

**Service des patrimoines
Bureau de la politique documentaire**

Personne chargée de la recherche :
Nathalie NIZAN

Effets de l'IA, de l'IA générative et des deepfakes sur la santé, notamment la santé mentale

Destinataires : Maud GODIGNON, Ingrid BOINET
(DGS/SP4)

Projet : liste de références documentaires

Version : 03/02/2025

Cette recherche documentaire sur les effets de l'IA et des deepfakes sur la santé, dont la santé mentale, est demandée par la DGS (Bureau SP4). Elle porte sur les publications en France et à l'étranger (ouvrages, rapports, recommandations, articles scientifiques) depuis 5 ans¹.

Certains articles indiqués ne sont pas en texte intégral, le Bureau DOC peut le fournir si besoin.

Effets sur la santé mentale

Publications en France

- ❖ Ifop. *Les Français et les jeunes face aux deepfakes*. 2024.

A l'heure où les clichés de Kate Middleton ou d'un « président boxeur » suscitent de nombreuses réactions sur les réseaux sociaux, la question de la véracité des photos des personnalités, trafiquées par deepfake ou photoshoppées, se pose de plus en plus, mettant en lumière la capacité des Français à détecter une image retouchée, ou tout simplement fausse. Plus largement, quelle place occupent aujourd'hui les images et vidéos générées par des Intelligences Artificielles dans l'espace public, et suscitent-elles des appréhensions chez les Français ? L'Ifop, pour Alucare.fr, s'est penché sur le sujet en réalisant une étude sur les deepfakes auprès d'un échantillon de 2191 individus.

[Voir.](#)

- ❖ Institut CSA. *Ados, IA & deepfakes*. 2024.

Quelle perception les 13-17 ans ont-ils de l'intelligence artificielle ? La craignent-ils ou bien au contraire, y voient-ils l'avenir ? L'ont-ils déjà

¹ Références concernant les apports de l'IA en santé et santé mentale en vert

expérimentée ? Ont-ils déjà été exposés à des deepfakes ? Pour le savoir, Milan presse a sondé un échantillon représentatif de plus de 1 000 adolescents. L'objectif de l'étude réalisée avec l'institut CSA est de mieux connaître les lecteurs, notamment ceux de ses magazines Julie et GEO Ado. Les résultats montrent des adolescents partagés entre crainte et fascination.

[Voir.](#)

Publications à l'étranger

Australie

- ❖ eSafetyCommissioner. *Deepfake trends and challenges - position statement*. 2024.

Les deepfakes peuvent causer des dommages considérables. Ils ont été utilisés pour créer de fausses informations, de fausses vidéos pornographiques et des canulars malveillants, ciblant généralement des personnes connues comme des politiciens et des célébrités. Les deepfakes peuvent potentiellement être utilisés comme un outil de vol d'identité, d'extorsion, d'exploitation sexuelle, d'atteinte à la réputation, de ridicule, d'intimidation et de harcèlement.

[Voir.](#)

- ❖ Australian Institute of Health and Safety. *AI in the workplace : balancing productivity and psychosocial risks*.

Selon le RMIT, la mise en œuvre d'outils d'IA ou d'IA générative sur les lieux de travail peut avoir un impact sur la perception et l'existence de risques psychosociaux au travail. Par exemple, l'IA peut automatiser la collecte de données sur les performances des travailleurs en temps réel, réduire potentiellement les biais dans l'évaluation et automatiser différents types de retours d'information. Cependant, les outils d'IA peuvent également supprimer la composante d'interaction sociale, qui est essentielle pour aligner les perceptions humaines entre les travailleurs et les employeurs sur leur productivité et leurs conditions de travail, a déclaré le Dr Emmanuelle Walkowiak, chercheuse principale du vice-chancelier au RMIT à l'École d'économie, de finance et de marketing. « Il est probable que l'intégration de l'IA sur le lieu de travail entraînera également des risques psychosociaux, de santé et de bien-être pour les travailleurs », a déclaré le Dr Walkowiak.

[Voir.](#)

Canada

- ❖ Service canadien du renseignement de sécurité. *Évolution de la désinformation : un avenir « hypertruqué »*. 2023.

