L’idée que l’IA pourrait incarner la perfection du raisonnement et de la décision est une profonde illusion.

Avertissement *: Souveraine Tech revendique par vocation une approche transpartisane. Seule nous oblige la défense des intérêts supérieurs de notre pays. Nous proposons ainsi un lieu de « disputatio » ouvert aux grandes figures actives de tous horizons. La parole y est naturellement libre et n’engage que ceux qui la prennent ici. Cependant, nous sommes bien conscients des enjeux en présence, et peu dupes des habiles moyens d’influence plus ou moins visibles parfois mis en œuvre, et dont tout un chacun peut faire l’objet, ici comme ailleurs. Nous tenons la capacité de discernement de notre lectorat en une telle estime que nous le laissons seul juge de l’adéquation entre le dire et l’agir de nos invités.*

Vendredi 24 janvier 2025

Nous sommes heureux de publier aujourd’hui un entretien avec Mathieu Guillermin, qui est enseignant-chercheur en philosophie et éthique des sciences et des technologies, membre de l’Unité de Recherche CONFLUENCE : Sciences et Humanités (EA 1598) de l’UCLy (Université Catholique de Lyon). Il est docteur en philosophie et en physique. Il coordonne le projet international « Nouvel Humanisme au temps des Neurosciences et de l’IA

» (NHNAI). Ses travaux portent sur les articulations entre démarche éthique, recherche scientifique et développement technologique. Il s'intéresse tout particulièrement aux questions éthiques et sociétales soulevées par nouvelles technologies numériques.

1/ Pouvez-vous svp décrire en quoi consiste la démarche internationale du NHNAI ?

Le projet part de deux idées centrales :

Premièrement : Pour s'orienter dans la multitude d'enjeux et de défis éthiques soulevés par les technologies d'IA, il est nécessaire de s'interroger, en arrière-plan, sur ce que signifie « être humain » au temps de l'IA (qui sommes-nous en tant qu'humain ? qui voulons-nous être et devenir ? …). Sans se substituer à la réflexion éthique in situ, sujets par sujets, cette réflexion d'arrière-plan nous semble intéressante pour fournir une sorte de boussole, une sorte de nord magnétique sur lequel on pourra s'appuyer, tenter de s'aligner, dans chaque situation.

Deuxièmement : Il nous semble de plus que cette question « que signifie être humain au temps de l'IA ? » n'est pas une question à laquelle les experts (quelle que soit leur discipline) peuvent répondre directement, seuls dans leur coin. C'est une question éminemment politique et existentielle qui nous concerne, et nous engage, toutes et tous. Néanmoins, cette question peut difficilement s'étudier sérieusement sans s'appuyer sur les connaissances scientifiques et les expertises disponibles (informatique, anthropologie, sociologie, psychologie, neurologie, philosophie, théologie …). De manière complémentaire, la régulation de technologie comme celles de l'IA ne peut se faire seulement dans une démarche descendante avec des lois et des règles établies par les institutions politiques (ou par d'autres institutions) s'appuyant sur l'expertise scientifique. Cette forme de régulation est nécessaire, indispensable, mais risque de ne

pas être suffisante. Déjà, il y a un défi de mise en œuvre, d'application : il faut mettre des moyens sur la table pour faire respecter les réglementations. De plus, des règles générales auront besoin de s'implémenter concrètement, d'évoluer au contact des situations réelles, de se confronter à la réalité singulière. Ainsi, le mouvement de régulation descendant ne peut être couronné de succès que s'il rencontre un effort ascendant en provenance des citoyennes et des citoyens, des associations, des acteurs socio-économiques … Il est impératif d'encourager une culture de la réflexion et de l'orientation éthique très horizontale, chez toutes les personnes concernées par les enjeux de l'IA. Ainsi pourra-t-on espérer que les réglementations descendantes entreront en résonance avec les choix du quotidien de tout un chacun. Notamment les choix de consommation qui peuvent créer un appel d'air pour une activité économique plus vertueuse. (La plupart des entrepreneurs que je rencontre sont très soucieux des enjeux éthiques … encore faut-il que ce souci soit partagé par leurs clients).

