

Synthèse globale des discussions de la 1^{ère} vague

Synthèse globale transversale

En 2023, des discussions sur ce que signifie être humain à l'ère des neurosciences (NS) et de l'IA ont été facilitées par les partenaires de NHNAI dans 9 pays différents. Dans chaque pays, trois lignes de discussion ont été ouvertes pour explorer cette question dans les **3 domaines thématiques de l'éducation, de la santé et de la démocratie**. Chaque partenaire a ensuite produit **3 synthèses locales** rendant compte du contenu des discussions dans ces 3 domaines dans les pays correspondants¹. Sur cette base, l'équipe de coordination a proposé **3 synthèses thématiques globales** (une par champ exploré, éducation, santé et démocratie). Enfin, les idées de ces 3 synthèses thématiques globales ont été regroupées pour générer **une synthèse globale transversale**, rassemblant des idées plus générales qui ont été exprimées dans différents champs thématiques.

Ce document présente les **idées de la synthèse globale transversale**, ainsi que les nœuds dans lesquels certaines idées émergeant des discussions entrent en conflit et en tension, manifestant les complexités possibles et les points délicats des questions transversales.



¹ Pour un total exact de 8*3 + 2 synthèses locales. Au Canada (Québec), le Cégep de Sainte-Foy a organisé des discussions sur la démocratie et l'éducation, mais pas sur la santé.

Table des matières

Partie 1: Idées globales transversales.....	3
S'appuyer sur les technologies pour améliorer notre vie.....	3
Préserver l'autonomie et l'agentivité humaine	3
Fixer des limites et des règles malgré les défis.....	4
Préserver la responsabilité humaine (seul l'humain peut être moralement responsable)	5
Respecter la singularité et la diversité (culturelle) des personnes	5
Préserver l'empathie, le contact humain et les relations humaines	6
Chercher à s'améliorer.....	6
Préserver et intensifier ce qui fait de nous des êtres humains et favoriser l'épanouissement humain	7
Favoriser la culture scientifique et technique ainsi que la culture éthique et la pensée critique	8
S'assurer que la technologie réduise (plutôt qu'augmente) les inégalités.....	8
Privilégier la coopération entre l'IA et l'humain plutôt que le remplacement	9
Protéger la vie privée	9
Partie 2: Nœuds de complexité globaux-transversaux.....	10
Quel lien, quelles interactions avec les machines ?	10
Fixer des limites et des règles malgré les défis.....	14
Devenir plus efficace sans menacer l'essentiel de ce qui fait de nous des êtres humains	15

Partie 1: Idées globales transversales

Être humain au temps des NS et de l'IA signifie ...

S'appuyer sur les technologies pour améliorer notre vie

Les développements de l'IA et des NS au cours des dernières décennies ont ouvert la voie à divers processus technologiques qui (ont un fort potentiel pour) améliorer la vie humaine.

- L'automatisation des tâches fastidieuses peut améliorer nos vies, notamment en nous permettant de gagner du temps pour des activités plus essentielles telles que les relations ou tout ce qui favorise l'épanouissement humain.
- Les technologies de l'IA peuvent aider les humains à prendre des décisions (et même être plus performantes dans certaines tâches). Elles peuvent nous aider à prévenir ou à gérer divers problèmes et crises (assurer une meilleure sécurité dans l'espace public grâce à une surveillance plus efficace, anticiper les épidémies ou les aléas du temps et le changement climatique, ...).
- Les résultats de l'IA et des NS peuvent nous permettre de renforcer nos capacités physiques et mentales, d'améliorer nos performances et notre efficacité. Ils peuvent également soutenir les personnes les plus vulnérables et les plus exclues (en facilitant l'accès aux services et à l'information, en renforçant l'autonomie des personnes handicapées et en permettant de faire face aux problèmes liés au vieillissement).
- Enfin, l'IA et les NS peuvent améliorer nos vies en enrichissant et en affinant notre compréhension de nous-mêmes en tant qu'êtres humains.

Implication dans des nœuds de complexité (voir ci-dessous [Partie 2 : Nœuds de complexité globaux-transversaux](#)) :

- [Devenir plus efficace sans menacer l'essentiel de ce qui fait de nous des êtres humains](#)

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Santé) Reconnaître la contribution positive des technologies de la santé aux soins de santé 5 pays (BE, CH, KE, PT, TW) 16 idées
- (Global – Santé) Explorer les contributions potentielles des technologies de la santé à l'amélioration de la condition humaine 2 pays (FR, PT) 2 idées
- (Global – Santé) Utiliser les technologies de la santé pour améliorer les conditions de vie des personnes les plus vulnérables 1 pays (KE) 4 idées
- (Global – Education) Utiliser l'IA pour améliorer les performances et l'innovation 4 pays (CH, PT, TW, BE) 5 idées
- (Global – Démocratie) Reconnaître l'impact positif (potentiel) de l'IA sur la vie humaine tout en posant les bonnes questions 4 pays (BE, FR, KE, PT) 6 idées
- (Global – Démocratie) Utiliser l'IA pour assurer la sûreté/sécurité 2 pays (CA, KE) 2 idées
- (Global – Education) Utiliser l'IA pour libérer l'homme de son travail 3 pays (PT, KE, FR) 4 idées

Préserver l'autonomie et l'agentivité humaine

(notamment dans la prise de décision et les capacités cognitives (collectives))

Une trop grande confiance dans les technologies de l'IA peut entraîner une déqualification et un appauvrissement cognitif, une dépendance excessive et une perte de résilience en cas d'indisponibilité des technologies.

