



AAIE - IHEDN

ASSOCIATION DES ANCIENS AUDITEURS
DE L'IHEDN - **INTELLIGENCE ECONOMIQUE**



« QUELQUES RÉFLEXIONS SUR LE SAVOIR & LA MODERNITÉ... »

COMMISSION « SAVOIR & MODERNITÉ »
Février 2025

L'association nationale des auditeurs en intelligence économique de l'IHEDN fait partie de l'Union-IHEDN — de la communauté IHEDN.

L'association fête ses 25 ans en 2025,
elle fait vivre une communauté de plus de 200 auditeurs.

Fondée en 1999, l'association des auditeurs en intelligence économique (AAIE-IHEDN) contribue aux missions de l'Institut des Hautes Études de Défense Nationale. Elle rassemble près de deux cents praticiens, cadres et dirigeants d'entreprises, fonctionnaires de l'état et relais d'opinion représentant un éventail très large de métiers et de secteurs.

« QUELQUES RÉFLEXIONS SUR LE SAVOIR & LA MODERNITÉ... »

Une publication de la Commission « Savoir et Modernité » de l'AAIE - IHEDN
Edition février 2025

Illustration couverture : © Geralt - pixabay.com

www.auditeurs-ihedn-ies.org

La vocation de L'AAIE-IHEDN est de :

- ▶ Maintenir et développer les liens entre les auditeurs ;
- ▶ Promouvoir l'intelligence économique, sociétale et stratégique dans tous les secteurs d'activités de la société et en diffuser une véritable culture au plan national ;
- ▶ Apporter son concours à l'IHEDN pour l'accomplissement de sa tâche dans le domaine de l'intelligence économique, de la défense et de la sécurité nationale ;
- ▶ Apporter des méthodes et outils d'aide à la décision aux instances gouvernementales, notamment en matière de sensibilisation et de formation ;
- ▶ Contribuer à la réflexion sur les enjeux liés à la défense, la sécurité et la souveraineté nationale.

La Commission « Savoir & Modernité » a été animée pendant 2 ans par Dominique LAMOUREUX et Louis COLIN. Ses membres sont :

- Jérémie ABOIRON
- François ARNOULD
- Emmanuel BLOCH
- Louis COLIN
- Yves de MESTIER
- Pierre ENGELBACH
- Alexandre GAMBUTO
- Pierre JALLEY
- Claire JULIEN-VAUZELLE
- Dominique LAMOUREUX
- Alexandre LAURENT-ATTHALIN
- Igor MAZABREY
- Mathilde NIANG
- Delphine PICAVET
- Yannick REBOUL
- Arnaud SERS
- Juliana THIERRY
- Xavier TOUZEAU
- Floran VADILLO

L'intelligence économique et stratégique est fondée sur l'information, le savoir qui en découle et la raison critique qui permet de la situer. À l'aube d'une nouvelle modernité et dans un monde hyperconnecté, le volume de connaissance s'accroît à mesure que sa vitesse de diffusion s'accélère. Si l'humanité n'a probablement jamais disposé d'autant de savoir, elle semble parfois démunie pour organiser et défendre ce fruit de la raison, faisant courir le risque d'une crise de la connaissance sans précédent.

Entre études scientifiques, articles de presse, tribunes d'opinion, publication sur les réseaux sociaux, communication marketing ou politique, les messages coexistent, s'entremêlent et parfois se concurrencent. Comment les distinguer ? Les faire dialoguer ? Comment retrouver la raison critique, organiser le savoir et penser une épistémologie du XXI^e siècle ?

« Ni optimisme ni pessimisme, juste les yeux ouverts sur le monde »

Marguerite Yourcenar

Avant de s'interroger sur la question du « *savoir face à la modernité* », il est sage de tenter de caractériser le concept de modernité ou mieux encore de post- ou hyper-modernité comme les sociologues définissent les temps présents.

La célèbre citation d'Antonio Gramsci semble résumer ce début de millénaire.

« *Le vieux monde se meurt, le Nouveau Monde nouveau tarde à apparaître et dans ce clair-obscur émergent des monstres* »¹. Le citoyen découvre avec angoisse un monde qu'il maîtrise de moins en moins et qu'il comprend de plus en plus difficilement.

Les grands discours structurants du siècle passé ont menti que ce soit celui des forces politiques ou des courants religieux. Ils ont servi de catalyseur à des épisodes qui font honte au genre humain.

Le socialisme n'a pas abouti au « *Grand Soir* », mais a déployé des goulags. Le libéralisme n'a pas offert une croissance sans limite, mais un réchauffement planétaire. Les grandes religions ont sombré soit dans une crise (liée aux scandales d'agressions sexuelles) soit dans des radicalismes, source de terrorisme.

Ce clair-obscur inquiète légitimement les citoyens, trouble leurs esprits et obsède leurs échanges. Avec juste raison, car ce qui caractérise la postmodernité (ou hyper-modernité) c'est qu'elle ne s'inscrit plus – comme lors des décennies précédentes – dans des séries de crises ou de mutations conjoncturelles, mais dans des ruptures structurelles et radicales qui suscitent un nouvel espace-temps.

Les ruptures géopolitiques

Ainsi, la France n'est plus au centre de cette carte du monde qui – à côté du tableau noir – trônait au milieu de la classe d'école. La planète s'organise dorénavant autour de l'Océan Pacifique avec les États-Unis, la Russie et la Chine qui forment les trois puissances impériales. D'autres grandes puissances comme l'Inde, le Brésil ou l'Australie deviennent également des acteurs incontournables. Sans oublier ce nouveau « *Global South* » qui veut aussi jouer sa partition sur cet échiquier planétaire.

¹ Cahiers de prison (1938)

Or dans cette nouvelle cartographie, la France est à l'extrémité de la carte, tout au bord de la marge, comme prête à sortir de l'Histoire. Francis Fukuyama avait tort et il y a bien une histoire après la fin de l'Histoire, mais dont notre patrie des Lumières ne tient plus ni la plume ni le sabre ! Il faut accepter cette douloureuse blessure narcissique.

Les conférenciers d'aujourd'hui aiment utiliser le sigle « VUCA » pour définir le monde moderne. Ce sont les premières lettres des mots Versatilité, Incertitude (uncertainty en anglais), Ambiguïté et Complexité. Ces qualificatifs sont désormais familiers et imprègnent quotidiennement les réflexions et les débats occidentaux.

Les ruptures technologiques

Par ailleurs, la transformation numérique en cours de déploiement va conduire à des bouleversements encore plus considérables des systèmes de référence. La révolution qui s'engage aujourd'hui sera-t-elle comparable à celle causée par l'invention de l'imprimerie ou même par la découverte du feu ?

