

Que se cache-t-il d'essentiel derrière l'anecdote ChatGPT ?

par

■ Thierry Rayna ■

Professeur de l'École polytechnique, directeur de la chaire *Technology for Change*

En bref

Une fois de plus, la technologie nous offre une superbe bulle médiatique autour de l'intelligence artificielle (IA) 2.0 et du trublion du moment, ChatGPT. Sans surprise, les commentaires couvrent tout le spectre habituel : de ceux qui ont testé et prétendent que c'est une révolution qui va tout emporter sur son passage – à commencer par les emplois qualifiés – à ceux qui ont testé et tentent de prouver que cela ne marche pas, en passant par ceux qui ne connaissent pas, mais en parlent quand même. Thierry Rayna présente l'état des lieux pour mieux le dépasser et montrer ce qui se cache d'essentiel derrière ce buzz, non pas une menace sur les emplois qualifiés, mais une nouvelle vague d'ubérisation des entreprises, GAFAM comprises, menée cette fois-ci par des hordes de travailleurs indépendants lourdement armés de modules préentraînés dont personne ne sait rien, véritables boîtes noires.

Compte rendu rédigé par Yann Verdo

L'Association des Amis de l'École de Paris du management organise des débats et en diffuse les comptes rendus, les idées restant de la seule responsabilité de leurs auteurs. Elle peut également diffuser les commentaires que suscitent ces documents.

Parrains & partenaires de l'École de Paris du management :

Algoé¹ • Chaire etilab • Chaire Mines urbaines • Chaire Phénix – Grandes entreprises d'avenir • ENGIE • Groupe BPCE • Groupe CHD • GRTgaz • IdVectoR² • L'Oréal
• La Fabrique de l'industrie • Mines Paris – PSL • RATP • Université Mohammed VI Polytechnique • UIMM • Ylios¹

1. pour le séminaire Vie des affaires / 2. pour le séminaire Management de l'innovation

Je suis professeur de l'École polytechnique, au département Management de l'innovation, et chercheur au CNRS à l'Institut interdisciplinaire de l'innovation (i3). Je dirige également la chaire *Technology for Change* de l'Institut polytechnique de Paris. Économiste de formation, j'ai un côté geek qui me conduit à m'intéresser à ce qui va advenir plutôt qu'à ce qui s'est déjà produit. Mes recherches portent principalement sur l'impact des technologies numériques, qu'il s'agisse de la blockchain, dont je suis venu vous parler ici même il y a quelques années¹, ou de l'intelligence artificielle (IA) générative, sujet de cette séance. Quand une nouvelle technologie fait irruption dans le débat public, on se heurte toujours à un double écueil : ceux qui connaissent bien la technologie y voient une myriade d'applications – là où seules quelques-unes sont vraiment pertinentes –, ce qui tient du fantasme; ceux qui connaissent bien les marchés et les usages, mais mal la technologie, vont faire fi des limites de cette dernière – un détail! – et envisager, là aussi, moult usages, qui tiennent de la science-fiction. Face à ce double écueil, mon rôle est de guider et d'éclairer les entreprises, surtout au moment critique où elles baissent la garde parce que, trop prévenues d'une rupture frontale qui ne vient pas, elles en ignorent les ruptures latérales qui vont les déstabiliser.

IA générative : rupture ou continuité?

Le battage technologique se concentre aujourd'hui sur l'IA générative. Néanmoins, qu'y a-t-il de vraiment neuf là-dedans? Les débats enflammés sur l'IA sont apparus en même temps que l'IA elle-même, il y a plusieurs décennies! Cependant, il y a quelques mois, on a vu subitement émerger un nouvel acteur, OpenAI, le créateur de ChatGPT, dont on peut se demander s'il ne va pas devenir, à l'instar d'un Google en son temps, ce que les Anglosaxons appellent un *game changer*. Or, la technologie sur laquelle sont basés ChatGPT ou les générateurs d'images Dall-E et Stable Diffusion date de plusieurs décennies, les ajouts les plus récents remontant à 2017. Elle est connue de tous, notamment des GAFAM. Est-ce alors la taille des bases d'entraînement, la puissance de calcul, qui aurait fait un saut spectaculaire? Pas davantage, car comme le montre un remarquable graphique publié par le magazine *The Economist*, ils ont évolué à peu près conformément à la loi de Moore, avec une inflexion au début des années 2000, mais pas de manière à expliquer le soudain engouement. Est-ce alors le côté génératif qui est neuf? Eh bien non! L'IA générative repose sur les mêmes algorithmes qu'une IA "habituelle", mais utilisés dans le sens inverse. Pour la reconnaissance d'image, on nourrit habituellement l'algorithme de milliard d'images similaires – des chats, par exemple –, qui ne sont pour lui que de la "bouillie" de pixels, au sein desquelles il va falloir reconnaître des motifs de pixels particuliers – des patterns – qui seront associés à un mot clé – *chat*. C'est l'IA discriminative que vous utilisez tous les jours. L'IA générative fonctionne exactement de la même manière – même entraînement – sauf qu'au lieu de fournir en entrée une image et d'obtenir en sortie un mot clé, on fait l'inverse : on fournit en entrée un mot clé, que l'algorithme combine à des pixels aléatoires pour façonnner une image de chat. La génération de texte, avec ChatGPT, repose sur le même principe.

