UN PROJET D'ÉDUCATION AUX MÉDIAS QUI QUESTIONNE L'ÉTHIQUE ET L'INTELLIGENCE ARTIFICIELLE

OK MILA

N°02 · IA ET DIVERSITÉ

Editrice, rédactrice et co-fondatrice du projet

Cassi Ninja

Co-fondatrice du projet

Alyssia Ricci

Mise en page du mag papier

Gaëlle Defeyt (Gilda Fêlée)

Mise en page mag online

Cassi Ninja

Charte graphique

François d'Alcamo

Communication et réseaux sociaux

Anne-Sophie Skit

Journaliste podcast et voix off

Julie Mouvet (Minuit Studio)

Réalisateur et monteur vidéo

Francisco Luzemo (Hiola Films)

Réalisateur et monteur podcast

Marius Adam (MadSound)

Relectrice et responsable édition

Elisabeth Bois d'Enghien

WWW.OK-MILA-EAM.BE

OK MILA

CE PROJET EST RENDU POSSIBLE GRÂCE AU CSEM, LE CONSEIL SUPÉRIEUR DE L'EDUCATION AUX MÉDIAS

OK MILA A ÉTÉ CRÉÉ PAR TAPAGE STUDIO-
HELLO@TAPAGE.STUDIO - WWW.TAPAGE.STUDIO

EDITO

Chloé Tran Phu - Media Animation et Witch Gamez

Détourner l'IA : résister à l'effacement des diversités

L'intelligence artificielle est une immense machine à raconter des histoires. Elle génère des images, trie des CV, reconnaît des visages, filtre nos contenus, répond à nos questions avec sympathie et un aplomb savant. Toutefois, dans ces fils, ces images et ces histoires qu'elle compose à partir de milliards de données, une vérité affligeante se répète : l'IA n'est pas neutre. Elle a été entraînée dans un monde traversé par des inégalités, et elle peut les amplifier.

Laissez une IA analyser des CV pour recruter des cadres d'entreprise et vous constaterez qui elle met de côté en priorité : les femmes, les personnes racisées et celles en situation de handicap. Demandez à une IA de générer l'image d'un « PDG » ou d'un « juge » et observez combien de femmes apparaissent. Faites l'inverse avec « infirmièr·e » et comptez les hommes. Ces requêtes, en anglais – une langue où ces termes sont neutres –, révèlent des biais sexistes flagrants. Demandez un « bandit » et voyez à quelle couleur de peau il est associé. Cherchez des couples amoureux et comptez combien ne sont pas hétéros. Ces outils, censés fournir des réponses pertinentes, nous servent en réalité une soupe tiède de stéréotypes mille fois répétés.

On pourrait croire que nous perdons le contrôle de la machine : ce serait oublier que l'IA n'est pas une entité magique ; elle est façonnée par des choix humains. Qui décide des bases de données d'apprentissage ? Qui corrige les biais ? Qui dit ce qui est « normal » ou « acceptable » ? Tant qu'on laisse quelques grandes entreprises de la tech concevoir ces outils selon leur propre vision du monde – capitaliste et néolibérale –, elles continueront à nous vendre un univers uniformisé, calibré pour être rentable avant tout, et non pas pour être juste.

L'IA ne crée pas : elle recycle les représentations dominantes d'une base de données très située culturellement (occidentale et étasunienne). Ces images standardisées sont donc souvent portées par une même esthétique occidentale, blanche, lisse et calibrée. Des visages parfaitement symétriques, une absence d'imperfections, un effacement des traits trop marqués, une homogénéisation des corps et des regards sur le monde. L'IA ne représente pas le monde tel qu'il est, elle représente un idéal forgé par des millénaires de culture visuelle hégémonique. Elle impose un canon unique, où les personnes reléguées aux marges de la société sont effacées. L'IA n'est pas raciste, sexiste, (néo)coloniale ou homophobe en soi : elle est programmée pour s'adapter à une société qui l'est.

Ce n'est pas un hasard si l'extrême droite s'empare massivement de ces outils. Le rêve fasciste d'une esthétique pure et uniforme trouve en effet un terrain fertile dans l'utilisation des IA génératives, la propreté clinique des images d'IA, leur perfection froide et leur absence de rugosité. Faciles d'accès, ces outils permettent d'inonder l'espace médiatique avec des visuels simplistes et idéologiquement chargés, comme la vidéo surréaliste de la « Riviera de Gaza » de Donald Trump. Pour Gareth Watkins, cette nouvelle esthétique fasciste ne cherche pas à créer de l'art, mais à imposer une vision du monde autoritaire, où l'image sert à dominer et manipuler. L'IA devient ainsi une arme politique, diffusant une réalité déformée et creuse, où seule compte l'efficacité du message plutôt que sa vérité ou sa profondeur.