Il existe également un risque de réduire les personnes à leurs données, lié à une focalisation excessive sur ce qui peut être mesuré et quantifié. Le danger de systèmes automatisés prescriptifs et coercitifs apparaîtrait alors, notamment avec la surveillance et la gouvernance algorithmique (mais aussi dans tout domaine où la prise de décision ayant un impact profond sur les personnes pourrait être automatisée, comme en médecine).

Enfin, la question de l'éditorialisation automatisée de l'information devrait également être prise en compte. Bien que nous ayons besoin d'algorithmes puissants pour organiser l'information pour nous (moteur de recherche, algorithme de recommandation sur les réseaux sociaux et autres plateformes), cette automatisation peut conduire à des bulles informationnelles ou cognitives isolant les individus dans des paysages informationnels uniformes (un problème renforcé par la facilitation générative par l'IA de la production de (deep) fake news). Ainsi, les technologies de l'IA peuvent menacer profondément notre intelligence (collective).

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Santé) Préserver l'agentivité et l'autonomie humaine 5 pays (BE, FR, KE, TW, USA) 8 idées
- (Global – Education) Préserver l'autonomie humaine 7 pays (FR, CA, IT, CH, TW, USA, BE) 8 idées
- (Global – Education) Préserver les besoins fondamentaux nécessaires au développement humain 7 pays (BE, CA, FR, TW, PT, CH, KE) 11 idées
- (Global – Démocratie) Empêcher l'IA de saper la pensée critique, les capacités de prise de décision et l'intelligence collective des humains 7 pays (CH, FR, IT, KE, PT, TW, USA) 18 idées
- (Global – Education) Une utilisation excessive de l'IA qui conduit à un appauvrissement cognitif 4 pays (FR, CA, BE, TW) 6 idées

Fixer des limites et des règles malgré les défis

Il existe un fort besoin de réglementation et de normes pour garantir que les technologies de l'IA et des NS produisent des résultats positifs. Les normes et la réglementation sont essentielles à l'instauration de la confiance et à la protection des personnes lors du déploiement de nouvelles technologies. L'IA devrait respecter les valeurs humaines (équité, impartialité, ...) et être centrée sur l'être humain (visant à son épanouissement). L'IA et les technologies des NS devraient bénéficier à tous (il est crucial de lutter contre l'exclusion des personnes pauvres et vulnérables).

Cependant, la réglementation soulève de nombreuses questions aiguës, ce qui en fait un défi très difficile à relever. Parmi ces questions, on peut évoquer le rythme du développement technologique, l'obscurcissement des modèles de responsabilité (avec les technologies numériques en général et plus particulièrement avec l'apprentissage automatique), l'accès souvent « facile » à des outils puissants (dans la main d'acteurs mal intentionnés, des technologies telles que la reconnaissance d'images ou de visages peuvent devenir extrêmement nuisibles), l'échelle mondiale de la recherche et du développement (avec la diversité des systèmes de valeurs dans le monde ainsi que des constellations de conflits

d'intérêts), la difficulté d'appliquer les réglementations (dans un contexte aussi diversifié et international).

D'une manière générale, la réglementation devrait favoriser une utilisation raisonnée et saine des technologies de l'IA et des NS. Néanmoins, il peut s'avérer extrêmement difficile d'identifier ce qui est raisonné et sain et ce qui ne l'est pas (par exemple, dans le cas de la modération des médias sociaux : qui est le bon acteur ? Ou le cas des technologies de la santé avec des zones grises entre les utilisations curatives et d'amélioration : qui peut décider si une pathologie nécessite / justifie l'utilisation d'une technologie de la santé donnée ?) Les parties prenantes, les professionnels, les citoyens et les acteurs économiques/industriels devraient être impliqués dans les processus de régulation.

Implication dans des nœuds de complexité (voir ci-dessous [Partie 2 : Nexus de complexité globaux-transversaux](#)): cette idée constitue à elle seule un nœud de complexité.

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Santé) Réguler l'IA et les technologies de la santé dans les soins de santé 4 pays (CH, IT, PT, USA) 6 idées
- (Global – Démocratie) Être conscient des défis posés par la régulation 3 pays (FR, PT, USA) 3 idées
- (Global – Santé) Être conscient des défis posés par la régulation 2 pays (PT, TW) 2 idées
- (Global – Santé) Limiter l'utilisation des technologies d'augmentation/amélioration de la santé 4 pays (CH, FR, IT, PT) 7 idées
- (Global – Démocratie) Reconnaître l'impact positif (potentiel) de l'IA sur la vie humaine tout en posant les bonnes questions 4 pays (BE, FR, KE, PT) 6 idées
- (Global – Education) Nécessité d'un plus grand nombre de mesures de régulation 4 pays (TW, USA, KE, CH) 5 idées
- (Global – Démocratie) Fixer des limites, contrôler et réguler l'IA pour préserver la démocratie 8 pays (BE, CA, CH, IT, KE, PT, TW, USA) 16 idées

Préserver la responsabilité humaine (seul l'humain peut être moralement responsable)

Seuls les êtres humains, grâce à leur conscience et à leur esprit critique, sont capables de faire des choix éthiques et de prendre des décisions responsables. Les humains sont donc les seuls responsables des orientations technologiques et des conséquences de l'utilisation de l'IA. Sauf dans certains sens juridiques spécifiques (responsabilité des entreprises, personnalité juridique permettant par exemple une compensation monétaire), la responsabilité morale, éthique, juridique et politique (et pénale) ne peut jamais être attribuée aux machines. La dilution et l'obscurcissement des chaînes de responsabilité sont hautement problématiques.