La capacité à recueillir des milliards d'informations et à les traiter en temps réel grâce à d'opaques algorithmes qui disent le monde et formatent les comportements va probablement changer la place de l'Homme dans l'univers, en bouleversant les relations qu'il entretient avec ses congénères, les institutions et son écosystème.

Faut-il aller jusqu'à accepter les prédictions du scientifique britannique Stephen Hawking qui annonce : « *le développement d'une intelligence artificielle complète pourrait demain mettre fin à l'humanité* » ?

Les fantasmes prémonitoires de Stanley Kubrick ne se sont pas réalisés en 2001 dans son « *l'Odyssée de l'espace* ». Toutefois, les SALA ou systèmes d'armes létales autonomes sont déjà en développement dans les entreprises de défense. L'intelligence artificielle s'immisce déjà dans les environnements de travail, et demain de plus en plus, dans tous les processus décisionnels. Des algorithmes sont déjà en cours d'écriture pour figer des comportements normés qui écarteront tous ceux qui seront désignés comme « *déviants* », au nom de la sécurité ou de l'efficacité économique.

Les ruptures institutionnelles

Ce qui est particulièrement préoccupant, c'est que cette révolution technologique majeure s'accompagne parallèlement d'une crise structurelle des institutions de référence. Dans un inventaire d'apocalypse – réunissant croissance démographique, réchauffement climatique, criminalité économique, pandémies meurtrières et

terrorisme sans frontière – émerge un espace planétaire qui laisse sans pouvoir les instances de gouvernances traditionnelles.

Dans ce pandémonium de l'incertitude, il faut, en effet, constater l'inadaptation irrémédiable des institutions laborieusement et méticuleusement édifiées par les démocrates du passé. Les systèmes de représentation politiques ou syndicaux ont perdu leur légitimité institutionnelle. Les structures hiérarchiques et pyramidales occupées par les élites ne savent plus fonctionner, produire, décider et convaincre. Chaque jour se confirme que les pyramides du pouvoir sont d'abord des tombeaux.

Les grandes institutions consacrent dorénavant toute leur énergie – non plus à créer de la valeur ou de nouvelles idées – mais à chercher à justifier leurs raisons d'être. Engagées dans des combats désespérés pour exister ou même subsister, les élites refusent de voir que la majorité de leurs mandants leur ont retiré toutes leurs confiances.

Pour illustrer ce constat : le « *baromètre de la confiance* »² indique que 74 % des Français ont un sentiment négatif vis-à-vis de la politique, 81 % considérant que les responsables politiques ne prennent pas en compte l'avis des citoyens.

Dans ce baromètre, on découvre, par ailleurs, que les syndicalistes suscitent autant de confiance que les banquiers avec un taux de seulement 20 % ! Quelle ironie !

La confiance dans les entreprises – petites et moyennes – en revanche est de 78 %. Le monde des valeurs démocratiques est renversé.

Face à cette vacuité, s'ouvre la « *Société Liquide* », décrite par Sigmund Baumann. Ce vide, cette absence tragique de panneaux indicateurs créent des angoisses profondes et laisse les citoyens face à des constats définitivement anxiogènes. Se développent à loisir des mouvements idéologiques réactionnaires. Il ne s'agit pas seulement de l'Amérique de Donald Trump, mais d'un nombre croissant de pays européens où l'extrême droite est au pouvoir. Pendant ce temps, des religions réactionnaires et archaïques imposent leur joug à des peuples entiers.

Les concitoyens désorientés se méfient de la philosophie des lumières et de la foi dans le progrès et la science qui a marqué depuis plus de 400 ans l'Occident. Ils ont surtout cessé de croire que leurs enfants auront un meilleur avenir que le leur. C'est le « *crépuscule de l'innocence* »³.

L'avenir ne peut plus se concevoir à partir de la seule extrapolation des certitudes ou des connaissances passées et actuelles, mais dans une projection d'idées et de concepts radicalement novateurs, y compris s'ils paraissent a priori improbables. Ce constat impose d'admettre que les ruptures du monde sont inéluctables, sans possible retour nostalgique à un passé idéalisé. Il impose, au contraire, de s'engager avec courage et confiance dans une réflexion absolument radicale et utopique qui appelle à des formes de gouvernance renouvelées.

² de CEVIPOF/Opinionway de février 2024

³ Jean Toutain

Et dans cette démarche, il ne faut évidemment pas se laisser engluier par les bureaucraties stérilisantes, les instructions infantilisantes et les principes de précaution déshonorants. Il est critique de s'inscrire dans le désordre du vivant et de l'utopie, tout en tenant d'une main ferme la raison et le rationalisme pour ne pas sombrer dans de nouvelles barbaries. Plus encore, il est nécessaire de s'appuyer sur une idée forte qui veut que « *moins on a de certitudes, plus il est nécessaire d'avoir des convictions* ». Or, à l'évidence, s'il est de plus en plus hasardeux d'avoir des certitudes dans un monde liquide, il est d'autant plus urgent et indispensable d'avoir des convictions infrangibles.

Cette démarche exige du courage. Beaucoup de courage. Mais comme nous y invitait Gandhi « *Nous devons être le changement que nous voulons voir dans le monde* ».

Le savoir dans la modernité

L'intelligence économique et stratégique se fonde sur l'accès à des données qui, une fois contextualisées, produisent de l'information, dont découle parfois de la connaissance : la compréhension de règles, de modèles, de logiques qui permettent d'appréhender son environnement, voire d'anticiper en partie son évolution.

À l'aube d'une nouvelle modernité et dans un monde hyperconnecté, le volume de données s'accroît à mesure que sa vitesse de diffusion s'accélère. Et si l'humanité n'a probablement jamais disposé d'autant de connaissances, elle semble parfois démunie pour organiser et défendre ce fruit de la raison, faisant courir le risque d'une crise du savoir sans précédent.

Entre études scientifiques, articles de presse, tribunes d'opinion, publication sur les réseaux sociaux, communication marketing ou politique, les messages coexistent, s'entremêlent et parfois se concurrencent. Comment les distinguer ? Les faire dialoguer ? Comment retrouver la raison critique, organiser le savoir et penser une épistémologie du XXI^e siècle ?

Des remises en question

Fake news, défiance généralisée envers les experts et les institutions qui les hébergent, polarisation des débats... D'où viennent ces maux caractéristiques de notre modernité ? Peut-on encore les attribuer à des aveuglements individuels lorsqu'ils touchent tant de nos contemporains et n'épargnent plus les esprits les plus brillants ?

