

Les intelligences artificielles génératives (IAg) est devenue une force incontournable dans notre monde moderne, mais elle soulève des questions éthiques et réglementaires cruciales. L'AILouvain réunit en ce 9 novembre 2023 trois intervenants afin d'explorer ces préoccupations complexes : **Mieke de Ketelaere** (Ingénieure), **Mark Hunyadi** (Philosophe), **Mathieu Michel** (Secrétaire d'état).

L'introduction par **Steve Tumson** (coprésident @ ALLouvain) met en avant l'utilisation omniprésente des IAg dans notre vie quotidienne et appuie l'importance d'adopter une perspective multidisciplinaire pour aborder ces questions liées à l'IA, perspective qui se traduit entre autres dans la composition du panel.

Durant la discussion modérée par **Benoit Simon** (secrétaire général @ ALLouvain), plusieurs points clés sont abordés :

1. Définition de l'IA : **Mieke** souligne que les IAg sont mal définies. Leur objectif est flou, et leur impact sur les êtres humains est insuffisamment testé. **Mark** ajoute que le terme "intelligence" est problématique. Il ne faut pas assimiler une machine cognitive à l'intelligence, car cela réduit l'intelligence à une simple manipulation de données. L'intelligence humaine se manifeste dans les questions qu'on se pose et dans notre relation à ce qui n'est pas ... et non ce qui est. C'est par exemple ce qui nous permet d'imaginer un autre monde et d'orienter nos actions en fonction de nos valeurs (les valeurs n'étant pas une donnée formalisable). En considérant l'IA comme de l'intelligence, il existe un risque de dévalorisation de l'intelligence humaine en la réduisant à un concept purement fonctionnel.

2. Régulation de l'IA et transdisciplinarité : **Mieke** souligne la rapidité de l'évolution des IAg et leur omniprésence dans de nombreux domaines, ce qui rend la régulation extrêmement complexe. Actuellement des groupes de travail interdisciplinaires existent afin d'évaluer les impacts de l'IA mais ce n'est pas toujours suffisant. Par exemple, même des psychologues et psychiatres sont pris de court sur les développements de l'IA et leur capacité à manipuler l'esprit. **Mark** insiste sur l'importance de réfléchir aux IAg avant qu'elles n'atteignent le marché, afin d'éviter des situations d'urgence et de panique, comme lors du déploiement de ChatGPT. Pour lui il y a une différence fondamentale entre cette innovation et toutes celles d'avant : les IAg ont un impact direct sur le cerveau humain. **Mathieu** n'adhère pas complètement à ce point de vue : il préconise de réguler les applications des IAg plutôt que la technologie elle-même. Pour lui, l'IA est une innovation comme les autres qui apporte des dangers et des bienfaits. Il faut des lois qui canalisent les dérives potentielles de l'IA, et une confiance des citoyens envers ces lois. Etant donné la vitesse à laquelle change l'IA, il suggère une approche de gouvernance réactive "just in time". **Mathieu** propose la création d'une agence européenne de l'algorithme pour relever les défis de la régulation de l'IA avec des groupes de travail transdisciplinaires.

Les intervenants réfléchissent également à comment équilibrer les droits à l'emploi sur un marché du travail en évolution, étant donné une robotisation et automatisation croissante. Bien que chaque innovation crée et fait disparaître des jobs, **Steve** rappelle que les derniers rapports de l'OCDE sur l'emploi mettent en avant une polarisation grandissante du marché du travail.

Mathieu suggère d'adapter la fiscalité et de ne plus taxer directement le travail.

4. Éducation et Protection de l'Esprit Humain : Les intervenants mettent l'accent sur l'importance de l'éducation. Pour **Mieke** il faut apprendre aux utilisateurs que l'IA n'est pas comme une calculatrice, qu'on ne peut pas partir du principe qu'elle nous donne une réponse juste. **Mark** insiste sur l'importance d'apprendre la réflexion au-delà des données. C'est nécessaire afin de sortir du cockpit numérique. Il faut laisser respirer l'esprit qui est étouffé par les désirs factuels que le numérique permet d'exécuter. **Mark** propose de protéger l'esprit humain en tant que patrimoine commun de l'humanité, similaire à la protection des fonds marins.

Cette discussion souligne que les IAg sont un outil puissant, mais que leur utilisation soulève des questions profondes sur comment nous nous définissons en tant qu'humains, comment nous définissons l'IA, comment nous la régulons, et comment nous nous éduquons face à cette technologie. **Pourrons-nous adapter les utilisations de la technologie à notre humanité, avant que notre humanité ne s'adapte à la technologie ?**

L'humanité a toujours pu s'adapter à la technologie, pourquoi ne pourrait-elle pas le faire maintenant ?

