



Critique de l'IA centrée sur l'humain : plaidoyer pour un cadre d'IA féministe (FAIF)

Tanja Kubes¹

Reçu : 17 mars 2025 / Accepté : 29 juillet 2025

© L'auteur(s) 2025

Abstrait

Alors que l'intelligence artificielle (IA) transforme la science, l'industrie, le discours public et nos vies personnelles, il apparaît de plus en plus clairement que les avancées non réglementées dans ce domaine présentent des inconvénients. L'IA centrée sur l'humain (IAC) a été saluée comme une avancée importante vers l'équité, la justice et l'inclusion. Cependant, une analyse féministe critique montre que l'IAC ne parvient pas à atteindre ces objectifs. L'IAC, telle qu'elle est pratiquée actuellement, reste largement eurocentrique, androcentrique et anthropocentrique, et est guidée par une logique de marché capitaliste sans engagement envers la durabilité. Elle ignore et exacerbe les injustices mondiales et les impacts écologiques. L'article critique l'IAC d'un point de vue féministe, posthumaniste et néo-matérialiste et propose un Cadre d'IA féministe (CAIF) qui intègre également les résultats de l'anthropologie plus qu'humaine et de la phénoménologie corporelle. Le FAIF vise à réévaluer la relation entre les humains, les autres formes de vie et la technologie, et explore le potentiel d'une conception, d'une utilisation et d'un contrôle collaboratifs et non hiérarchiques de l'IA. Plutôt que de se concentrer principalement sur la limitation des dégâts, le FAIF constitue une tentative proactive pour garantir le bien-être coconstitutif des humains et des non-humains, en rompant avec les discours traditionnels sur l'IA et le progrès.

Mots-clés IA féministe · IA centrée sur l'humain · éthique de l'IA · gouvernance de l'IA · posthumanisme · IA responsable

1 Les effets perturbateurs de l'IA

L'intelligence artificielle a fondamentalement transformé nos façons de vivre, de travailler, de faire des affaires, de communiquer, de nous organiser, d'apprendre, d'écrire, de créer des œuvres d'art, de faire de la politique, de faire la guerre, de penser et de nouer des relations. Aucune technologie dans l'histoire n'a eu autant d'impact sur la vie d'autant de personnes en si peu de temps. L'IA fait désormais partie intégrante de notre quotidien et transforme la vie sociale dans son ensemble. Si beaucoup espèrent encore que ces changements seront positifs, les risques d'une confiance naïve dans la technologie deviennent de plus en plus évidents. À ce jour, les intelligences artificielles restent confinées à des tâches spécifiques. Dans les domaines du traitement des données, de la reconnaissance de formes et de l'automatisation, elles sont déjà supérieures aux humains. Dans d'autres domaines, comme la compréhension du contexte et les décisions morales, les humains ont (encore) une longueur d'avance. Cependant, si les prédictions de nombreux experts en technologie se confirment, les IA génératives pourraient considérablement élargir leur champ d'action et

Les systèmes d'intelligence artificielle (IA) pourraient devenir des superintelligences artificielles (ISA) ou des systèmes d'intelligence artificielle générale (IAG) (Bostrom 2016). Si la situation se complique vraiment, une intelligence artificielle pourrait développer un moyen de se répliquer de manière autonome et de fixer et poursuivre des objectifs de manière indépendante. Que nous soyons à des décennies ou à quelques années d'un tel scénario dystopique (Magee & Tessaleno 2011, Guangyin et al. 2025), sa simple possibilité nécessite une reconfiguration substantielle de la place de l'humanité dans le cadre planétaire. Il est donc essentiel de prendre en compte les ramifications sociales et politiques de l'IA et de répondre aux questions fondamentales : comment, à quelles fins et pour qui nous entendons la développer à l'avenir.

Actuellement, la prise de décision en matière d'IA est limitée à un petit nombre d'individus et d'entreprises aux États-Unis, en Chine, en Europe et en Inde, avec une tendance décisive vers une plus grande monopolisation (Reuters 2025, Engster/Moore 2020).¹

Cependant, comme les effets de l'IA ne s'arrêtent pas aux frontières géopolitiques, il est impératif de prendre en compte ses risques profonds (et véritablement existentiels) et de développer des mécanismes de régulation capables de stopper efficacement les avancées excessivement irréfléchies. Les décisions prises aujourd'hui détermineront notre façon de vivre dans le futur.

* Tanja Kubes
tanja.kubes@fu-berlin.de

¹ Département de physique, études scientifiques, Freie Universität Berlin, Berlin, Allemagne

¹ Voir, par exemple, l'offre de rachat d'OpenAI par Elon Musk (Reuters 2025).

L'avenir, dont les intérêts seront représentés et dont les décisions seront prises au-dessus de qui. L'IA peut ouvrir la voie à un monde plus juste et égalitaire avec une multiplicité d'acteurs (humains et non humains), ou elle peut surveiller, manipuler, réprimer et discipliner les humains et détruire les écosystèmes.

Pour orienter les choses dans la bonne direction, nous avons besoin d'une coopération mondiale et de réglementations transnationales (Batool/Zow-ghi/Bano 2025). Dans cet objectif, l'ONU a créé en 2020 un Conseil consultatif de haut niveau pour l'IA. Dans un rapport publié l'année dernière, ce Conseil a recommandé la création d'un groupe scientifique international permanent et d'un Secrétariat de l'ONU pour l'IA (ONU 2024). Le rapport a également souligné un problème très préoccupant : à ce jour, 118 pays, principalement du Sud, n'ont participé à aucun des grands efforts internationaux visant à réglementer l'IA (ONU 2024 ; UNRIC 2024).

Il existe un risque considérable qu'une grande partie de la population mondiale soit affectée par les progrès de l'IA sans avoir son mot à dire sur ses développements. Le rapport poursuit en affirmant que « pour que l'IA puisse véritablement bénéficier à toute l'humanité et ne laisser personne de côté, les droits de l'homme doivent être au cœur de la gouvernance de l'IA, garantissant une responsabilisation fondée sur les droits dans toutes les juridictions » (ONU 2024).

Dans cet article, je soutiens qu'un tel « engagement fondamental en faveur des droits humains » est insuffisant. Il ne suffit pas de cesser de développer des IA qui servent principalement les intérêts d'une élite technologiquement avancée, majoritairement nordique, et de les orienter vers le bien-être de l'humanité tout entière. Il ne suffit pas non plus de mettre un terme à l'accumulation du pouvoir et de l'argent entre les mains d'une minorité et de viser une répartition équitable.

des bénéfices de l'IA et de la préservation des ressources. Il ne suffit pas de démocratiser la recherche et la conception dans un esprit de science citoyenne et de s'engager dans une conception technologique participative. Une approche durable de l'IA doit aller plus loin et nécessite une révision fondamentale de ses prérequis ontologiques et épistémologiques.

Cette réévaluation intègre un large éventail de perspectives et les synthétise dans ce que j'appelle un Cadre d'IA féministe (FAIF). S'appuyant sur le féminisme « classique », il examine le rôle des femmes et d'autres groupes marginalisés dans le secteur technologique. Elles sont sous-représentées dans les équipes de développement et leurs besoins particuliers en matière d'utilisation des nouvelles technologies sont peu pris en compte. S'appuyant sur le postcolonialisme, il aborde les inégalités mondiales et les relations de pouvoir. Issu des études scientifiques et technologiques, il adopte l'idée de la situation du savoir, et des études environnementales, l'attention portée à la consommation excessive de ressources par l'IA actuelle. Le FAIF va au-delà de ces critiques particulières et s'interroge sur les lacunes fondamentales des discours actuels sur l'IA centrée sur l'humain.

Dès 1993, Bruno Latour affirmait que la modernité occidentale repose sur un profond paradoxe. Dans leur tentative de séparer nature et société pour créer des domaines « purs », les modernes ont constamment généré des hybrides, brouillant la frontière entre

les deux et créant de nouvelles formes d'intrication.² Le problème est structurel et ne peut être résolu par le progrès technologique, aussi sophistiqué soit-il. FAIF en tient compte et développe une perspective sur l'IA qui ne s'arrête pas à des améliorations progressives, mais s'interroge sur les visions fondamentalement différentes de l'humain et de l'IA qui émergent lorsque nous complétons la position anthropocentrique par une perspective plus qu'humaine.

En anthropologie, il existe de nombreux exemples d'ontologies non occidentales prêtes à offrir aux acteurs non humains d'autres options de connaissance et d'action. L'étude de ces formes alternatives de « création de mondes » (Descola 2024) peut donner une impulsion décisive à une utilisation responsable de l'IA. Non seulement parce qu'elle dépasse la simple considération de l'IA comme un simple « outil » (ou objet) facilement configurable d'une manière ou d'une autre, mais surtout parce qu'elle réinterroge la nature de l'existence humaine et les assemblages multi-espèces (Kubes/Reinhardt 2022) au sein desquels nous coexistons et évoluons.

