talentgal

Engineering & Technology

Watermarking, Provenance Tracing, and Integrity of Generative AI and Multimedia Content

Description

This research develops comprehensive statistical and explainable AI (XAI) methodologies for multimedia forensics that are inherently robust against adversarial manipulation. Investigation will center on targeted research directions, which may include: 1) developing secure watermarking strategies and countering extraction techniques for generative models and their outputs, 2) advancing forgery detection utilizing both adversarially-trained models and robust hashing, 3) establishing tamper-resistant end-to-end content provenance tracing, and 4) achieving robust source and camera attribution, even when faced with sophisticated counter-forensic measures.

Infrastructures

The project will leverage the research group's on-site high-performance computing (HPC) infrastructure. Our computing resources consist of multi-core, high-memory servers with over 2 TB of aggregate RAM, optimized for intensive computational tasks. Crucially, this infrastructure includes a strong multi-GPU capability, featuring multiple distinct nodes equipped with a total of 32 high-performance NVIDIA GPUs. This specialized hardware is ideally suited for accelerating large-scale model training and inference.

Keywords

Adversarial Robustness | Content Provenance Tracing | Counter-Forensics | Deepfake Detection| Explainable AI (XAI) | Generative AI Security | Multimedia Forensics | Robust Hashing | Robust Watermarking

CIGUS Network
Research Centres of the Galician University System
Co-funded by the European Union
Fondos Europeos
GALICIA CALIDADE
Xacobeo 2027