

L'intelligence artificielle au service de la production de contenu : l'expérience Syllabs

par

■ **Claude de Loupy** ■

Cofondateur et *CEO* de Syllabs

En bref

Passionné par l'informatique et l'intelligence artificielle (IA) appliquée à la sémantique et aux langues, Claude de Loupy cofonde Syllabs en 2006, avec Helena Blancafort. Cette entreprise développe une technologie à la fois probabiliste et symbolique, associant ainsi deux mondes traditionnellement opposés : celui du *machine learning* et celui des modèles à base de règles paramétrées par l'humain. La valeur ajoutée de ce rapprochement est la capacité à produire automatiquement et en un temps record des volumes considérables de contenus de qualité, publiables sans relecture préalable, comme des comptes rendus de résultats d'élections, des annonces immobilières ou des descriptifs produits. Ses choix techniques, stratégiques et éthiques ont fait de Syllabs une référence dans le domaine de la génération de contenus en France. Après un fort ralentissement de son activité lié à la pandémie de Covid-19, Syllabs fait désormais face – comme la société tout entière – au “bulldozer marketing” ChatGPT.

Compte rendu rédigé par Florence Berthezène

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

J'ai commencé à m'intéresser à l'intelligence artificielle (IA) à l'âge de 14 ans, avec le film *2001 : l'odyssée de l'espace*. En dépit d'un final assez sombre, le côté très humain de l'IA HAL 9000 m'a subjugué. Rétrospectivement, je trouvais fascinant qu'alors qu'on ne peut pas disséquer un cerveau pour comprendre comment fonctionne la pensée, l'IA rende cette opération possible en simulant des comportements humains.

L'IA et le traitement des langues

En parallèle de mes études en ingénierie informatique, j'ai fait un DEA (diplôme d'études approfondies) en IA, équivalent du master de recherche actuel. Lorsqu'un professeur a évoqué le traitement automatique des langues par l'IA, j'ai eu un véritable coup de foudre – peut-être parce qu'étant le seul matheux dans une famille de littéraires, c'était le moyen de travailler aussi sur des textes? J'ai ensuite passé deux ans au CNRS, en tant qu'ingénieur d'études, dans le cadre d'un projet de traitement des langues, avant de rédiger une thèse consacrée aux moteurs de recherche et à la façon d'améliorer leur fonctionnement avec des connaissances sémantiques. J'ai commencé cette thèse dans l'entreprise Bertin Technologies, avant de la terminer dans mon université d'encadrement à Avignon.

Même si j'ai vraiment apprécié cette expérience à l'université, la dimension de l'usage des outils que nous développions me manquait. J'ai alors rejoint l'entreprise Sinequa, qui commercialise un moteur de recherche fondé sur des connaissances sémantiques, pour y créer un laboratoire de recherche interne et encadrer une équipe de linguistes. Un fonctionnement bicéphale, avec un moteur de recherche principalement probabiliste et un traitement sémantique, m'a permis de découvrir une approche symbolique de la linguistique : les linguistes paramètrent le système pour reconnaître certains éléments, comme des noms de personnes, d'entreprises ou de lieux, et pour les signaler au moteur de recherche afin d'accroître ses performances dans une optique de satisfaction des utilisateurs. Cela m'a permis de faire évoluer ma perception de l'IA et de comprendre qu'un système purement probabiliste n'est pas la panacée. Il est essentiel que l'humain l'entraîne et lui apprenne des règles – par exemple, que le mot *est* est nécessairement un verbe, et non un substantif, s'il est précédé du pronom interrogatif *qui*.

Par la suite, j'ai vécu la moitié du temps au Caire, pour des raisons familiales. J'ai d'abord continué à travailler pour Sinequa. Néanmoins, il m'est rapidement apparu que quelque chose me manquait, et j'ai décidé de créer ma propre entreprise.

L'aventure Syllabs

J'ai cofondé Syllabs en 2006, avec une associée qui n'est autre que ma femme!

Une vision et trois piliers

Alors qu'il existe souvent une opposition dure entre les probabilistes et les "symboliques", la combinaison de ces deux disciplines est notre premier pilier. En l'occurrence, mon associée est linguiste et je suis informaticien probabiliste.

Notre deuxième pilier est le dialogue, à l'encontre du monopole des informaticiens, encore trop souvent répandu.

