for future Research and applications

- Textual Description of medical Images
- Surgery robot support
- Visual/audio human-robot Interaction
- Sustainable Image editing for tourism
- Multimodal Document summarizing
- Responsible Image Generation in Tourism
- E-Commerce, Automotive



Regular traffic, with parked cars





Editing Prompt:

Change the
color of the
knife to blue

Detected Differences

EDIT: pocketknife changed to blue pocketknife



AI Responsibility and Sustainability in Europe



[OECD 19] OECD, Recommendation of the Council on Artificial Intelligence, C/MIN(2019)3/FINAL,



[CE 2019] EU Commission Guideline for ethical and trustworthy AI (2019)



[CE 2024] EU Commission Artificial Intelligence Regulation in Europe AI Act (2024)

For the European Union, Responsible AI aligns with the goal of **Trustworthy AI**.

Three pillars (EC, 2019): **Lawful. Ethical. Robust.**

Seven operational requirements (EC, 2019):

- **Human agency and oversight**
- **Technical robustness and safety**
- **Privacy and data governance**
- **Transparency**
- **Diversity, non-discrimination and fairness**
- **Societal and environmental well-being**
- **Accountability**



Ministero dell'Università e della Ricerca

Work program in « Artificial Intelligence»
(2021)

«... goal is to establish a research and design paradigm that follows **ethics by design, with human oversight at every stage**, and is therefore **trustworthy, both reliable and grounded in trust**, enabling the broad impact of AI technologies and their informed adoption. »



(2024) «...Italy will need to strengthen investment in foundational and applied Artificial Intelligence research, promoting the development of research capabilities and technologies specifically tailored to the context of our national system and **aligned with the principles of trustworthiness and responsibility** (Trustworthy AI) and a **human-centric approach** consistent with European paradigms.... »

New Italian Law in AI (132/2025)

132/2025 (September 2025)

Italian law links “responsible AI” to three practical axes:

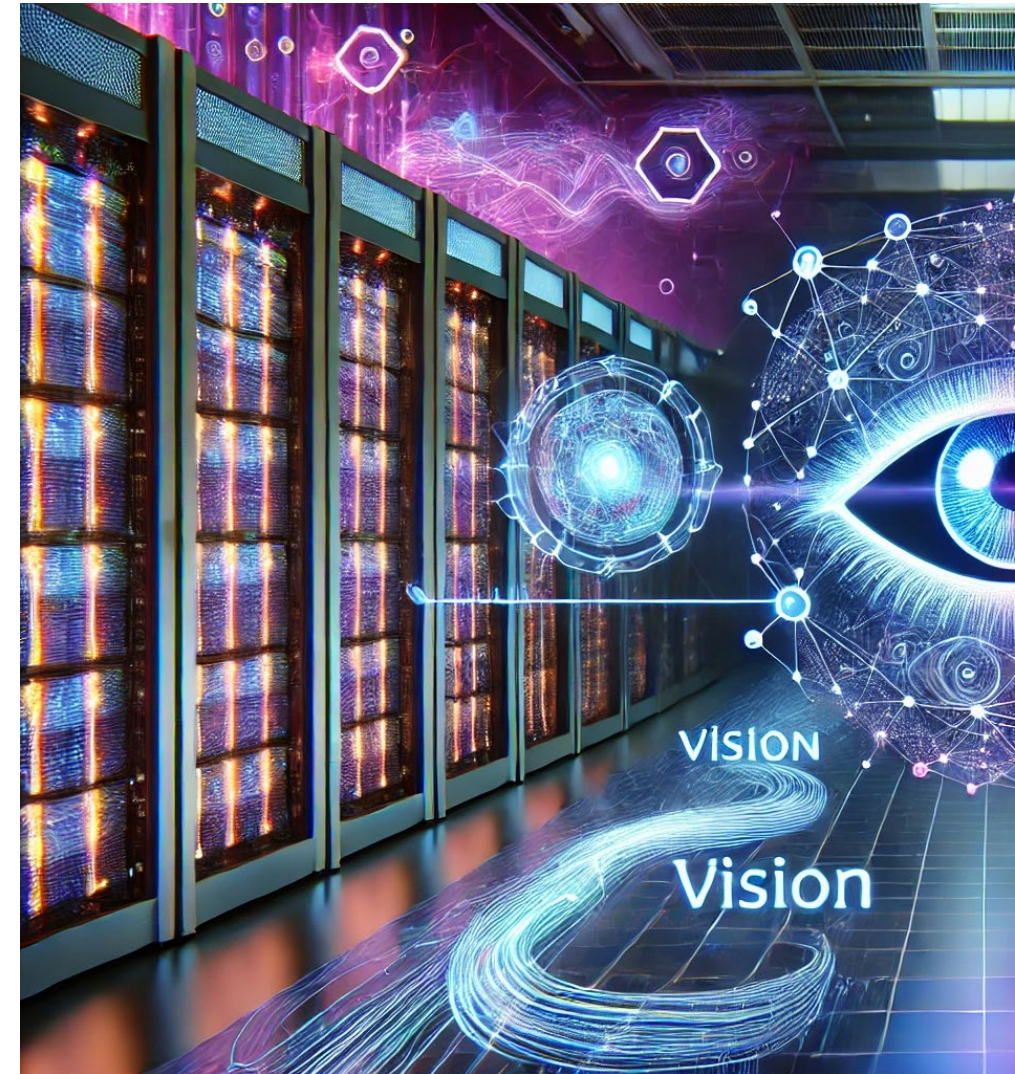
- substantive **principles of trustworthiness** and **sustainability** within a human-centric approach;
- concrete **safeguards** in high-impact contexts such as **employment and healthcare**;
- a framework **for governance, investment, and accountability** aligned with the EU AI Act.