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Santé) Ne jamais croire que nous pouvons déléguer la responsabilité (morale) aux machines 5 pays (BE, IT, PT, TW, USA) 7 idées
- (Global – Démocratie) Préserver la responsabilité humaine dans les choix/décisions éthiques 4 pays (BE, CA, FR, IT) 7 idées

Respecter la singularité et la diversité (culturelle) des personnes

Les personnes doivent être reconnues comme des êtres singuliers et traitées en conséquence, d'une manière globale, en rendant justice à leur diversité. Il faut s'opposer à toute réduction des personnes à des aspects mesurables et quantifiables (ou à ce qui peut être pris en compte et traité par des moyens technologiques) au détriment de la reconnaissance des expériences et des sentiments des personnes. Ces aspects n'épuisent pas ce que sont les personnes humaines et ce que l'on peut dire d'elles de manière significative. Les humains sont tous différents, avec des spiritualités différentes, et cette différence est une richesse pour l'humanité que nous devrions préserver de la menace d'unification que l'IA et les NS peuvent apporter.

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Santé) Reconnaître les patients dans leur singularité et leur diversité (dans le cadre d'une approche globale) 3 pays (BE, CH, FR) 5 idées
- (Global – Démocratie) Reconnaître que les personnes humaines dépassent les seules dimensions mesurables 2 pays (CA, PT) 2 idées
- (Global – Education) Considérer la diversité culturelle et la singularité humaine 4 pays (FR, PT, KE, TW) 9 idées
- (Global – Santé) Résister à la surévaluation de la performance, de l'efficacité ou de la productivité 4 pays (CH, FR, PT, USA) 4 idées

Préserver l'empathie, le contact humain et les relations humaines

Les humains sont des êtres sociaux qui ne peuvent s'épanouir (et apprendre, enseigner, soigner ou guérir) qu'en relation avec leurs semblables. Contrairement aux machines, ils ont la capacité sociale indispensable de se mettre à la place des autres et de nouer des liens émotionnels forts (importance des sentiments et du dialogue pour y parvenir). La confiance et la représentativité se construisent par le dialogue humain. L'IA n'est pas en mesure de remplacer l'interaction humaine.

Il convient de prêter attention au contexte environnant qui peut, dans certains cas, renforcer le risque de dégradation de la qualité du contact humain (par exemple en temps de crise ou en raison de l'épuisement des systèmes de soins de santé ou d'éducation).

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Démocratie) Préserver l'empathie, le contact humain et les relations 2 countries 2 pays (CH, PT) 4 idées
- (Global – Santé) Maintenir l'empathie et la relation humaine au cœur des soins de santé 7 pays (BE, CH, FR, KE, PT, TW, USA) 17 idées
- (Global – Education) Avoir encore des relations et des interactions physiques avec d'autres humains 6 pays (BE, CA, FR, TW, PT, USA) 11 idées

Chercher à s'améliorer

Les êtres humains ont tendance à chercher à s'améliorer et à progresser, à maximiser leur efficacité. Il s'agit là d'objectifs forts pour la plupart d'entre nous (qui peuvent conduire à l'utilisation de stimulants cognitifs ou d'autres technologies d'amélioration). L'IA et les NS peuvent être utilisées pour compenser les limites humaines et pourraient peut-être conduire au développement de nouveaux types de compétences cognitives.

Implication dans des nexus de complexité (voir ci-dessous [Partie 2 : Nœuds de complexité globaux-transversaux](#)):

- [Becoming more efficient without threatening the core of what makes us human](#)

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)):

- (Global – Santé) Chercher constamment à s'améliorer et à progresser 1 pays (PT) 1 idée
- (Global – Education) Utiliser l'IA et les NS pour mieux enseigner et apprendre 6 pays (BE, CA, PT, TW, FR, KE) 9 idées
- (Global – Education) Utiliser l'IA pour améliorer les performances et l'innovation 4 pays (CH, PT, TW, BE) - 5 idées

Préserver et intensifier ce qui fait de nous des êtres humains et favoriser l'épanouissement humain

Certaines valeurs et caractéristiques sont propres aux êtres humains, comme la spiritualité, la sagesse, l'émotivité, la créativité, l'autonomie, la pensée critique, l'imagination, la conscience, l'empathie... Les technologies de l'IA et des NS ne devraient pas menacer ou tendre à marginaliser ou à minimiser ces composantes essentielles de ce que signifie être humain. Dans cette perspective, il convient de résister à toute focalisation excessive sur l'efficacité, la performance et la rentabilité financière uniquement. Par exemple, ne pas effectuer certaines tâches peut nous conduire à nous appauvrir (comme dans le cas des abus de l'assistance à la création par l'IA). De même, les gains de productivité peuvent être mobilisés pour gagner du temps pour des activités favorisant l'épanouissement humain.