Sur la base de ces deux axes structurants, le projet NHNAI entend donc mettre à la disposition de la réflexion collective horizontale (avec toutes les personnes se sentant concernées) les ressources d'un réseau d'universités catholiques. Sous la bannière de la Fédération Internationale des Universités Catholiques, le réseau NHNAI rassemble aujourd'hui 12 partenaires dans 9 pays différents (Chili, USA, Canada, Kenya, Taïwan, Belgique, Portugal, Italie, France). L'action du réseau se déploie selon deux dimensions principales : d'une part, l'organisation, la restitution et la synthèse de grands ateliers de discussion, dans les différents pays participants, sur la question de l'humain au temps de l'IA ; d'autre part, l'apport d'expertise pour nourrir la réflexion collective (avec environ 70 chercheuses et chercheurs d'horizons disciplinaires très variés qui s'efforcent de partager, en fonction du contenu des discussions collectives, les éléments les plus saillants de leur discipline qui peuvent aider à

approfondir les échanges).

La démarche et les résultats des ces discussions peuvent être consultés sur le site web du projet.

2/ Faut-il entendre par « nouvel humanisme » que les temps que nous vivons annoncent l'avènement d'un « homme nouveau » ?

L'humain est très probablement toujours en évolution, en particulier avec les technologies dont il se dote et qui, souvent, le transforment en profondeur. Néanmoins, l'idée de « nouvel humanisme » dans le cadre de notre projet désigne plutôt la réflexion sur l'humain, le besoin de rouvrir cette réflexion. Le terme « humanisme » est très souvent associé à la modernité et la période des Lumières qui a vu des penseurs comme Rousseau ou Kant mettre l'accent sur la légitimité (et même le devoir) pour chaque individu de penser par lui-même, de faire usage de sa raison pour faire ses propres choix et guider sa vie de manière plus autonome. Force est de constater que ce mouvement d'émancipation et de responsabilisation des personnes a pour une part pris une tournure beaucoup plus discutable (individualisme, dénigrement des dimensions non objectivables de la vie, vision de l'humain comme propriétaire de la nature vue comme un moyen au service des fins humaines …) que ce que véhiculaient les intuitions originales. Pourtant la solution n'est pas forcément de rejeter cet héritage en bloc. Il est peut-être intéressant de remettre tout cela sur le métier à tisser, de repenser ce que veut dire être une personne autonome et faisant un bon usage de sa raison, de repenser ce qu'est notre spécificité en tant qu'humain et ce que cette spécificité signifie pour notre rapport à autrui et aux autres vivants … Dans cet effort, l'héritage de l'humanisme des Lumières est indispensable. Mais comme une composante de l'exploration, un point d'appui pour se poser les bonnes questions …

3/ Comment expliquez-vous le fait que l'Homme soit le grand absent du brouhah-IA actuel ?

Peut-être que, précisément, il nous est difficile de mettre l'humain en avant dans la mesure où nous n'avons pas les idées claires ce que signifie « être humain ». Il est tentant, face aux grands problèmes du monde actuel, de croire à la possibilité de faire table rase du passé. Si humain est synonyme d'égoïsme et d'individualisme, de cupidité, d'une intelligence biaisée … peut-être qu'il faut devenir autre chose ou bien accepter de s'éclipser de la scène ? De laisser la place à des intelligences supérieures ?

Ces idées me semblent profondément erronées. D'une part, ce n'est regarder que le mauvais côté de l'humain, ne s'attacher qu'aux échecs, et en plus en noircissant le tableau (notamment sur le dénigrement de l'intelligence humaine). D'autre part, ce serait aussi se méprendre profondément sur ce qu'est l'intelligence artificielle. Il est important d'être très clair ici (dans l'idéal il faudrait même développer beaucoup plus) : l'IA n'est pas une nouvelle forme d'intelligence dans le sens de quelque chose qui serait autonome comme nous, que l'on pourrait placer à côté de nous, comme un autre être vivant. Il faut résister à la petite musique ambiante qui se fait de plus en plus pressante à ce niveau. Avec l'IA, on reste dans le domaine de l'informatique, c'est-à-dire dans le domaine de la construction de mécanismes programmables, d'artefacts qui vont, de manière contrôlée (merveilleusement bien contrôlée d'ailleurs), transformer des données d'entrée en données de sortie. On parle bien ici de machines, de mécanismes, d'automates … Et transformer des données, c'est quelque chose de très concret, rien de magique : il s'agit de passer d'une configuration matérielle (une série d'aimants orientés d'une certaine manière par exemple) à une autre (une autre série d'aimants avec des orientations différentes), et nous, les humains, nous associons de la signification aux orientations de ces aimants (des 1 et des 0, des nombres, des mots, des images …). Mais tout repose sur ce pouvoir