2 L'IA centrée sur l'humain et la perspective sociotechnique

Depuis quelques années, des efforts sont déployés aux niveaux national et européen pour contrer les risques liés à l'IA par le biais de réglementations et de normalisations, et garantir que les avancées technologiques répondent aux attentes et aux exigences de leurs utilisateurs humains. L'idée a une histoire qui s'étend sur plus de quatre décennies dans le domaine de l'interaction homme-machine (IHM).

L'application du concept à l'IA est relativement nouvelle, mais s'est avérée productive. Une recherche Google Scholar, menée début mars 2025, a répertorié environ 27 700 publications scientifiques sur le thème de l'IA centrée sur l'humain.³ La loi européenne sur l'IA de 2024 (EUR-Lex 2024), les Nations Unies (UN 2024), l'Organisation internationale de normalisation (ISO 2023) et l'Institut allemand de normalisation (DIN 2022) abordent tous ce sujet. Un thème récurrent dans toutes ces publications est l'impératif que les systèmes d'IA fiables soient conçus en accordant une importance primordiale aux humains et à leur bien-être à toutes les étapes du développement, de la mise en œuvre et du contrôle.

² Latour (1993) cite le trou dans la couche d'ozone comme exemple. Aujourd'hui, je ferais plutôt référence au changement climatique ou à l'intelligence artificielle.

³ J'ai utilisé la chaîne de recherche suivante pour rechercher simultanément toutes les orthographes possibles en anglais : ALL ("HCAI") OU ALL ("human-cen-tered AI") OU ALL ("human-centred AI") OU ALL ("human-centric AI") OU ALL ("human-centered artificial intelligence") OU ALL ("human-centred artificial intelligence") OU ALL ("human-centric artificial intelligence") OU ALL ("human centered AI") OU ALL ("human centred AI") OU ALL ("human centric AI").

L'adoption d'une perspective sociotechnique holistique est essentielle au succès de telles mesures. Celle-ci va au-delà de la simple faisabilité technologique et se concentre sur les questions éthiques et culturelles et les interactions complexes entre la technologie et l'individu, ou la technologie et la société (Suchman 2007 ; Zweig et al. 2021, DIN DKE 2022). Quelles normes culturelles et sociales, par exemple, façonnent nos interactions avec les chatbots ? Devons-nous dire « s'il vous plaît » et « merci » dans les chats ? Est-il acceptable d'aller droit au but lorsque nous posons une question ? Dans quelle mesure une réponse peut-elle être directe sans être perçue comme impolie et impolie par les utilisateurs ? Ou, concernant les logiciels de reconnaissance faciale : quelles méthodes d'identification personnelle automatisée peuvent être utilisées dans une société où les gens sont voilés dans l'espace public ou transportent des charges sur la tête ? Et comment la technologie interfère-t-elle avec les conceptions locales de la vie privée et de l'espace personnel ? Une perspective sociotechnique de l'IA intègre des éléments techniques (matériel, données, algorithmes, etc.) à des composantes psychologiques et sociales (individus, société, économie, marché du travail, politique, etc.) et vise à permettre au plus grand nombre de bénéficier du progrès technologique. Voilà pour la théorie.

En pratique, cependant, l'histoire de la connaissance et de la science est profondément imprégnée de sexisme, de racisme et d'eurocentrisme (Barad 2007, Braidotti/Lykke 1996, Fox-Keller 1986, Haraway 1991, Haraway 1988, Harding 1998, Harding 1986, Harding 1991, Harding 2006, Vallor 2024, Wajcman 2004).

Des auteures féministes comme Ruha Benjamin, Sascha Constanza-Chock, Safiya Noble ou Joy Buolamwini ont démontré à maintes reprises comment les logiciels de reconnaissance faciale ne reconnaissent pas de manière fiable les personnes de couleur, comment les systèmes de crédit désavantagent les personnes présentant certaines caractéristiques ethniques et démographiques, comment les scanners corporels classent mal les personnes queer et non binaires, comment les outils de diagnostic médical ne fonctionnent pas correctement pour les femmes ou les personnes de couleur, comment les logiciels de recrutement désavantagent systématiquement les femmes et les personnes de couleur, et comment les moteurs de recherche peuvent amplifier le sexisme, le racisme et la misogynie (Benjamin 2019, Buolamwini 2018, Constanza-Chock 2020).

Par conséquent, les défenseurs d'une perspective féministe sur l'IA appellent à plus de justice, de responsabilité et de transparence dans l'IA et soulignent à juste titre la nécessité de perspectives diverses et intersectionnelles afin de rendre justice aux interactions complexes entre la technologie, la société et l'individu.

Ce faisant, ils ouvrent le concept de centrage sur l'humain à des perspectives intersectionnelles plus diverses. En fait, la critique féministe – et par extension, mon Cadre féministe pour l'IA (FAIF) – va encore plus loin. Non seulement elle identifie les pratiques concrètes de discrimination et appelle à davantage de diversité, d'équité et d'inclusion (au sein des conseils d'administration, dans les STEM, dans le design, etc.), mais elle identifie également les facteurs structurels à l'origine des pratiques discriminatoires. Le concept de « savoirs situés » développé par Donna Hara-way (1988) est ici central. Il rejette les notions d'objectivité universelle et de connaissance ultime, et privilégie la connaissance et

Vérité au pluriel. Selon Haraway, l'« objectivité » ne peut être pensée que dans le contexte d'actes de production de connaissances réflexifs, responsables et contextualisés. Elle n'est pas une réalité extérieure aux individus, mais plutôt une compréhension performativement incarnée – ou située – de ce qui est vrai dans les limites de structures de pouvoir concrètes et valable uniquement d'un point de vue spécifique.

Pour les innovations techniques, cela signifie que l'illusion répandue d'une « technologie neutre » se révèle n'être qu'une illusion. Savoir et pouvoir sont indissociables. Toute connaissance, toute technologie, toute innovation s'enracinent dans des contextes culturels, sociaux et historiques spécifiques, et les inégalités qui y sont établies se fondent dans des structures intersectionnelles.

La connaissance demeure donc nécessairement perspectiviste, pluraliste, incomplète – ou, selon Haraway (1992) et Braidotti et Lykke (1996), « monstrueuse ». Elle ne peut plus être réduite à de simples chaînes de causes et d'effets ou de réflexion et de réplication, mais se constitue par des interférences intersectionnelles. Elle devient « diffractive » (Haraway 1992).

Il existe un large consensus sur le fait que cela est vrai pour les savoirs culturels et sociaux, mais au sein de la FAIF, cela s'applique également aux connaissances supposément axiomatiques des mathématiques, des sciences naturelles et de l'informatique. Le perspectivisme et l'approche performative du féminisme classique et des STS féministes constituent l'un des piliers de la FAIF. Leurs revendications sont importantes et justes, et leur interprétation large peut également inclure des questions clés de la critique postcoloniale. Cependant, pour contrer substantiellement le discours capitaliste technodéterministe, qui met l'accent sur la maximisation du profit et l'expansion du marché (Engster/Moore 2020 ; Moore 2020 ; Fox 2023), il est nécessaire de s'attaquer non seulement à ses manifestations patentes, mais aussi à ses racines. C'est là qu'entre en jeu le posthumanisme critique, le nouveau matérialisme et l'anthropologie plus qu'humaine.

3 Les problèmes faciles et difficiles de l'IA

Le discours dominant sur l'IA centrée sur l'humain s'appuie sur une forme spécifique de dualisme de substance, qui place l'humain dans une position unique au sein de l'ordre du monde. S'appuyant sur la distinction de Descartes entre la conscience « immatérielle », capable de réflexion rationnelle (res cogitans), et son environnement matériel, incapable de penser (res extensa), l'esprit humain a la primauté sur son substrat physique. Cela ignore le « problème difficile » (Chalmers 1995), à savoir la relation entre les phénomènes physiques (par exemple, les activités neuronales) et les expériences mentales. Cela écarte également le « paradoxe de la modernité » et sa « tension insurmontable entre sujet et objet » (Latour 1993). Au lieu de cela, une grande partie de la recherche en IA repose sur un matérialisme tacite qui considère la conscience comme une fonction des processus physiques et suppose qu'il sera tôt ou tard possible de

Répliquer le cerveau humain dans un ordinateur. Dans sa forme extrême, cette pensée mène à la vision transhumaniste de l'émulation cérébrale globale et du téléchargement de l'esprit.

Dans une telle fusion de l'homme et de la technologie, ce qui nous rend humains aux yeux de beaucoup (notre esprit) serait totalement détachable de son substrat matériel (notre corps). Par conséquent, la condition humaine, encore soumise à de multiples contraintes physiques et impitoyablement limitée par notre mortalité, pourrait être remplacée par une conscience immortelle se manifestant sous forme d'information pure, indépendante de tout support particulier. Une telle conscience « pure » est-elle seulement concevable ? Probablement oui. L'avoir (ou l'être) serait cependant difficilement compatible avec l'être humain. En s'appuyant sur le philosophe Thomas Nagel (2018), il est raisonnable de supposer qu'il y a quelque chose que l'on ressent à être une conscience désincarnée. En même temps, cependant, cette expérience serait totalement incommensurable avec notre existence corporelle.

