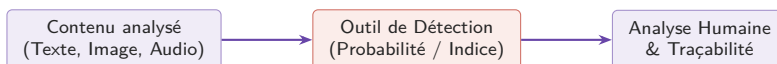


Détection de contenus générés par IA

Classificateurs ■ Multimodalité ■ Provenance et filigranes



1. PRINCIPE ET ATTENTION MÉTHODOLOGIQUE

Un détecteur produit une estimation à partir de caractéristiques d'un contenu, mais ne dispose généralement pas de la preuve de l'historique réel de rédaction ou de création.

- **Variables de performance** : Les résultats varient selon la langue, le modèle génératif, la longueur du texte, les transformations et le corpus de test.
- **Règle pédagogique** : Un score de détection doit être considéré comme un **indice**. Pour une décision importante, il doit être confronté à d'autres éléments (brouillons, historique de versions, métadonnées, entretien).

2. DÉTECTION DE TEXTE

Les outils de cette famille estiment la probabilité qu'un texte soit généré par IA.

- **Outils spécialisés et académiques** : **GPTZero** (analyse globale et par phrase), **Winston AI** (avec OCR et français), **Turnitin** (intégration académique LMS), **Scribbr** et **Compilatio**.
- **Plateformes professionnelles et API** : **Originality.ai**, **Copyleaks**, **Pangram** (distingue génération et assistance) et **Sapling AI**.
- **Outils grand public et rapides** : ZeroGPT, QuillBot, Writer, Quetext, ContentDetector.AI, Isgen, SciSpace, Undetectable AI, AIDetectPlus et Detecting-AI.

3. DÉTECTION MULTIMODALE (IMAGES, VIDÉOS, AUDIO)

- **Images** : Analyse des pixels et des structures visuelles avec **Hive AI**, **AI or Not**, **Illuminarty**, WasItAI, Optic et Fake Image Detector.
- **Vidéos et Deepfakes** : Recherche d'anomalies de manipulation avec **Reality Defender** (multimodal), **Deepware Scanner** et le projet universitaire **DeepFake-o-meter**.
- **Audio et Voix** : Identification de voix de synthèse et de clonage vocal via **ElevenLabs Voice Detector** et **Resemble Detect**.

4. PROVENANCE ET AUTHENTICITÉ (À DISTINGUER DE LA DÉTECTION)

- **Standards ouverts (C2PA / Content Credentials)** : Associent des informations cryptographiquement signées sur l'origine et les modifications. L'absence de métadonnées ne prouve pas une origine artificielle.
- **Filigranage invisible (Google SynthID)** : Marque intégrée directement par certains modèles (images, audio, texte). Résiste à certaines transformations, mais son absence ne garantit pas l'origine humaine.

💡 L'IDÉE ESSENTIELLE

« La détection de contenus IA repose sur des probabilités et des indices techniques, jamais sur une certitude absolue. Face à un enjeu critique, croisez toujours les scores avec une analyse humaine et des preuves de traçabilité. »

Comprendre la différence entre classificateurs probabilistes et protocoles de provenance est indispensable pour évaluer correctement un contenu numérique.