d’association entre des choses autour de nous (parfois des choses très très petites) et des significations. Ensuite, il s’agit de mécanique pour construire des automates qui transforment cette matière organisée pour signifier … Sans un être humain capable de faire le lien avec les significations, il n’y a que de la mécanique (quand bien même il s’agit d’une merveilleuse mécanique très miniaturisée) … Ainsi, il n’y a aucun sens à mon avis à parler de radicalement laisser la place à l’IA. De ce point de vue, la deep ecology a au moins le mérite de la cohérence en parlant de laisser la place aux autres vivants … Laisser la place à l’IA, ça revient seulement à laisser la place à des machines qui tourneront « à vide » …

4/ La fascination pour la puissance de calcul procède t-elle selon vous de notre « volonté de puissance » ? Et le cas échéant, sait-elle bien à quoi elle est exactement ordonnée ?

Il y a en effet quelque chose de l’ordre de l’hybris dans la fascination pour le calcul logico-mathématique, encore plus lorsqu’il est combiné à la data, à la donnée empirique brute … Cela rejoint l’idée de l’enquête pure chère aux cartésiens : si, dans une démarche de construction de connaissance, nous ne nous appuyons que sur des choses inévitables, si nous restons parfaitement neutres, que nous ne mobilisons pas notre jugement subjectif, alors les conclusions produites sont absolument certaines. Nous produisons des « vérités ». Il est assez courant de considérer ainsi l’activité scientifique : elle ne mobilise que le donné empirique et le calcul logico-mathématique, elle est donc parfaitement neutre et produit donc des vérités.

Il est extrêmement important de bien comprendre que ceci est une caricature de la science et que l’idée de l’enquête pure est une illusion. Il y a toujours du jugement, de l’arbitrage … C’est inévitable. Par exemple de manière très fondamentale : qui nous dit que le donné empirique est fiable ? Descartes en avait douté d’ailleurs. Peut-on réellement prouver que nous ne

vivons pas dans une illusion ? Dans la caverne de Platon ? Peut-on absolument démontrer que nous ne sommes pas victimes d'illusions provoquées dans nos cerveaux par des machines (comme dans le film Matrix) ou que nos résultats expérimentaux ne sont pas altérés par des extra-terrestres malveillants et très avancés technologiquement (qui souhaitent entraver notre développement scientifique) ? Nous savons déjà qu'il est impossible de prouver la validité des mathématiques … Ultimement, tout repose sur une question de confiance, de jugement sur ce qui est raisonnable ou non. Je suis bien entendu convaincu que les scientifiques ont raison de faire confiance à leur expériences (ou a minima au principe du test expérimental). Mais je suis conscient qu'il s'agit là d'un jugement, d'un arbitrage de ma part … Mettre en place ou suivre une procédure (un algorithme ou une recette) pour, notamment, construire des connaissances, ce n'est jamais éliminer le jugement et l'arbitrage (en tous cas ça ne devrait pas) … Une procédure n'a de légitimité qu'en tant qu'elle a été jugée valide, fiable, adéquate au problème auquel elle répond …
En résumé : de la puissance de calcul oui, mais seulement dans le cadre d'arbitrages soigneux et de jugements approfondis …

5/ Vivons-nous des temps éminemment philosophiques qui voient s'affronter la quantité et la qualité ?

Je crois que nous vivons des temps marqués par notre incapacité de plus en plus grande à prendre sérieusement en compte le qualitatif (probablement en partie en raison de l'illusion de l'enquête pure). Observer et calculer c'est très utile, mais il faut aussi (en premier lieu ?) savoir arbitrer et juger.

6/ Si l'on se place dans une perspective chrétienne, le développement fulgurant de l'IA vous paraît-il de nature à la placer dans l'ordre de la « pro-création » ? Ou sommes-nous plutôt en plein mythe faustien ou prométhéen ?

L'exposition rapide proposée en amont de la manière dont je comprends ce qu'est l'IA permet de répondre à cette question.