Mais prenons un peu de recul. Même sans faire allusion au « totalisme cybernétique » (Schnetker 2019) et à l'évasion transhumaniste, les avancées actuelles de l'IA soulèvent un certain nombre de questions, dont la cause commune peut être identifiée comme une forme spécifique de création du monde. C'est par là que nous devons commencer si nous voulons réellement construire une IA plus juste. Les défis peuvent être regroupés en deux catégories. La plupart concernent les conséquences politiques et sociales des actions de l'IA. Les autres abordent les fondements ontologiques et épistémologiques de ses actions. Les discussions concernant les progrès non réglementés de l'intelligence artificielle se sont jusqu'à présent principalement concentrées sur la première série de questions. En m'appuyant sur Chalmers (1995), je les qualifie de « problèmes faciles ».

Ces problèmes doivent clairement être traités – le plus tôt sera le mieux. Et soyons clairs : les qualifier de « problèmes faciles » de l'IA ne signifie pas qu'ils soient réellement simples à résoudre.

Résoudre. Ce n'est pas le cas. Je suis également conscient que les conséquences pour toute personne concernée peuvent aller de simples désagréments à des menaces vitales. Tableau 1

Les systèmes d'IA ont des difficultés avérées à reconnaître les personnes de couleur. En 2018, l'American Civil Liberties Union a testé le logiciel Rekognition d'Amazon en comparant les images des 535 membres du Congrès américain à une base de données de 25 000 photos d'arrestation. Le logiciel a généré 28 fausses identifications positives (environ 5 %), les personnes de couleur étant affectées de manière disproportionnée (Singer, 2018). Ce biais – il est important de le souligner – provient de données d'entraînement faussées, et non d'une intention malveillante. La « méthodologie I » (Oudshoorn et al., 2004) et le « regard codé » (Buolamwini, 2024) ont produit des IA qui excluent et discriminent systématiquement.

L'informaticienne Joy Buolamwini a fait l'expérience directe de cette discrimination involontaire mais apparemment inévitable lorsqu'elle a créé « Aspire Mirror », un programme destiné à superposer le visage de l'utilisateur avec des images inspirantes.

Les premiers tests furent encourageants. Cependant, une fois que Buolamwini eut intégré un logiciel de reconnaissance faciale open source, le programme ne parvint plus à identifier son visage. Il détecta parfaitement un simple masque d'Halloween blanc et même un visage dessiné schématiquement dans la paume de la main de Buolamwini, mais échoua si systématiquement à reconnaître les traits sombres de la programmeuse qu'elle dut effectuer les derniers tests.

logiciel « codage en whiteface » (Buolamwini 2024).⁴

Répondre aux besoins des minorités, des personnes de couleur ou des populations du Sud global et freiner le colonialisme numérique rampant (Gray/Suri 2019, Young 2019)⁵ Cela exige clairement plus qu'une simple volonté politique, une évaluation complète des besoins et un engagement à conclure des accords internationaux contraignants. Si je continue à qualifier tous ces problèmes de « faciles », c'est parce qu'ils peuvent être abordés sans dépasser les limites des schémas de pensée établis.

Tableau 1 Les « problèmes faciles » de l'IA

Conséquences des avancées non réglementées de l'IA : les « problèmes faciles »	
eurocentrisme	ignore les perspectives et les besoins des membres des sociétés non occidentales
Racisme	ignore les perspectives et les besoins des personnes de couleur
Androcentrisme	privilégie les perspectives et les besoins masculins
Concentration du capital	les intérêts fondés sur le marché dominant le développement
Concentration du pouvoir	pouvoir, richesse et pouvoir d'interprétation concentrés entre les mains de quelques nations, industries et individus
Exclusion du Sud global	marginalise et exploite les pays du Sud global ; les exclut des initiatives réglementaires

⁴ Cette expérience a incité Buolamwini à fonder l' Algorithmic Justice League qui attire l'attention sur l'impact social de l'IA et des biais algorithmiques et à lutter pour des technologies plus justes (AJL 2024).

⁵ Le colonialisme numérique signifie ici non seulement l'exploitation des matières premières et des conditions commerciales défavorables, mais aussi la délocalisation d'activités telles que le travail fantôme ou le travail au clic vers le Sud global.

Aussi complexes et compliqués soient-ils, le chemin vers leur solution s'inscrit toujours dans le cadre des pratiques culturelles que Weber a décrites un jour comme « un segment fini de l'infinité insignifiante des événements du monde qui a été imprégné de sens et de signification » (Weber 1985, ma traduction). Ils peuvent être difficiles à résoudre, mais nous n'avons pas besoin de nous aventurer dans l'inconnu pour les surmonter. Ils restent bien dans les limites au-delà desquelles, comme le disait Foucault, les procédures disciplinaires d'une époque particulière poussent « toute une tératologie de la connaissance » (Foucault 1981). Ce n'est pas vrai pour les problèmes difficiles. Ils nous poussent vers l'inconnu et nécessitent de profondes réorientations épistémologiques et ontologiques Tableau 2.

Pour s'attaquer aux problèmes difficiles, nous devons creuser plus profondément et nous préparer au monde de monstres qui se profile à l'horizon des écrits de Foucault, Haraway, Braidotti et Lykke. Heureusement, il existe des témoignages de mondes où les frontières entre humains et non-humains sont tracées différemment de celles des pays du Nord. Elles proviennent principalement de la théorie queer-féministe et de l'anthropologie. Prises ensemble, elles forment ce que l'on appelle un tournant ontologique (Holbraad/Pedersen 2017). La FAIF reprend certaines de ses idées et les applique à l'IA.

4. Diffracter les ontologies et les anthropologies plus qu'humaines

La modernité se caractérise par des dualismes : nature/culture, sujet/objet, masculin/féminin, original/copie, occidental/non occidental. Cette formulation ontologique va de pair avec la croyance épistémologique selon laquelle le monde peut être connu et représenté selon une perspective neutre et désincarnée – le « god trick » de Donna Haraway, un « saut hors du corps marqué vers un regard conquérant venu de nulle part » (Haraway 1988). L'épistémologie moderniste affirme que les représentations reflètent objectivement la réalité. Tels des miroirs courbes, elles peuvent nous présenter une image déformée, mais nous pouvons toujours retracer les transformations optiques pour voir la réalité .

Cette vision simpliste de la relation entre les phénomènes et leurs représentations est critiquée de multiples façons. Les physiciens rejettent depuis longtemps les notions naïves-matérialistes de la réalité. Heisenberg affirmait déjà que la « réalité » quantique – l'effondrement des fonctions d'onde de la superposition à un état probabiliste classique – n'émerge que par l'observation. Si l'observation ne requiert pas de conscience, « l'enregistrement, c'est-à-dire la transition du "possible" au "réel", est absolument nécessaire et ne peut être omis de l'interprétation » (Heisenberg 1958). Un tel effet d'observateur transcende les positions réalistes et constructivistes.

Le concept de « connaissances situées » de Haraway suit une trajectoire similaire. Haraway rejette à la fois l'illusion divine consistant à « tout voir de nulle part » et la réduction des objets physiques à des « éphémères de production discursive et de construction sociale » (Haraway 1988). Comme Heisenberg, elle cherche des alternatives au réalisme et au constructivisme radical.

Plutôt que d'adopter une réalité indépendante de l'observateur ou un relativisme poststructuraliste, Haraway prône « des connaissances partielles, localisables et critiques soutenant la possibilité de réseaux de connexions appelés solidarité en politique et conversations partagées en épistémologie » (ibid).

Malgré cela, Haraway ne cautionne pas un pluralisme acritique. Elle insiste sur la nécessité de séparer « les bonnes brebis scientifiques des mauvaises chèvres des biais et des abus » (ibid.). Autrement dit : certaines interprétations de la réalité sont tout simplement « meilleures » que d'autres. L'idée d'objectivité de Haraway évite à la fois l'observation désincarnée du « god trick » et les affirmations essentialistes de la théorie du point de vue sur le privilège épistémique des groupes marginalisés (Hartsock 1983). Elle propose plutôt une ouverture radicale , allant jusqu'à remettre en question l'identité même du chercheur. Selon Haraway, le sujet connaissant et sa connaissance du monde n'émergent que dans et à travers le processus de recherche.