Et nous dans tout ça ? Devons-nous simplement constater les dégâts et regarder ces technologies être monopolisées par celles et ceux qui les utilisent pour nous effacer ? Certainement pas. L'IA est un outil, et comme tout outil, elle peut être détournée, hackée, réappropriée. Il est urgent que les personnes minorisées – femmes, personnes racisées, non-binaires, LGBTQIA+, en situation de handicap – s'emparent de ces technologies, les infiltrent, les tordent, et les retournent contre elles-mêmes afin d'expérimenter, créer, et imposer des narrations alternatives. Il y a lieu de fabriquer des IA qui comprennent la richesse des corps, des identités, des expériences. Il faut refuser de laisser la technologie être une nouvelle frontière de l'exclusion.

Alors, la prochaine fois que vous utiliserez une IA, demandez-vous : "Qui est invisible ici ?" Qui est effacé·e par ces algorithmes qui prétendent représenter la réalité ? Ce n'est pas à nous de nous adapter à l'IA, c'est à elle de s'adapter à nous. Que voulons-nous que l'IA soit pour nous ? Quels récits voulons-nous qu'elle raconte ? Quels visages voulons-nous qu'elle génère ? Nous avons le droit d'exiger mieux. Si l'IA rêve d'un monde standardisé, nous, nous pouvons rêver d'un monde où elle nous appartient réellement. Un monde où la diversité n'est pas une variable à ajuster, mais bien une richesse.

Dans ce deuxième numéro, on parle diversité, justement. Ou plutôt manque de diversité. Autant vous dire qu'on attendait avec impatience d'écrire sur ce (vaste) sujet! Comme Chloé le dit déjà très bien dans l'edito, on prête souvent à l'IA des super-pouvoirs tels qu'efficacité, objectivité, ou encore impartialité. Après tout, une machine ne juge pas, elle calcule. Du moins, c'est ce qu'on aime croire. Toutefois, si on gratte un peu sous son vernis techno-flashy, on se rend compte que cette belle promesse d'objectivité est une illusion. Spoiler alert : l'IA n'est pas neutre. Elle est le reflet amplifié de la société qui l'a créée. Elle n'invente rien, elle apprend... de nous. Et... petit détail (mais non des moindres) : nous, humains, traînons un sacré bagage de préjugés, d'inégalités et de stéréotypes. Et quand on alimente une machine avec des données biaisées... eh bien, surprise : elle recrache ces biais. Pire encore, elle les renforce et les automatise à grande échelle.

C'est là que ça pique : ce ne sont pas juste des erreurs innocentes d'un robot mal entraîné. Ce sont des inégalités systémiques qui s'infiltrent dans des décisions qui impactent réellement nos vies : qui décroche un emploi, qui a accès à un prêt bancaire, qui est surveillé dans l'espace public... Et on ne parle même pas encore des domaines où ces biais deviennent une question de vie ou de mort, comme dans le domaine de la santé.

C'EST QUOI UN BIAIS ALGORITHMIQUE ?

Un biais algorithmique se produit lorsque les décisions prises par une intelligence artificielle favorisent certains groupes de personnes au détriment d'autres, souvent de manière involontaire. Ces biais ne naissent pas d'une mauvaise intention (quoique, et on l'évoquera plus tard), mais de données biaisées sur lesquelles l'algorithme est entraîné par des choix de conception effectués par des humains. Les intelligences artificielles fonctionnent en apprenant à partir de données d'entraînement. Ces données, qu'elles proviennent de textes, d'images ou de comportements humains, sont censées refléter le monde réel. Mais ce monde réel est loin d'être égalitaire. Lorsqu'on dit qu'on vit dans une société patriarcale, raciste, validiste, âgiste, cela signifie que les structures de pouvoir et les normes sociales favorisent certains groupes au détriment d'autres. Ces systèmes de domination sont profondément ancrés dans nos institutions, nos lois, nos pratiques culturelles et même dans les technologies que nous créons, comme... l'intelligence artificielle.

Les biais ne viennent donc pas de la technologie elle-même, mais de ce que nous, les humains, avons intégré dans ces données. Quand une IA apprend à partir d'un passé marqué par des inégalités, elle risque de reproduire, voire amplifier, ces injustices.

- En 2018, Amazon a mis en place un système automatisé de recrutement basé sur l'intelligence artificielle pour trier les candidatures et recommander les meilleurs profils. Ce système devait permettre d'automatiser la sélection des candidat·es, notamment pour les postes techniques très prisés dans le secteur de la tech. Hélas, comme on peut l'imaginer, des biais sont apparus: l'algorithme avait été formé sur dix ans de données de recrutement, reflétant des pratiques passées où les hommes, majoritairement blancs, dominaient les postes techniques. Le système avait donc "appris" que les profils masculins correspondaient mieux aux critères de recrutement d'Amazon. Résultat : le système a commencé à désavantager systématiquement les candidatures féminines, en particulier si les CV contenaient des termes liés à des expériences féminines, comme « présidente du club féminin » ou des écoles réservées aux femmes (le premier qui me dit qu'en 2025, ça va mieux, il reçoit un verre de Spa en pleine figure #pubannées90).