Les menaces pour la sécurité et la démocratie qui sont associées à la désinformation sont jugées importantes et d'actualité, voire permanentes. Nés des progrès de l'intelligence artificielle (IA), les hypertrucages sont considérés comme une version moderne de la désinformation qui pose de nouvelles difficultés aux gouvernements, aux personnes et aux sociétés. Protéger l'intégrité de l'écosystème informationnel est primordial pour la démocratie, mais aussi pour l'ensemble de la société.

[Voir.](#)

Pays-Bas

- ❖ Virtual afterlife : the use of deepfake technology in prolonged grief treatment - an examination of bereaved people's and clinicians' opinions (thèse). 2024.

Environ un adulte sur dix en deuil développera des niveaux cliniquement significatifs de symptômes de trouble du deuil prolongé. La technologie Deepfake peut combler le manque actuel de traitement dans les soins du deuil. Cependant, les connaissances sur les opinions des personnes en deuil et des cliniciens concernant son utilisation dans le traitement du deuil font défaut. Cette étude transversale a exploré comment les caractéristiques de base, les caractéristiques liées à la perte et les facteurs d'adoption de la technologie sont liés à l'intention d'utiliser la technologie Deepfake dans le traitement du deuil parmi les personnes en deuil et les cliniciens.

[Voir.](#)

Plusieurs pays

- ❖ UN. *Governing AI for humanity*. 2024.

Ce rapport présente un plan directeur pour faire face aux risques liés à l'IA et partager son potentiel de transformation à l'échelle mondiale, notamment en : exhortant l'ONU à jeter les bases de la première architecture mondiale inclusive et distribuée pour une gouvernance de l'IA fondée sur la coopération internationale ; proposer sept recommandations pour combler les lacunes dans

les dispositifs actuels de gouvernance de l'IA ; appeler tous les gouvernements et toutes les parties prenantes à travailler ensemble pour régir l'IA afin de favoriser le développement et la protection de tous les droits de l'homme.

[Voir.](#)

- ❖ OECD. *Facts not fakes : tackling disinformation, strengthening information integrity.* 2024.

La désinformation croissante a des conséquences profondes dans de nombreux domaines politiques, allant de la santé publique à la sécurité nationale. Elle peut jeter le doute sur les faits, compromettre la mise en œuvre des politiques publiques et saper la confiance des citoyens dans l'intégrité des institutions démocratiques. Ce rapport examine comment répondre à ces défis et renforcer la démocratie.

[Voir.](#)

- ❖ Amnesty International. *Global : TikTok's 'For You' feed risks pushing children and young people towards harmful mental health content.* 2023.

« Ces résultats révèlent les pratiques manipulatrices et addictives de TikTok, qui visent à fidéliser les utilisateurs le plus longtemps possible. Ils montrent également que le système algorithmique de recommandation de contenu de la plateforme, qui a permis l'essor rapide de la plateforme à l'échelle mondiale, expose les enfants et les jeunes adultes souffrant de troubles mentaux préexistants à de graves risques de préjudice », a déclaré Lisa Dittmer, chercheuse à Amnesty International.

[Voir.](#)

- ❖ OMS. *Artificial intelligence in mental health research : new WHO study on applications and challenges.* 2023.

L'utilisation de l'intelligence artificielle (IA) dans les services de santé mentale et la recherche a du potentiel, mais une nouvelle étude révèle des lacunes importantes qui pourraient indiquer une promotion trop accélérée de nouveaux modèles d'IA qui n'ont pas encore été évalués comme viables dans le monde réel.

[Voir.](#)

- ❖ OMS. *Infodemics and misinformation negatively affect people's health behaviours, new WHO review finds.* 2022.

Les interprétations erronées des informations sanitaires, qui se multiplient lors d'épidémies et de catastrophes, ont souvent un impact négatif sur la santé mentale des personnes et accroissent l'hésitation à se faire vacciner, et peuvent retarder la fourniture de soins de santé, montre une nouvelle étude de l'OMS.

[Voir.](#)

- ❖ Think Tank European Parliament. *Tackling deepfakes in European policy.* 2021.

