

# Détection de l'usage d'IA générative. Analyse du discours sur l'authenticité

Projet de recherche doctoral – SCAI – Sorbonne Center for Artificial Intelligence – Sciences humaines et sociales & Intelligence artificielle

## Contexte et problématique

La diffusion massive des IA génératives textuelles depuis 2022 a provoqué une crise de l'authenticité dans de multiples sphères sociales : académique (controverses sur la détection d'IA dans l'enseignement supérieur), journalistique (articles générés automatiquement, désinformation, *fact-checking*), et esthétiques (*deepfakes*, art génératif, manipulation de l'image). Le noyau problématique commun est que l'authenticité n'est pas une propriété intrinsèque des textes ou des images, mais un processus social de validation et de contestation, historiquement situé et institutionnellement encadré. Les outils techniques déployés pour la détecter, l'attester ou l'infirmer – que ce soient des détecteurs automatiques à base d'IA ou des dispositifs humains de vérification – ne mesurent pas seulement l'authenticité : ils la produisent en instaurant de nouveaux régimes de preuve. Ce qui est intéressant est que certains effets inégalitaires commencent à être signalés : surreprésentation des locuteurs non natifs dans les faux positifs des détecteurs, vulnérabilité des profils neuroatypiques dont les particularités stylistiques déclenchent les alertes illusoire, ou pénalisation des usages légitimes de reformulation et d'assistance à l'écriture.

Si les travaux en sciences de l'information et de la communication et en STS (Science and Technology Studies) commencent à informer ces controverses secteur par secteur, il n'existe pas d'analyse comparative systématique des régimes de preuve qui émergent selon les contextes institutionnels, ni des mécanismes de circulation des arguments, des cadres interprétatifs et des figures rhétoriques entre sphères. Le champ académique et le champ journalistique partagent pourtant des enjeux structurellement similaires – la question de la paternité, de la vérifiabilité, de la confiance dans la parole – qui méritent d'être mis en regard. Ce projet vise à combler ce manque en proposant une approche communicationnelle comparative et multi-terrains.

## Objectif scientifique et questions de recherche

Le projet entend comprendre comment est construite l'authenticité des contenus textuels et visuels au moment où se déploie massivement l'usage de l'IA générative. Il vise à mettre en évidence les régimes de preuve – techniques, argumentatifs, identitaires – dans trois champs : l'enseignement supérieur, le journalisme, et les contenus visuels et vidéographiques. Il s'agit d'identifier les mécanismes de validation de l'authenticité, d'analyser les inégalités qu'ils engendrent, et d'explorer les éventuelles différences en contexte francophone ou anglophone.

La question principale porte sur les régimes de preuve élaborés pour valider ou contester l'authenticité des contenus textuels et visuels dans un contexte de généralisation de l'IA générative, et comment ces régimes circulent entre terrains institutionnels et espaces géographiques.

QR1 – *Régimes de preuve* : quelles formes de preuve (techniques, argumentatives, identitaires) s'imposent selon les contextes, comment s'articulent-elles, et quelles figures d'acteurs – détecteuristes, contrevenants, faussement accusés – structurent les controverses ?

QR2 – *Inégalités* : dans quelle mesure les dispositifs de détection produisent-ils ou amplifient-ils des inégalités linguistiques (locuteurs non natifs), cognitives (profils neuroatypiques) et de genre (impacts différenciés des *deepfakes*) ? Quelles pratiques de résistance et quelles économies de la preuve d'authenticité émergent en réponse ?

QR3 – *Circulation* : comment les cadres argumentatifs, les normes et les « preuves » circulent-ils entre terrains (académie, médias, image) et entre contextes francophones et anglophones ?

## Approche scientifique et interdisciplinarité

Le projet s'inscrit au croisement des sciences de l'information et de la communication et de l'intelligence artificielle, en mobilisant les outils de l'analyse du discours, de la sociologie des controverses et de l'analyse critique des algorithmes. L'expertise du porteur principal en matière de rumeur, de viralité et de vérité médiatique constitue un ancrage théorique solide pour aborder les nouvelles formes de contestation de l'authenticité générées par l'IA. La co-direction interdisciplinaire associe cette expertise SHS à des compétences en informatique ou en linguistique informatique permettant de compléter l'analyse critique des outils techniques.

Sur le plan méthodologique, le projet nécessite : (1) la constitution et l'analyse de corpus numériques multi-plateformes (forums, réseaux sociaux, presses professionnelles, débats académiques) selon une approche comparative en contextes francophone et anglophone ; (2) la cartographie des positions argumentatives et des acteurs au sein des controverses ; (3) l'analyse critique des outils de détection d'IA et de leurs limites épistémologiques ; (4) des entretiens avec des acteurs (enseignants, étudiants, journalistes, développeurs d'outils). La constitution des corpus sera assurée dans le respect des contraintes éthiques et juridiques liées aux données issues de plateformes numériques.

### **Encadrement et compétences scientifiques**

**Directeur principal** : Pascal Froissart (GRIPIC, CELSA – Sorbonne Université, ED 433 Concepts et langages). Spécialiste de l'analyse des rumeurs, de la viralité et des questions de vérité médiatique. Publications en lien avec le projet :

2024. *L'invention du fact-checking. Enquête sur la "Clinique des rumeurs". Boston, 1942-1943.* Paris : Presses universitaires de France, 344 pages

2025, « Informer, c'est d'abord vérifier ». *La fabrique de l'opinion. Fake news, propagande, complotisme... d'hier à aujourd'hui.* Le monde & La vie, n° hors série, mai, pp. 98–100

2011. « Mesure et démesure de l'emballement médiatique. Réflexions sur l'expertise en milieu journalistique ». *MEI – Médiation et information.* N° 35 (« Scientisme(s) et communication », Aurélie Tavernier, coord.). Paris : L'harmattan, pp. 143-160

**Co-directeur-riche** : [à venir]. Compétences apportées : Analyse informatisée des textes, images et sons générés par des outils d'intelligence artificielle. Publications en lien avec le projet : [à venir].

### **Adéquation aux objectifs de l'Institut SCAI**

Ce projet s'inscrit pleinement dans la mission de SCAI d'articuler intelligence artificielle et sciences humaines et sociales. Il propose une analyse critique et sociale des outils d'IA – détecteurs, systèmes de vérification – en interrogeant leurs effets sur les pratiques sociales, les inégalités et les régimes de vérité. En cela, il contribue à une compréhension interdisciplinaire des impacts sociétaux de l'IA générative, au croisement de l'informatique et des sciences de la communication. Le projet s'ancre dans la société civile via des partenariats envisagés avec des organisations de vérification de l'information (AFP Factuel, Les Décodeurs) et des méthodes participatives dans la collecte des données.

### **Profil recherché**

Master en Sciences de l'information et de la communication, sociologie des médias, humanités numériques. Expérience en analyse de corpus numériques ou en méthodes qualitatives (entretiens, ethnographie numérique). Intérêt marqué pour les enjeux d'IA, d'équité et de vérification de l'information. Anglais courant recommandé.