La vraie nouveauté : l'ouverture

Il n'y a donc absolument rien de nouveau qui explique ce buzz, sinon une cause de rupture fondamentale : l'ouverture! Là où Google, Apple, Amazon ou Facebook, pour ne citer qu'eux, conservaient jalousement leurs algorithmes d'IA sous cloche, OpenAI a décidé de faire l'inverse, en disant à chacun de nous : « *Vous voulez vous servir de notre IA? Allez-y! Et comme bon vous semble.* » Cela change tout! Nous avons basculé d'une situation dans laquelle l'IA était entre les mains d'une poignée d'entreprises, qui essayaient à tâtons d'en découvrir les usages pertinents, à une autre où tout un chacun peut s'en servir pour en faire l'usage de son choix. OpenAI

1. « Tout ce que vous devriez savoir sur les vrais usages de la blockchain », séminaire Transformations numériques et entrepreneuriales, séance du 13 mai 2019.

a fait sortir le génie de la bouteille. Là encore, ce phénomène n'est pas neuf, car nous avons déjà connu plusieurs fois des ruptures causées par l'ouverture d'une technologie numérique – Web 2.0, blockchain... – au plus grand nombre. L'IA générative n'est donc qu'une technologie numérique comme les autres.

Une nouvelle forme de capitalisme

Ces multiples précédents peuvent nous aider à prédire ce qui va se passer à moyen et long terme pour les entreprises, mais cela nécessite de prendre un peu de recul. De fait, toutes ces technologies numériques ont en commun de nous faire passer d'un monde – celui du capitalisme industriel – où les entreprises avaient le monopole des moyens de production (conception, production/fabrication, distribution, communication...), à un monde dans lequel tout individu, s'il le souhaite, a à sa disposition des moyens de design, de conception, de création/fabrication, de distribution et de communication. Et ce "simple" changement conduit lui-même à deux tendances absolument fondamentales.

Impact des technologies numériques pour les entreprises : l'ère de la prosommation...

La première tendance est ce que l'on appelle la *prosommation*. Un *prosommateur*, pour reprendre le mot-valise (combinaison de *producteur* ou *professionnel* et de *consommateur*) inventé dès 1980, avec une remarquable prescience, par le sociologue et futurologue américain Alvin Toffler, est un consommateur qui joue un rôle actif dans le processus de production, en participant au design, à la fabrication et à la distribution. Avant l'arrivée des technologies numériques, la prosommation n'existe pas – l'assemblage de meubles Ikea ou les magasins d'usine en sont deux des rares cas. Depuis l'avènement des technologies numériques, la prosommation ne cesse de s'étendre. Dans les années 2000, elle a d'abord touché les contenus : encyclopédies collaboratives en ligne, piratage de films ou de musiques, banques de photos comme Flickr, médias sociaux... En bref, tout ce que l'on a appelé à l'époque la disruption était de la prosommation. Dans les années 2010, la prosommation s'est répandue hors ligne et a touché les services, affectant les transports (Uber), l'hôtellerie (Airbnb), les services à domicile (Deliveroo), etc. C'est ce que l'on a appelé l'*ubérisation*. Maintenant, la prosommation s'étend aux objets : impression 3D, *dropshipping*²...

Les technologies numériques ne causent des révolutions que lorsqu'elles sortent des mains des entreprises pour être librement utilisées par les particuliers. Quand la rupture se produit, la technologie numérique en question est déjà utilisée depuis longtemps par les entreprises, mais dans la continuité et cela ne donne que des évolutions. Une fois entre les mains des particuliers, qui en découvrent les "vrais" usages, elles deviennent des révolutions. Cela s'est vérifié hier avec le Web, la musique numérique, la blockchain, etc., et cela se vérifie aujourd'hui avec l'IA. En effet, les entreprises, même les plus à la pointe, qui l'ont utilisée n'ont fourni que des évolutions de choses existantes, comme Alexa, Cortana, Google Assistant, ou encore Siri. Puis OpenAI a lancé son interface et les individus s'en sont emparés pour des applications auxquelles aucune entreprise n'avait pensé. Ce n'est qu'à partir de ce moment-là que l'IA est devenue révolutionnaire.