L'émergence d'une nouvelle génération de médias manipulés numériquement – également connus sous le nom de deepfakes – a suscité de vives inquiétudes quant à leur possible utilisation abusive. En réponse à ces inquiétudes, ce rapport évalue les aspects techniques, sociétaux et réglementaires des deepfakes. Le développement et la propagation rapides des deepfakes s'inscrivent dans le contexte plus large d'un système médiatique en pleine mutation. Une évaluation des risques associés aux deepfakes montre qu'ils peuvent être de nature psychologique, financière et sociétale, et que leurs impacts peuvent aller du niveau individuel au niveau sociétal. Le rapport identifie cinq dimensions du cycle de vie des deepfakes que les décideurs politiques pourraient prendre en compte pour prévenir et traiter les effets négatifs des deepfakes.

[Voir.](#)

Articles scientifiques

- ❖ The public mental representations of deepfake technology : an in-depth qualitative exploration through Quora text data analysis. 2024.

L'avènement de la technologie deepfake a soulevé d'importantes inquiétudes quant à son impact sur les processus cognitifs et les croyances des individus, compte tenu des relations omniprésentes entre la technologie et la cognition humaine. Cette étude se penche sur la littérature psychologique entourant les deepfakes, en se concentrant sur la représentation publique de cette technologie émergente et en mettant en évidence les thèmes, opinions et émotions dominants.

[Voir.](#)

- ❖ Psycho-social impact of deepfake content in entertainment media. 2024.

Cette recherche examine l'impact de la technologie deepfake sur les médias de divertissement, en particulier ses effets sur la relation entre les artistes, les fans et la perception de la réalité. Elle vise à explorer les conséquences psychologiques et sociales du contenu deepfake, en se concentrant sur la modification des images de célébrités et d'individus vénérés.

[Voir.](#)

- ❖ Impact of deepfake technology on children : risks and consequences. 2024.

L'un des problèmes les plus urgents d'aujourd'hui est la sécurité de nos enfants dans le cyberspace. Le monde d'Internet recèle un certain nombre de menaces graves qui peuvent nuire à la santé émotionnelle, sociale et mentale

des enfants. Cet article se concentre sur les technologies connues et émergentes de contrefaçon.

[Voir.](#)

- ❖ **Promising for patients or deeply disturbing ? The ethical and legal aspects of deepfake therapy. 2024.**

Les deepfakes sont des vidéos hyperréalistes mais fabriquées à l'aide de l'intelligence artificielle. Dans le contexte de la psychothérapie, les premières études sur l'utilisation de la technologie deepfake émergent, avec des applications potentielles incluant le soutien au deuil et le traitement des traumatismes liés à la violence sexuelle. Cet article explore ces applications du point de vue de l'éthique médicale et du droit de la santé.

[Voir.](#)

- ❖ **After deception : how falling for a deepfake affects the way we see, hear, and experience media. 2024.**

Avec l'émergence de l'intelligence artificielle, les deepfakes ont rendu possible la manipulation de la représentation audiovisuelle de n'importe qui et de n'importe quoi, ajoutant de l'huile sur le feu au débat sur la crédibilité de ce que nous entendons et voyons dans les nouvelles. Cependant, nous ne savons pas encore si les deepfakes peuvent réellement avoir un impact sur la crédibilité attribuée aux médias audiovisuels en général, ainsi que sur l'auto-efficacité perçue pour distinguer les vrais des faux médias. En outre, il n'est pas clair si différents formats de deepfakes peuvent affecter les citoyens à des degrés divers.

[Voir.](#)

- ❖ **Mental health in the virtual world : challenges and opportunities in the metaverse era. 2024.**

L'intelligence artificielle, la réalité augmentée et la réalité virtuelle créeront des environnements en ligne totalement immersifs, expérientiels et interactifs dans le métavers. Les gens utiliseront un avatar unique pour faire tout ce qu'ils font dans leur vie « réelle », y compris rechercher et recevoir des soins de santé mentale. Dans cette revue d'opinion, nous réfléchissons à la manière dont le métavers pourrait remodeler la façon dont nous fournissons des traitements de santé mentale, ses opportunités et ses défis (voir éléments sur deepfakes).

[Voir.](#)

- ❖ **Inability to detect deepfakes : deepfake detection training improves detection accuracy, but increases emotional distress and reduces self-efficacy. 2024.**

Les deepfakes sont des contenus médiatiques générés par une intelligence artificielle dans le but d'être perçus comme réels. Les deepfakes représentent un risque croissant pour la société en raison de leur caractère trompeur répandu et efficace. Bien que plusieurs stratégies aient été développées pour augmenter

les performances de détection des deepfakes par les humains, les recherches utilisant la technologie actuelle sont encore rares. De plus, les effets de l'exposition aux deepfakes sur la détresse personnelle sont sous-étudiés.