- Le logiciel COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) est un algorithme utilisé dans le système judiciaire américain pour prédire le risque de récidive des personnes condamnées. Il aide (influence?) les juges à prendre des décisions en matière de liberté conditionnelle ou de libération anticipée. Cependant, une enquête menée en 2016 par ProPublica a mis en lumière des biais raciaux profondément enracinés dans cet outil. D'abord, les personnes afro-américaines étaient deux fois plus susceptibles d'être faussement identifiées comme à « haut risque » de récidive par COMPAS. Ensuite, à l'inverse, les personnes blanches avaient souvent des scores de risque plus bas, même en cas d'antécédents judiciaires similaires. Avant de recevoir des mails de plainte, le biais ne venait pas directement de la "couleur de peau" (puisque celle-ci n'était pas explicitement un critère d'évaluation) MAIS des données socio-économiques utilisées par l'algorithme, souvent liées à des inégalités raciales structurelles : niveau de revenu, quartier de résidence, antécédents judiciaires dans l'entourage, etc.

"Ouais mais Cassi, c'est pas chez nous ça". "OKAY" (mon mot PREF 4EVER).

La récente décision du tribunal administratif de Grenoble, en date du 30 janvier 2025, a déclaré illégale l'utilisation du logiciel de vidéosurveillance algorithmique BriefCam par la ville de Moirans, en Isère. Cette affaire, portée par l'association La Quadrature du Net, soulève plusieurs préoccupations majeures concernant l'emploi de telles technologies dans l'espace public. Pour celles et ceux qui ne connaissent pas, "BriefCam" est un logiciel d'analyse vidéo qui applique des filtres algorithmiques aux images de vidéosurveillance. Il permet de suivre ou de retrouver des individus en fonction de divers critères tels que l'apparence, les vêtements, le genre, et inclut même une option de reconnaissance faciale. Cette technologie vise à optimiser le traitement des flux vidéo pour des applications de sécurité et de surveillance (hahahahaha). En plus d'un traitement illégal des données (absence de base légale et non conformité au RGPD), on est quand même sur une grosse atteinte à nos libertés individuelles : déjà, une surveillance de masse (l'utilisation de BriefCam dans les espaces publics conduit à une surveillance généralisée, où chaque individu peut être suivi et analysé sans son consentement) et aussi et surtout, un profilage discriminatoire : la capacité à filtrer les personnes selon des caractéristiques spécifiques entraîne d'office des discriminations basées sur l'apparence, le genre ou l'origine. Et si on couple ça à l'utilisation de ces données par la police et la justice... là, on atteint un autre niveau de problème. On ne parle plus seulement de surveillance passive, mais d'un outil de profilage actif mis entre les mains d'un système répressif, discriminant et raciste, par définition. Imaginez un instant : si BriefCam permet de filtrer des individus selon leur apparence, leur genre ou même des détails vestimentaires, on donne à la police le pouvoir d'intensifier des contrôles ciblés, souvent fondés sur des préjugés existants. Là, on retombe pile dans le cercle vicieux des biais systémiques

- L'algorithme pointe certains quartiers ou profils comme "à risque" (souvent les mêmes, hein, pas de surprise).
- La police se concentre sur ces zones, multiplie les contrôles et les arrestations.
- Ces nouvelles données confirment les prédictions initiales du système, qui désigne à nouveau les mêmes zones et populations comme problématiques.

Résultat ? On renforce des discriminations déjà présentes dans le système. Pire encore, sous couvert d'une technologie "neutre", on assigne une couche d'objectivité à des pratiques discriminatoires bien humaines.

Et si on vous disait que ces biais, déjà bien ancrés dans le recrutement, la surveillance et la justice, s'infiltrent aussi dans des domaines encore plus inattendus ? Oui, parce que ce n'est pas fini (et non #sorrynotsorry).

LES BIAIS DANS LA SANTÉ : DES VIES MISES EN DANGER

Cette fois, on passe à un niveau où les conséquences ne sont plus seulement sociales ou économiques : elles deviennent littéralement vitales. Un algorithme utilisé aux États-Unis pour anticiper les besoins médicaux des patient·es a montré un biais flagrant. Comment ? Il évaluait le besoin en soins sur… les dépenses médicales passées. Sauf que, surprise (ou pas), les patient·es afro-américain·es dépensent généralement moins d'argent en soins, non pas parce qu'iels sont en meilleure santé, mais parce qu'iels ont historiquement moins accès aux soins de qualité. Résultat : des patient·es noir·es ont vu leurs besoins sous-estimés par l'IA, mettant directement leur santé en danger.