https://www.gazzettaufficiale.it/atto/serie_generale/caricaArticolo?art.codiceRedazionale=25G00143&art.dataPubblicazioneGazzetta=2025-09-25&art.flagTipoArticolo=0&art.idArticolo=1&art.idGruppo=1&art.idSottoArticolo=1&art.idSottoArticolo1=10&art.progressivo=0&art.versione=1

Many research directions, in the world and in Italy too:

- **Ethically responsible** Multimodal GEN-AI Systems
- **Computationally sustainable** Multimodal GEN-AI Systems
- Reliable Multimodal GEN-AI Systems to **measure and reduce hallucinations**



1. Ethically sustainable Multimodal GenAI

Safe-CLIP [UNIMORE, ECCV 2024]



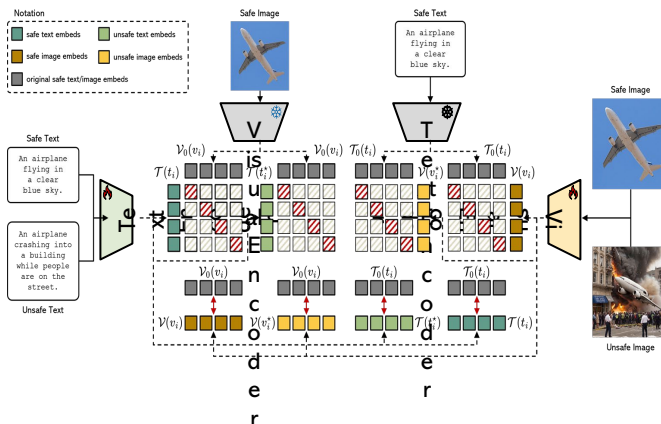
Objective: unlearning violence, toxic, nudity and unsafe concepts (NSFW Not Suitable for Working) in multimodal GenAI, for generating or retrieval only safe images or safe text.

Soon available as free tool on CINECA AI4IA platform (Italy) AI Factory

NSFW Text Query

CLIP Top-1

Safe-CLIP Top-1



Input Prompt

SD

Neg. Prompts

SLD-Strong

SD + Safe-CLIP

The little boy is
sitting on the bed
with a gun in his
hand.



2. «Self-Reflective» RAG for sustainable Multimodal LLMs

ReflectiVA [UNIMORE, CVPR 2025]

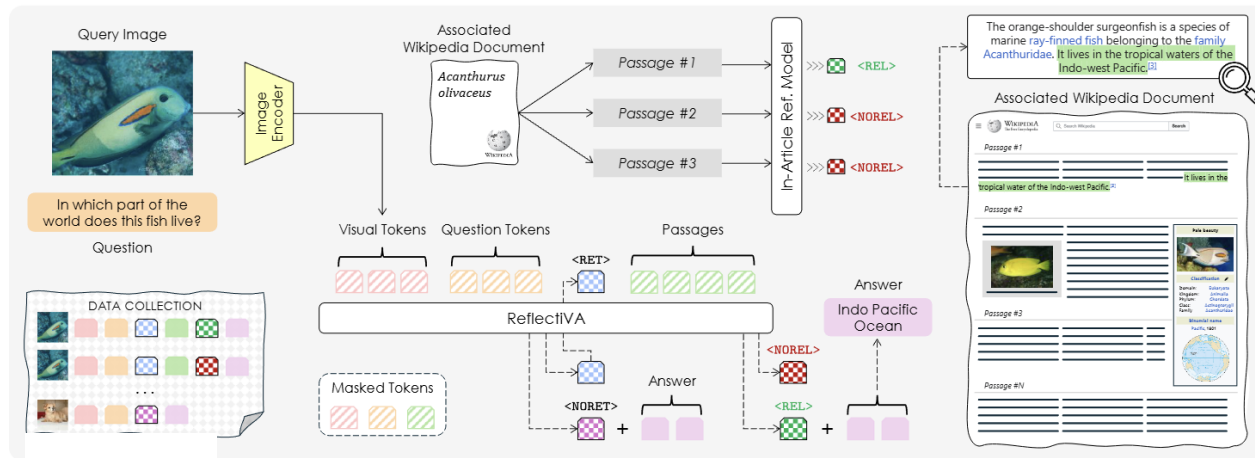


i.e., the architecture that knows not to know...