Il semble au contraire qu'il existe une pluralité de facteurs qui ont mené à cette situation de remise en question du savoir, des experts et des institutions qui le produisent. Pour lutter contre le mouvement actuel, il convient de les mentionner ici, sans prétention d'exhaustivité ni ordre de prévalence :

- ▶ Hausse de l'éducation et du degré d'individualisation : La hausse continue ces dernières années du niveau de diplôme moyen a permis une augmentation générale de la connaissance et une plus forte individualisation⁴ qui conduit les individus à affirmer leur autonomie intellectuelle et leur esprit critique. Cette tendance s'accompagne d'une remise en cause plus vive des autorités, institutions et prescriptions sociales ou morales.
- ▶ La « trahison » des experts : Mode de gouvernance désormais répandue, la « technocratie », la mise en œuvre d'un gouvernement des experts, accompagne la première révolution industrielle. Défendue d'abord par Saint-Simon, cette doctrine semble aujourd'hui avoir gagné toutes les enceintes : des parlements aux plateaux de télévision. Or, les « promesses » des experts se heurtent aujourd'hui à des constats anxiogènes : le fantastique progrès technique et la croissance économique illimitée ne neutralisent pas le vertige éprouvé face à la vulnérabilité écologique et les risques existentiels qui pèsent désormais sur l'Humanité⁵. Dans ces conditions, comment croire des experts dont les compétences n'ont manifestement pas suffi à prévenir les catastrophes ?
- ▶ Un « progrès » dévoyé : De manière quelque peu similaire, le progrès des Lumières, entendu comme l'accumulation de connaissances scientifiques comme condition de l'émancipation humaine, et des libertés collectives s'est vu progressivement dévoyé au point qu'aujourd'hui le progrès est surtout technique, visant l'amélioration de l'efficacité de systèmes, sans considération véritable pour ses éventuelles implications sociales ou environnementales. Dans ce mouvement, la science et la recherche semblent de plus en plus soumises à des logiques de rentabilité économique, et ce parfois jusqu'à certaines formes d'instrumentalisation. Certains scandales ont démontré que la méthode scientifique pouvait également servir à produire des faits alternatifs visant à semer la confusion sur l'état des connaissances actuelles⁶. Si la science peut être détournée et les scientifiques corrompus, comment les croire ?
- ▶ Des médias partiaux : Source principale d'accès à l'information, la presse, les journaux, les médias télévisuels influencent largement les opinions et participent à la production de savoirs. Or, la très importante concentration des médias français fait naître des doutes légitimes sur leur impartialité et leur neutralité faisant ombrage à l'entière confiance de la profession et altérant la confiance en ses travaux.
- ▶ La « révolution » numérique : L'accès aux réseaux sociaux modifie profondément l'accès à l'information et la hiérarchisation des connaissances. Si les causes sont relativement bien identifiées, les effets peuvent être discutés. Sans arrêter ici de positions trop hypothétiques, il est possible de noter que les réseaux sociaux permettent l'accès à un volume de données et à un nombre de sources diverses

⁴ Voir par exemple : https://www.lemonde.fr/idees/article/2023/08/28/pierre-brechon-politiste-l-individualisme-a-nettement-regresse-en-europe_6186782_3232.html

⁵ Kemp and Al., Climate endgame : Exploring catastrophic climate change scenarios, PNAS, 2022; disponible ici : <https://www.pnas.org/doi/epdf/10.1073/pnas.2108146119>

⁶ Un exemple particulièrement bien renseigné est celui de l'industrie du tabac : <https://cnct.fr/ressource/actualites/lacademie-du-mensonge-de-lindustrie-du-tabac-inventeur-de-faits-alternatifs/>

immenses, qui pourrait permettre, tout chose égale par ailleurs, l'augmentation du savoir global des utilisateurs. Néanmoins, certains mécanismes réduisent la validité de cette affirmation. Premièrement, les architectures des réseaux sociaux reposant souvent sur des modèles publicitaires, elles ont vocation à prioriser le contenu sensationnaliste et conforme aux attendus réels ou supposés de l'utilisateur, ce qui a pour effet d'enfermer ce dernier dans des « chambres d'échos » au contenu pauvre. En second lieu, il n'existe pas sur les réseaux sociaux de hiérarchisation des sources et du savoir – articles scientifiques, communication marketing ou politique, commentaire particulier – toutes les opinions se valent et leur mise en avant ne dépend pas de leur qualité objective, mais de leur potentiel de viralité. Enfin, force est de constater que les réseaux sociaux alimentent la pratique du zapping rapide, ce qui favorise un rapport à l'information très superficiel.

Quelles pistes pour demain ?

Face aux problématiques énoncées, il convient de s'interroger sur les différentes mesures à privilégier pour lutter contre la mésinformation et la désinformation, renforcer la confiance dans les institutions et permettre une mobilisation du savoir favorable à l'émancipation humaine. La diversité des problématiques identifiées induit le recours à de nombreux outils, aucun n'étant par lui-même suffisant. Sans prétention d'exhaustivité et parmi d'autres éléments, il conviendrait donc :

- ▶ De renforcer l'offre de formation à l'esprit critique et à l'épistémologie, notamment depuis le collège jusqu'à l'université, de façon à permettre aux individus d'être mieux à même d'évaluer la qualité d'une information. Cette formation ne peut aller sans un approfondissement de l'enseignement aux sciences humaines – notamment la philosophie, la sociologie et l'histoire – anciennement défrayées comme les Humanités, et qui doivent permettre de mieux contextualiser les informations auxquelles un individu est confronté. De manière similaire, il semble important de mettre en lumière les controverses scientifiques dans les enseignements scientifiques.
- ▶ Soutenir les politiques visant à réduire la concentration des médias et le respect par les médias bénéficiant d'une exposition importante des obligations de pluralisme. Ainsi que l'enjoignait le programme du Conseil National de la Résistance : « *assurer la liberté de la presse, son honneur et son indépendance à l'égard de l'État, des puissances d'argent et des influences étrangères* ».
- ▶ Favoriser des actions spécifiques destinées au grand public afin de l'alerter et de l'acculturer aux menaces de manipulations et de désinformations.
- ▶ Favoriser la transparence des éventuels liens d'intérêts des experts ayant une exposition médiatique importante en rendant obligatoire la mention de ces liens dans les médias.

- ▶ Démocratiser la décision publique, à l'échelle locale et nationale, par exemple au moyen de conventions citoyennes. Ces démarches participatives doivent faire dialoguer experts et citoyens, sans placer les premiers au-dessus des seconds. De la même manière, il convient de démocratiser l'accès à l'Office Parlementaire des Choix d'Évaluation Scientifiques et Technologiques (OPECST).
- ▶ Favoriser le dialogue entre sciences dures et sciences humaines et l'intelligence collective en instaurant des institutions accessibles aux citoyens en charge d'organiser des débats sur les controverses scientifiques.
- ▶ Soutenir les politiques visant à favoriser la transparence et l'encadrement des algorithmes de recommandation utilisés par les plateformes numériques. Favoriser les réglementations visant à rendre obligatoire pour les plateformes d'offrir la possibilité à l'utilisateur de définir en tout ou partie ces logiques de fonctionnement ou à tout du moins, rendre obligatoire le fait de proposer un mode de consultation des réseaux sociaux purement chronologiques ou basés sur l'intérêt scientifique des publications évaluées par les utilisateurs.