[Voir.](#)

- ❖ **Generative AI and deepfakes : a human rights approach to tackling harmful content. 2024.**

La loi sur l'intelligence artificielle (AIA) de l'UE introduit des réglementations nécessaires en matière de deepfake. Cependant, celles-ci pourraient porter atteinte aux droits des fournisseurs et des déployeurs ou utilisateurs d'IA, ce qui pourrait être en conflit avec la vie privée et la liberté d'expression en vertu des articles 8 et 10 de la Convention européenne des droits de l'homme et du Règlement général sur la protection des données (UE) 2016/679 (RGPD). Cet article examine de manière critique comment une AIA non modifiée pourrait permettre la manipulation des électeurs, le chantage et la génération de contenus sexuel abusifs, facilitant la désinformation et pouvant nuire à des millions de personnes, tant sur le plan émotionnel que financier.

[Voir.](#)

- ❖ **AI technology panic-is AI dependence bad for mental health? A cross-lagged panel model and the mediating roles of motivations for AI use among adolescents. 2024.**

En se basant sur le cadre de la dépendance technologique, cette étude visait à déterminer la prévalence de la dépendance à l'IA, à examiner les effets croisés entre les problèmes de santé mentale (anxiété/dépression) et la dépendance à l'IA et à explorer le rôle médiateur des motivations d'utilisation de l'IA.

[Voir.](#)

- ❖ **Enhancing mental health with Artificial Intelligence : current trends and future prospects. 2024.**

L'intelligence artificielle (IA) est devenue une force transformatrice dans divers domaines, et son application dans le domaine de la santé mentale ne fait pas exception. Par conséquent, cette revue explore l'intégration de l'IA dans les soins de santé mentale, en élucidant les tendances actuelles, les considérations éthiques et les orientations futures dans ce domaine dynamique.

[Voir.](#)

- ❖ **AI chatbots for mental health : a scoping review of effectiveness, feasibility, and applications. 2024.**

Les troubles de santé mentale sont l'une des principales causes d'invalidité dans le monde et il existe une pénurie mondiale de professionnels de la santé mentale. Les chatbots IA sont apparus comme une solution potentielle, offrant des interventions de santé mentale accessibles et évolutives. Cette étude visait

à mener une revue de la portée pour évaluer l'efficacité et la faisabilité des chatbots IA dans le traitement des problèmes de santé mentale.

[Voir.](#)

- ❖ **Understanding the role of fear of missing out and deficient self-regulation in sharing of deepfakes on social media : evidence from eight countries. 2023.**

Nous avons mené une étude d'enquête transnationale dans huit pays pour examiner le rôle de la peur de rater quelque chose, de l'autorégulation déficiente et des capacités cognitives dans le comportement de partage de deepfakes.

[Voir.](#)

- ❖ **The potential influence of AI on population mental health. 2023.**

Nous proposons trois considérations pour définir comment l'IA peut influencer la santé mentale de la population : par l'amélioration des soins de santé mentale ; par la modification des contextes sociaux et économiques ; et par les politiques qui façonnent l'adoption, l'utilisation et l'abus potentiel des outils améliorés par l'IA.

[Voir.](#)

- ❖ **Fake news et deepfakes : une approche cyberpsychologique. 2022.**

Le développement massif des réseaux sociaux couplé à un certain nombre d'événements démocratiques de grande ampleur ces dernières années ont sans doute contribué à une plus grande sensibilisation du grand public aux dangers posés par les fake news. Cet article propose un bref état de l'art des recherches en cyberpsychologie – notamment des travaux relevant d'une approche expérimentale – sur les facteurs influant sur la crédibilité des fake news.

[Voir.](#)

- ❖ **Deepfakes, deep Harms. 2022.**

Les deepfakes sont des enregistrements vidéo et audio modifiés par des algorithmes qui projettent l'apparence d'une personne sur celle d'une autre, créant ainsi un enregistrement apparent d'un événement qui n'a jamais eu lieu. De nombreux universitaires et journalistes ont commencé à s'intéresser aux risques politiques de la tromperie des deepfakes. Nous étudions ici d'autres façons dont les deepfakes peuvent causer des dommages plus graves que ceux que l'on pensait.

[Voir.](#)