- **Mark** : différence fondamentale – impact sur le cerveau humain
- **Mieke** : il serait dangereux de s'adapter à la technologie de l'IA

Question pour Mathieu : vous dites que personne n'aurait pu prédire le suicide suite à une mauvaise utilisation d'ELISA. Où étaient les psychologues dans votre entourage qui eux voient ce genre de problème venir depuis bien longtemps ?

- **Mathieu** : on est en relation avec plusieurs disciplines mais les impacts sont dans tous les domaines → on crée des groupes de travail pour évaluer l'ensemble des impacts mais c'est compliqué, c'est du travail à pouce levé.
- **Mieke** : même dans une équipe transdisciplinaire, des psychologues et psychiatres sont pris de court sur les développements de l'IA et leur capacité à manipuler l'esprit

Questions pour Mark et Mieke : comment éduquer ?

- **Mieke** : pas forcément au niveau des enfants
- **Mark** : d'abord il faut éduquer les éducateurs. Education à la transcendance, aller au-delà du donné, éduquer à sortir de la bulle numérique

Mark : définir esprit humain et comment le protéger ?

- Traditionnellement opposer esprit et nature
- Tout ce qui n'est pas causale même si c'est causé
- Capacité à se relier à ce qui n'existe pas
- Que faut-il protéger ? Il faut laisser respirer l'esprit qui est étouffé par la factualité du désir que le numérique permet d'exécuter

Quid des outils qu'on développe mais qui pourraient tomber entre des mains moins sympathiques / régimes problématiques

- **Mathieu** : en effet c'est angoissant. Même en Belgique les politiciens se laissent déjà avoir au piège du « c'est possible donc c'est bien », par exemple le ministre de la mobilité souhaitait détourner les caméras de vitesse pour reconnaître les personnes qui sont sur leur gsm au volant.

Mieke : Comment établir un équilibre par rapport au droit à l'emploi et aux taxes ?

- Chaque innovation crée et fait disparaître des jobs
- Robot tax ? Si la main d'œuvre est remplacée par les machines : moins de revenus fiscaux sur le travail
 - **Mathieu** : taxer le capital plutôt que le travail
- **Steve Tumson** : notons que sur base du dernier rapport de l'OCDE(*) sur l'avenir du travail, on constate que (1) le marché de l'emploi se polarise, (2) les inégalités se creusent et (3) la qualité des emplois soulève de plus en plus de préoccupations.
(*) OECD, [Perspectives de l'emploi de l'OCDE 2019 : L'avenir du travail](#)

Ne pourrait-on demander à une IAG de proposer un texte de régulation efficace ?

- **Mathieu** : on a déjà essayé, mais le résultat est actuellement très décevant

Questions du public posées à Mathieu Michel

Nous n'avons pas reçu de réponse à ce jour.

1. Question à Mr Michel : lorsqu'il a été évoqué le potentiel manque de diversité des « Digital Minds » (issus principalement d'industriels, de fédérations et de groupes d'intérêts), vous avez évoqué l'existence d'autres groupes de travaux pluridisciplinaires sur certaines thématiques pour évaluer les impacts. Peut-on avoir plus de précisions sur ces groupes : quelles sont les thématiques, qui sont les personnes qui les composent, quelles sont les questions traitées, quelles sont les résultats de leurs travaux ? Et pourquoi ces informations ne semblent pas disponibles publiquement ?
2. Dire qu'il faut éduquer à l'esprit critique comme rempart des IAG, c'est une façon de décharger la responsabilité chez les utilisateurs (comme évoqué par Mark pour le consentement) ! Quid de la part de responsabilité des entreprises ou des Etats ?

Questions du public posées à Mark Hunyadi

1. Quid de la réalité qui devient de plus en plus floue à l'aide des deep fake / fake news ? Quel nouveau rapport au monde cela engendra-t-il ?

MH : ChatGPT ne peut pas faire la différence entre probabilité et vérité : pour lui, les deux se confondent. Et ce, parce qu'il repose sur ce qu'on appelle la "sémantique distributionnelle", qui ne repose que sur la probabilité d'apparition de séquences de mots. Mais le problème, c'est que ce qu'il dit fait autorité, et donc semble prétendre à la vérité (alors que ce n'est qu'une probabilité). Et donc, effectivement, cela fait problème dans notre relation au monde : l'IA accentue cette tendance déjà à l'œuvre dans les réseaux sociaux, de ne plus faire la différence entre corrélation et causalité. Dans cet univers de probabilisme généralisé, les individus ont tendance à n'entendre que ce qu'ils veulent entendre, et donc à renforcer leurs convictions dans leur bulle numérique.