POURQUOI PAS UN « PISA » DE L'ESPRIT CRITIQUE ?

Contribution : Igor Mazabrey

Croire n'est pas savoir. Pourtant, nous nous en remettons bien souvent à notre « croyance » dans le savoir d'experts. Avant la conquête de l'espace et ses merveilleuses images, nous « savions », depuis plus de 2000 ans, que la terre n'est pas plate et même, grâce à Eratosthène, que sa circonférence est d'environ 40 000 km. Or combien d'élèves ingénieurs sauraient le démontrer comme leurs glorieux anciens ?

Connaissance n'est pas vérité. Nous approchons la réalité par nos sens, par nos théories. Celles-ci confrontées à l'expérience, au progrès sont, hélas pour notre confort intellectuel, souvent vouées à être remises en cause. « La nature n'a pas pour le vide une horreur indicible », conclut Pascal, bouleversant le crédo des académies de son temps.

La recherche de la vérité, l'appétit de comprendre, de savoir ont donc toujours été un parcours semé d'embûches pour l'humain, souffrant de ses propres limites (d'instruction par exemple), de sa difficulté d'accès à l'information, de ses biais cognitifs...

La modernité vient en apparence résoudre cette frustration. Internet apporte les connaissances, l'intelligence artificielle apporte les raisonnements, fermez le ban ! Tout cela en quelques minutes et pour un effort et un coût marginal. Une victime collatérale : l'esprit critique.

Dans ce maelstrom, le citoyen libre doit être capable de croiser les sources d'information, d'identifier les motivations et intérêts, louables ou non, des émetteurs, d'admettre la dissonance dérangeante de certaines approches disruptives, bref de profiter de la modernité sans rendre les armes de sa liberté.

L'esprit critique est là pour maintenir la raison éveillée. Or celui-ci n'est pas inné. Pire, notre nature grégaire et la pression de notre environnement nous poussent vers des groupes à la pensée homogène, souvent exclusive. La collectivité doit porter une part de son enseignement. C'est clairement une mission centrale du système éducatif français. Présent dans de nombreux programmes d'enseignement, la formation des élèves à l'autonomie critique existe.

Est-elle à la hauteur des enjeux, adaptée à la puissance des nouveaux outils ? Cette question est cardinale pour l'avenir de notre pays.

Nous pourrions l'apprécier en comparant la formation française à l'esprit critique à celles d'autres pays. Un benchmark pourrait être pertinent. Allons plus loin. Si l'on évalue le niveau en mathématiques, en langue maternelle, pourquoi ne pas évaluer l'esprit critique des jeunes Français aux différents stades de leur scolarité ?

Rêvons même un peu et lançons l'idée d'un « PISA » de l'esprit critique des enfants du monde !

« C'EST VRAI ÇA », UNE INITIATIVE CITOYENNE QUI LUTTE CONTRE LA DÉSINFORMATION SUR LE RÉSEAU LINKEDIN.

Yoan BLANC est co-dirigeant de « C'est vrai ça ? », une initiative citoyenne indépendante qui réunit de simples particuliers animés par l'envie de lutter contre les fake news. Il est également professionnel dans le renseignement en sources ouvertes (OSINT) qu'il exerce à son compte. Nous l'avons interviewé pour mieux comprendre leur action.

Propos recueillis par Alexandre GAMBUTO le 8 septembre 2023

Bonjour, Yoan, pouvez-vous nous raconter la genèse de « C'est vrai ça » et nous expliquer son fonctionnement ?

« C'est vrai ça » est une association qui a été fondée par Sylvain Tillon (spécialiste en pédagogie numérique) et par Guy Nagel (avocat) pendant le premier confinement. Ils ont d'abord participé au projet « CopyPost » qui avait pour but de débusquer les copiés-collés. Puis le concept a été amélioré avec le projet actuel. La période du COVID a engendré une forte vague de désinformation et l'envie de lutter contre les fake news leur est apparue naturellement.

Le principe est extrêmement simple. Si une publication LinkedIn semble contenir des informations trompeuses ou erronées, vous citez « C'est vrai ça » en commentaire et nos bénévoles vont « débunker » ce qui a été dit. Cet anglicisme signifie « démystifier » et s'apparente à du fact-checking ou de la vérification d'informations. Le but est de dire si l'information est vraie ou fausse en apportant des arguments objectifs et des sources fiables. Notre posture est toujours la plus neutre possible. On s'autorise parfois quelques traits d'humour dans nos réponses aux spécialistes en « neurobullshit », à savoir, les coachs qui profitent de la faiblesse des gens pour dire des âneries.

Votre association est constituée de bénévoles uniquement ?

Tout à fait. CVC a pris de l'ampleur, car LinkedIn s'éloigne de plus en plus du réseau professionnel et que les publications deviennent un grand fourre-tout. Avec le temps, nous sommes devenus plus nombreux pour faire face. Sur la vingtaine de bénévoles actuelle, seulement trois personnes ne sont pas anonymes. Par ailleurs, le management interne est assez horizontal.

Vous attaquez-vous à tous les sujets ?

D'abord, on ne répond pas quand c'est l'auteur de la publication qui nous taque. Certains auteurs essaient d'augmenter la visibilité de leurs publications par ce subterfuge.

Ensuite, on évite d'intervenir sur un article politique ou un message d'opinion pur. En revanche, si cette opinion s'appuie sur des faits qui sont faux, on intervient.

Quels sont les profils des désinformateurs ?

Il y a plusieurs niveaux. Pendant le confinement nous avons surtout des désinformateurs candides et c'est encore aujourd'hui la plus grande masse. Dans ce cas on désamorce les messages avec une posture pédagogique.

Vous avez ensuite les désinformateurs stratégiques derrière lesquels vous retrouvez des entreprises, des états, ou des organismes paragouvernementaux. Ceux-là produisent de la désinformation suffisamment bien faite pour que les gens la partagent. Ces désinformateurs avancés s'appuient sur le fonctionnement des algorithmes qu'ils connaissent très bien et qu'ils exploitent, ils misent aussi sur la viralité de repartage. Ces acteurs ont clairement un but d'influence de l'opinion. « C'est vrai ça » n'a pas l'ambition de s'attaquer à cette désinformation plus sensible qui nécessite un niveau de sécurité opérationnelle supérieure. Sur ces niveaux stratégiques, il faut être très prudent, car il y a des enjeux géopolitiques très forts, et que l'on s'expose à des menaces. Sur LinkedIn néanmoins, il n'y a pas beaucoup de grands stratèges.