Dans « Meeting the Universe Halfway », Karen Barad (2007) va encore plus loin. Titulaire d'un doctorat en physique des particules et en théorie quantique des champs, elle est également professeure de théorie féministe au département d'histoire de la conscience de l'UCSC depuis 2006, le même institut qui

Tableau 2 Les problèmes difficiles de l'IA

Causes des avancées non réglementées de l'IA – les « problèmes difficiles »	
L'exceptionnalisme humain et l'anthropocentrisme	Les humains sont surévalués ; on leur accorde une position privilégiée ; les acteurs non humains et leur bien-être ne sont pas pris en compte.
Dualisme	Penser par paires d'opposés (sujet/objet, nature/culture, humains/non-humains) à partir d'entités discrètes qui peuvent être clairement séparées les unes des autres
séparation corps-esprit	la res cogitans est privilégiée par rapport à la res extensa ; l'expérience corporelle et la perception affective sont ignorées
Les régimes occidentaux de la connaissance	Les sciences naturelles et techniques sont élevées au-dessus de toutes les autres disciplines
ntologie naturaliste	La distinction occidentale entre les sphères ontologiques (c'est-à-dire les humains et les non-humains) est exprimée en termes de absolu

L'ouvrage, animé par Donna Haraway et Anna Tsing, dont nous parlerons plus loin, s'appuie davantage sur l'interprétation de Copenhague. Suivant Bohr (et donc finalement Kant et Hegel), elle considère « les phénomènes, plutôt que les objets indépendants dotés de limites et de propriétés inhérentes » comme « unité ontologique primaire » (Barad 2007). Sa théorie des « agentivités intra-agissantes » remplace les identités prédiscursives par des enchevêtrements ontologiques – des relations sans relata préexistants. Barad se concentre sur les schémas relationnels.

plutôt que des différences ontologiques (Haraway 1997).

Du point de vue de la FAIF, cela a deux conséquences majeures. Premièrement, les acteurs humains et non humains ne doivent plus être classés dans des sphères ontologiques distinctes, mais forment des réseaux ou collectifs d'acteurs communs où personnes et choses se rassemblent dans des articulations spécifiques en tant qu'entités instituées (Latour 2001). Il en résulte une ouverture et une incomplétude fondamentales des classes ontologiques. Deuxièmement, le fait que les soi cristallisés à partir de telles articulations soient mutuellement constitutifs ne signifie pas que la relation soit nécessairement non hiérarchique ou organisée sur un pied d'égalité. Reconnaître l'agentivité des acteurs non humains ne signifie pas leur accorder le même degré d'intentionnalité, de détermination et de volonté qu'à la part humaine du collectif. Nous avons l'habitude de privilégier l'agentivité humaine dans les constellations inter-espèces et, en fin de compte, de considérer l'agentivité des animaux ou des choses comme une simple forme de « résistance » (Därman 2014).

Les développements actuels de l'IA menacent cependant de bouleverser cette relation et, pour la première fois, de placer les humains dans la position la plus passive de l'articulation. Cela comporte des risques considérables pour la pérennité de l'existence humaine. Si, un jour, l'IA pouvait prendre des décisions et poursuivre ses propres objectifs, elle pourrait décider d'utiliser l'énergie nucléaire pour satisfaire ses besoins énergétiques sans tenir compte des effets des radiations radioactives sur la vie. Pour une IA, la contamination potentiellement mortelle de l'environnement pour les humains et de nombreux autres êtres vivants ne constituerait tout simplement pas une « différence décisive » (Bateson 1985). La question de savoir comment les collectifs humains et l'IA négocient un monde commun est donc une question existentielle qui ne doit pas être laissée à la main invisible du marché.

Même si l'IA n'est pas encore au point, nous devrions prendre le temps de repenser notre relation au monde. Ces dernières décennies, il est devenu évident que la modernité souffre d'un problème structurel. Une antinomie qui ne peut être résolue en s'appuyant sur les répertoires conventionnels de critique. Nos tentatives incessantes de séparer nature et société ont conduit à une prolifération rapide de « quasi-objets » hybrides (Latour 1993). L'humanité a peut-être atteint un degré remarquable de « maîtrise de la nature ». La géo-ingénierie, les voyages spatiaux, la procréation assistée, les vaccins, etc. en témoignent. Cependant, chaque nouvelle solution a également créé de nouveaux problèmes : changement climatique, points de bascule, débris spatiaux, infertilité « héréditaire », allergies, etc. Il semble que les quasi-objets

Leur complexité augmente. Pire encore, aucune d'entre elles ne peut plus être attribuée sans ambiguïté à la nature ou à la culture. La fabrication continue de quasi-objets rend donc toute tentative de séparation absurde. De plus, l'insistance sur l'exceptionnalisme humain – fondement ontologique et épistémologique de la modernité – s'avère inefficace. Pourtant, comment pourrions-nous reconceptualiser la relation entre (wo)

l'homme et le monde ?

S'appuyant sur la réorientation épistémologique à travers les savoirs situants (Haraway) et la reformulation ontologique de la subjectivité émergente (Barad), il convient de regarder au-delà de l'horizon ontologique du Nord global. Les anthropologues ont maintes fois soutenu que les sociétés qu'ils étudient ne se contentent pas d'imposer d'autres catégories au monde.

Au contraire, ils créent d'autres mondes. De plus, la création de mondes n'est en aucun cas une occupation exclusivement humaine. Anna Tsing souligne avec insistance que chaque organisme, de la plus petite bactérie à l'arbre géant, façonne activement son environnement. Tous contribuent de manière dynamique à la rencontre ouverte et situationnelle des acteurs les plus divers que nous appelons le « monde ». Non seulement par les relations collaboratives qu'ils instaurent, mais aussi par les perturbations qu'ils provoquent (Tsing 2015).⁶ Tsing décrit un monde qui refuse d'être unifié et hiérarchisé. Il ne se révèle au spectateur que par des « parcelles » localement limitées. Si nous voulons comprendre l'Anthropocène, conclut Tsing, nous devons nous intéresser aux acteurs locaux. « Non pas parce que les conditions planétaires ont des effets et des instanciations locaux, comme le suggèrent de nombreuses études ; au contraire, (...) des parcelles façonnent l'Anthropocène, y compris sa forme planétaire » (Tsing et al. 2024).

S'appuyant sur cette perspective, l'Ontologie Relationnelle des Assemblages Multi-espèces (ROMA) (Kubes/Reinhardt 2022) remet en question l'ontologie moderniste en déconstruisant son dualisme et son anthropocentrisme sous-jacents. Alors qu'il existe déjà une abondante littérature sur les réseaux d'acteurs et les structures multi-espèces (Kneer 2008, Ogden/Hall/Tanita 2013), ROMA va plus loin en soulignant la nature co-constitutive et co-créative des acteurs – biologiques, techniques et symboliques – en les intégrant terminologiquement et conceptuellement dans des formes de création du monde interférentes et diffractives. ROMA propose une perspective sur la liaison entre acteurs humains et non humains, axée radicalement sur la relationnalité plutôt que sur l'identité. Elle souligne explicitement que tous

les acteurs (humains, non humains, biologiques, techniques et symboliques) sont tissés dans des tissus de relations co-constitutifs et co-créatifs.

La relationnalité ne sert donc pas seulement de facteur épistémique ou social parmi d'autres, mais de base fondamentale de

⁶ Une approche similaire est poursuivie par les travaux d'autres représentants de l'anthropologie multi-espèces, qui attirent l'attention sur les multiples façons dont les acteurs humains et non humains (champignons, arbres, animaux, microbes, etc.) sont interconnectés (voir Hastrup [éd.] 2015, Kohn 2013, Morizot 2022, Viveiros de Castro 2017).

L'existence elle-même. Dans ROMA, il n'existe pas d'entités distinctes (contrairement à d'autres modèles ontologiques), mais le monde est appréhendé entièrement comme une structure relationnelle, en constante évolution et émergente . Cela s'applique à l'existence en général, mais plus particulièrement à ce que nous entendons par vie. Dans ROMA, la vie ne peut être réduite à une forme spécifique de matérialité ou à un état binaire, mais est définie comme un processus relationnel progressif.

S'appuyant sur ROMA, FAIF applique spécifiquement ce concept à l'IA. Aujourd'hui déjà, les intelligences artificielles partagent certaines propriétés de la « vie » au sens large : elles participent à des processus intraactifs de cocréation et de coévolution, façonnent l'existence des acteurs avec lesquels elles interagissent et sont actives de multiples façons. Contrairement à Barad, qui décrit la vie comme une dynamique matérielle , ROMA la définit comme une propriété émergente des structures relationnelles. Il ne s'agit pas de l'un des deux états binaires (vivant ou non vivant), mais d'un processus graduel et dynamique de cocréations de réalités, d'où émergent non seulement la connaissance, mais aussi l'existence elle-même – sans distinction claire entre l'homme (ou la femme), la nature, la matière ou la technologie.