Cette vague de prosommation n'a rien d'anecdotique. En 1980, dans son livre *La Troisième Vague*, Alvin Toffler voyait dans l'arrivée de l'ordinateur personnel – seulement quatre ans auparavant – les prémisses d'un nouvel âge de l'humanité. Après l'âge agricole, puis l'âge industriel, nous voilà entrés dans l'âge de l'information, dans lequel l'organisation clé n'est plus l'entreprise, mais les communautés de prosommateurs armés de PC.

... et des plateformes reines

Parallèlement à ce phénomène de prosommérisation est apparue la deuxième tendance fondamentale évoquée précédemment : la plateformisation. Dans les années 2000 ont émergé des acteurs économiques d'un type radicalement nouveau, qui nous ont tout d'abord laissés ébaubis : une entreprise de contenus qui ne produit

2. Un *dropshipper* est un particulier qui, au lieu de monter une entreprise, précommande auprès d'une usine en Chine un produit qu'il met en vente sur Amazon Marketplace. Une fois les ventes réalisées, il paye l'usine qui expédie le produit directement : son investissement est nul, si ce n'est en temps.

pas de contenus (Facebook, en 2004), une entreprise d'hôtellerie qui ne possède pas d'hôtels (Airbnb, en 2008), une entreprise de taxis qui ne possède pas de taxis (Uber, en 2009), etc. Ces acteurs du troisième type sont les plateformes, dont l'émergence est étroitement liée à celle de la prosommation : sans elles, la prosommation serait un cauchemar! Imaginez trouver une location à Rome sans Airbnb. Il vous faudrait identifier vous-même les dizaines de milliers de sites Web – contenant tous des informations différentes – des prosommateurs offrant un logement à Rome, puis négocier et contractualiser individuellement avec chacun d'entre eux, sachant qu'une fois sur place, si ça se passe mal, tout recours est illusoire... Sans plateformes, la prosommation est impossible. C'est cette absorption des coûts de transaction (coûts de recherche, de négociation et d'exécution) qui donne leur force aux plateformes et qui leur permet d'obtenir la vraie ressource rare, le capital social, c'est-à-dire la confiance, les liens sociaux, les valeurs et les normes partagées, la réciprocité, la coopération, etc., que leur accordent des utilisateurs de plus en plus nombreux. Ces derniers attirent à leur tour d'autres utilisateurs, réduisant ainsi les coûts de transaction et augmentant le capital social, et conduisant à une concentration inéluctable : on ne veut pas trois Airbnb ou cinq Netflix, on n'en veut qu'un.

Or, cette notion de plateforme n'apparaît nulle part dans l'analyse, par ailleurs remarquablement juste, de Toffler. En effet, l'information, promulguée par Toffler comme ressource clé de ce nouvel âge, est – contrairement aux ressources clés de l'âge agricole (la terre) et de l'âge industriel (ressources non renouvelables) – tout sauf rare. Elle est, au contraire, tellement surabondante qu'elle n'a plus aucune valeur. Nous vivons dans "l'ère post-vérité" : ce qui compte n'est pas l'information elle-même – qui peut être factuellement fausse, bienvenue dans l'ère des fake news –, mais *qui* donne l'information. Ce pouvoir vient du capital social : c'est la confiance que nous avons envers l'acteur concerné, les liens sociaux que nous avons à travers lui, les valeurs et les normes que nous partageons avec cette personne ou organisation, la coopération et la réciprocité que cela entraîne... C'est ce capital social et non l'information qui est la vraie rareté et la vraie ressource clé de ce nouvel âge de l'humanité. Les organisations qui contrôlent ce capital social ne sont autres que les plateformes. L'âge de l'information décrit par Toffler n'est qu'une brève étape transitoire qui a eu lieu au milieu des années 2000 et a laissé sa place au vrai nouvel âge : l'âge numérique, contrôlé par les plateformes – là où les entreprises régnaient sur l'âge industriel – tirant leur force d'un capital social rare – là où les ressources non renouvelables étaient la clé du pouvoir dans l'âge industriel. Nous passons donc progressivement – il a fallu deux cent ans pour passer de l'âge agricole à l'âge industriel! – d'une ère de capitalisme financier à une ère de capitalisme social : c'est la concentration de capital social qui fait désormais toute la différence. À titre d'exemple, pensez au pouvoir des marques.