Comment faites-vous pour être garant d'une crédibilité indiscutable ?

La première des choses que je dis aux bénévoles qui nous rejoignent, c'est que l'on peut se tromper dans le travail de débunkage. Cela peut arriver, car on est humains et l'on ne peut pas être des spécialistes de tout. Dans ce cas, on s'excuse. Notre posture est claire, on débunke toujours une histoire, mais jamais une personne, on s'attache à des faits objectifs, nous ne devenons pas des adversaires systématiques. Sur certains sujets cela nous arrive de ne pas savoir ; on ne peut pas être affirmatif dans notre réponse dans ce cas. Pour garantir des réponses de qualité sur des sujets précis, on peut aussi compter sur nos experts bénévoles (agronomes, médecins, chercheurs en biologie, avocat, etc.) dont c'est le métier. Enfin, nous avons un processus interne de validation de nos réponses qui est collectif et qui implique plusieurs niveaux d'expérience au sein des bénévoles. Ce processus permet de garantir une bonne qualité de réponse.

Quelle est la réaction des personnes qui postent des informations fausses, lorsque vous les débunkez ?

C'est variable. Il y a des gens qui nous ignorent, d'autres qui suppriment nos commentaires (dans ce cas-là nous avons des astuces pour pouvoir les republier sans qu'ils le sachent), d'autres qui initient une discussion avec nous. Certains s'excusent et corrigent leur publication. Il est amusant de constater que certaines personnes nous ont bloqués de manière préventive...

Pourquoi n'êtes-vous présent que sur LinkedIn ?

On a déjà du mal à traiter la désinformation sur LinkedIn donc cibler Facebook c'est partir en dépression en 15 jours. Le volume de désinformation est tel que l'on ne pourrait pas suivre.

Est-ce que vous considérez que c'est normal que ce soit à des citoyens d'effectuer ce travail bénévolement ? Est-ce que vous ne palliez pas le travail défaillant des plateformes ?

Oui c'est anormal. Mais au même titre que « Les Restos du cœur » pallient la solidarité nationale. Nous pourrions trouver de nombreux exemples dans le milieu associatif. Malheureusement les plateformes se désengagent de toute modération...

Est-ce que les services publics, les cellules de fact-checking des grands médias ne peuvent pas contribuer ?

Malheureusement, les États sont occupés par ailleurs et les cellules de fact-checking des quotidiens nationaux sont déjà monopolisées pour traiter de gros sujets d'actualité... et elles sont très peu présentes sur LinkedIn (on les trouve plus facilement sur X, Facebook et TikTok).

Selon vous, est-ce que le règlement européen Digital Service Act (DSA) qui vient d'être mis en application va changer quelque chose ?

Je n'y crois pas vraiment, il va y avoir quelques ajustements, mais à la marge. Le seul réseau qui risque vraiment des sanctions c'est X (ex-Twitter), car Elon Musk fait n'importe quoi. Ça n'engage que moi, mais ma position c'est qu'il y a un manque de courage politique sur ces sujets. Les réseaux sociaux sont une exception à la loi, ils ne sont pas considérés comme des éditeurs de contenus, donc ils ne sont pas responsables de la production qui y est faite. Mais n'oublions pas qu'il n'y a pas que les GAFAM qui ont une responsabilité dans la diffusion de la désinformation par une mauvaise modération. Les grands groupes produisent aussi beaucoup de désinformation.

Pour ceux qui voudraient se lancer, comment vous rejoint-on ?

On lance des campagnes de recrutement régulièrement, sinon il faut contacter quelqu'un de « C'est vrai ça » pour proposer ses services. Nous sommes maintenant assez vigilants, car nous avons déjà eu de l'entrisme. Il existe désormais différents de « responsabilité » au sein de l'association, en fonction des débunks réalisés et de la participation à l'association.

Je précise qu'il n'y a pas de contrôle de l'implication des bénévoles. On fait comme on peut et quand on a le temps, à côté de nos vies professionnelles. À l'arrivée, on forme les personnes sur les techniques de débunkage. Gardons en tête que plus on sera nombreux, plus notre action sera efficace.

Quel message voudriez-vous faire passer aux décideurs qui nous liront ?

Il faut absolument éduquer et sensibiliser les enfants dans les écoles à rechercher de l'information et à lutter contre la désinformation : qu'est-ce que c'est, comment la détecter, qu'est-ce que les acteurs cherchent à faire ? Cela pourrait être une intervention dès le niveau collège avec une récurrence de 3 ou 4 fois dans l'année. Personne n'est à l'abri de la désinformation, mais avec l'éducation, on limite les risques.

La désinformation, ça désagrège le tissu sociétal, c'est une arme pour fracturer les sociétés. Je vais prendre une image pour être plus clair. On est dans un énorme feu de forêt de désinformation et « C'est vrai ça » ce n'est qu'une couverture pour éteindre les braises. Le seul moyen de lutter à la racine, c'est d'éduquer les gens. L'esprit critique a réellement disparu de la circulation surtout chez les jeunes générations, et c'est très inquiétant...

NE FAUT-IL PAS QUESTIONNER LES SYSTEMES D'IA GENERATIVE ?

Raja CHATILA est roboticien, professeur émérite d'intelligence artificielle et d'éthique des technologies à Sorbonne Université. Il est l'auteur de plus de 170 publications internationales sur ces sujets. Au regard de notre sujet de réflexion, il est en particulier coprésident du groupe de travail sur l'IA responsable au sein du Partenariat mondial sur l'IA (GPAI) et membre du Comité national français de pilotage pour l'éthique numérique (CNPEN). Il a été également membre du groupe d'experts de haut niveau en IA de la Commission européenne (HLEG-AI). En mars 2023, il figure parmi les signataires d'une pétition publiée par le think tank Future Of Life Institute. Cette pétition demande de mettre temporairement en pause le développement des systèmes d'IA générative les plus avancés afin de prendre du recul sur la fiabilité de ces technologies, de mesurer les impacts sur la société, et de mettre en place des mesures de prévention du risque. Cet entretien a été réalisé pour mieux comprendre les problèmes sociétaux posés par ces systèmes conversationnels au regard de l'exactitude et de la fiabilité des réponses générées.

Propos recueillis par Alexandre GAMBUTO le 07 septembre 2023

NDLR : depuis la réalisation de l'entretien, plusieurs des outils les plus avancés ont ajouté des capacités à faire référence aux sources, et le texte de l'IA Act a été publié au Journal officiel (JO) de l'Union européenne le 12 juillet 2024.

Raja Chatila, bonjour. Pouvez-vous nous expliquer pourquoi les IA génératives suscitent autant de commentaires cette année ?