Enfin, notre troisième pilier est la vitesse. En créant notre entreprise, nous souhaitions faire du *text mining*, c'est-à-dire de l'analyse de textes pour extraire de l'information et catégoriser les contenus. À l'époque, les meilleurs dans ce domaine se vantaient de traiter 17 mégaoctets de texte pur à l'heure, ce qui n'était pas

beaucoup. Lorsque nous avons embauché notre premier ingénieur, l'objectif était de traiter 1 gigaoctet de texte pur à l'heure, ce que nous avons presque réussi à faire en atteignant 750 mégaoctets.

Du text mining à la génération de texte

Au départ, nous n'avions rien à vendre sinon nos cerveaux. Aussi avons-nous commencé par du *consulting*. Notre démarche commerciale reposait sur un principe simple : trouver des subventions pour nos clients afin qu'ils puissent nous payer. Si un projet déposé par Syllabs était refusé, il n'était pas facturé. C'est donc Syllabs qui prenait tous les risques. Finalement, nous avons vendu de nombreux projets, ce qui nous a permis de recruter 12 personnes en deux ans et demi.

Cependant, force a été de constater qu'aucun projet n'a véritablement abouti, dans la mesure où les priorités des entreprises concernées avaient changé au fil de l'eau. Or, mon leitmotiv était, encore et toujours, de développer des outils qui servent. Nous avons alors mis un terme à cette première activité pour nous concentrer sur notre technologie et nous orienter vers la génération de texte, qui commençait à se développer aux États-Unis. En trois mois, grâce au travail d'un de nos ingénieurs, notre prototype était déjà suffisamment efficace pour être vendu.

Le principal défi, sur le plan commercial, était de ne pas tomber amoureux de notre technologie, laquelle doit toujours rester un outil. Seule la satisfaction du client compte. Cet apprentissage n'allait pas de soi, compte tenu de mon parcours.

Une autre condition de succès résidait dans la façon de présenter notre projet : une machine capable de produire des contenus. Alors que nous comptions de nombreux médias parmi nos clients, j'ai envisagé de promouvoir notre technologie auprès d'eux. Tous étaient enthousiastes, mais craignaient que les journalistes ne soient vent debout face à ce qu'ils pourraient percevoir comme la menace de voir leur métier disparaître. J'ai alors passé trois ans à faire de l'évangélisation, pour tenter de changer les mentalités.

À la fin de l'année 2014, le directeur des rédactions du *Monde* a pris le risque de travailler avec nous pour les élections départementales. Nous avons alors réussi une première mondiale en produisant en temps réel les résultats des élections de 35 000 communes en mars 2015. Le lendemain, le buzz a été phénoménal, avec un millier de tweets consacrés à cet exploit et des dizaines d'articles dans les médias ! Cela a permis de montrer que c'était possible : puisque *Le Monde* publiait nos textes sans les relire, d'autres entreprises pouvaient faire appel à notre technologie. Lors des dernières élections, nous avons produit 350 000 articles différents en quatre heures, publiés automatiquement et sans relecture dans 15 médias, soit un tiers de ce que publie l'AFP en un an. Notre solution est une véritable machine, qui fonctionne.

Nous avons également investi le secteur de l'immobilier, pour produire les annonces de 2 800 agences immobilières. Il suffit qu'un négociateur immobilier entre les caractéristiques du bien qu'il souhaite vendre et, en un clic, une annonce est rédigée dans une langue parfaite.

Nos clients attendent une production de textes ne nécessitant pas de relecture, ce qui requiert de faire du symbolique. Ainsi, aux côtés de nos informaticiens et de nos ingénieurs data, notre équipe de linguistes est chargée de paramétrer le moteur de rédaction en fonction des données fournies en entrée. Le travail sur les outils prend du temps et a un coût. Il est donc primordial de disposer d'un grand volume de contenus du même type à produire pour que l'opération soit rentable – c'est le cas avec les résultats sportifs ou électoraux, les annonces immobilières, les contenus de la Bourse ou les annonces météorologiques de toutes les communes françaises.

Un besoin crucial de fonds pour continuer

Notre technologie a considérablement évolué au fil du temps. En 2019, nos ventes ont progressé de 120 % et nous avons investi dans le tourisme. Cependant, en mars 2020, avec la pandémie de Covid-19, tout s'est arrêté et nous avons ensuite passé deux années très difficiles.