L'IA ouverte, un iceberg dont ChatGPT n'est que la partie émergée

C'est dans ce contexte de prosommérisation et de plateformisation que se déroule cette nouvelle ère de l'IA ouverte. Néanmoins, pour comprendre le rôle d'accélérateur et d'amplificateur que l'IA – générative, notamment – va jouer vis-à-vis de ces deux phénomènes, il faut revenir sur cette notion d'ouverture. Je vous ai expliqué que le seul élément véritablement novateur apporté par OpenAI et ChatGPT était leur ouverture. Or, aujourd'hui, OpenAI serait plutôt "ClosedAI", car toutes les caractéristiques techniques (sources des données, coefficients de poids de l'algorithme, stratégies d'apprentissage...) sont désormais tenues secrètes et l'apprentissage est fermé. Seul reste ouvert l'usage, mais déjà avec certaines limites imposées. La vraie IA ouverte n'est pas ChatGPT, mais la myriade de modules d'IA préentraînés, téléchargeables et exécutables localement qui est disponible en ligne. Ces modules open source d'IA peuvent être ensuite réentraînés et personnalisés par tout un chacun : un simple PC et quelques heures suffisent! Là où énormément de ressources sont nécessaires à l'entraînement initial d'un GPT, il se trouve – ce que personne n'avait anticipé – qu'une fois entraîné dans une optique généraliste, spécialiser un algorithme d'IA générative dans un domaine ou un autre – le droit, par exemple – ne nécessite que très peu de temps et de ressources.

Quelques considérations pratiques

Abordons maintenant quelques considérations pratiques répondant aux questions que l'on me pose le plus souvent. L'IA générative est-elle "bonne" : fournit-elle des réponses fiables? Pas vraiment... ou du moins pas

toujours ! Il n'est pas exagéré de voir dans ChatGPT un générateur de fake news, car l'algorithme énonce de manière préremptoire des choses qui peuvent être manifestement fausses. L'IA générative est totalement dépourvue de connaissances symboliques puisque ce ne sont que des probabilités. Les fameuses "hallucinations" de ces algorithmes ne sont pas des bugs, mais le produit naturel de leur fonctionnement même : des mots, des pixels tirés aléatoirement au sort selon des statistiques issues de leur apprentissage. L'IA générative n'étant jamais que de l'IA discriminative dont les algorithmes ont été inversés, elle se heurte aux mêmes limites, qui proviennent de l'absence de connaissances symboliques. Ces limites font que, par exemple, se faire comprendre de Siri tient souvent de la gageure, ou qu'il suffit qu'il y ait beaucoup de blanc en arrière-plan d'une image pour que l'algorithme voie dans n'importe quel chien un husky ! Néanmoins, elles posent moins de problèmes pour l'IA générative qui est faite pour créer – création certes imparfaite, telle une ébauche, mais qui peut être améliorée –, que pour l'IA discriminative. En effet, le but de cette dernière est de sélectionner, ce qui ne donne pas le droit à l'erreur : Siri ne peut pas faire tout à fait autre chose que ce qui lui a été demandé.

L'imperfection de l'IA générative ne l'empêchera cependant pas de causer des ruptures, tout comme, par exemple, le MP3 en son temps : sa qualité inférieure de restitution du son le rendait impropre pour les chaînes de salon, mais, dans un environnement bruyant (la rue, le métro...), cette limite devenait imperceptible et la taille faible de ses fichiers était alors hautement avantageuse. Toute déficiente qu'elle était, la technologie du MP3 s'est imposée, en induisant de nouveaux usages pour lesquels le référentiel de performance était différent. De tels nouveaux usages, au départ déviants, assureront de la même manière le succès de l'IA, et ce malgré son imperfection.

L'IA générative n'aura donc pas besoin d'être bonne pour déclencher une révolution. La vraie rupture viendra de la combinaison de cette IA imparfaite avec une intelligence humaine elle-même imparfaite, mais qui lui est complémentaire. Ce qui distingue ChatGPT de la concurrence en matière de pertinence et de qualité, c'est l'adjonction à un premier réseau de neurones entraîné sur un très grand corpus de textes d'un second réseau de neurones entraîné, lui, par une horde de travailleurs kényans – sous-payés – chargés de sélectionner les réponses les plus plaisantes parmi celles fournies par le premier réseau de neurones.

On peut se demander pourquoi les GAFAM n'ont pas proposé leur propre *LLM* (*Large Language Model*) avant OpenAI ? De fait, elles auraient pu : le T de ChatGPT se réfère d'ailleurs à Transformer, un modèle d'apprentissage profond mis au point par Google en 2016. Les GAFAM se sont abstenues de lancer leur propre interface ouverte justement parce que leur immense capital social leur interdisait de rendre publique une IA aussi inintelligente ! Les GAFAM ne peuvent pas être aussi bonnes qu'OpenAI parce qu'elles ne peuvent pas être aussi mauvaises : tout biais, toute erreur, et les voilà clouées au pilori.

Craintes pour l'emploi : l'arbre qui cache la forêt ?