Cela nécessite un cadre éthique qui permette une col-Conception et contrôle de l'IA de manière collaborative, non hiérarchique et équitable. Compte tenu de l'effacement des frontières et de la reconnaissance des acteurs non humains – animaux, plantes, écosystèmes, objets, technologies et êtres spirituels – comme participants sur un pied d'égalité avec les humains dans les processus sociaux, culturels et écologiques, la FAIF met en avant les conclusions de l'anthropologie plus qu'humaine et du posthumanisme critique – en particulier leur critique de l'anthropocentrisme (Braidotti 2013, 2021 ; Ferrando 2013, 2020). De nombreuses raisons justifient le dépassement de l'idée moderniste de l'humain comme sujet rationnel et autonome. Rosi Braidotti soutenait que l'être humain rationnel et autonome au cœur de l'humanisme classique n'était pas seulement rationnel et autonome, il était également masculin et européen. Les femmes étaient exclues de cette définition, tout comme les membres des sociétés non occidentales et les êtres non humains. Ainsi, chaque fois que « nous, les humains » était évoqué, tout ce qui était non masculin, non européen, non humain était systématiquement marginalisé. Ou, comme le dit Braidotti : « nous ne sommes pas une seule et même personne » (Braidotti).

Au lieu d'un humanisme aussi exclusif, Braidotti défend une compréhension relationnelle de la subjectivité dans laquelle humains, animaux, technologie et environnement sont étroitement liés de multiples façons. Elle emprunte à Deleuze et Guattari un terme clé pour sa théorie posthumaniste du sujet – nomade – : « Le sujet nomade est un assemblage collectif, affectif et relationnel, matériellement ancré et incarné, point de relais d'un réseau de relations complexes qui déplacent la centralité des notions d'identité indexées sur l'ego » (Braidotti 2011). La FAIF et son concept de subjectivité comme processus rejettent l'exceptionnalisme humain et appellent à une éthique affirmative des enchevêtrements dans laquelle les soi non humains et leurs besoins sont pris en compte.

On peut observer dans de nombreuses sociétés non européennes à quoi peut ressembler une éthique non anthropocentrique des enchevêtrements. L'anthropologue français Philippe Descola, par exemple, décrit avec vivacité les schémas collectifs utilisés par les Achuar, un sous-groupe des Jivaros habitant la zone frontalière entre le Pérou et l'Équateur, pour structurer les relations entre humains et non-humains dans un ordre symbolique (Descola, 2013). Les modes d'identification que Descola distingue apportent des réponses à deux questions apparemment simples : Qui suis-je ? (respectivement : qu'est-ce qu'un humain ?) et qu'est-ce que l'autre ? (respectivement : que sont les non-humains ?). En tant que « moyen de spécifier les propriétés des êtres existants » (ibid.), répondre à ces questions permet de distinguer deux modes d'existence : un niveau intérieur (que Descola appelle : intériorité) et un niveau physique (extériorité, physicalité). L'intériorité désigne « la croyance universelle qu'un être possède des caractéristiques qui lui sont internes ou qui la prennent pour source » – en

En d'autres termes, une sorte de (auto-)conscience, soi ou âme – tandis que la physicalité fait référence à « la forme extérieure, la substance, les processus physiologiques, perceptifs et sensorimoteurs, voire la constitution d'un être et sa manière d'agir dans le monde, dans la mesure où ceux-ci reflètent l'influence exercée sur les modèles de comportement et un habitus par les humeurs corporelles, les régimes alimentaires, les caractéristiques anatomiques et les modes particuliers de reproduction » (ibid).

Contrairement à ce que suggère la vision humaniste, le seuil entre soi (ou l'humain) et l'autre (le non-humain) n'est en aucun cas universel. Il dépend plutôt de la personne à qui ou à quoi est attribuée ou refusée telle ou telle qualité d'intériorité et de physicalité. L'« autre » devient ainsi graduel et relatif.

Descola appelle les ruptures, ou plutôt les schémas de fracture, qui surviennent dans ce processus de différenciation des modes ou des schémas de création du monde.

Il n'y en a que quatre. Pour définir ce que signifie être humain, chaque société humaine connue distingue d'une manière ou d'une autre les qualités intérieures et extérieures (intériorité et physicalité). Celles-ci peuvent être différentes ou identiques à celles des autres êtres vivants. Les combinaisons possibles peuvent être représentées dans un simple tableau croisé. Descola a appelé les « ontologies » ou modes d'identification qui en résultent animisme, totémisme, analogisme et naturalisme (ibid.). Dans le contexte de cet article, il suffit d'examiner de plus près deux d'entre elles : le naturalisme et l'animisme (Tableau 3).

Les choses deviennent intéressantes lorsque nous nous demandons quelles similitudes et quelles dissemblances (et quelles possibilités de communication) apparaissent en relation avec les êtres vivants non humains. Les modernistes,

Tableau 3 Les quatre ontologies selon Descola (2013)

	Dissemblables en physique	Similaire dans Physicalité
Similaire en intériorité	Animisme	Totémisme
Dissemblables en intériorité	Analogisme	Naturalisme

Dans la tradition de Descartes, ils opéreraient évidemment pour le schéma naturaliste. Ils pourraient soutenir que, bien que tous les êtres vivants partagent une base physique commune, seuls les humains ont la capacité de penser de manière complexe, de formuler des jugements moraux et de créer une culture. Les animaux seraient dépourvus de cette intériorité complexe. Étant ainsi considérés comme dépourvus de conscience et d'intentionnalité, ils peuvent être rétrogradés au rang de ressources. Selon l'ontologie naturaliste, même les excès les plus cruels de l'élevage industriel ou de l'expérimentation animale peuvent être légitimés. Cependant, il existe des alternatives viables pour décrire la relation entre humains et non-humains.

Au cours de ses deux années de travail de terrain auprès des Achuar, Descola a remarqué que ses hôtes traitaient certains éléments de leur environnement (mais pas tous !) comme des personnes, « dotées de qualités cognitives, morales et sociales analogues à celles des humains, permettant ainsi la communication et l'interaction entre des classes d'êtres qui, à première vue, semblent très différentes » (ibid.). Quant à l'intériorité, les Achuar ne supposent pas de dissemblance fondamentale entre eux-mêmes et les membres d'autres groupes jivaros ou européens, et, par exemple, un jaguar ou un pécarí. S'ils diffèrent, c'est dans leur physicalité. Le fait que seuls les chamanes en transe parviennent à combler le fossé communicationnel entre les humains et, par exemple, un jaguar n'est pas dû au fait que les jaguars n'ont pas de soi, mais aux « processus physiologiques, perceptifs et sensorimoteurs » différents qui façonnent la « constitution et la manière d'agir d'un être dans le monde » (ibid.).

Il peut être tentant de rejeter la vision du monde achuar comme une simple façon exotique de déformer les relations « réelles » entre humains et non-humains, entre sujet et objet, entre culture et nature. Mais pouvons-nous vraiment être sûrs que nos idées naturalistes sur le monde sont « vraies » au sens non arbitraire du terme ? Et si notre vision du monde était déterminée par des schémas cognitifs qui créent un angle mort dans notre pensée ? J'ai déjà souligné le paradoxe de la modernité, la séparation non réalisée entre nature et société, et les conséquences dramatiques de la production constante de quasi-objets hybrides sur la vie sur notre planète.

Se pourrait-il que l'insistance sur une ontologie naturaliste marque précisément un tel angle mort ?

À première vue, cela peut paraître un ethnoromantisme naïf. Cependant, compte tenu des problèmes que l'Anthropocène pose à la vie sur notre planète, il serait tout simplement insensé d'ignorer l'opportunité offerte par une ouverture à des conceptions ontologiques alternatives du monde que nous habitons.

La FAIF appelle donc à une réévaluation ontologique ouverte des options de création de mondes. Si les progrès de l'IA se poursuivent à leur rythme actuel, nous pourrions bientôt être contraints de le faire de toute façon. Aujourd'hui, la plupart des gens hésiteraient probablement encore à affirmer que l'IA possède une « conscience ». Mais que se passerait-il si les années à venir produisaient une IAG dotée d'une intériorité comparable à celle de l'humain ? Dans le croisement des ontologies, nous pourrions soudainement nous retrouver dans l'animisme.

domaine, partageant un monde avec une nouvelle forme de conscience dont seule notre corporéité spécifique nous distingue. Ce n'est pas rien, mais cela ne cadrerait plus avec la croyance selon laquelle l'humain est le summum de la création.

5 FAIF – Un cadre d'IA féministe

L'idée générale d'une « IA centrée sur l'humain » est bien intentionnée, mais elle repose sur des prémisses erronées. L'IA centrée sur l'humain, telle qu'elle est pratiquée actuellement, est eurocentrique et androcentrique. Elle ne prend pas en compte la diversité culturelle d'un monde interconnecté à l'échelle mondiale ni les multiples façons dont les humains et les non-humains ont coévolué au fil du temps. Elle est guidée par une logique de marché froide et ignore les effets écologiques et sociaux néfastes d'une accumulation effrénée de pouvoir et de capital. Les appels à une « IA plus juste » se limitent généralement à lutter contre les symptômes des avancées non réglementées – ce que j'ai appelé (de manière quelque peu provocatrice, certes) les problèmes faciles de l'IA. Les problèmes difficiles – la pensée dichotomique, nos tentatives infructueuses de séparer l'homme et l'environnement, le sujet et l'objet, la nature et la culture – restent sans réponse.