Par ailleurs, nous avons levé de l'argent fin 2017, mais – erreur de débutants – auprès d'acteurs qui ne nous correspondaient pas et qui ne connaissaient pas suffisamment les entreprises technologiques. Qui plus est, notre produit n'était pas suffisamment prêt. Durant un an, il nous a donc fallu continuer à le travailler au lieu d'en développer le commerce.

De 2020 à 2022, nous avons dû compter le moindre sou. C'est encore le cas, dans une moindre mesure toutefois. Alors que nous sommes la référence en France dans le secteur le plus couru du moment, et alors que notre outil permet de produire des contenus sûrs, nous manquons de fonds pour nous développer.

La révolution de ChatGPT

Le 30 novembre 2022, l'apparition de ChatGPT a engendré un véritable ouragan dans les médias. Jamais autant de personnes ne s'étaient inscrites pour utiliser un outil.

Un réseau de neurones artificiels et une patine de discussion humaine

Les réseaux de neurones et les réseaux de neurones profonds existent de longue date, mais la technologie des *LLM* (*large language models*), qu'utilise ChatGPT, est apparue il n'y a que quelques années. GPT-3 compte 96 couches de neurones artificiels, alimentées par une quantité phénoménale de contenus – 250 milliards de mots ! L'objectif est de produire du contenu de tout type, à partir de textes.

La machine apprend en continu. Elle analyse à la fois chaque mot d'une phrase, mais aussi les contextes dans lesquels celle-ci apparaît, pour constamment mettre à jour ses paramètres. Si ce système m'a d'emblée impressionné, il ne répondait néanmoins pas aux besoins de mes clients, qui attendent des contenus parfaits n'ayant pas besoin d'être relus. Nous avons donc attendu 2022 pour sortir un produit utilisant cette technologie. Il est destiné à l'immobilier, et nous utilisons notre propre technologie pour effectuer divers contrôles. Alors que notre technologie historique demande deux semaines de configuration et permet de publier des textes sans relecture, il suffit ici de poser intelligemment une question pour qu'un article soit produit... mais il faut le relire.

En outre, une "patine de discussion", notamment avec des formules de politesse, a été intégrée au modèle par-dessus les *LLM*. Ainsi, le système ChatGPT a appris, à partir d'échanges entre humains, dans l'objectif de répondre à toutes les questions qui lui sont posées – quitte à indiquer qu'il ne connaît pas la réponse ou qu'il s'est trompé, et à s'en excuser. Les rédacteurs lui ont aussi appris à finir ses réponses par « *en conclusion* », ou « *pour résumer* ». Cette dimension nourrit les polémiques classiques que soulève l'IA : qui est intervenu pour façonner cette patine, dans quelles conditions, etc. ?

En somme, l'entreprise OpenAI, créatrice de ChatGPT, a franchi un pas que les autres géants du secteur n'avaient pas osé franchir. Après une démarche d'ouverture et de partage, en se fondant sur les travaux d'acteurs comme Google ou Meta et en observant la façon dont les utilisateurs exploitaient son outil, OpenAI commence à se refermer, afin que d'autres ne puissent pas la dépasser et mettre à mal son succès commercial.

Quelles limites ?

D'un point de vue sociétal, nous marchons sur le fil. Personne ne sait vraiment ce qui peut arriver lorsque de tels outils se trouvent entre les mains de tous. Si les premiers outils nécessitaient d'être un peu geek, n'importe qui peut utiliser ChatGPT et lui poser n'importe quelle question. Certains y passent des heures et se confient même à lui, le considérant comme un compagnon, voire un thérapeute. Et pour cause, le système leur répond en montrant de l'empathie. Les utilisateurs peuvent avoir le sentiment que "quelqu'un" leur répond et interagit avec eux pour les aider. De la même façon qu'à la fin de 2001 : *l'odyssée de l'espace*, l'IA HAL 9000 est émouvante quand elle déclare : « *J'ai peur, j'ai peur, j'ai peur !* », ChatGPT peut laisser penser que nous sommes en présence d'une conscience, d'une vie avec laquelle il est possible d'échanger. Un utilisateur qui se trouve dans une situation périlleuse peut ainsi se laisser emporter.