Une autre antienne du moment est que les gens vont perdre leur travail à cause de l'IA générative. Cela n'a rien d'évident, étant donné ses limites. ChatGPT peut être un bon apprenti, mais c'est un apprenti qui n'apprend pas, ou plutôt qui n'apprend plus, car une fois que l'utilisateur a fermé sa fenêtre de dialogue, l'algorithme "oublie" les informations contextuelles qui lui ont été fournies et repart de zéro. En effet, OpenAI a trop peur que des groupes malveillants utilisent en masse cette capacité d'apprentissage pour lui faire apprendre des contrevérités ou des choses tendancieuses. Le risque est donc moins, pour ces experts humains que sont les employés, de voir leurs fonctions disparaître que de les voir évoluer et changer de nature. Une portion de leur travail, celle qui est la plus bêtement répétitive et ne requiert pas de "discriminer", sera déléguée à l'IA tandis que le reste, soit la majeure partie, restera entre leurs mains. Il ne faut pas oublier que la qualité de l'IA est asymétrique, car elle est très bonne pour créer et mauvaise pour sélectionner. On peut déjà anticiper que la quantité de contenus et d'information (e-mails, etc.) va exploser, mais que l'on aura de ce fait besoin de beaucoup plus de cerveaux pour faire le tri. Aussi paradoxal que cela puisse paraître, plus l'on numérise, plus l'humain acquiert de la valeur.

Ce débat sur l'emploi est cependant l'arbre qui cache la forêt. Ce ne sont pas les individus qui vont perdre leur travail, ce sont les entreprises elles-mêmes ! Le mouvement qui a vu tomber, sous les coups de boutoir du numérique, des éditeurs d'encyclopédies, des maisons de disques, des titres de presse, les sociétés de service, etc., va s'accélérer avec l'IA générative. La réaction des entreprises, quand elles se sentent menacées par

les technologies numériques, est de recourir à une utilisation toujours plus massive de ces mêmes technologies, pour améliorer leur offre existante – ce qui est nécessairement incrémental – ou de réduire leurs coûts, notamment en licenciant. Or, la rupture ne vient pas de l'amélioration de l'existant, mais de la découverte d'usages et de produits entièrement nouveaux, et de prime abord déviants. C'est de là que vient la révolution, portée par de nouveaux entrants. Par ailleurs, dans ce monde de prosommation, les personnes licenciées perdent leur emploi, mais pas leur travail. Les technologies et les plateformes leur permettent de continuer à "faire leur job", mais de manière indépendante. En effet, c'est clairement ce qui s'est passé avec les journalistes licenciés qui ont lancé leur propre média. En licenciant, les entreprises créent in fine de nouveaux concurrents. Les outils de décision stratégique utilisés par les entreprises (forces de Porter, marges, SWOT, etc.), conçus dans l'âge industriel, sont devenus caducs et doivent être remplacés par des outils adaptés à l'âge numérique.

L'épineuse question de la responsabilité et de la régulation

Voici une autre considération pratique très importante : l'aspect *no code* de l'IA générative. Une interface comme ChatGPT ne nécessite aucune compétence en matière de programmation pour générer du code informatique fonctionnel. Il est très facile de programmer une IA générative à faire quelque chose – enlever un filigrane sur une image, par exemple – sans avoir besoin, pour cela, d'écrire la moindre ligne de code : il suffit de lui montrer de façon répétitive un avant et un après pour obtenir le résultat désiré. C'est formidable pour les individus, mais il en résulte que les entreprises vont au-devant de problèmes majeurs en matière de responsabilité juridique. Autant un programme écrit en code peut être vérifié et ses modifications tracées, autant cela est impossible avec ces algorithmes d'IA. En effet, il n'est pas de boîte noire plus hermétique que ces matrices de milliards de lignes et de milliards de colonnes ne contenant que des nombres compris entre -1 et 1 ! Dans un monde où rien n'est plus aisément accessible que de télécharger un module d'IA open source et de s'en servir pour réentraîner l'algorithme en vue de lui faire accomplir une tâche spécifique, comment une entreprise pourrait-elle s'assurer du bon fonctionnement ou de l'innocuité de tel matériel, tel logiciel ou tel service qui lui est vendu par un fournisseur ? Qu'est-ce qui lui dit que l'algorithme d'IA ne va pas se mettre tout à coup à avoir un comportement totalement erratique ou dangereux ? Elle aura beau le tester et le retester des milliers de fois, elle ne pourra jamais en être certaine ! Tout cela soulève bien sûr la question de la nécessité de réguler. Cependant, si une plateforme comme OpenAI peut être tenu juridiquement responsable de ce que produit ChatGPT, c'est bien évidemment impossible dans le cas des modules d'IA open source qui constituent la partie immergée de l'iceberg, et sur lesquels le régulateur n'a et ne peut avoir aucune prise...

Quels impacts sur l'humain et les connaissances ?

Pour finir, voici deux dernières considérations. Comme mentionné précédemment, l'impact de l'IA sur l'emploi se traduira moins au niveau quantitatif – pertes d'emploi – que qualitatif – emplois changeant de nature. Si les tâches purement répétitives sont déléguées à la machine, ne laissant à l'humain que celles cognitivement intensives, ne s'expose-t-on pas à un risque encore plus élevé de burn-out ? l'IA ne va-t-elle pas faire de la semaine de quatre jours une nécessité ? La question est posée.