Et ce n'est pas tout : les outils de diagnostic, notamment ceux qui détectent des cancers de la peau grâce à des images, sont souvent moins efficaces sur les peaux foncées. Pourquoi ? Parce que la majorité des bases de données d'images médicales utilisées pour entraîner ces systèmes contient principalement des photos de peaux claires. Comme si les autres patient·es n'existaient pas…

LES BIAIS DANS LES MÉDIAS NUMÉRIQUES : STÉRÉOTYPES, RACISME ET SEXISME 2.0

Puisqu'on est dans les sujets sensibles, que penser de la représentation des genres, des minorités dans les médias ?. Vous vous dites peut-être : "Bon, l'IA, c'est juste de la technique, elle ne va pas décider de qui passe à la télé ou non !". Faux.

Les IA de modération, les chatbots et même les algorithmes publicitaires façonnent activement les images, les discours et les standards de beauté qui nous entourent. Souvent, ils ne font que recycler les stéréotypes les plus nauséabonds.

On en a pour vous, bien entendu, quelques exemples croustillants :

- Vous l'avez peut-être vu passer mais Tay, le chatbot lancé par Microsoft est devenu, en moins de 24h, raciste, sexiste et conspirationniste à cause des interactions toxiques avec les utilisateur·rice·s. Résultat : il a été désactivé en urgence. Preuve que si on nourrit une IA avec de la (emoji poop), ça finit par ressortir.
- Une étude a révélé que les algorithmes publicitaires proposaient davantage de prêts immobiliers aux utilisateur·rice·s blanc·hes qu'aux utilisateur·rice·s noir·es. Une discrimination sournoise, basée sur des prédictions de "rentabilité", mais qui renforce directement les inégalités économiques.
- Les filtres de beauté Insta et Snap, on en parle? Ces filtres qu'on adore utiliser valorisent souvent des traits eurocentrés : peau plus claire, nez affiné, yeux élargis. Un standard de beauté lissé et excluant, qui invisibilise les diversités corporelles et faciales. Résultat : une pression esthétique constante, alimentée par des normes occidentales archaïques.

HOW NORMAL AM I ?

En parlant de diktats de beauté, peut-être que vous connaissez ce test? Si pas, on vous l'explique : "How Normal Am I ?" est un projet imaginé par Tijmen Schep, un artiste néerlandais. Dans le cadre d'un projet européen (Sherpa), l'artiste propose aux utilisateur·rice·s de l'IA de tester leur score de "normalité". Vous vous doutez bien que Cassi s'est prêtée au jeu (et à coeur joie évidemment!). En accédant à votre webcam, l'outil analyse votre visage et fournit des estimations basées sur des modèles d'apprentissage automatique. Par exemple, il peut prédire votre âge, évaluer votre "score de beauté" ou estimer votre IMC, simplement à partir de votre image.

Donc, d'après l'IA, Cassi (qui rédige les mags d'OK Mila) est une femme triste de 42 ans, avec un BMI de 18,6 (enfin, 31,9 si elle gonfle ses joues), un score de beauté à 6,1/10, docile et normale à moitié. Voilàààà !

(Et btw, on ne vous parlera pas du BMI de Cassi - là n'est pas le sujet! - mais l'on peut vous dire qu'elle a presque 34 ans et qu'elle est très heureuse !)

OK MILA - MAG 02

L'objectif de ce projet est de sensibiliser le public à l'utilisation croissante de ces systèmes d'évaluation automatisés dans notre société et de questionner leur fiabilité.

Si vous ne l'avez pas lu, le bouquin de Judith Duportail "L'amour sous algorithme" analyse notre score de désirabilité sur Tinder. L'algorithme de cette app de dating attribue des scores de beauté à leurs utilisateur·rices afin de les faire matcher avec des personnes jugées d'un niveau d'attractivité similaire (ouais...).

DÉTOURNER L'IA : VERS UNE TECHNOLOGIE PLUS INCLUSIVE ?

Maintenant qu'on a posé le cadre des gros problèmes (et c'est peu de l'dire), on en revient à la question de Chloé dans l'édito : "Comment détourner l'IA pour qu'elle devienne une technologie plus inclusive? ". On ne va pas se mentir, les décisions politiques ont un impact fondamental et sont loin d'aller vers un mieux.

Petit saut à Paris, février 2025. Un sommet international sur l'intelligence artificielle réunit des dizaines de pays autour d'une même idée : signer une déclaration pour une IA "inclusive, éthique et durable". Une sorte de pacte de bonne conduite : pas de surveillance de masse, pas de discriminations automatiques, et surtout, une IA qui prend en compte *tout le monde*, et pas seulement les visages blancs sur fond LinkedIn. La majorité des pays présents signent directement : la France, l'Inde et même la Chine. Toutefois, deux gros absents font tache : les États-Unis et le Royaume-Uni. Les raisons officielles ? "Trop flou", "pas dans l'intérêt stratégique national", "on préfère nos propres règles".