Avec l'IA, nous ne sortons pas du domaine de la création d'outils … Des outils puissants, possiblement dangereux et difficiles à maîtriser, mais des outils tout de même (il y a d'ailleurs quantité d'autres technologies qui, nous le réalisons parfois un peu trop tard, présentent des dangers et dont les conséquences ne sont pas facilement prédictibles ou maîtrisables … on pourrait parler de toutes les technologies liées aux hydrocarbures bien évidemment). Il est donc clair que nous ne sommes pas en train de réellement créer une nouvelle espèce ou quelque chose de cette sorte. Si on le croit (de même que si on s'illusionne sur la puissance épistémique réelle des procédures ou des algorithmes), on se rapproche en effet des registres faustien et prométhéen.

7/ On avait déjà du mal à décrire le réel jusqu'à aujourd'hui, est-ce que la noétique, notre manière de percevoir et de penser le « réel », ne risque pas de devenir le sujet de demain en matière d'IA ? Comment saurons nous discerner alors quels phénomènes, sentiments, crédits, autorités, témoignages sont bel et bien réels ? Et quels impacts prévoir s'agissant de la confiance qui doit imprégner notre commerce avec nos semblables ?

Le danger est réel et pressant. Dans ce registre de réflexion, il est intéressant d'évoquer la question de l'organisation et de la gestion des informations que nous recevons au quotidien grâce à nos outils informatiques. Cette question est cruciale et épineuse. Cruciale, car on ne peut de nos jours pas espérer tirer quelque chose d'intéressant de nos systèmes d'information, d'internet, sans de puissants algorithmes de recommandation (que ce soit pour les moteurs de recherche ou pour les réseaux sociaux). Épineuse, car on doit se demander les principes selon lesquels ces algorithmes fonctionnent, les objectifs qu'ils servent. Sont-ils fait pour nous donner les informations qui vont attiser ce qu'il y a de meilleur en nous ? Les informations qui vont nous faire grandir ? On peut en douter. Le modèle économique de la gratuité est ainsi à

interroger. Puisque nous ne payons pas (ou très peu) pour les services numériques comme les moteurs de recherche et les réseaux sociaux, il faut bien que de la valeur soit créée autrement. Et cela se fait via la collecte de données, personnelles ou non, qui va servir au ciblage publicitaire (c'est de plus en plus connu de toutes et tous), mais aussi au développement d'algorithmes possiblement très puissants par les techniques d'apprentissage machine (c'est peut-être un peu moins mis en avant). En simplifiant, la rentabilité des services numériques est donc étroitement associée à la capture de l'attention des utilisateurs afin de maximiser la quantité de données collectées. Et comme l'a bien décrit Gérald Bronner dans son ouvrage Apocalypse Cognitive, les contenus qui captent notre attention sont loin d'être ceux qui favorisent notre épanouissement …

8/ Si l'IA promet la perfection du raisonnement et par suite de la décision, qu'est-ce qui nous empêchera, à terme, de lui confier l'organisation de la Cité en devenant les abonnés comblés (ou presque) de ses impeccables services ?

Comme je l'évoquais, les programmes informatiques (en incluant ceux relevant de l'IA) sont des algorithmes, des procédures guidant une transformation de données. Et réduire le raisonnement à la mise en œuvre de recettes, c'est une grande mutilation … L'idée que l'IA pourrait incarner la perfection du raisonnement et de la décision est donc une profonde illusion. Tout au plus, et c'est déjà merveilleux, l'IA peut nous aider à prendre de meilleures décisions, au même titre que d'autres outils … Mais prendre une décision c'est précisément tout sauf mettre en œuvre une procédure … C'est arbitrer et juger …
Le grand danger me semble donc être de se méprendre sur ce qu'est l'IA et de nous automutiler en refusant, en dénigrant ou en ignorant la dimension vivante et informelle de jugement et d'arbitrage qui réside au cœur de l'intelligence.

9/ Quel avenir prédisez-vous à la fragilité, au hasard, à la lenteur ?

Tant qu'il y aura des vivants et des vivants humains, la fragilité sera au centre. La fragilité ou la vulnérabilité c'est bien sûr parfois une mauvaise chose qui nous apporte de la souffrance, de la frustration, lorsqu'on se blesse ou qu'on échoue … Mais la vulnérabilité c'est aussi la possibilité d'être affecté, de ressentir, de vivre ! Et donc de vivre des choses positives et épanouissantes comme des choses terribles et difficiles … S'il faut bien sûr réduire la souffrance dans la mesure du possible, cela ne doit pas revenir à ne plus vouloir être affecté, être affectible, à renoncer à ce merveilleux mystère du fait d'être quelqu'un, de ressentir, d'être touché par le monde et par autrui.