Il y a un autre risque qui est celui d'un abêtissement généralisé. L'humain est l'expert, l'IA l'apprenti, disons-nous. Néanmoins, pour qu'un humain puisse devenir un expert dans un domaine donné, il faut qu'il ait lui-même été un apprenti. Si ce rôle est entièrement dévolu aux machines, comment effectuera-t-il cet apprentissage censé le conduire à l'expertise ? À cela s'ajoute un dernier sujet de préoccupation formulé par certains chercheurs. Les *LLM* comme ChatGPT ou LLaMA ne produisent pas de nouvelles connaissances, mais ne font que naviguer – bien plus rapidement que nous – sur la vaste mer des connaissances existantes, qui sont toutes d'origine humaine. De ce fait, ces *LLM*, qui sont les premiers de leur génération, ne sont-ils pas appelés à être aussi les derniers, dans la mesure où l'immense corpus de textes sur lesquels ils ont été entraînés est d'ores et déjà abondamment pollué par les contenus – que l'on ne peut filtrer – générés par ces mêmes *LLM* ?

Débat



L'IA générative : ses limites, ses défis, ses opportunités

Un intervenant : *À vous écouter, il m'est revenu en tête cette phrase de Gaston Bachelard : « Pour un esprit scientifique, toute connaissance est une réponse à une question. » Quand je vois ChatGPT affirmer sans sourciller que 1 kilogramme de viande est plus lourd que 1 kilogramme d'air, je me dis que nous ne sommes pas près d'avoir des machines se hissant au niveau de l'intelligence et des connaissances humaines...*

Thierry Rayna : C'est vrai. Les algorithmes comme ChatGPT ne sont jamais qu'un jeu de dés. Tant qu'ils seront dépourvus de connaissances symboliques, ils resteront loin du compte. Cependant, j'insiste sur le fait que du point de vue des applications, cela n'a aucune espèce d'importance. Le niveau atteint par ChatGPT ou un autre LLM du même acabit est largement suffisant pour quantité d'usages. Si vous cherchez une recette pour faire une bonne mousse au chocolat, recourir à ChatGPT est une stratégie très efficace ! Pourquoi ? Parce qu'il existe une sorte de consensus sur les ingrédients et les quantités qui aboutissent à une bonne mousse chocolat. Quand un tel consensus existe, le jeu consistant à tirer au dé à chaque mot pour deviner le mot suivant, même s'il peut produire quelques hallucinations ici ou là, fonctionne dans l'ensemble assez bien.

Int. : *Face aux limites de l'IA générative, les entreprises ne vont-elles pas revenir à l'IA discriminative ?*

T. R. : De ce que j'ai pu observer, non. L'IA discriminative souffre toujours des mêmes faiblesses. On essaie bien d'améliorer les assistants personnels intelligents comme Alexa et consorts, mais leur progression est très lente, la principale difficulté étant l'absence totale de connaissances symboliques, qui explique que, contrairement à un humain, une IA puisse confondre un chiwawa et un muffin. De toute façon, même si les entreprises revenaient en arrière, l'IA générative est là, entre les mains des individus, et la rupture est inévitable. Encore une fois, le fait que l'IA générative ne soit pas une très bonne technologie n'est pas un problème : plein de nouveaux usages peuvent voir le jour avec une IA générative même pas très bonne, à partir du moment où elle est largement distribuée.

Int. : *La nécessaire validation des contenus, pour s'assurer qu'ils ont bien été produits par des humains, pourrait-elle s'effectuer via des blockchains ?*

T. R. : Je reste extrêmement dubitatif quant à cet usage de la blockchain. La validation des contenus est tout de même plus sûre quand elle est le fait d'intelligences humaines. En revanche, je vois pour la blockchain, outre le bitcoin, un autre usage possible, le seul que j'aie pu lui trouver en fait. Ces modules d'IA open source ne contiennent pas de code explicite. Ainsi, contrairement aux logiciels open source classiques, il n'est pas possible d'expliciter les changements au "code" – les fameuses matrices. Des acteurs malveillants pourraient réentraîner l'algorithme pour y introduire un biais, un comportement malveillant, voire une porte dérobée, et distribuer cette nouvelle version sans que ceci ne soit visible. On pourrait donc stocker ces fameuses matrices dans une blockchain. Chaque changement pourrait, de ce fait, être visible séquentiellement et être associé à un contributeur – ou groupe de contributeurs – particulier. Dans un tel modèle décentralisé, l'utilisation des blockchains peut donc avoir un intérêt.