Outre ses incidences sociétales, ChatGPT peut aussi avoir des conséquences sur l'emploi.

Sur le plan technologique, ChatGPT donne le sentiment d'être face à un "être omniscient" qui a tout lu, qui connaît tout et qui partage tous ses savoirs. C'est à la fois génial et phénoménal, tant du point de vue de l'apprentissage que de celui de la capacité de production. Et même s'il commet des erreurs, il apprend à les résoudre.

Pour autant, certaines personnes en parlent comme d'un simple système probabiliste, qui prévoit le mot suivant. C'est sans doute vrai... mais que se passe-t-il dans le système pour qu'il arrive à calculer la probabilité du mot suivant. Par exemple, si on lui demande : « *Que se passe-t-il quand on retourne un vase contenant des fleurs?* », il sait interpréter ce que signifie retourner un vase et décrire ce qui se produit : « *L'eau coule, les fleurs tombent et peuvent s'abîmer.* » De surcroît, le système se souvient de ce qui a été échangé précédemment, ce qui est très rarement le cas des chatbots. À cet égard, affirmer que ChatGPT ne fait *que* prévoir le mot suivant paraît dangereux, car cela minimise l'ampleur du phénomène à l'œuvre.

Même si je ne suis pas en mesure de le prouver, car c'est une boîte noire, je suis persuadé qu'au sein des 96 couches de neurones de GPT-3, certaines zones servent à la conceptualisation. De ce point de vue, ChatGPT constitue une véritable révolution technologique. Il y a encore dix ans, je n'aurais jamais cru y assister de mon vivant.

Non seulement ChatGPT parvient à conceptualiser, à déduire, à jouer et à reconnaître ses erreurs, mais il sait aussi inventer – des langues, des couleurs, des sentiments. Par exemple, si je lui demande d'inventer un sentiment humain, sa réponse est la suivante : « *La vélicitude, mélange d'un sentiment d'anticipation excitante avec une pointe de mélancolie et une profonde appréciation de la beauté éphémère de la vie.* » Certes, cela ne signifie pas qu'il ressent lui-même des sentiments ou qu'il sait de quoi il parle, mais il donne vraiment bien le change!

Et demain ?

Les outils de type ChatGPT font des erreurs et continueront à en faire, car ils sont optimisés pour produire du contenu et faire plaisir à l'utilisateur, et n'ont aucune prise sur la réalité : expliquer la couleur rouge à un aveugle de naissance restera toujours compliqué.

Si leur système est actuellement alimenté par d'innombrables contenus textuels et visuels, l'étape suivante consistera à le nourrir avec des données pour qu'il puisse effectuer des vérifications et moins se tromper – et, ce faisant, accroître la confiance des utilisateurs.

Ces outils s'approcheront donc d'un Wikipédia. Certains sont déjà placés dans des moteurs de recherche – comme Bing –, dans lesquels le système cite ses sources et donne des liens pour approfondir les sujets concernés. C'est extrêmement intéressant! Dans le même temps, tous les utilisateurs ne cliqueront pas nécessairement sur ces liens. Or, nous l'avons vu, le système peut se tromper et raconter n'importe quoi. Il peut même être orienté. L'avantage des moteurs de recherche, en comparaison, est qu'ils proposent plusieurs réponses.

En connectant tous ces systèmes entre eux, un niveau d'efficacité très puissant pourrait être atteint. Des analyses et des garde-fous sont indispensables. Le législateur doit s'en saisir, car la maîtrise de ce type d'outils ne saurait être laissée aux seuls entrepreneurs.

Débat



Syllabs

Un intervenant : *Pouvez-vous nous en dire plus sur la situation de votre entreprise ?*

Claude de Loupy : Depuis la pandémie de Covid-19, notre effectif est passé de 25 à 19 personnes. Les affaires reprennent et nous avons retrouvé une croissance de 20 %, sans avoir les moyens d'investir dans le volet commercial. Si nous ne sommes pas encore à l'équilibre, notre objectif est de l'être en fin d'année.

Nos principaux marchés sont ceux de l'immobilier et des médias. Nous travaillons aussi pour l'e-commerce et le *retail*. Cela fonctionne bien, mais nous ne pouvons pas embaucher pour nous développer.