L'intelligence artificielle est mon domaine depuis toujours et je constate aujourd'hui une évolution. Que se passe-t-il ? Depuis plusieurs mois, vous avez des systèmes d'IA générative très puissants, qui sont facilement disponibles pour le grand public. L'outil ChatGPT d'OpenAI fait partie de ces nouveaux systèmes. Historiquement et jusqu'à maintenant, les systèmes d'IA étaient plutôt utilisés pour interpréter des données, c'est-à-dire extraire des connaissances qui étaient enfouies. À titre d'exemple, la compréhension automatique des images est un cas tout à fait classique, pouvant être appliquée à la vidéosurveillance urbaine ou au contrôle automatisé sur une ligne de production industrielle. Ça, c'est le domaine de l'interprétation. Si l'on cherche à déterminer « est-ce qu'un évènement futur risque d'arriver en fonction d'évènements passés ? », nous sommes dans un domaine très classique de l'IA statistique, celui de la prédiction.

Aujourd'hui, l'IA générative est d'une autre nature. Elle ne se borne pas à interpréter des données, mais elle produit du contenu en grand volume. Quelle est l'idée ? Lors d'une requête (prompt), le système va produire des données qui sont statistiquement les plus probables au regard de la demande de l'utilisateur et de ce que l'IA a appris. Le contenu produit a toujours pour caractéristique de ressembler au corpus d'apprentissage, puisque l'IA est basée sur un modèle statistique obtenu en phase d'apprentissage, et ce qui se corrèle le mieux avec ce modèle sera sa réponse.

Le premier changement majeur introduit par les IA génératives de dernière génération, c'est que les domaines d'apprentissage sont devenus extrêmement vastes, cela représente tout le contenu que les concepteurs du système ont pu récolter sur Internet, ce qui donne un contexte très large pour interpréter les requêtes. Cette notion « d'extension du contexte » est centrale puisque cela rend possibles des réponses très larges sur à peu près n'importe quel sujet.

Le deuxième changement majeur avec des outils tels que ChatGPT, c'est que l'on dialogue en langage naturel et avec une qualité rédactionnelle de très bonne qualité du côté de la machine, ce qui est assez nouveau. Or la langue, c'est le véhicule de notre pensée et de notre logique humaine. C'est à travers elle que l'on échange des informations, des connaissances, des vérités entre humains et que l'on partage des significations communes du monde.

Quels sont les problèmes avec les IA génératives ?

Ils sont nombreux et je vais en évoquer cinq principaux...

PROBLÈME N° 1 : LA PRODUCTION DE FAUSSES INFORMATIONS

À cause du fonctionnement corrélatif du système, la réponse générée sous forme de texte va pouvoir mélanger sans distinction du contenu vrai et du contenu faux. Ce mélange est souvent indiscernable aux yeux du lecteur, car tout apparaît comme « plausible ». Le texte généré est souvent de très bonne facture, car il est bien rédigé et bien construit, ce qui renforce le crédit qu'on lui prête. Si le lecteur humain ne connaît pas la réalité, il est pour lui impossible de discerner le vrai et le faux. Cela représente un réel danger pour nos sociétés, car la vérité partagée est essentielle, c'est notre fondement commun. Sa remise en cause engendre forcément des conflits, et malheureusement, la généralisation de ces outils risque d'engendrer une diffusion de contre-vérités de plus en plus large. C'est une menace directe sur notre cohésion sociale, et la démocratie ! Si vous vous connectez à l'interface d'OpenAI vous pouvez d'ailleurs lire la mise en garde « ChatGPT may produce inaccurate information about people, places, or facts », mais cela n'arrête pas pour autant les gens de l'utiliser.

Je vais maintenant évoquer un exemple pour illustrer mon propos. J'ai demandé à ChatGPT d'écrire une notice biographique sur moi-même. Après lecture de sa réponse, j'estime que le texte comportait environ 50 % de vrai et 50 % de faux. L'erreur est qu'il me prêtait des fonctions erronées qui concernent des personnes que je côtoie, mais pas moi. Comment est-ce possible ? Parce que mon nom apparaît régulièrement aux côtés de ces personnes et que le système mélange mes attributs avec les leurs.

Ce type de confusion est rendu possible, car le système est basé sur l'analyse d'occurrence de mots dans des espaces à très grandes dimensions. Pour le dire autrement, ChatGPT constate la probabilité que certains mots se retrouvent à proximité d'autres groupes de mots (c'est en réalité au niveau des composants constitutifs du codage des mots que sont les « tokens »). C'est en suivant ces observations statistiques que le système génère des phrases entières : « je prédis que le prochain mot le plus probable est celui-ci, et je l'écris ». Par exemple, si j'étais ami avec Einstein et que beaucoup de textes font mention de son prix Nobel à proximité de mon nom, le système peut me coller l'étiquette « prix Nobel ». Si l'on signale manuellement ces erreurs au système,

lorsque les utilisateurs les détectent, il peut se rétracter... Ainsi ces systèmes peuvent inventer des faits qui n'ont aucune réalité en effectuant des corrélations fallacieuses, ce qui est appelé « hallucinations » par certains chercheurs.

PROBLÈME N° 2 : LA PROJECTION DE SÉMANTIQUE

Le second problème est lié au premier, toutefois il concerne le fonctionnement même du système, et non la qualité de sa production. Comme évoqué plus haut, le système utilise des chaînes de caractères (token) qui vont être corrélées statistiquement. Le système va ainsi produire du texte en se basant sur ces observations statistiques, il n'utilise à aucun moment le sens, la signification, des mots qu'il ne peut même pas appréhender. C'est très différent de la manière d'écrire d'un humain, car nous produisons du sens lorsque l'on écrit et exprimons des concepts dont nous comprenons la signification profonde. Tout ce que ce que ChatGPT sait, c'est que tel mot est associé statistiquement à un contexte. Il ne comprend pas le sens de ce qu'il produit !

PROBLÈME N° 3 : LES BIAIS LIÉS À LA BASE D'ENTRAÎNEMENT

Rappelons que ces systèmes d'IA générative sont entraînés sur de très grandes bases de données de texte, en plusieurs langues, issues des informations disponibles sur Internet. L'anglais est prépondérant dans ces bases d'entraînement. Pourquoi est-ce un problème ? Lorsque l'on pose une requête dans une langue peu représentée, la réponse va éventuellement passer par la langue pivot qui est l'anglais, et la réponse va être empreinte de la culture de la langue dominante utilisée pour produire la réponse avant sa traduction dans la langue de la requête. Ce problème sera d'autant plus marqué que la langue d'interaction sera minoritaire. Les réponses vont être marquées par le modèle de pensée de la culture anglo-saxonne, dont la culture de l'utilisateur peut être assez éloignée, et ainsi influencer cette culture.