Int. : *Je vous avoue mon scepticisme lorsque je vous entendez dire que, avec l'IA ouverte, il ne suffit que de quelques heures à un particulier équipé d'un simple portable pour réentraîner un algorithme à effectuer une tâche spécifique...*

T. R. : Il faut bien comprendre que la communauté open source ne fait que mettre en ligne des tableaux, qui ne sont rien d'autre que des fichiers Excel. Tout un chacun peut les télécharger et les mettre dans l'algorithme pour le faire se réentraîner. Ce qui était et qui reste très coûteux, c'est l'entraînement initial, à portée généraliste. Cependant, une fois que le génie est sorti de la bouteille, les résultats montrent que l'on peut réentraîner l'algorithme à faibles coûts.

Int. : Je souhaiterais vous faire revenir sur l'aspect no code. Il est vrai que ChatGPT peut apparaître comme le graal en la matière, puisque l'on s'adresse à lui dans un langage naturel. ChatGPT et les autres LLM ne vont-ils pas donner une deuxième chance à tout un tas de technologies restées bloquées en matière d'usage, car nécessitant des compétences de programmation pour être utilisées ?

T. R. : Je pense effectivement que cela peut avoir un impact positif sur certaines technologies présentant une barrière à l'entrée assez élevée, l'une d'entre elles étant à mon avis l'impression 3D. Le vrai coût d'entrée de cette dernière réside dans le manque d'interface, qui fait que seuls les geeks sont capables de scanner et d'imprimer un objet. L'accès plus simple offert par l'interface d'IA générative pourrait permettre à une technologie de niche comme celle-ci de se déployer.

Int. : Le caractère ouvert de l'IA générative constitue à mes yeux une formidable lueur d'espoir. Cela étant, toutes les cultures, toutes les parties du monde ne font pas forcément jeu égal en la matière. Comment l'IA générative va-t-elle s'accommorder de ces inégalités ?

T. R. : On peut raisonnablement penser que certains régimes autoritaires vont vouloir contrôler, d'une manière ou d'une autre, ces modules téléchargeables formant l'IA ouverte. Cela reste toutefois, techniquement parlant, extrêmement compliqué à réaliser. Il est clair que tous les pays qui se montreront particulièrement restrictifs en matière d'expérimentation ou qui essaieront de tout réguler, tout contrôler – sachant que c'est quasi impossible –, subiront un désavantage compétitif par rapport aux autres. On observe le même phénomène en sécurité informatique : faute de laisser tous les hackers du monde "jouer dans le même bac à sable", c'est dans les pays les moins régulés qu'ils le font... et qu'ils apprennent. Il en ira de l'IA générative comme du piratage : plus l'on cherchera à contrôler et limiter les usages, moins cela marchera. En réalité, il faut accompagner ces ruptures technologiques. Ce ne sont pas la Hadopi (Haute Autorité pour la diffusion des œuvres et la protection des droits sur internet) ou les dispositifs anti-copie qui ont fini par faire baisser le piratage, mais l'émergence du bon *business model*, avec Spotify, Netflix, etc.

Enseignement, entreprises... : le monde à l'heure de ChatGPT

Int. : Quel va être l'impact de ChatGPT et de l'IA générative sur l'enseignement et sur les écoles de commerce en particulier ?

T. R. : Très bonne question. L'IA générative aura sur l'enseignement le même impact que la numérisation, les moocs, etc. Ces derniers n'ont pas renversé la table parce que, dans le monde de l'enseignement, il faut qu'il y ait des acteurs de confiance, lesquels, pour le moment, continuent d'être les institutions d'enseignement supérieur. Cependant, le modèle de l'enseignant fournissant à ses étudiants de l'information descendante est de moins en moins d'actualité. Personnellement, face à mes propres étudiants, je me considère comme un coach. Je suis là pour les accompagner dans leur méta-apprentissage, les aider à apprendre à apprendre. L'IA générative s'intègre parfaitement bien dans ce tableau.

Int. : Vous venez d'évoquer l'évolution du rôle de l'enseignant. Comment va évoluer celui du manager ?

T. R. : Excellente question à nouveau ! Tous les débats que nous avons eus ces dernières années autour des notions d'*entreprise libérée* et d'*holocratie* – une forme de management consistant à donner le pouvoir à l'organisation elle-même plutôt qu'à ses membres – ne tombent pas du ciel. Ils résultent de l'application, à l'intérieur de l'entreprise, de la prosommation. Il est clair que l'irruption de l'IA générative va nous obliger à repenser le rôle du manager, qui se pratiquera sans doute de moins en moins dans sa forme descendante traditionnelle. Je pense que l'avenir est aussi à un management plus humain, moins pressurisant : plus on pressurise les salariés, plus on les pousse à aller sur ChatGPT ou autre pour produire vite du résultat.