10/ Comment imaginez-vous que notre souveraineté pourrait réunir au sein d'une même nation, des personnes qui voudraient vivre ensemble autour de la même vision de l'IA ? Idéalement au service de l'Homme.

De mon point de vue de philosophe des sciences ou de la connaissance, je crois que nous pourrions progresser en nous clarifiant les idées sur ce que veut dire connaître, savoir, être rationnel, produire des affirmations qui « tiennent la route ». Puisque l'enquête pure est une illusion, produire des connaissances, ça ne peut pas vouloir dire produire des « vérités ». On ne peut (malheureusement ?) jamais être totalement sûr, il y a toujours une part de confiance, de jugement et d'arbitrage informel. Cela est peut-être traumatisant car il aurait été fantastique de pouvoir saisir la vérité, une sorte de pouvoir prométhéen de connaissance … Mais il est important de reconnaître la réalité de notre situation : connaître, produire des connaissances, des affirmations qui tiennent la route, ça ne peut probablement pas vouloir dire grand-chose d'autre que d'essayer de faire le travail de réflexion du mieux possible, avec humilité, en utilisant les guides et les outils à disposition. Les

approches scientifiques ne sont pas infaillibles, mais sont a priori dignes de confiance … Il est légitime, sauf à détecter un problème d'intégrité ou une autre difficulté (comme un enfermement dogmatique) dans une communauté scientifique, de faire confiance aux (de supposer vrais les) résultats produits par cette communauté (ce qui fait consensus dans cette communauté).

Mais du coup, bien faire le travail, plus ou moins bien faire le travail, c'est quelque chose qui n'appartient pas qu'aux sciences … On devrait réaliser que nous sommes toutes et tous appelés à cela, à faire du mieux possible le travail lorsqu'on affirme quelque chose, et ce quel que soit le domaine. Cela s'applique en particulier aux domaines politique et éthique, à l'éthique des technologies, aux réflexions sur ce que nous devons faire à propos de l'IA. Dans ces domaines, on ne peut pas juste dire « c'est ton avis, tu es libre de penser ce que tu veux … c'est mon avis et je le partage … chacun a bien le droit de penser ce qu'il veut ». Bien sûr, et conformément à l'esprit des Lumières, chacun doit se faire son idée … Mais cela ne veut pas dire que chacun peut penser ce qu'il veut dans son coin … Cela me semble plutôt vouloir dire que chacun a le devoir de faire le travail de réflexion de la meilleure des manières possibles, le plus honnêtement et sincèrement possible. Et quand le travail est jugé bien fait, les affirmations qui en résultent sont considérées comme « tenant la route ».

En appliquant cette manière d'aborder les questions politiques et éthiques, on pourrait vite se rendre compte, si on prend le temps, qu'il y a de nombreuses choses qui tiennent la route à dire, en particulier sur le thème de l'IA. Et parmi toutes ces choses qui tiennent la route, certaines sont parfois en contradiction ou en tension. Mais cela ne veut pas forcément dire que les tensions reflètent des oppositions entre les personnes, que quelqu'un a forcément tort ou que chacun peut bien penser ce qu'il veut dans son coin. Une tension peut

aussi refléter un dilemme ou une complexité inhérente à la question explorée … Si c'est le cas, il est possible de se mettre d'accord sur l'existence de la tension … C'est même il me semble un devoir … On pourra ensuite éventuellement diverger quant aux priorités à établir face à de telles tensions. Mais il est crucial de partir d'une base commune qui reconnaît les tensions, de faire commun sur l'ensemble de ce qui tient la route … un commun faillible mais essentiel si l'on veut avancer ensemble …

C'est cet esprit que nous avons essayé de mettre en pratique dans le projet NHNAI. Les synthèses des échanges disponibles sur le site web du projet font la part belle à l'exposition de points de complexité et de tension qui ont émergé des discussions, afin de permettre de construire du commun pour avancer ensemble dans une exploration, non pas de la question de la place de l'humain à l'ère de l'IA, mais bien plutôt de la question de ce que veut dire être humain, afin de mieux être capable demain de décider quelle IA nous voulons, quelle doit être la place de l'IA dans le monde des humains et des autres vivants.