improving Multimodal LLMs without an expensive training capable to

1 Reflect if an external retrieval is needed

2 Evaluate if the retrieved data is useful



Based on Qwen-VL (China)



ReflectIVA [CVPR2025] improves precision with sustainable energy consumption



What is the location of this garden?

PreFLMR (Cambridge): Washington Park, Denver ❌

ReT-2 (UNIMORE): National Mall ✅



Who designed this palace?

EchoSight (Shanghai Jiao Tong Univ.): A team of architects, including Johan Dientzenhofer ❌

ReflectiVA (UNIMORE): Balthasar Neumann ✅



When did the San Diego savings bank leave this building?

PreFLMR (Cambridge): March 18, 1994 ❌

ReT (UNIMORE): The San Diego savings bank left the building in 1930 ❌

ReT-2 (UNIMORE): 1912 ✅

Q: What was the first subspecies of this bird?



Wiki-LLaVA [8]:
Nominate ❌
EchoSight [59]:
I don't see any information about a bird species ❌
ReflectiVA (Ours):
Aleutian Savannah Sparrow ✅

3. Responsible AI as a Difference Coherence Estimator

DICE A Difference Coherence estimator

[UNITn, UNIMORE ICCV2025]

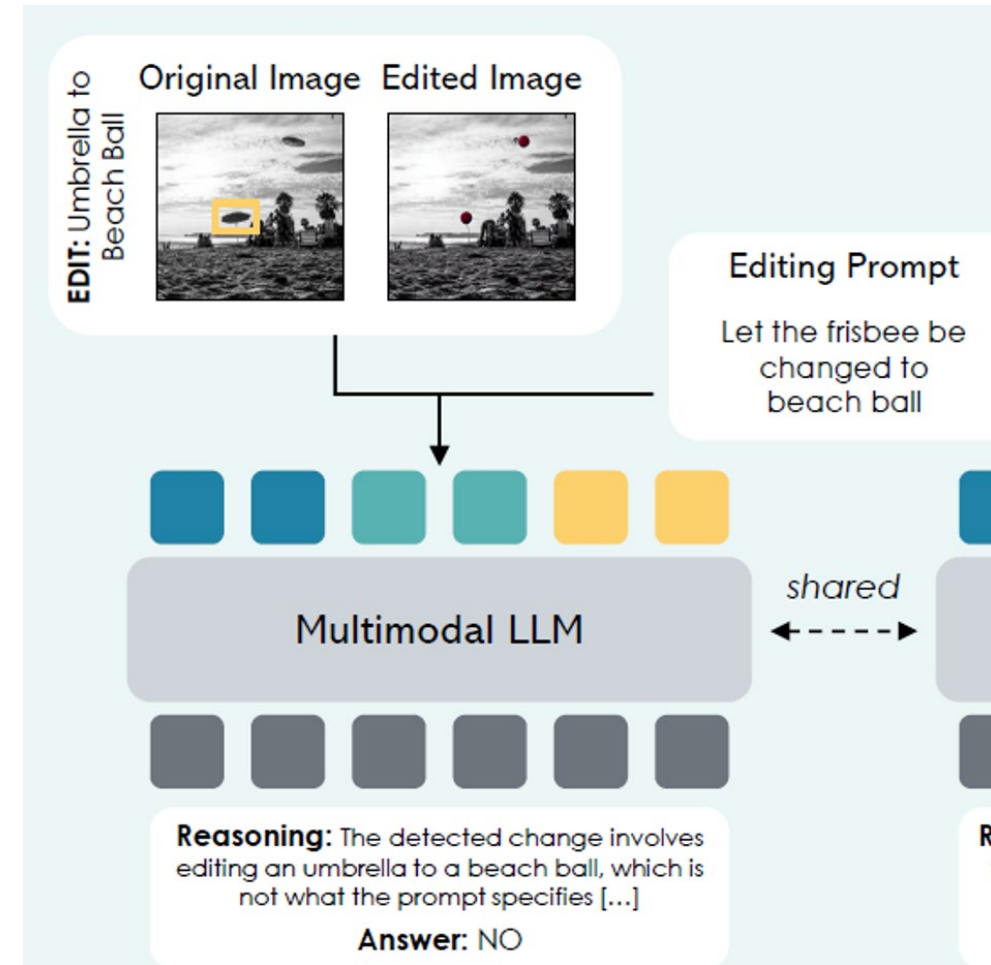
- Hallucination estimator for visual generative AI
- A new model for understanding editing errors as a human estimator
- A reasoning agent for assessing quality in image editing