Int. : Vous avez évoqué le problème du tri d'une information devenue pléthorique et l'importance corrélative du capital social pour les différentes plateformes. Ce capital social n'a rien à voir avec la technologie, il traduit la confiance que l'on a dans une marque (qui peut tout aussi bien être la marque d'une grande banque). Cela ne revient-il cependant pas à donner une sorte de privilège conservateur aux grandes institutions ayant déjà et depuis longtemps pignon sur rue ?

T. R. : Vous avez entièrement raison quant à la nature du capital social. J'ajoute que cette confiance a une composante affective (« *Je vous fais confiance parce que je vous apprécie* ») et une composante cognitive (« *Je ne vous aime pas, mais je m'en remets quand même à vous parce que je sais que le travail sera fait* »). Néanmoins, j'attire votre attention sur le fait que ce n'est pas parce qu'une entreprise – une grande banque, pour reprendre votre exemple – dispose à un moment donné d'un capital social important, que ce capital social va se maintenir à l'avenir. Je remarque d'ailleurs que, d'ores et déjà, nombreux sont ceux qui font davantage confiance, en matière de paiement, à Apple ou à Google, qu'à HSBC ou à la Société Générale. Il en est de même dans l'automobile, où les constructeurs se sont faits complètement "plateformiser" par Apple CarPlay et Android Auto. Le danger de ringardisation est particulièrement prégnant pour les entreprises qui ont un capital social au zénith, mais qui pourraient le perdre très vite à cause de décisions maladroites, d'innovations irresponsables, etc.

Int. : *Toutes les entreprises sont-elles vraiment condamnées à suivre une seule et même stratégie, c'est-à-dire devenir des plateformes ?*

T. R. : Ce qui me paraît clair, c'est que le modèle de l'entreprise qui *fait pour*, au lieu de *faire avec*, est de plus en plus dépassé. C'est la raison pour laquelle j'ai tendance à voir dans cette transformation une nécessité stratégique pour les entreprises. Cela dit, il est probable que toutes n'y parviendront pas, parce que, dans cette nouvelle forme de capitalisme caractérisée par la primauté des plateformes, il existe bien une forme de concentration, liée au capital social justement.

Int. : *En cette époque où la question de la sobriété est devenue si décisive, l'IA générative peut-elle aider les entreprises à aller vers la low tech ?*

T. R. : Je ne parierais pas là-dessus... Ce qui est clair, en revanche, c'est que la débauche technologique à laquelle s'adonnent actuellement les entreprises vient de ce qu'elles utilisent les mauvais outils de gestion. Avec les bons outils de gestion, on aboutit à des solutions beaucoup plus *low tech* que celles initialement envisagées. On l'a vu dans le ferroviaire, dans les télécommunications, etc. Si l'on veut aller vers plus de sobriété, il faut se concentrer sur la création de valeur et non sur la marge.

Un risque civilisationnel ?

Int. : *Quand on constate la baisse générale du niveau de connaissance, d'une part, et les risques que représente l'IA générative, d'autre part, peut-on encore être optimiste pour l'avenir ?*

T. R. : Je ne suis pas très optimiste de nature... Nul doute que le challenge, pour les générations à venir, est absolument énorme. Il va falloir faire en sorte de préserver – mais comment ? – tout un tas de savoirs et de connaissances.

Int. : *Justement ! Ce que vous nous avez raconté à propos de l'IA m'a fait songer au déclin des grandes civilisations de l'Antiquité – Mésopotamie, Grèce, Rome... –, dont les savoirs et connaissances ont disparu, au point que l'humanité a fini par sombrer dans l'obscurantisme moyenâgeux. En sommes-nous là ? L'IA ne va-t-elle pas précipiter la fin de notre civilisation ? C'est une question presque philosophique...*

T. R. : Votre question n'est pas philosophique, elle est très pratique au contraire ! Je suis absolument effaré par le nombre de personnes qui utilisent ces algorithmes d'IA générative sans comprendre du tout comment ils fonctionnent. Cela nous fait courir un vrai risque. L'enjeu de perte de connaissances (de savoirs et de savoir-faire) est réel, et ce, au moment critique où le nombre de fake news s'apprête à exploser et où la capacité à démêler le vrai du faux sera plus cruciale que jamais...

■ Présentation de l'orateur ■

Thierry Rayna : Professeur de management de l'innovation à l'École polytechnique et chercheur au CNRS au sein du laboratoire i3-CRG (Institut interdisciplinaire de l'innovation, Centre de recherche en gestion, UMR CNRS 9217), il est également directeur de la chaire *Technology for Change* de l'Institut polytechnique de Paris. Sa recherche porte sur l'impact des technologies numériques sur les modèles d'affaires, les modes de management de l'innovation et les politiques publiques.

Diffusion en octobre 2023

**Retrouvez les prochaines séances et dernières parutions
du séminaire Transformations numériques et entrepreneuriales sur notre site www.ecole.org.**