À l'inverse, le Cadre d'IA féministe (FAIF) proposé ici s'inscrit dans la tradition de la critique féministe et se concentre particulièrement sur les problèmes complexes. Rejetant les fantasmes humains de supériorité et d'exceptionnalisme, il rassemble une multitude de perspectives, de visions du monde et d'ontologies autour d'une même table pour relever les défis de l'IA avec une « forte objectivité » (Harding 2024).

L'IA recèle un potentiel énorme pour façonner l'avenir de manière positive, non seulement celui des humains, mais aussi celui des autres êtres vivants. Nous disposons déjà aujourd'hui d'une multitude d'outils technologiques, utilisables pour répondre à des enjeux scientifiques et sociaux. Du diagnostic des maladies à la prévision météorologique, de la protection du climat à une meilleure accessibilité, de l'identification des fausses informations à l'étude de l'ADN et à la création de nouveaux types de protéines : dans tous ces domaines, l'IA est déjà extrêmement utile et nous aide à résoudre nos problèmes les plus urgents.

Si nous le faisons correctement, bien sûr. Le FAIF nous fournit un cadre épistémologique pour identifier les problèmes concrets et déterminer les acteurs à prendre en compte pour les résoudre. Son ancrage dans les études de genre nous aide à identifier les relations de pouvoir, les structures racistes et sexistes, et leur imbrication intersectionnelle, facilitant ainsi l'audit des biais intersectionnels dans les dimensions raciales, de genre et de classe, par exemple dans les algorithmes de recrutement et l'IA médicale. À travers un post-humanisme critique, le FAIF remet en question la centralité de l'humain et se concentre sur des modes d'existence hybrides, guidant la conception d'IA multi-espèces pour la conservation de la faune sauvage et l'agriculture durable. Avec un nouveau matérialisme, il met l'accent sur l'action des acteurs non humains et exige l'abandon d'une perspective exclusivement anthropocentrique, garantissant ainsi la matérialité.

Évaluation d'impact pour les systèmes de cloud computing et d'économie circulaire soucieux de leur empreinte carbone. En s'appuyant sur des ontologies autochtones, elle remet en question l'anthropocentrisme rationaliste de la production de connaissances naturalistes. Parallèlement, elle promeut le contrôle communautaire des données dans le cadre de projets soutenant des objectifs tels que la préservation des langues et la surveillance écologique traditionnelle. Grâce à une anthropologie plus qu'humaine, elle fait progresser notre compréhension de la diversité des interconnectivités entre espèces tout en formulant des recommandations culturellement sensibles. Enfin, le lien de FAIF avec la phénoménologie du corps vivant attire l'attention sur la corporéité de toute expérience du monde et « réancré » le sujet dans le monde physique. Parallèlement, elle attire notre attention sur des domaines qui (pour l'instant) restent fermés à la froide logique rationaliste de l'IA : la perception corporelle de son existence et l'expérience individuelle (et la création) d'un « environnement » particulier (Uexküll 2023). Cette dimension phénoménologique garantit que le développement de l'IA reconnaît la nature incarnée de l'expérience humaine et remet en question les approches purement informatiques de la compréhension et de la représentation de soi et de la réalité.

Si notre objectif est véritablement une IA plus équitable et inclusive, nous devons intégrer ces approches à nos processus de développement. La FAIF appelle donc à une réévaluation multi-perspectives de la relation entre humains et technologie, qui pose la question, à chaque étape du processus de développement : qu'implique cette étape au-delà des logiques du « ce serait cool si... » (Hollis 2023) et du « aller vite et casser des choses » (Baer).

2014) Tableau 4. ⁷

Le mathématicien John von Neumann, l'un des pionniers de l'informatique moderne, s'est un jour interrogé sur les conséquences du développement des nouvelles technologies pour l'humanité. Il voyait un réel danger pour la survie de la civilisation humaine dans le fait que les gens pourraient tout simplement se désintéresser de leur environnement : « Les intérêts de l'humanité pourraient changer, la curiosité scientifique actuelle pourrait cesser, et des choses entièrement différentes pourraient occuper l'esprit humain à l'avenir » (Usam 1958). Un autre mathématicien, Michael Atiyah, a soulevé la question de savoir si les mondes des êtres vivants ayant des horizons d'expérience physique différents avaient au moins un fondement logique commun. Pendant longtemps, le réalisme mathématique a tenu pour acquis l'universalité des vérités mathématiques : « Quand deux dinosaures ont rencontré deux dinosaures...

Dans les saurs, il y avait quatre dinosaures. Dans ce tableau préhistorique, « $2+2=4$ » était fidèlement modélisé par les animaux, même s'ils étaient trop stupides pour le savoir et même si aucun humain n'était présent pour l'observer » (Gardner 1981). C'est peut-être vrai.

Mais que se passerait-il, demande Atiyah, si l'intelligence supérieure ne s'était pas développée chez les humains, mais chez des êtres vivants qui ont des expériences corporelles complètement différentes et dont la vie se déroule dans

Des environnements complètement différents ? Chez une méduse, par exemple, vivant dans les profondeurs de l'océan Pacifique. « Tout dans son expérience serait continu, du flux de l'eau environnante aux fluctuations de température et de pression. Dans un tel environnement, dépourvu d'objets individuels ou de toute autre chose distincte, le concept de nombre apparaîtrait-il ? S'il n'y avait rien à compter, les nombres existeraient-ils ? »

(Attiyah 1995).

Toute connaissance authentique du monde repose sur la perception. Les corps vivants ne sont pas des réceptacles contenant un esprit ; ils constituent et façonnent l'environnement même dans lequel un être évolue et auquel il participe par ses actions. S'il est impossible de prédire quel sera l'environnement d'une intelligence artificielle générale, une chose est sûre : toute véritable IAG représentera nécessairement une forme d'intelligence incarnée.

L'enjeu est on ne peut plus important. Sommes-nous vraiment prêts à nous soumettre aux règles d'une intelligence artificielle générale dont nous ne pouvons même pas prédire le comportement ? Six ans et demi avant que ChatGPT ne bouleverse le monde, nous avons eu un aperçu de ce que pourrait être l'interaction avec une intelligence supérieure.

Le 10 mars 2016, au 37^e coup de sa quatrième partie contre Lee Sedol, l'un des meilleurs joueurs de Go au monde, le programme informatique AlphaGo a surpris tous les experts. Ce coup ressemblait à une erreur. D'un point de vue humain, il n'avait aucun sens. Ce n'est qu'après une analyse rétrospective de la partie qu'il s'est avéré décisif pour la victoire de la machine sur son adversaire humain. AlphaGo avait développé de manière autonome une stratégie que ses créateurs pouvaient observer, mais qu'ils ne pouvaient pas reconstituer par rétro-ingénierie. Efficace, elle manquait de transparence.

Aujourd'hui déjà, les intelligences artificielles peuvent imiter le comportement humain à un point qui était longtemps considéré comme impossible.

Ils gagnent aux jeux de stratégie, négocient des actions, composent de la musique dans le style de Bach ou de Nirvana, créent des œuvres d'art d'une simple pression sur un bouton, écrivent des textes, communiquent et servent de compagnons et de partenaires amoureux. Presque tout ce qui a été perçu comme véritablement humain pendant des millénaires en est exempt – et dans bien des cas, les résultats générés par l'IA sont au moins équivalents à ceux de l'activité humaine.

Après les trois blessures narcissiques (Freud 1917) de l'héliocentrisme, de l'évolution et de la psychanalyse, voici un nouveau coup porté à l'exceptionnalisme humain. Cela peut assurément être positif. Il faut cependant rester prudent.

Car, que la singularité survienne l'année prochaine ou dans quelques années, elle nous impose des changements auxquels il vaut mieux réfléchir en amont plutôt qu'après coup, comme les ingénieurs d'AlphaGo. Le FAIF nous en donne les moyens.

⁷ Jusqu'en 2014, « bouger vite et casser des choses » était la devise semi-officielle de Facebook.

Tableau 4 Perspectives et atouts du FAIF

Perspectives et atouts du FAIF	
Perspective	Points forts
Études de genre	identifie systématiquement les structures d'exclusion, les préjugés, les racismes et les sexismes dans une perspective intersectionnelle ; analyse les structures de pouvoir et de domination
Posthumanisme critique	remet en question la centralité de l'humain et souligne l'importance des modes d'existence hybrides
Le nouveau matérialisme	décentralise l'humain et met l'accent sur l'action des acteurs non humains
Indigène Ontologies	rejette l'universalité de l'ontologie naturaliste et son anthropocentrisme ; inclut les formes de connaissance non rationalistes
Plus que-Anthropologie humaine	analyse les modes d'identification et de relations et renégocie la question de ce que signifie être « humain »
Phénoménologie corporelle	met l'accent sur la corporéité de l'expérience et de la connaissance ; repositionne le sujet (émergent) dans le monde

6 Un appel au changement radical

La FAIF critique les tentatives actuelles d'IA centrée sur l'humain, affirmant que si nous voulons réellement une IA plus juste pour tous, l'IAHC ne suffit pas. Les dernières avancées technologiques et politiques ont clairement démontré la nécessité de changer radicalement notre approche de l'IA. L'accumulation apparemment illimitée d'argent et de pouvoir entre les mains de quelques-uns n'apporte rien à la plupart des gens ni à notre environnement. La régulation et la gouvernance sont donc indispensables.

