FAIR AI Attribution (FAIA)

Flagging AI-generated content for people and machines

The Proliferation of AI-Generated Content

The proliferation of AI-generated content on the internet, fueled by near-zero production costs, poses significant challenges. Current estimations indicate that AI accounts for 30-50% of online content, with predictions suggesting a rise to 90% by 2026. This surge, combined with the growing indistinguishability between AI and human-created content and the absence of standardized AI flagging, heightens the risk of misinformation (including AI hallucinations) and intentional disinformation. Without standard mechanisms of AI attribution, misinformation and manipulated content can spread unchecked, eroding public trust in media and published knowledge. In response to the risks posed by AI, the European AI Act mandates the explicit flagging of AI-generated content, highlighting AI transparency as a crucial public policy concern. However, effective implementation requires technical solutions for standardized, performant, and machine-actionable disclosure of AI involvement.

FAIR AI Attribution

The FAIR AI Attribution (FAIA) framework (https://faia.liccium.com/) addresses the increasing need for transparency in AI-generated content creation by providing a lightweight solution that enables researchers, creators, publishers, and content platforms to easily document and disclose AI's role in content creation and distribution. The FAIA framework defines standardised human-AI contribution flags, types of AI contribution and descriptive metadata about the AI system in machine-readable formats. When integrated into digital publishing workflows through tools like Liccium, the framework offers a practical way to certify unambiguous AI involvement at scale, and ensure traceability throughout the content lifecycle. By embedding transparency directly into the publishing process, FAIA fosters public trust, scientific integrity, and informed constructive engagement with AI-generated content.

FAIA
FAIR AI ATTRIBUTION

FAIA Framework

A structured vocabulary for disclosing the role of AI in content creation and workflows

FAIA Flags

Human-Created Content (HCC)

AI-Assisted Content (AAC)

AI-Generated Content (AIG)

AI Contribution

IPTC

- Trained Algorithmic Media
- Pure algorithmic media
- Created using Generative AI
- Composite including generative AI elements
- Edited using Generative AI
- Original digital capture sampled from real life
- Human-edited media
- Etc.

STM

- Language and Grammar Correction
- Text Generation
- Translation
- Data/Table Formatting
- Image or Illustration Generation
- Data Visualization
- Code Formatting or Refactoring
- Citation Assistance
- Synthetic Content Generation

FAIA

- Generation
- Contribution
- Enhancement
- Transformation
- Analysis
- Refinement

Examplary activity types. FAIA supports integration with domain-specific standards and practices

System Attribution

Tool	Model	Version	Provider
"ChatGPT"	"GPT-4o"	"4.0"	"OpenAI"

Universiteit Leiden

GO FAIR foundation

Liccium

https://faia.liccium.com – Contact: info@faia.io

https://faia.liccium.com – Contact: info@faia.io