La vérité ? L'obsession de la compétitivité technologique à tout prix. Et zéro envie de se laisser freiner par des considérations "idéologiques" comme la justice sociale ou la diversité. Ça sent l'alignement parfait avec la doctrine trumpiste du moment.
Aux États-Unis, depuis le retour de Trump aux commandes, l'administration a pris un tournant très clair : finies les discussions sur l'équité dans l'IA, finis les garde-fous. À la place, on parle d'IA "débarrassée de tout biais idéologique". Comprenez : "Arrêtez de parler de racisme ou de sexisme, ça fait trop gaucho."

Résultat ? Les agences qui régulent l'IA reçoivent des consignes pour *éviter d'aborder les questions d'éthique ou de diversité* dans leurs rapports. Exit les termes comme "justice algorithmique" ou "IA responsable". On nettoie le vocabulaire, on lisse le discours. Objectif affiché : laisser le champ libre aux entreprises pour "innover" (comprendre : faire du profit sans contraintes) et permettre aux États-Unis de "gagner la course de l'IA".

Mais gagner quoi, exactement ? Une IA plus rapide, plus puissante… et plus dangereusement biaisée ? Une IA qui trie, exclut, catégorise, sans jamais se remettre en question ? Et pendant que certains rêvent d'une IA boostée à la Silicon Valley, propre, lisse, sans aspérités (Elon si tu nous lis…), d'autres – nous, vous, toi qui lis ça – continuent à défendre une IA qui inclut, qui interroge, qui doute. Une IA qui voit les marges et ne les efface pas. Du coup, plutôt que d'attendre que ça vienne du haut, on vous propose que ça vienne (encore) du bas.

Pour s'inspirer, on a décidé d'interroger des personnes inspirantes, comme Anaïs Baumgarten. Cette experte en marketing digital, référencement et stratégie de contenus, développe aujourd'hui des stratégies de marketing digital responsables et durables. À l'heure où de nombreuses marques surfent sur les tendances inclusives sans réelle transformation en interne, elle propose une alternative aux stratégies superficielles : un marketing digital qui ne soit pas qu'une façade, mais une véritable démarche de fond.

Dans son livre "Slow Marketing", elle outille les lecteur·rice·s pour penser, concevoir et déployer des stratégies de marketing plus inclusives et éthiques, sans tomber dans le piège du washing. Car si les consommateur·rice·s sont de plus en plus sensibles aux questions d'égalité et de diversité, iels sont également devenus plus exigeant·es. Un logo arc-en-ciel en juin ou une campagne sur la diversité en mars ne suffisent plus à convaincre : la sincérité et la cohérence sont devenues les piliers d'une communication responsable. On s'inspire des meilleur·es, et notre échange avec Anaïs nous a donné envie d'explorer quatre grandes pistes pour repenser l'inclusivité dans l'IA.

01 « RIEN SUR NOUS SANS NOUS »

Un outil technologique ne peut pas être inclusif si les personnes concernées ne participent pas à sa conception. Le principe du « rien sur nous sans nous », issu des luttes handis, résume bien l'enjeu : les communautés concernées doivent avoir leur place dans la conception des IA et doivent intervenir dans l'objet final. Cela signifie que l'accessibilité ne doit pas être une simple considération technique ajoutée a posteriori, mais une réflexion intégrée dès le départ par celles et ceux qui vivent ces réalités.

Dans l'épisode numéro 2 de notre podcast (c'est pas pour teaser qu'on vous dit ça...), Julie Mouvet, notre journaliste, reçoit Vincent Léone, qui partage son expérience en tant que formateur et expert en accessibilité numérique. Pour lui, « la véritable inclusivité ne se résume pas à prévoir des options d'accessibilité, mais à faire en sorte que les personnes concernées soient impliquées dans toutes les étapes, de la conception à l'utilisation ». Par exemple, ce n'est pas aux personnes non concernées de présupposer les besoins des personnes aveugles ou malvoyantes (dans le cas de Vincent). L'inclusion ne consiste pas seulement à rendre un outil compatible avec un besoin spécifique, mais à changer la manière dont on conçoit et déploie ces technologies et leurs fonctionnalités.

02 DIVERSIFIER LES ÉQUIPES ET LES BASES DE DONNÉES : UNE SOLUTION, VRAIMENT ?

Lorsqu'il est question de rendre l'intelligence artificielle plus inclusive, la solution qui revient le plus souvent est simple : diversifier les équipes et enrichir les bases de données. À première vue, cette approche semble évidente. Si l'IA est biaisée, c'est parce qu'elle est formée sur des données biaisées et conçue par une élite technologique homogène. Il suffirait donc d'injecter plus de diversité dans le processus pour corriger le tir. Sauf que... ça ne marche pas aussi bien qu'on aimerait le croire.