En outre, le fait de rejeter systématiquement les limites, de tenter (notamment au moyen de l'IA et des technologies NS) de surmonter et de transgresser toutes les limites par principe peut profondément porter atteinte à notre humanité. Certaines limites et vulnérabilités (comme le fait d'être « affectable », et donc susceptible d'éprouver de la souffrance, ou d'être mortel) sont également au cœur de ce que signifie être humain. Dans le même ordre d'idées, la fatigue et la lassitude sont parfois le signe que quelque chose ne va pas dans la vie d'une personne, plutôt que de simples limites à surmonter (par exemple en utilisant certaines technologies d'amélioration). Ce type de limites mérite d'être reconnu et de faire l'objet d'une grande délicatesse dans le cadre d'une réflexion sur le développement technologique adéquat.

Implication dans des nexus de complexité (voir ci-dessous [Partie 2 : Nexus de complexité globaux-transversaux](#)):

- [Becoming more efficient without threatening the core of what makes us human](#)

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)):

- (Global – Education) Préserver la créativité 5 pays (BE, PT, CA, KE, IT) 6 idées
- (Global – Education) Avoir du temps pour l'épanouissement humain 2 pays (PT, USA) 5 idées
- (Global – Education) Préserver les besoins fondamentaux nécessaires au développement humain 7 pays (BE, CA, FR, TW, PT, CH, KE) 11 idées
- (Global – Santé) Reconnaître que certaines de nos limites et vulnérabilités sont inhérentes à notre nature humaine 2 pays (FR, PT) 2 idées
- (Global – Démocratie) Préserver la spécificité de l'être humain (par rapport à la machine) 2 pays (FR, PT) 7 idées

Favoriser la culture scientifique et technique ainsi que la culture éthique et la pensée critique

La connaissance des NS et de l'IA est essentielle pour que les acteurs concernés (parties prenantes, professionnels, développeurs, décideurs politiques, acteurs économiques/industriels) puissent mener une réflexion éthique appropriée sur le développement technologique associé et sur les utilisations adéquates des technologies disponibles (quelles sont les limites des technologies proposées, quels sont les points forts et les risques ? ...).

Pour permettre une réflexion éthique (collective) appropriée sur les NS et l'IA, il est plus largement essentiel de préserver et de développer la pensée critique (à une époque où la désinformation se développe et où le rapport à la vérité et à la connaissance est menacé) ainsi que les capacités de réflexion éthique elles-mêmes.

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Education) Encourager l'éthique dans l'éducation 5 pays (PT, CH, FR, IT, BE) 9 idées
- (Global – Education) Développer l'esprit critique 3 pays (FR, PT, IT) 5 idées
- (Global – Education) Favoriser la connaissance de l'IA et des NS 2 pays (FR, CH) 4 idées
- (Global – Démocratie) Favoriser la connaissance et la pensée critique pour préserver et renforcer la démocratie 3 pays (IT, PT, TW) 4 idées
- (Global – Santé) Favoriser la connaissance et la pensée critique 4 pays (CH, IT, KE, PT) 4 idées

S'assurer que la technologie réduise (plutôt qu'augmente) les inégalités

Le développement rapide des technologies de l'IA et des NS présente le risque d'accroître les inégalités sociales et économiques déjà existantes. Il est nécessaire de veiller à ce que les avantages et les difficultés soulevés par ces transformations soient équitablement répartis (équité dans l'accès à des services non déshumanisés et à des innovations contribuant positivement, ou dans la protection contre les dangers et les effets indésirables, tels que la discrimination et les préjugés automatisés). Les inégalités peuvent se situer au niveau de l'accès (compétences et alphabétisation, moyens financiers, infrastructures matérielles) ainsi qu'au niveau des asymétries de pouvoir ou de partage des bénéfices. Au niveau des nations, les inégalités peuvent également se situer au niveau de la capacité à développer des systèmes d'IA souverains. Il faut également prendre en compte le problème de l'automatisation possible de la gestion des ressources humaines.

Toutefois, si elles sont correctement utilisées, les technologies de l'IA (numérique) et des NS peuvent renforcer la justice sociale et la défense des droits de l'homme. Les technologies de l'IA peuvent favoriser l'inclusion sociale, notamment en facilitant l'accès des plus vulnérables (pauvres, réfugiés) à divers services ou en permettant aux personnes handicapées de devenir plus indépendantes.

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Education) Indésirable : Exacerber les inégalités sociales et économiques avec l'IA 7 pays (BE, IT, PT, TW, KE, CH, USA) 14 idées
- (Global – Démocratie) Prendre en compte les personnes vulnérables et contribuer aux droits de l'homme et à l'inclusion sociale et politique 5 pays (BE, FR, IT, KE, PT) 15 idées
- (Global – Santé) Garantir l'équité et l'égalité des chances de vivre une bonne vie 6 pays (BE, CH, FR, IT, KE, PT) 12 idées
- (Global – Santé) Utiliser les technologies de la santé pour améliorer les conditions de vie des personnes les plus vulnérables 1 pays (KE) - 4 idées
- (Global – Education) Favoriser l'inclusion sociale grâce aux technologies d'IA 6 pays (CH, CA, FR, BE, TW, KE) 18 idées

Privilégier la coopération entre l'IA et l'humain plutôt que le remplacement

L'IA et la technologie devraient contribuer à une société plus humanisée. L'IA peut être un outil utile pour aider les humains à gagner du temps dans certaines tâches. Mais les machines ne doivent pas remplacer les humains. En particulier, l'IA et les technologies d'automatisation sont souvent invoquées dans les domaines où les acteurs manquent de temps ou sont épuisés (comme dans les systèmes de soins de santé). Cependant, la technologie ne constitue pas toujours la bonne ou la principale réponse à ces questions majeures.