Mais même si nous prenons en compte les besoins du plus grand nombre dans nos efforts de réglementation, même si nous tentons de représenter les individus dans toute leur diversité, se concentrer uniquement sur les humains pose toujours un problème majeur. Si cela peut atténuer certaines des conséquences les plus drastiques de l'eurocentrisme, de l'androcentrisme et du colonialisme numérique pour améliorer durablement les conditions de vie du plus grand nombre d'acteurs (humains et non humains), il ne suffit pas de raisonner dans le cadre techno-optimiste de la modernité naturaliste. Il est au contraire nécessaire de réorienter fondamentalement nos façons de penser et d'agir.

Le Cadre d'IA féministe (FAIF) appelle donc à une réévaluation radicale de la relation entre les humains, l'environnement et la technologie. L'existence humaine ne peut plus être le centre « naturel » de toutes les considérations.

Il faut plutôt le considérer comme une partie d'une corne d'abondance d'enchevêtrements rhizomatiques, dans lesquels les agences des acteurs humains et non humains sont négociées et renégociées dans des contextes d'action concrets.

L'intégration du FAIF dans le développement et la mise en œuvre de l'IA dans les institutions de recherche, la technologie

L'intégration des entreprises et des institutions publiques, ainsi que des cadres réglementaires tels que la loi européenne sur l'IA, nécessite à la fois des méthodologies de gouvernance innovantes et une transformation culturelle fondamentale dans la manière dont nous conceptualisons la relation entre la technologie, la société et la justice.

Le « progrès » ne doit pas nécessairement être orienté vers le profit et vers tout ce qui est technologiquement faisable.

Le progrès, interprété en termes féministes, peut aussi signifier apprécier les liens, les interactions et les dépendances de chacun avec tout le reste, et assumer sa responsabilité envers les autres et le monde. Même si cela implique de sortir de sa zone de confort. Le FAIF est un premier pas dans cette direction.

Contribution de l'auteur TK auteur unique.

Financement Financement en libre accès rendu possible et organisé par Projekt DEAL.

Disponibilité des données Aucun ensemble de données n'a été généré ou analysé au cours de l'étude actuelle.

Déclarations

Conflit d'intérêts Les auteurs ne déclarent aucun conflit d'intérêts.

Libre accès. Cet article est sous licence Creative Commons Attribution 4.0 International, qui autorise l'utilisation, le partage, l'adaptation, la distribution et la reproduction sur tout support et sous tout format, à condition de citer l'auteur(e) original(e) et la source, de fournir un lien vers la licence Creative Commons et d'indiquer les éventuelles modifications. Les images et autres éléments tiers de cet article sont inclus dans la licence Creative Commons de l'article, sauf mention contraire dans une ligne de crédit. Si un élément n'est pas inclus dans la licence Creative Commons de l'article et que l'utilisation que vous envisagez n'est pas autorisée par la réglementation ou dépasse les limites autorisées, vous devrez obtenir l'autorisation directement auprès du détenteur des droits d'auteur. Pour consulter une copie de cette licence, rendez-vous sur <http://creativecommons.org/licenses/by/4.0/>.

Références

AJL (2024) Ligue de justice algorithmique. <https://www.ajl.org>

Atiyah MF (1995) Critique du livre Conversations on Mind, Matter and Mathematics par Jean-Pierre Changeux et Alain Connes, dans : Times Higher Educational Supplement, 29 septembre 1995.

Baer D (2014) Mark Zuckerberg explique pourquoi Facebook ne « va plus vite et ne casse plus rien ». Business Insider, 03/05/2014. <https://www.businessinsider.com/mark-zuckerberg-on-facebooks-new-motto-2014-5>

Barad K (2007), À mi-chemin de l'univers : la physique quantique et l'intrication de la matière et du sens. Duke University Press, Durham et Londres.

Bateson G (1985) Ökologie des Geistes : Perspectives anthropologique, psychologique, biologique et épistémologique. Suhrkamp, Francfort

Batool A, Zowghi D, Bano M (2025) Gouvernance de l'IA : une revue systématique de la littérature. Éthique de l'IA. <https://doi.org/10.1007/s43681-024-00653-w>

- Benjamin R (2019) *La course après la technologie : des outils abolitionnistes pour le nouveau Jim Code*. Polity Press, Cambridge
- Boström N (2016) *Superintelligence : chemins, dangers, stratégies*. Oxford University Press, Oxford
- Braidotti R (2021), *Féminisme posthumain*. Polity Press, Cambridge
- Braidotti R (2019), *Connaissances posthumaines*. Polity Press, Cambridge
- Braidotti R. (2013), *Le posthumain*. Polity Press, Cambridge
- Braidotti R (2011), *Sujets nomades*. Columbia University Press, New York
- Braidotti R, Lykke N (1996) *Entre monstres, déesses et cyborgs : confrontations féministes avec la science, la médecine et le cyberspace*. Zed Books, Londres.
- Buolamwini J (2024) *Démasquer l'IA : ma mission : protéger l'humain dans un monde de machines*. Random House, New York.
- Buolamwini J, Gebru T (2018) Nuances de genre : disparités de précision intersectionnelles dans la classification commerciale par genre. *Actes de Machine Learning Research* 81 : 1–15
- Chalmers DJ (1995) Face au problème de la conscience. *J Con-scious Stud* 2 : 200–219
- Costanza-Chock S (2020) *Design Justice : Pratiques communautaires pour construire le monde dont nous avons besoin*. MIT Press, Cambridge
- Därmann I (2014) *Kraft der Dinge : Phänomenologische Skizzen*. Pad-erborn : Wilhelm Fink.
- Descola P. (2024) *La composition des mondes : entretiens avec Pierre Charbonnier*. Polity Press, Cambridge.
- Descola P. (2013) *Au-delà de la nature et de la culture*. Presses universitaires de Chicago, Chicago
- DIN (2022) *Zweite Ausgabe der Deutschen Normungsroadmap KI*. <https://www.din.de/forschung-und-innovation/themen/kuens-tliche-intelligenz/fahrplan-festlegen>
- DIN, DKE (2022) : *Deutsche Normungsroadmap Künstliche Intel-ligenz (Ausgabe 2)*. <https://www.din.de/resource/blob/916792/20bf33d405710a703aa26f81362493bb/nrm-ki-deutsch-2022-final-web-250-data.pdf>
- Engster F, Moore PV (2020) La recherche de l'intelligence (artificielle) dans le capitalisme. *Cap Class* 44(2):201. <https://doi.org/10.1177/0309816820902055>
- EUR-lex (2024) *Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées en matière d'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (loi sur l'intelligence artificielle)*
- (Texte présentant de l'intérêt pour l'EEE). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=EUR-Lex:32024R1689>
- Ferrando F (2020) *Posthumanisme philosophique*. Bloomsbury, Londres
- Ferrando F (2013) : *Posthumanisme, transhumanisme, antihumanisme, métahumanisme et nouveaux matérialismes : différences et relations*. *Existenz* 8/2 (2013), 26–32.
- Foucault M (1981), *L'Ordre du discours*. Dans : Young R (éd.), *Délier le texte : Un lecteur post-structuraliste*. Routledge, Kegan & Paul, Boston, p. 51–78.
- Fox, NJ (2023) *Capitalisme et État : une nouvelle perspective matérialiste*. *Eur J Soc Theory* 28(1) : 126. <https://doi.org/10.1177/13684310231204540>
- Fox-Keller E (1986) *Liebe, Macht und Erkenntnis. Männliche ou weibliche Wissenschaft?* Carl Hanser, Munich
- Freud S (1917) Une difficulté de la psychanalyse. *Revue Imago pour l'application de la psychanalyse aux sciences humaines*, V : 1–7
- Gardner M (1981) *Les mathématiques sont-elles réelles ?* *New York Review*, 13. Août 1981. <https://www.nybooks.com/articles/1981/08/13/is-mathematics-for-real/>
- Gray ML, Suri S (2019) *Travail fantôme : comment empêcher la Silicon Valley de créer une nouvelle sous-classe mondiale*. Houghton Mifflin Harcourt, Boston.
- Guangyin J, Xiaohan N, Kun W, Jie Z, Haoming Z, Leiming J (2025) La singularité technologique sera-t-elle pour bientôt ? Modélisation de la dynamique de développement de l'intelligence artificielle via un processus de croissance multi-logistique. *Physica A* 664(15) : 130450. <https://doi.org/10.1016/j.physa.2025.130450>
- Haraway D, Haraway D (1997), *Modest_Witness@Second_Millennium.FemaleMan@_meets Oncomouse*. *Modest_Witness@Second_Millennium.FemaleMan@_meets Oncomouse : féminisme et technoscience*. Routledge, New York, p. 21–121.
- Haraway D (1992), *Les promesses des monstres : une politique régénératrice pour les autres inappropriés*. Dans Grossberg L, Nelson C, Treichler, PA (dir.) (1992), *Cultural Studies*, New York : Routledge, p. 295–337. https://monoskop.org/File:Haraway_Donna_1992_Les_promesses_des_monstres_Une_politique_régénératrice_pour_Inappropriated_Others.pdf
- Haraway D (1991), *Simiens, cyborgs et femmes : la réinvention de la nature*. Routledge, New York.
- Haraway D (1988) *Savoirs situés : la question scientifique dans le féminisme et le privilège de la perspective partielle*. *Femin Stud* 14 (n° 3 (automne 1988)) : 575–599. <https://doi.org/10.2307/3178066>
- Harding S (2024) *Nouvelle théorie du point de vue : Était-ce un « objectif star » ?* Dans : Hoppe, Katharina ; Vogelmann, Frieder. *Épistémologie féministe : un lecteur | Eine kompakte Einfüh-rung in die feministische Erkenntnis- und Wissenschaftstheorie* (p. 305). (Fonction). Édition Kindle. ? dans : Hoppe, K, Vogelmann F (2024) *Épistémologies féministes : un lecteur*. Suhrkamp, Berlin, pp. 305-347.
- Harding S (2006), *Science et inégalités sociales : enjeux féministes et postcoloniaux*. University of Illinois Press, Urbana.
- Harding S (1998), *La science est-elle multiculturelle ? Postcolonialismes, féminismes et épistémologies*. Indiana University Press, Bloomington
- Harding S (1991), *À qui appartient la science ? À qui appartient le savoir ? Réflexions à partir de la vie des femmes*. Cornell University Press, Ithaca.
- Harding S (1986) *La question scientifique dans le féminisme : la politique industrielle en Europe*. Cornell University Press, Ithaca
- Hartsock NCM (1983), *Le point de vue féministe : préparer le terrain pour un matérialisme historique spécifiquement féministe*. Dans : Harding S, Hintikka MB (dir.), *À la découverte de la réalité : Perspectives féministes sur l'épistémologie, la métaphysique, la méthodologie et la philosophie des sciences*. Kluwer Academic Publishers, Dordrecht, p. 283-310.
- Hastrup, K. (éd.) (2015), *Anthropologie et nature*. New York : Rout-ledge ; Kirksey, E. (2015), *Écologies émergentes*. Duke University Press, Durham
- Heisenberg W (1958), *Physique et philosophie : la révolution de la science moderne*. Harper & Brothers Publishers, New York.
- Holbraad M, Pedersen MA (2017), *Le tournant ontologique : une exposition anthropologique*. Cambridge University Press, Cambridge
- Hollis B (2023) « Ce serait génial, non ? » et autres mauvaises approches de conception. NDC Londres 2023, <https://www.youtube.com/regarder?v=GGUqyb6mzDw>
- ISO (2023) *Technologies de l'information — Intelligence artificielle — Système de management*. SO/IEC 42001:2023. <https://www.iso.org/stand/81230.html>
- Kneer G, Schroer M, Schüttelpe E (2008) *Bruno Latours Kollektive : Kontroversen zur Entgrenzung des Sozialen*. Suhrkamp, Francfort
- Kohn E (2013), *Comment les forêts pensent : Vers une anthropologie au-delà de l'humain*. University of California Press, Berkeley.
- Kubes T, Reinhardt T (2022) *Techno-espèces en devenir : vers une ontologie relationnelle des assemblages multi-espèces (ROMA)*. *Nan-oEthics* 16 : 95–105. <https://doi.org/10.1007/s11569-021-00401-y>
- Latour B (2001) *Das Parlament der Dinge*. Suhrkamp, Francfort
- Latour B (1993) *Nous n'avons jamais été modernes*. Harvard University Press, Cambridge