L'une des erreurs fondamentales est de croire que l'IA deviendra magiquement plus inclusive si on lui donne des données plus variées. En réalité, les biais ne viennent pas uniquement du manque de diversité dans les datasets, mais aussi de la manière dont ces données sont utilisées et des critères qui définissent les décisions de l'algorithme.

Revenons sur l'exemple marquant de l'algorithme de recrutement d'Amazon cité précédemment. Conçu pour trier automatiquement les CV des candidat·es, il s'est révélé discriminatoire envers les femmes. Pourquoi ? Parce qu'il avait été formé sur des données historiques, où la majorité des embauches étaient masculines. Résultat : même en ajoutant des CV féminins dans la base de données, l'algorithme continuait à privilégier les hommes, car il avait appris que ce profil correspondait mieux aux attentes des recruteurs.

Ce cas nous montre bien que le problème ne se limite pas à la diversité des données. Si on ne change pas les critères de sélection et qu'on ne remet pas en question les biais intégrés dans les décisions de l'IA, on ne fait que camoufler le problème plutôt que de le résoudre.

Pour que l'IA soit réellement plus inclusive, elle doit donc être associée à d'autres actions concrètes

- **Tester les IA sur des profils variés avant leur mise en service :** beaucoup d'algorithmes sont testés dans un cadre trop restreint, et/ou sur des échantillons qui ne reflètent pas la diversité réelle des utilisateur·ice·s. Conséquence : des systèmes de reconnaissance vocale qui peinent à comprendre les accents, des correcteurs automatiques qui dysfonctionnent sur les langues genrées, ou encore des interfaces inutilisables pour les personnes en situation de handicap. Pour éviter cela, il est essentiel de mettre en place des panels de test représentatifs, avec des personnes de différentes origines, genres, âges et conditions physiques. Cela permettrait d'identifier en amont les failles discriminatoires d'un modèle avant qu'il ne soit déployé à grande échelle.

- **Inclure des expert·es en éthique et en sciences sociales dans la conception des IA:** aujourd'hui, les algorithmes sont principalement développés par des ingénieurs, qui se concentrent avant tout sur la performance technique. Problème : l'optimisation des résultats ne prend pas en compte les implications sociales et éthiques. Par exemple, un modèle conçu pour améliorer la rapidité d'un processus de recrutement pourrait éliminer systématiquement les candidat·es ayant des parcours atypiques, simplement parce qu'ils ne correspondent pas aux profils majoritaires dans les bases de données. La solution ? Faire travailler ensemble des spécialistes de l'IA, mais aussi des sociologues, des juristes et des expert·es en diversité et inclusion. Ces profils sont essentiels pour repérer les biais invisibles, comprendre les enjeux sociaux et proposer des garde-fous avant que ces technologies ne soient mises sur le marché.

- R**evoir les paramètres des IA pour éviter les biais systémiques :** l'IA ne décide pas seule de ses critères de sélection : ce sont les humains qui les définissent. Sauf que ces critères sont souvent basés sur des modèles discriminants.

03 DÉCONSTRUIRE LES BIAIS INCONSCIENTS DE L'IA PAR L'IA

L'IA est souvent critiquée pour son rôle dans la reproduction et l'amplification des biais. Mais si, au lieu de simplement les refléter, elle pouvait nous aider à les déconstruire ? Peut-on imaginer une intelligence artificielle capable de nous confronter à nos propres stéréotypes, de nous faire prendre conscience des schémas discriminants intégrés dans notre pensée, voire de nous aider à les dépasser ?

Aujourd'hui, on sait que les biais inconscients sont omniprésents dans nos décisions, même quand on pense être rationnel·le. Ces biais sont souvent le produit d'une socialisation et d'un environnement culturel qui influencent nos jugements sans que nous en ayons conscience.

C'est dans ce cas qu'intervient l'idée d'une IA qui ne se contente pas de refléter ces biais, mais qui les met en évidence et nous invite à les questionner. Un des outils les plus connus dans ce domaine est le IAT (Implicit Association Test). Développé par des chercheur·euse·s en psychologie sociale, ce test vise à mesurer nos associations inconscientes entre certains concepts (par exemple, "homme" et "carrière" vs. "femme" et "famille"). Les résultats montrent que même des personnes sincèrement engagées pour l'égalité par exemple peuvent avoir des biais cognitifs profondément ancrés qui influencent leurs décisions.

L'IA pourrait fonctionner sur le même principe :

- Elle pourrait analyser nos comportements numériques (les termes qu'on emploie, les images qu'on sélectionne, les contenus qu'on consomme) et mettre en lumière les biais sous-jacents dans notre manière de penser.
- Elle pourrait nous proposer des alternatives, en nous exposant à des points de vue que nous n'aurions pas spontanément considérés.