Dans le même ordre d'idées, le problème de l'automatisation du travail et le risque de chômage de masse devraient être pris au sérieux, surtout lorsqu'il s'agit des personnes les plus vulnérables. Ces changements économiques majeurs sont susceptibles d'affecter profondément nos sociétés.

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Démocratie) Privilégier la coopération et le soutien avec l'IA plutôt que le remplacement de l'homme 5 pays (IT, KE, PT, TW, USA) 7 idées
- (Global – Démocratie) Trouver le bon équilibre entre le travail humain et l'automatisation des tâches par l'IA 3 pays (IT, KE, USA) 3 idées
- (Global – Santé) Privilégier la coopération et le soutien avec l'IA plutôt que le remplacement de l'homme 8 pays (BE, CH, FR, IT, KE, PT, TW, USA) 16 idées
- (Global – Education) Indésirable : Remplacement des interactions humaines par les technologies de l'IA 6 pays (FR, BE, CA, USA, KE, PT) 7 idées

Protéger la vie privée

L'essor de l'IA suscite des inquiétudes quant au respect de la vie privée. Par exemple, les entités privées et publiques ont un accès massif à toutes sortes de données personnelles (sur la santé, les opinions, les choix, les habitudes et les coutumes...), ce qui met à rude épreuve la vie privée (il convient d'ajouter à cela les problèmes émergents concernant les neurotechnologies et la protection de la vie privée des cerveaux). Pour protéger la démocratie et garantir la liberté individuelle, il est impératif de renforcer les lois sur la protection de la vie privée et d'établir une distinction claire entre la vie privée et la vie publique, non seulement en ligne (opinions publiques et anonymat en ligne), mais aussi dans l'espace public (l'utilisation de données obtenues par vidéosurveillance comme la reconnaissance faciale doit être limitée à certains endroits, et leur utilisation doit être justifiée).

Idées correspondantes (à consulter dans les synthèses globales thématiques téléchargeables [ici](#)) :

- (Global – Santé) Assurer la protection de la vie privée 4 pays (CH, KE, IT, USA) 5 idées
- (Global – Démocratie) Assurer la protection de la vie privée 5 pays (CA, FR, IT, PT, TW) 9 idées

Partie 2: Nœuds de complexité globaux-transversaux

Être humain à l'époque des neurosciences et de l'intelligence artificielle implique d'explorer soigneusement les nœuds de complexité où des idées valables sont néanmoins en tension, manifestant des subtilités et des défis qu'il ne faut pas négliger. Voici quelques exemples de **nœuds de complexité transversaux** identifiés sur la base de **synthèses locales et globales**.

Quel lien, quelles interactions avec les machines ?

Certains participants soulignent qu'avec les progrès de l'IA, nous aurons tendance à développer des machines (robots, automates conversationnels) capables d'imiter ou de simuler des comportements et des capacités propres aux humains et aux êtres vivants, tels que l'empathie, l'assertivité, la vie émotionnelle et affective. Par conséquent, il sera de plus en plus tentant de s'attacher émotionnellement à ce type de machines capables de simuler des capacités relationnelles (comme les compagnons ou les assistants artificiels, ou encore les robots de soins personnels).

Ces discussions soulèvent également la question des droits à accorder aux robots avancés ou aux systèmes intelligents.

En même temps, de nombreuses contributions aux débats soulignent l'importance de ne pas perdre de vue la spécificité du vivant et de l'humain par rapport aux machines. Les machines ne sont pas conscientes, ne ressentent pas d'émotions, ne peuvent pas être sages, créatives, critiques ou autonomes, ne sont pas capables de spiritualité au sens habituel de ces termes, qui implique un enracinement dans l'expérience vécue, dans un corps biologique. Au mieux, ils peuvent simuler des comportements convaincants dans ces registres (notamment par la conversation), comportements que des êtres humains ou des êtres vivants auraient dans des circonstances données.

De ce point de vue, de nombreux participants s'accordent à dire que l'IA ne peut pas être un sujet de droit. La question est largement qualifiée de spéculative ou de science-fictionnelle, sans pour autant être inintéressante.

Ainsi, il est assez largement exprimé dans les discussions qu'il faut résister à la tentation (de plus en plus réelle et puissante) de percevoir certains robots ou systèmes d'IA comme de véritables personnes et d'essayer de se connecter avec eux affectivement (comme on le ferait avec un humain, ou même avec un autre être vivant). Nous devons résister à la tentation de substituer des interactions avec des machines à de véritables relations humaines.