PROBLÈME N° 4 : LES RESSOURCES NÉCESSAIRES POUR DÉVELOPPER CES TECHNOLOGIES

Nous sommes dans une situation un peu particulière. Historiquement, les nouvelles idées et les nouveaux systèmes naissent souvent dans le creuset de la recherche académique. Mais en intelligence artificielle telle qu'elle est aujourd'hui (statistique, générative), les moyens nécessaires pour développer ces technologies sont excessivement importants et dépassent ceux qui sont disponibles dans le monde académique. Ces technologies se développent ainsi dans le monde industriel, et encore, seulement dans les laboratoires de certains industriels qui ont les chercheurs, et ressources nécessaires en calcul et en données pour être à la pointe de l'innovation. On les retrouve surtout dans la Silicon Valley ou en Chine. Ce n'est donc pas la recherche scientifique (avec d'abord la motivation de faire avancer les concepts et les connaissances et éventuellement de diffuser une solution novatrice) qui stimule le développement de ces technologies, ce sont certains industriels privés. Il faut peut-être s'interroger sur comment donner plus de moyens à la recherche académique pour qu'elle puisse occuper une place plus importante dans le développement de ces technologies.

Un autre sujet que je ne ferai que mentionner ici est l'impact environnemental de ces technologies en général (émissions CO₂, consommation de ressources) qui nécessite un cadrage particulier.

PROBLÈME N° 5 : LA CITATION DES SOURCES

Les systèmes actuels d'IA générative sont alimentés par des corpus multiples présents sur internet, mais sont incapables, par construction, de mentionner les sources utilisées pour construire leurs réponses.

On a donc au final un agent conversationnel qui ne sait pas de quoi il parle, qui peut commettre d'importantes erreurs, qui semble cependant plausible, et dont on ne peut pas vérifier les sources...

Est-ce que le problème de la production de fausse information ne va pas se résoudre naturellement avec l'amélioration des systèmes ? Si c'est le cas : combien de temps faudra-t-il ?

Le problème initial c'est simplement la corrélation de choses différentes qui va produire du vrai et du faux mélangé. Ce n'est pas parce qu'il y a du faux au départ que ça produit du faux, c'est parce que le système fait des associations erronées. Et ce modèle de production de faux ne va pas se résoudre, car c'est inhérent à la nature corrélatrice du système. Il faudrait donc soit changer complètement de paradigme, soit ajouter des filtres pour vérifier la validité des sorties après coup, ce qui serait une révision majeure.

Le second problème c'est la qualité de la base d'entraînement. Plus elle contient des inexactitudes, plus les réponses seront potentiellement fausses. Or, on peut estimer qu'une partie de la production de ces systèmes va se retrouver sur Internet avec leurs erreurs. Les systèmes comme ChatGPT vont continuer à apprendre sur ce corpus numérique qui contiendra ce « faux supplémentaire » de leur propre fait. En définitive, le problème du faux va potentiellement être démultiplié, car ces systèmes vont avoir leur apprentissage utiliser des choses fausses qu'ils auront eux-même produites...

À ce sujet, il faut rappeler que le système d'OpenAI n'apprend pas en continu. L'entraînement du modèle actuel [NDLR : à date de l'entretien] a été arrêté en septembre 2021, donc le système ne connaît rien de postérieur. C'est lorsqu'ils vont mettre à jour le modèle que ce « faux supplémentaire » va entrer dans son apprentissage et va multiplier le risque d'associations fausses.

En dehors des retours des utilisateurs, OpenAI corrige aussi les choses, mais après coup. C'est un processus de « renforcement supervisé par l'humain ». En d'autres termes, ce sont des employés qui affinent, entraînent et corrigent. Certaines de ces personnes sont payées moins de 2 dollars de l'heure et sont basées au Kenya...

La seule manière d'avoir un niveau de réponse satisfaisant reposerait sur une démarche différente, ce que l'on ne sait pas faire aujourd'hui. Il faudrait pour cela sélectionner les sources les plus fiables, avoir des systèmes causaux (id est qui établissent des implications logiques entre des corpus et les réponses du système) et pas uniquement basés sur la corrélation (id est une analyse statistique des corpus), distinguer et valider les sources qui contribuent à l'apprentissage, le tout plus ou moins filtré par des êtres humains.

Malgré les efforts déployés par OpenAI et d'autres concepteurs de ces systèmes, ils sont encore largement imparfaits, et je pense qu'il n'y a pas d'issue pour résoudre les problèmes évoqués avec la technologie actuelle...

Pourquoi avoir signé la pétition du Future Of Life Institute ?

Pour toutes ces raisons que je viens de mentionner. Je ne suis pas forcément d'accord sur toutes les prises de position de ce think tank, ni sur tous les paragraphes de cette pétition, en particulier celui qui précise que l'IA pose un risque existentiel à l'humanité. Je ne le crois pas. En revanche : quand on m'a proposé de signer, j'ai été très intéressé par la proposition demandant l'arrêt du développement de technologies plus puissantes que GPT4 (le plus avancé à l'époque) pendant 6 mois, afin de mettre en place des mécanismes de protection, un moratoire donc, et que les États se mettent en avant pour légiférer s'il n'est pas respecté. Je suis d'accord sur la nécessité de ne pas se ruer dans une course effrénée dans laquelle chaque organisation développe un système encore plus puissant.

J'ai signé ce texte publié le 31 mars 2023 dans un contexte particulier. L'opinion publique et les médias se posaient peu la question des problèmes, on était plutôt dans une séquence d'émerveillement. Il fallait lever un carton rouge pour interpeller les médias, les politiques, le grand public et les rendre conscients de ces problèmes. C'est ça qui m'a motivé et je pense que j'ai eu raison. On parle beaucoup plus de ces sujets dorénavant.

Pourquoi les systèmes deviennent-ils populaires et performants seulement aujourd'hui ?

Le domaine traditionnel de l'IA est l'approche symbolique, celle qui avait le vent en poupe avant l'approche statistique. Cette dernière était peu performante et ne l'est que depuis les années 2010, permettant d'atteindre les résultats que nous connaissons maintenant.

L'idéal serait de remariager ces deux approches pour pallier un certain nombre de problèmes évoqués. Ce défi est extrêmement difficile, peut-être même pas faisable.

Selon vous, quel va être l'impact de la multiplication de ces assistants pour les professionnels ?

Ce genre de question est très difficile. Il y a tellement de données de contexte touchant aux volets économiques, géopolitiques et technologiques que l'on se trompe souvent. Il est clair que ces outils sont utiles, mais pour des spécialistes qui peuvent en apprécier la qualité. L'autre problème, c'est que ces outils ne produisent en fait pas grand-chose ! Les réponses de ChatGPT sont généralement plates et compter sur lui pour apporter une véritable connaissance, je n'y crois pas vraiment.