- Magee CL, Tessaleno CD (2011) Combien de singularités sont proches et comment vont-elles perturber l'histoire humaine ? Dans ; Prévisions technologiques et changement social 78(8) : 1365–1378. <https://doi.org/10.1016/j.techfore.2011.07.013>
- Moore PV (2020) Le miroir de l'intelligence (artificielle) dans le capitalisme. Classe de casquettes 200. <https://doi.org/10.1177/0309816820902040>
- Morizot B (2022) Philosophie der Wildnis – ou – Die Kunst, vom Weg abzukommen. Reclam, Ditzingen
- Nagel T (2018) À quoi ça ressemble d'être une chauve-souris ? À quoi ça ressemble d'être une chauve-souris ? Reclam, Stuttgart
- Ogden LA, Hall B, Tanita K (2013) Animaux, plantes, hommes et choses : une analyse ethnographique multispécifique. Environ Soc. <https://doi.org/10.3167/ares.2013.040102>
- Oudshoorn N, Rommes E, Stienstra M (2004) Configurer l'utilisateur comme tout le monde : genre et cultures de conception dans les technologies de l'information et de la communication. Sci Technol Hum Values 29(1) : 30–63. <https://doi.org/10.1177/0162243903259190>
- Reuters (2025) : Le conseil d'administration d'OpenAI rejette l'offre de 97,4 milliards de dollars de Musk. 15 février 2025 : <https://www.reuters.com/technology/openai-board-rejects-musks-974-billion-offer-2025-02-14/>
- Schnetker MFJ (2019) Mythologie transhumaniste : utopies justes d'une rédemption technologique par l'intelligence artificielle. Unrast Verlag, Münster.
- Shams RA, Zowghi D, Bano M (2025) L'IA et la quête de diversité et d'inclusion : une revue systématique de la littérature. AI Ethics 5 : 411–438. <https://doi.org/10.1007/s43681-023-00362-w>
- Singer N (2018) La reconnaissance faciale d'Amazon identifie à tort 28 législateurs, selon l'ACLU. New York Times, 26 juillet 2018. <https://www.nytimes.com/2018/07/26/technology/amazon-aclu-facial-recognition-congress.html#:~:text=In%20the%20test%2C%20%20Amazon,les%20personnes%20sur%20les%20photos%20de%20mug>
- Suchman L. (2007) Reconfigurations homme-machine. Plans et actions situées. Cambridge University Press, Cambridge
- Thompson D'AW (2016) Über Wachstum und Form. Suhrkamp, Frankfurt
- Tsing AL (2015) Le Champignon du bout du monde : Sur la possibilité de vivre dans les ruines du capitalisme. Princeton University Press, Princeton.
- Tsing AL, Deger J, Saxena AK, Zhou F (2024) Guide de terrain sur l'anthropocène fragmenté. Stanford University Press, Stanford
- Uexküll J (2023) Streifzüge durch die Umwelten von Tieren und Men-schen : Eine Bedeutungslehre. Matthes & Seitz, Berlin
- ONU (2024) Gouverner l'IA pour l'humanité : Rapport final. Publications des Nations Unies, New York. https://www.un.org/sites/un2.un.org/fichiers/governing_ai_for_humanity_final_report_fr.pdf
- UNRIC (2024) Secrétariat général de l'ONU pour les connaissances intellectuelles. UNRIC - Regionales Informationszentrum der Vereinten Nationen. <https://unric.org/de/un-wollen-eigenes-sekretariat-fuer-kuenstliche-intelligenz/>
- Usam S (1958) John von Neumann 1903–1957. Bulletin de l'American Mathematical Society. 64, n° 3. Partie 2 : 1–49
- Vallor S (2024) Le miroir de l'IA. Comment retrouver notre humanité à l'ère de la pensée automatique. Oxford University Press, New York
- Viveiros de Castro E (2017) Métaphysique cannibale. Presses de l'Université de Minnesota, Minneapolis
- Wajcman J (2004) Technoféminisme. Polity Press, Cambridge
- Weber M (1985) Die 'Objektivität' sozialwissenschaftlicher und sozi-alpolitischer Erkenntnis. Dans : Winkelmann, J (éd.) Max Weber : Gesammelte Aufsätze zur Wissenschaftslehre. Mohr, Tübingen. <http://www.zeno.org/nid/20011440120>
- Young JC (2019) La nouvelle politique de la connaissance du colonialisme numérique. Environ Plan A. <https://doi.org/10.1177/0308518X19858998>
- Zweig KA, Krafft TD, Klingel A, Park E (2021) Socioinformatique : un nouveau regard sur l'informatique et la société. Carl Hanser, Munich
- Note de l'éditeur Springer Nature reste neutre en ce qui concerne les revendications juridictionnelles dans les cartes publiées et les affiliations institutionnelles.