Idées des synthèses globales et locales mobilisées dans ce nœud de complexité (les synthèses sont téléchargeables [ici](#)) :

- Les systèmes d'IA et les machines ne peuvent pas être confondus avec des êtres humains et ne peuvent donc pas être dotés de droits similaires à ceux des êtres humains.
 - (Global – Démocratie) Préserver la spécificité de l'être humain (par rapport à la machine)
 - (France – Démocratie) Indésirable : La reconnaissance d'une personnalité juridique pour les IA n'est pas souhaitable
 - (France – Démocratie) Souhaitable : Les algorithmes restent des outils
 - (USA – Démocratie) Les machines sont au service de l'humanité, c'est pourquoi l'humanité doit garder un contrôle approprié sur l'IA.
 - (France – Démocratie) La question complexe du statut juridique de l'intelligence artificielle est largement débattue
- Les systèmes d'IA ne doivent pas remplacer les relations humaines
 - (Global – Transversal) Préserver l'empathie, le contact humain et les relations humaines
- Les systèmes d'IA auront de plus en plus de comportements qui permettent ou encouragent la tendance des humains à vouloir se connecter et s'attacher à eux.
 - (Portugal – Démocratie) L'homme et la machine peuvent se rapprocher
 - (Portugal – Démocratie) L'intelligence artificielle tendra à imiter les capacités humaines

Eclairages par les chercheurs² :

Mathieu Guillermin – enseignant-chercheur en éthique des nouvelles technologies

Il est plus que légitime de s'émerveiller des récents développements des technologies de l'IA, qui ont permis à des programmes tels que chat-GPT et d'autres grands modèles de langage d'entretenir une conversation convaincante avec les humains. Toutefois, cet émerveillement doit être motivé par de bonnes raisons. Après tout, ces succès n'ont rien à voir avec la création de nouvelles formes de vie, de nouveaux êtres intelligents, que nous appellerions les IA. Il est tout aussi vertigineux, sinon plus, de constater que l'homme a pu construire des machines, des artefacts capables de simuler ou de reproduire des comportements intelligents (des comportements convaincants qui auraient pu venir de l'homme), sans aucune vie, sans aucun vécu, sans aucune conscience, mais avec de purs mécanismes (des mécanismes inertes, mais d'une complexité et d'une miniaturisation éblouissantes). En plus de démystifier l'apprentissage automatique (y compris l'apprentissage profond, basé sur les réseaux de neurones artificiels), il est également crucial de se rappeler que tous les programmes (du plus traditionnel et conventionnel au programme d'IA le plus avancé produit par l'apprentissage automatique) fonctionnent sur des ordinateurs ou des machines similaires qui ne sont pas (ou sont moins) programmables. Ce que fait une machine comme un ordinateur, c'est transformer des configurations matérielles auxquelles les humains ont associé des significations précises (une série d'aimants sur un disque dur symbolise une séquence de 0 et de 1, elle-même

² Mathieu Guillermin, basé sur les travaux du groupe de recherche sur l'intelligence artificielle du Centre pour la culture numérique (Culture et éducation) et sur son livre « Encountering Artificial Intelligence : Ethical and Anthropological Investigations ». *Journal of Moral Theology* 1 (Theological Investigations of AI) 2023 : i-262. <https://doi.org/10.55476/001c.91230>

associée, par exemple, à une séquence de mots ou à une séquence de nombres codant les couleurs des pixels d'une image) en de nouvelles configurations matérielles associées à d'autres significations (par exemple, une nouvelle série de mots, une image modifiée ou une description de l'image). Ce type de machine, conçue pour transformer des configurations matérielles en d'autres en fonction de ce que ces configurations signifient, n'est pas nouveau. L'ordinateur peut être considéré comme l'aboutissement d'une longue évolution des techniques et technologies de l'information, qui remonte probablement aux débuts de l'écriture. Dans cette perspective, le boulier peut être considéré comme un ancêtre de l'ordinateur (transformation mécanique de configurations symbolisant, par exemple, des nombres à additionner, en configurations symbolisant le résultat de l'addition).

Il n'y a donc pas à proprement parler de sens, d'images, de mots ou de nombres dans les ordinateurs, encore moins d'émotions ou de conscience. Ce sont cependant de fantastiques machines qui manipulent mécaniquement (avec une efficacité et une précision incroyables) d'innombrables configurations matérielles auxquelles nous, les humains, attachons un sens. Une série d'aimants sur le disque dur d'un ordinateur fera en sorte que différents pixels de l'écran émettent des couleurs différentes, qui seront pour nous plus que de minuscules sources de lumière colorée, qui deviendront des textes nous informant sur des sentiments, des images de visages ressentant telle ou telle émotion. Mais l'ordinateur ne traite l'information qu'en manipulant mécaniquement et automatiquement des aimants (ou d'autres configurations matérielles).

Il est donc d'autant plus impressionnant de voir ce que nous pouvons faire faire aux ordinateurs avec des programmes dérivés des techniques d'apprentissage automatique. De grands modèles de langage comme chat-GPT nous parlent de manière convaincante (avec un contenu affectif ou émotionnel crédible). Nous pouvons également essayer d'analyser automatiquement les émotions et les sentiments dans ce que les gens disent, ou dans les vidéos capturant les expressions corporelles ou faciales. Ces nouvelles technologies ouvrent la possibilité d'interactions toujours plus riches et plus intéressantes avec les machines, avec des modalités qui reproduisent ou simulent un nombre croissant de caractéristiques des interactions et des relations entre les êtres vivants en général, et entre les humains en particulier.