Est-ce utile pour la synthèse de textes dans certains métiers qui en manipulent beaucoup ? Par exemple pour les journalistes ou les professionnels du droit ?

L'histoire de cet avocat américain qui a suivi la synthèse juridique de ChatGPT, qui

contenait de nombreuses erreurs, pour construire sa plaidoirie parle d'elle-même. Parfois la vérification coûte plus d'effort que la production. Si un journaliste veut utiliser ChatGPT pour se documenter et générer un résumé d'un match de football, pourquoi pas. Mais sur une analyse politique ou un sujet de fond, je serais méfiant.

Et pour des étudiants ?

Surtout pas ! Etudier ce n'est pas recracher des textes. Il faut rapprocher, réfléchir, mémoriser, comprendre et ce mécanisme d'apprentissage est un cheminement. En tant qu'enseignant, je ne suis pas d'accord que mes élèves suivent cette voie. S'ils font cela, ils n'apprendront rien et souffriront des lacunes engendrées par l'utilisation de ces outils. Encore une fois, une utilisation par un spécialiste ou une personne avertie sur un sujet, pourquoi pas, mais pour une utilisation plus naïve, cela demande beaucoup de précautions.

Est-ce que la clé n'est pas dans l'éducation des utilisateurs ? Leur apprendre que ces technologies sont encore très faillibles et qu'elles ne remplacent pas un travail de bibliographie et d'étude. Mais qu'elles peuvent gagner du temps pour rechercher des ressources et synthétiser de l'information en première lecture ?

L'éducation est fondamentale pour utiliser ces outils avec discernement. Il faut apprendre aux gens quel est l'intérêt de ces outils, quelles sont leurs limites, comment cela fonctionne.

Mais l'éducation, aussi fondamentale qu'elle soit, ne fait pas tout. Si je vous donne un outil extrêmement puissant, mais que je vous mets en garde contre la plupart des utilisations. Que faites-vous ? Vous allez tout de même l'utiliser. On sait qu'il ne faut pas rouler trop vite, pourtant certains le font quand même. L'éducation ne protège pas de tout, mais contribue à diminuer le risque.

Il a fallu une quinzaine d'années pour protéger les données privées contenues dans les systèmes d'information (RGPD, 2016) et environ le même temps pour réglementer les réseaux sociaux (Digital Services Act, 2023). Sommes-nous partis pour réglementer plus vite l'IA générative et en particulier les agents conversationnels ?

Plus vite, je ne le dirais pas. Mais les choses ont commencé à bouger depuis un moment. Depuis 2020, l'Union Européenne ainsi que tous les États ont sorti leur plan pour encadrer l'IA.

En 2018, un groupe d'experts de haut niveau a été constitué, j'en faisais partie, pour réfléchir à un plan IA pour l'Union Européenne. Notre travail a débouché sur la production de deux documents : (1) des propositions d'orientations au regard du cadre éthique et (2) des recommandations pour l'investissement et la politique à mener.

La Commission s'est ensuite saisie de ce travail. Cela a abouti à la parution d'un livre blanc en 2020 et à la proposition d'une première version du règlement aux États et au Parlement, le 19 avril 2021. Fin 2022, le Conseil Européen l'avait amendée, et le Parlement a produit ses propres amendements le 14 juin 2023. Aujourd'hui, trois versions différentes coexistent donc et vont aboutir à une négociation tripartite (les trilogues)

entre la Commission, le Conseil et le Parlement. On peut espérer que ces négociations aboutissent à un texte définitif d'ici fin 2023. Vous avez ensuite généralement deux ans de délai entre la publication définitive et la mise en application. Nous pouvons donc espérer qu'en 2025 ou début 2026 nous aurons un texte applicable, sachant que tous les industriels vont probablement se préparer à être conformes à la réglementation en amont.

Est-ce que ce texte répond aux problématiques de l'IA en général, et celles évoquées plus haut en ce qui concerne l'IA générative ?

Plus ou moins selon mon point de vue. La question de la responsabilité des acteurs comme OpenAI qui produisent des modèles dits « de fondation » à partir desquels on peut construire des systèmes à usage général n'était initialement pas abordée dans le texte présenté en 2021. ChatGPT est arrivé après la publication du texte par la Commission. Dans la réflexion initiale, c'était l'usage prévu de l'IA qui définissait la notion de risque qui déterminait les exigences légales. Or, ChatGPT est un système non spécifique, car son usage est très général. Des modifications ont donc été introduites par le Conseil pour en tenir compte et par le Parlement qui a rajouté une surcote pour légiférer plus fortement sur ce type de systèmes et les modèles de fondation. La classe des « IA génératives » a ainsi été étayée. Si la Commission avait initialement estimé que les agents conversationnels présentaient un niveau de risque moyen, avec une obligation de transparence comprenant la nécessité de savoir que le système est une IA et pas un humain, le texte en discussion est plus exigeant maintenant, avec l'identification de niveaux de risque plus élevés.

Mais le texte final du règlement n'est pas encore connu et les négociations tripartites se poursuivent.

En tant que scientifique spécialisé du domaine, quel message voudriez-vous faire passer aux décideurs qui liront cet article ?

Je pense qu'il faut utiliser des outils avec une fiabilité garantie, reconnue, vérifiée, voire certifiée. Or, gardez bien en tête que ChatGPT ne fait pas encore partie de ces outils-là. Il faut aussi vérifier dans quelle mesure cet outil est utile et fait progresser l'intérêt de l'entreprise et des organisations. Il ne faut pas s'arrêter à de vagues promesses qui peuvent s'avérer fallacieuses ! Laisser la machine aller chercher la plus grande partie de l'information, je dis oui, à condition qu'on laisse l'humain en faire la synthèse et l'interprétation derrière. Je ne pense pas que beaucoup de métiers vont être remplacés par ces outils, car la compétence humaine pour comprendre et faire sens reste pour moi essentielle. Je pense qu'ils représentent un complément d'efficacité. Si ces outils doivent être utilisés en entreprise, j'incite à mettre en place des gouvernances appropriées dont le rôle est d'être garant de l'utilisation de ces systèmes, de la même manière, par exemple, que les délégués à la protection des données le sont vis-à-vis de la RGPD.

Raja Chatila, merci pour votre temps.

« QUELQUES RÉFLEXIONS SUR LE SAVOIR & LA MODERNITÉ... »

Une publication de la Commission « Savoir et Modernité » de l'AAIE - IHEDN
Edition février 2025

Illustration couverture : © Geralt - pixabay.com

www.auditeurs-ihedn-ies.org



En 2025, L'**UNION-IHEDN** fête ses **50** ans

Cette même année, L'**AAIE-IHEDN** fête ses **25** ans



Association des
Auditeurs en
Intelligence
Économique de
l'Institut des Hautes
Études de Défense
Nationale

www.auditeurs-ihedn-ies.org