2. Mark parle de la finalité de l'IA de pomper nos données. Mais les modèles collaboratifs ne sont-ils pas une forme d'intelligence collective ?

MH : Oui, Wikipedia en est l'exemple, et c'est un espoir: cela montre un usage possible du numérique. Mais le modèle actuellement dominant, celui du "capitalisme de surveillance" (Zuboff), est radicalement inverse de cette vision collaborative du numérique. Pour l'instant (mais cela pourrait changer, si on le voulait vraiment !), l'innovation technologique et surtout son emprise sur la société est aux mains du marché, et c'est un problème majeur. Voulons-nous laisser l'avenir de l'esprit humain aux puissances du marché ?

Questions du public posées à Mieke de Ketelaer

1. Si ChatGPT a connu un tel succès lors de son lancement, n'est-ce pas parce que la fonctionnalité répond à un besoin de la société ? Quel est-il? « C'est le marché qui lance ces technologies » Sans savoir/prédire ce que l'on peut en faire? Curieux, non?

MdK : ChatGPT 2 was open access, ChatGPT 3 was closed and behind an API, ChatGPT 3.5 was open to all ... The way ChatGPT3.5 was launched was a well planned approach which is part of a bigger competition between different AI Research Labs and teams. It has less to do with a need in the market, but has all to do with the competition which was launch in previous century triggered by Turing, claiming that human dialogues are the highest form of human intelligence. It is the AI game, nothing more, nothing less. Only, nobody seems to be able to define the endpoint of the game, nor the valid starting point of the game. We all know in the mean time that there is much more to human dialogues (a statement which goes way back in history by the way) to human intelligence.

2. Tester les IAG d'accord, mais comment ? Quels sont les points à tester et quelles personnes les définiraient ? Passer de quelques beta-testeurs à des milliers voire des millions a de fortes chances de mener aux mêmes situations que maintenant. De plus, l'entraînement de ces IA prend déjà en compte un certain feedback humain (RLHF pour chatGPT) faut-il plus contraindre ça quitte à diminuer le succès de ce type d'IA qui peuvent avoir énormément d'applications ?

MdK : Great question and it is indeed a very difficult step to cover. But if the alternative is no testing at all as it is today for systems such as ChatGPT, my preference would go to way more sandbox environments where systems can be tested under certain conditions in a controlled manner and on the different disciplines implicated: how is it used by younger generations, what is the potential risk of manipulation and addiction, what filters are needed towards certain historical facts which might be biased in the systems training data, etc.

3. D'un point de vue technique, la blockchain peut-elle être considérée comme un mécanisme renforçant la confiance, en particulier lorsqu'elle s'applique au domaine de l'intelligence artificielle ?

MdK : Although blockchain can help in create better tracing in the exchange of input/output between different AI systems in distributed environments for example, it will not help to increase the confidence into the outcome of a singular AI system.

COMMENTAIRES DU PUBLIC PENDANT L'ÉVÈNEMENT

- Sauf erreur de ma part l'explication scientifique (mathématiques) du fonctionnement de l'iag n'existe pas encore...l'iag serait une technologie en amont de la science et non issue de la science ... si c'est le cas ce serait une première qui interroge..
- L'utilisation de ChatGPT, mobilisé ici, est étroitement liée à la protection des données et à la cybersécurité. Bien que je trouve la réglementation du GDPR imparfaite et que j'émette des critiques à son égard, elle établit néanmoins une fondation sur laquelle nous pouvons construire et améliorer.
- J'ai noté le côté très humain de cet échange, avec une dose de malentendu, de traits de génies, de raccourcis, et d'enthousiasme, qui donnent le sel de ce genre débat.
- Pour être acceptée par la société civile, les technologies se doivent de créer la confiance en remplissant 2 conditions de base: a) les technologies doivent être fiable B) il existe un ensemble de régulation sur leurs usages C'est le ça par exemple pour l'aviation Aujourd'hui aucune de ces 2 conditions n'est remplie par l'iag...
- C'est écrit dans le popup sur chat.openai.com, il faut juste lire l'avertissement et l'utiliser judicieusement: Tips for getting started • Ask away ChatGPT can answer questions, help you learn, write code, brainstorm together, and much more. • Don't share sensitive info Chat history may be reviewed or used to improve our services. Learn more about your choices in our Help Center. • Check your facts While we have safeguards, ChatGPT may give you inaccurate information. It's not intended to give advice.

RETOUR EN IMAGE SUR L'ÉVÈNEMENT (VIDEO)