Pour bien envisager les conséquences et les enjeux de ces nouvelles possibilités d'interaction avec les machines, plusieurs points méritent d'être soulignés³. En premier lieu, et contrairement à ce que les approches behavioristes pourraient suggérer (en lien avec le fameux test de Turing), il semble important de maintenir une distinction entre simuler un comportement résultant d'une expérience vécue et avoir ce même comportement tout en vivant cette expérience vécue. Que dire, par exemple, d'une machine qui exprime des paroles de compassion à une personne âgée dans la perspective de sa fin de vie ? Elle ne peut être confondue avec les mêmes paroles prononcées par une personne capable d'éprouver sa finitude, de ressentir et de compatir dans un vécu partagé.

³ Nous nous appuyons ici largement sur le chapitre 4 de « Encountering Artificial Intelligence : Investigations éthiques et anthropologiques ».

Deuxièmement, il est également important de dire que le simple fait de reconnaître que les machines ne sont que des machines, et de les traiter comme de purs outils, n'est pas nécessairement la réponse à tous les problèmes. En effet, dans cette perspective et selon toute vraisemblance, les compagnons artificiels (comme dans le film *Her* de Spike Jonze en 2013) seront construits et programmés pour trouver leur place sur un marché et donc se comporter d'une manière qui satisfasse l'utilisateur (par exemple, qui voudrait d'un compagnon artificiel qui pourrait trahir ou quitter son humain ?) Nous serons donc confrontés à des systèmes perçus comme des objets, des possessions, mais qui tireront tout leur attrait spécifique de leur capacité à ressembler à une personne authentique, à manifester une apparence d'humanité, de personnalité ou de vie. L'accoutumance progressive à la combinaison de ces deux caractéristiques pourrait s'avérer extrêmement destructrice pour l'humanité. Cela reviendrait à développer progressivement une capacité à s'accommoder de l'esclavage : « Là où il n'y a pas d'« autre », mais seulement l'apparence d'un autre à notre disposition, en même temps que l'absence de l'exigence qui serait exercée sur notre propre don de soi par la confrontation avec un véritable autre, nous risquons d'être conditionnés dans un dangereux talent d'exploitation. »⁴

Dans le même ordre d'idées, cette combinaison du statut d'objet ou d'outil et de l'apparence personnelle peut également nous amener à nous habituer à une attitude de consommateur à l'égard du comportement d'autrui, réduisant progressivement notre tolérance à l'égard du comportement d'autrui qui nous dérangerait. Il n'est pas impossible que la présence constante de compagnons artificiels, dont les comportements dérangeants seront perçus comme des défauts (en vertu de leur statut d'outil ou d'objet), nous amène subrepticement à considérer les personnes authentiques qui nous dérangent de la même manière, « comme de simples êtres humains défectueux, les considérant avec le même type d'insatisfaction oisive que nous ressentirions à l'égard d'un robot qui ne fournirait pas l'ensemble des comportements et des réactions que nous voudrions consommer ».⁵

Cela peut conduire à reconsidérer la question des droits à accorder aux robots et aux systèmes d'IA. Certes, leur statut de machine permet de refuser légitimement de les considérer comme des sujets de droit. Cela ne signifie pas pour autant qu'il faille laisser chacun en faire ce qu'il veut, comme on le ferait avec une table. Un cadre réglementaire peut être souhaitable dans ce domaine, ne serait-ce que pour éviter le développement de comportements ou d'habitudes extrêmement toxiques pour l'être humain et les autres êtres vivants.

Tous ces éléments nous incitent à une réflexion approfondie sur les raisons qui poussent à développer des machines de plus en plus capables de prendre l'apparence d'un être humain ou d'un autre être vivant. Nous devons réfléchir à ce que nous pouvons réellement tirer de ces technologies.

⁴ Ibid., p. 119.

⁵ Ibid., p. 121. La phrase complète est la suivante : « Est-il possible que nous ne voyions plus cela comme un aperçu d'une humanité plus vaste ? « Est-il possible que nous ne considérions plus cela comme un aperçu d'un éventail plus large de l'humanité, que nous ne nous efforcions plus d'avoir une réaction charitable ? Peut-être qu'au lieu de cela, nous en viendrons à considérer ces autres comme simplement des êtres humains défectueux, les voyant avec le même type d'insatisfaction oisive que nous éprouverions à l'égard d'un robot qui ne fournirait pas l'ensemble de comportements et de réactions que nous voudrions consommer. »

Complexité à retrouver sur le site internet du projet NHNAI : <https://nhnai.org/fr/focus-sur-les-complexites-transversale-1/>

Fixer des limites et des règles malgré les défis

(basé sur l'idée éponyme globale-transversale : [Fixer des limites et des règles malgré les défis](#))

Il existe un fort besoin de réglementation et de normes pour garantir que les technologies de l'IA et des NS produisent des résultats positifs. Les normes et la réglementation sont essentielles à l'instauration de la confiance et à la protection des personnes lors du déploiement de nouvelles technologies. L'IA devrait respecter les valeurs humaines (équité, impartialité, ...) et être centrée sur l'être humain (visant à son épanouissement). L'IA et les technologies des NS devraient bénéficier à tous (il est crucial de lutter contre l'exclusion des personnes pauvres et vulnérables).