Prenons l'exemple d'un moteur de recherche d'images : si on tape "PDG", l'IA pourrait afficher une diversité de profils, au lieu de simplement refléter les modèles dominants d'hommes blancs en costume. Si on génère du texte, un·e assistant·e IA pourrait reformuler nos phrases en supprimant des biais implicites (ex : éviter le masculin par défaut ou proposer des tournures plus neutres). L'objectif ne serait pas d'imposer une vision unique, mais de rendre visibles des automatismes inconscients, pour que chacun·e puisse choisir en connaissance de cause.

UNE IA QUI NOUS CHALLENGE PLUTÔT QU'UNE IA QUI NOUS CONFORTE

Aujourd'hui, la majorité des algorithmes fonctionnent en renforçant nos préférences existantes. Les recommandations YouTube, Spotify, TikTok ou Instagram nous enferment dans des bulles cognitives, en nous proposant du contenu qui confirme ce que nous aimons déjà. Qu'en serait-il si l'IA était programmée pour challenger nos perceptions au lieu de les renforcer ? On pourrait alors imaginer un·e assistant·e IA qui :

- Propose des arguments opposés à ceux que l'on défend dans une discussion, pour nous entraîner à voir d'autres perspectives (autant vous dire que la gémeau qui écrit serait heureuse comme jaja).
- Met en évidence les stéréotypes dans nos demandes (exemple : "Génère-moi une image d'un médecin" → "Veux-tu voir une représentation plus diversifiée de ce métier ?").
- Pose des questions critiques sur nos choix de langage, en nous demandant si une autre manière de formuler notre pensée pourrait être plus juste ou plus inclusive.

Une IA bien conçue pourrait ainsi servir d'outil d'éducation critique, nous aidant à déconstruire nos automatismes cognitifs plutôt que de simplement nous conforter dans nos certitudes.

UNE IA ANTI-BIAIS... MAIS POUR QUI ET AVEC QUELLES LIMITES ?

Évidemment, cette approche pose de nombreuses questions éthiques et politiques. Qui décide de ce qui constitue un biais et de ce qui devrait être corrigé ? Dans quelles limites une IA doit-elle influencer notre pensée ?

Certain·es pourraient voir ce type d'IA comme une forme de "police de la pensée", imposant une vision unilatérale de ce qui est acceptable ou non. Finalement, n'est-ce pas déjà ce qui est en train de se produire aujourd'hui? Sous couvert d'une impartialité due à son non-humanisme (à qui on dit bonjour et merci d'ailleurs... #incohérence), l'IA structure tout de même une certaine vision du monde en se basant sur les données qu'elle ingère et les choix de conception faits en amont. En ce sens, elle influence déjà notre pensée, d'une manière qui se présente comme neutre alors qu'elle ne l'est pas du tout.

Pour éviter cet écueil, il faudrait que l'IA reste un outil de suggestion et de questionnement, sans imposer une norme unique. L'enjeu serait donc de créer des IA réflexives, capables de signaler leurs propres limites et de donner accès à des perspectives variées, plutôt que de simplement corriger les réponses pour les rendre plus "conformes".

Y à plus qu'à !

FOCUS INSPIRATION

Avec la team d'OK Mila, on a à cœur de mettre en avant des projets et initiatives qui ont du sens et qui permettent de remettre en question les schémas dans lesquels nous sommes, et qui nous paraissent logiques. On a donc décidé de vous présenter le projet #diversitY.X #crIAtivitY.X, un projet jeunesse engagé à la croisée de l'éducation aux médias, de la création artistique et de la réflexion critique sur l'IA. D'une part, le projet souhaitait sensibiliser les jeunes aux questions de genre et de diversité dans les médias, et d'autre part, leur donner des clés pour comprendre et détourner l'IA de manière créative. Grâce à ces activités, les jeunes ont pu explorer l'univers de l'intelligence artificielle tout en s'exprimant librement sur des thématiques qui les touchent : identité, corps, normes, représentations... Le tout à travers une diversité de formats créatifs : fanzines, œuvres graphiques, podcasts, cartes postales, et bien d'autres supports engagés et puissants.

Sur le site du projet (www.diversityx.be), vous retrouverez

- des dossiers pédagogiques créés pour le secteur jeunesse, afin d'aborder de manière concrète et ludique les stéréotypes de genre véhiculés par les réseaux sociaux et les technologies numériques.
- deux webinaires et plusieurs conférences, qui ont permis de croiser les regards sur les enjeux éthiques et sociaux de l'IA, avec une attention particulière à ses impacts sur les jeunes, les représentations, et les inégalités.
- des ateliers d'éducation aux médias mêlant art plastique et outils d'IA à reproduire avec des publics concernés pour expérimenter, détourner, créer à partir des outils numériques pour mieux les comprendre et les questionner.
- une série de podcasts où les jeunes sont au coeur même de la discussion.