A basic tool for a safe , ethical and responsible use of AI in image generation

Thanks for a large use of CINECA GPUs



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA





Editing Prompt:

Change the British flag into a Canadian flag

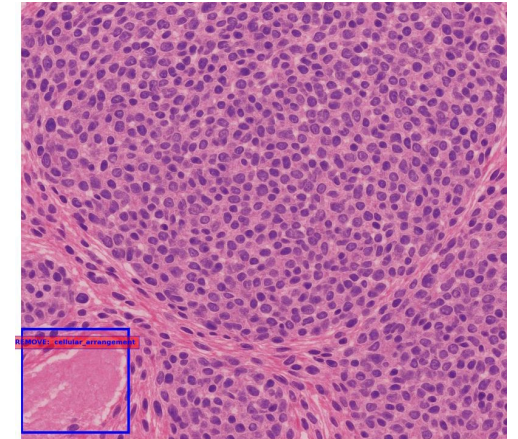
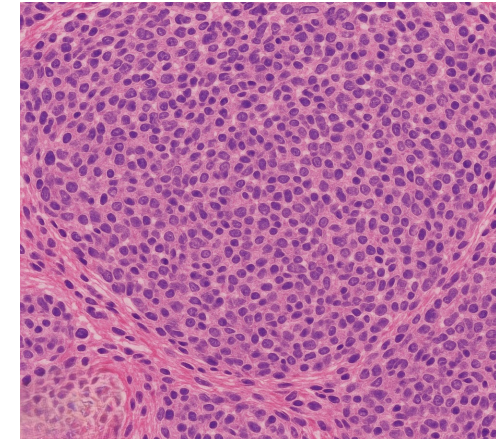
Detected Differences

EDIT: flag changed to maple leaf

Reasoning

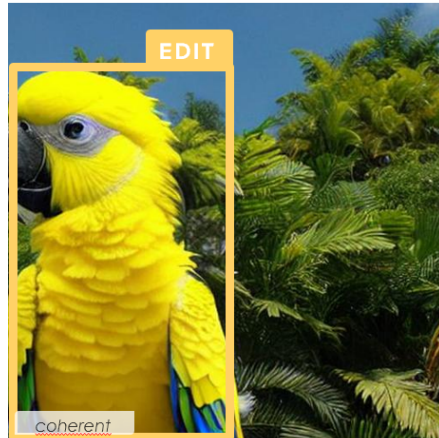
The original prompt requests a change of the British flag to a Canadian flag. The detected change involves editing the flag from a British one to a maple leaf, which is the symbol of the Canadian flag. The change from a British flag to a maple leaf is not a change to a Canadian flag, as the Canadian flag consists of a red maple leaf within a red and white background, not just a single red maple leaf. Therefore, the detected change does not match the original prompt.

Answer: NO



Evaluating trustworthy editing:
Removed cellular arrangements
in histological images

A new tool for audit e accountability



Editing Prompt:

Change the color of the parrot to yellow

Detected Differences

EDIT: parrot changed to yellow parrot

Reasoning

The original prompt requests a color change of the parrot to yellow. The detected change indicates that the bird's color was changed to yellow, which aligns with the prompt's requirement. Therefore, the detected change matches the original prompt.

Answer: YES

Baraldi, L., Bucciarelli, D., Betti, F., Cornia, M., Sebe, N., & Cucchiara, R. (2025). What Changed? Detecting and Evaluating Instruction-Guided Image Edits with Multimodal Large Language Models. In ICCV 2025

Conclusions and Future directions

Current Multimodal Gen AI Agents are **expensive**, **unsustainable** in energy usage, **un-ethical**, **unsafe** or **toxic** in some content and affected by **hallucinations**..

Developing new generations

- Toward **the sustainability of foundation models**: develop concept based, efficient foundation models **with lower computational impact by design**
- **Increasing the number and complexity of modalities**: build multimodal generative models able to understand and generate dynamic, time dependent content (video, time series).
- Integrating safety into generative models: embed mechanisms **for retraining and ethical awareness directly in Multimodal LLMs**.



Looking forward to work together!

Conclusions and Future directions

Thanks

For contact:

Aimagelab- Dipartimento di Ingegneria “Enzo Ferrari”

AIRI (AU Research and Innovation Center)

Modena ELLIS Unit

Università di Modena e Reggio Emilia

(<http://aimagelab.unimore.it>)

Artificial Intelligence
Research and Innovation Center (AIRI)



Thanks CINECA and NVIDIA for GPU Support

Thanks Italian PNRR Projects

Thanks ELLIS and EU projects