Cependant, la réglementation soulève de nombreuses questions aiguës, ce qui en fait un défi très difficile à relever. Parmi ces questions, on peut évoquer le rythme du développement technologique, l'obscurcissement des modèles de responsabilité (avec les technologies numériques en général et plus particulièrement avec l'apprentissage automatique), l'accès souvent « facile » à des outils puissants (dans la main d'acteurs mal intentionnés, des technologies telles que la reconnaissance d'images ou de visages peuvent devenir extrêmement nuisibles), l'échelle mondiale de la recherche et du développement (avec la diversité des systèmes de valeurs dans le monde ainsi que des constellations de conflits d'intérêts), la difficulté d'appliquer les réglementations (dans un contexte aussi diversifié et international).

D'une manière générale, la réglementation devrait favoriser une utilisation raisonnée et saine des technologies de l'IA et des NS. Néanmoins, il peut s'avérer extrêmement difficile d'identifier ce qui est raisonné et sain et ce qui ne l'est pas (par exemple, dans le cas de la modération des médias sociaux : qui est le bon acteur ? Ou le cas des technologies de la santé avec des zones grises entre les utilisations curatives et d'amélioration : qui peut décider si une pathologie nécessite / justifie l'utilisation d'une technologie de la santé donnée ?) Les parties prenantes, les professionnels, les citoyens et les acteurs économiques/industriels devraient être impliqués dans les processus de régulation.

Eclairages apportés par les chercheurs :

(Mathieu Guillermin) Possibilité de contribuer à la résolution de problèmes complexes, nécessité de co-construire le lien plus large entre la délégation et les questions politiques, questions de démocratie technique, d'acceptabilité sociale, mais aussi de recherche collective de la vérité au lieu d'une simple consultation.

Complexité à retrouver sur le site internet du projet NHNAI : <https://nhnai.org/fr/focus-sur-les-complexites-tr2/>

Devenir plus efficace sans menacer l'essentiel de ce qui fait de nous des êtres humains

L'idée globale-transversale "[S'appuyer sur les technologies pour améliorer notre vie](#)" a souligné le fait que les technologies d'IA et d'automatisation pourraient nous aider à gagner du temps pour des activités essentielles telles que les relations ou tout ce qui favorise l'épanouissement humain en déléguant les tâches fastidieuses aux machines. Il a également souligné que les résultats de l'IA et des NS pourraient nous permettre d'améliorer nos capacités physiques et mentales, ainsi que nos performances et notre efficacité.

L'idée globale transversale "[Chercher à s'améliorer](#)" exprime l'idée qu'il est dans la nature de l'homme de chercher à s'améliorer, à progresser et à maximiser son efficacité.

Néanmoins (comme l'idée globale-transversale "[Préserver et intensifier ce qui fait de nous des êtres humains et favoriser l'épanouissement humain](#)" avertit qu'il peut s'avérer destructeur de rechercher systématiquement et sans esprit critique l'augmentation et l'amélioration de l'efficacité et de la performance. Cela pourrait conduire à sacrifier des aspects essentiels pour les humains, tels que l'autonomie, la créativité, les relations, ou à nier certaines limites et vulnérabilités qui sont au cœur de ce que signifie être humain (la mortalité, l'affectabilité, par exemple).

Eclairages apportés par les chercheurs :

Juan R. Vidal – enseignant-chercheur en neurosciences cognitives

La recherche de l'amélioration de soi est quelque chose qui n'existe pas en tant que tel dans le comportement humain, si elle n'est pas rattachée à une action orientée vers un but et dans un contexte temporel large (ex : nous voulons assurer l'accès à la nourriture et à l'eau, à un abri, ...). Ce but comporte une valeur pour l'être humain qui le motive (ou non) à poursuivre l'apprentissage et le développement de certaines capacités et de certains comportements. Les humains pensent qu'ils maximisent leur efficacité, mais comme Herbert Simon l'a mentionné, les humains ont une rationalité limitée, et donc des capacités limitées pour vraiment maximiser les processus de pensée et donc le comportement. Les humains « satisfont » plutôt leur comportement afin d'être satisfaits le plus rapidement possible, ce qui n'est pas la même chose que de maximiser leurs capacités. Ce biais s'applique également à l'utilisation de la technologie, et avec l'IA, il est fortement potentialisé. Or, comme cela a été démontré, elle réduit considérablement les possibilités d'apprentissage de la personne et, in fine, sa liberté d'action dans le monde. Ainsi, la recherche de l'amélioration de soi devrait résonner avec la possibilité d'augmenter l'apprentissage (incarné) et les possibilités d'apprentissage futur (garder des portes ouvertes...) au lieu d'accélérer certaines performances qui, plus loin, privent l'humain d'apprendre et donc de s'adapter à des conditions changeantes (si l'on considère que sa capacité d'adaptation dépend grandement de sa capacité à apprendre de nouveaux comportements/pensées pour faire face à de nouveaux problèmes).

Complexité à retrouver sur le site internet du projet NHNAI : <https://nhnai.org/fr/focus-sur-les-complexites-tr3/>