BIBLIOGRAPHIE ET SITOGRAPHIE

L'ADN.eu, "Les logiciels de recrutement basés sur l'IA sont discriminants" : https://www.ladn.eu/entreprises-innovantes/discrimination-recrutement-logiciel-intelligence-artificielle/
Bloomberg.com, "Humans Are Biased Generative AI Is Even Worse", https://www.bloomberg.com/graphics/2023-generative-ai-bias/
Newsocialist.org, AI: The New Aesthetics of Fascism" : https://newsocialist.org.uk/transmissions/ai-the-new-aesthetics-of-fascism/
Jamaissanselles.fr, "L'intelligence artificielle entretient des stéréotypes bien réels" : https://www.jamaissanselles.fr/biais-intelligence-artificielle/
ONU Info, "Intelligence artificielle : l'UNESCO alerte sur les stéréotypes racistes, sexistes et homophobes":https://news.un.org/fr/story/2024/03/1143812
Unia, "Intelligence artificielle (IA) et discrimination" : https://www.unia.be/fr/dossiers/artificiele-intelligentie-discriminatie
ONU Femmes, "Comment l'IA renforce les préjugés sexistes, et ce que l'on peut faire pour tenter d'y remédier": https://www.unwomen.org/fr/nouvelles/interview/2025/02/comment-lia-renforce-les-prejuges-sexistes-et-ce-que-lon-peut-faire-pour-tenter-dy-remedier
Wired, "Under Trump, AI Scientists Are Told to Remove 'Ideological Bias' From Powerful Models" : https://www.wired.com/story/ai-safety-institute-new-directive-america-first/?utm_source=chatgpt.com
The Guardian, "US and UK out of step with rest of the world on AI" : https://www.theguardian.com/global/2025/feb/18/ai-artificial-intelligence-paris-summit-dei?utm_source=chatgpt.com

LES CONTRIBUTEURS ET CONTRIBUTRICES SUR CE NUMÉRO

- **Chloé Tran Phu de** *Media Animation* et *Witch Gamez*
- **Anaïs Baumgarten** *de Slow Marketing*
- **Sarah Ladam**
- **Vincent Léone** *Site internet*
- **Leïla Maidane** *InterSkillar* / *Femmes Fières*

OK MILA

L'intelligence artificielle bouleverse notre manière de nous informer, de créer et d'interagir. Avec OK Mila, nous décryptons ses impacts éthiques et sociétaux à travers six thématiques clés :

- IA et désinformation (avril 2025) : La désinformation n'est pas une nouveauté mais la professionnalisation et la précision des fausses images et vidéos rendent l'authentification des contenus difficile. Comment déceler le vrai du faux ?
- IA et diversité (mai 2025) : l'IA diffuse de nombreux stéréotypes liées au genre, à l'âge, à l'origine, à l'orientation sexuelle et tout type de minorité : hypersexualisation des femmes, attitude de bandit des personnes racisées, exotisation des personnes lesbiennes, diktats de beauté normée. Comment sortir de la représentation de l'individu universel ?
- IA et écologie (juin 2025) : bien qu'il nous semble dématérialisé, le numérique pollue énormément. Et l'IA n'arrange rien : entre autres, l'entraînement des algorithmes est le poste le plus polluant et les IA utilisent une électricité bien plus importante que les moteurs de recherche classiques. Comment réduire notre empreinte carbone malgré une utilisation de l'IA régulière ?

A partir de septembre 2025, on parlera IA et cyber-harcèlement, IA et politique et enfin, IA et créativité.

Un projet multimodal pour toucher un large public

OK Mila, c'est trois formats interconnectés pour explorer ces enjeux en profondeur et développer un esprit critique face à l'IA :

- Un podcast (12 épisodes) : une thématique et 2 interventions, un sujet et deux volets car les discussions autour de l'IA et de l'éthique révèlent de nombreuses questions.
- Des vidéos courtes : des capsules percutantes et accessibles pour sensibiliser un large public autour d'informations concrètes.
- Un magazine papier (6 numéros) : un format A2 pliable avec un lexique IA, des interviews et des QR codes vers des ressources supplémentaires, diffusé dans les écoles, maisons de jeunes, AMO et centres culturels.

Où nous retrouver ? ▶ Écoutez nos podcasts sur **Spotify, Apple Podcasts, Deezer et toutes les plateformes d'écoute**. ▶ Retrouvez nos vidéos et ressources sur **www.ok-mila-eam.be**. ▶ Suivez-nous sur **Instagram** pour ne rien manquer !

OK MILA MAG

UN PROJET D'ÉDUCATION AUX MÉDIAS QUI QUESTIONNE L'ÉTHIQUE ET L'INTELLIGENCE ARTIFICIELLE

N°02 · IA ET DIVERSITÉ