En somme, notre seule problématique est celle de l'argent. Nous avons la technologie, de grands noms comme clients et nous savons où il faudrait aller, mais nous n'en avons pas la possibilité. C'est très frustrant et parfois décourageant. Ce découragement ne dure pas longtemps, car il existe des raisons de continuer. Néanmoins, force est de constater que tout se joue maintenant et nous passe devant.

Int. : *Quel est le coût de vos services ?*

C. de L. : Un produit "sur étagère", comme les annonces immobilières, coûte 29 euros par mois sans limite de nombre de requêtes. Pour les développements spécifiques, il faut compter au minimum 25 000 euros. Notre modèle est plus coûteux que celui de ChatGPT, mais il garantit l'exactitude et la précision. Tout en étant différent, notre outil reste demandé, car il apporte un service enrichissant. Le choix de Syllabs est clair : nous ne vendons pas de logiciels, mais du contenu produit automatiquement et garanti parfait.

Int. : *Comment avez-vous conquis les agences immobilières avec lesquelles vous travaillez ?*

C. de L. : Nous faisons un peu de porte-à-porte, mais nous rencontrons surtout directement les franchiseurs, comme Orpi ou Guy Hoquet. Nous sommes également intégrés dans une douzaine de logiciels destinés aux agents immobiliers. Un onglet Syllabs permet de s'abonner à nos services.

Int. : *Jusqu'où allez-vous, dans la partie sémantique ? La structure d'une phrase n'étant pas la même dans toutes les langues, comment vous y prenez-vous ?*

C. de L. : Nous avons largement travaillé ces questions pour le *text mining*. En effet, extraire de l'information requiert d'analyser des phrases et de connaître leur fonctionnement sinon sémantique, au moins syntaxique. Dans mon entreprise précédente, j'avais également beaucoup travaillé la sémantique, par exemple pour apprendre au système qu'un article employant des termes d'astronomie comme *lune* ou *soleil* ne parle pas nécessairement d'astronomie, mais potentiellement aussi de météorologie, d'agriculture, d'astrologie, etc. En revanche, je n'ai pas travaillé sur le phénomène de conceptualisation, car je ne voyais pas comment faire avant l'apparition des réseaux profonds et du *deep learning*.

Par ailleurs, il est possible d'indiquer au système qu'il existe des liens forts entre deux termes. Cela améliore nettement sa performance, même si la façon dont il opère les regroupements de mots n'est pas toujours compréhensible.

ChatGPT

Int. : *Peut-on demander à ChatGPT de répondre en un nombre précis de signes ?*

C. de L. : Cela ne fonctionne pas encore bien.

Int. : *Quid de la régénération ?*

C. de L. : Il est possible de demander à ChatGPT d'écrire "à la façon de" tel ou tel auteur, ou dans tel ou tel format, par exemple celui du tweet.

Int. : Où en sont les logiciels de détection de l'utilisation de ChatGPT?

C. de L. : OpenAI a créé un outil de détection. Il estime que seuls 26% des textes produits par une IA sont détectés. Certaines situations ubuesques commencent aussi à s'observer, comme celle de contenus rédigés par des humains et considérés à tort comme produits par des outils d'IA. Dès lors que l'on demande aux systèmes de se comporter comme des humains pour produire des contenus, ce qu'ils apprennent à faire, la situation devient complexe. Le fait que nous souhaitons pousser la machine au maximum de l'humanité et, dans le même mouvement, conserver notre spécificité humaine engendre nécessairement des ambiguïtés, voire des conflits sociétaux.

En douze ans, aucun de nos contenus n'a été blacklisté ou signalé par Google, qui a même indirectement, mais sciemment, financé Syllabs lorsqu'il versait des subventions aux médias. Produire automatiquement des contenus n'est pas un problème en soi, tant qu'il s'agit d'un service. En revanche, cela peut en devenir un quand les contenus sont des spams ou de la désinformation.

Les études montrent que la disparition de médias locaux se traduit systématiquement par une montée des extrêmes. L'importance de l'information locale est donc indéniable pour la démocratie. En l'occurrence, s'ils sont correctement utilisés, les outils d'automatisation réduisent le coût de l'information locale et permettent aux journalistes de se concentrer sur des contenus à valeur ajoutée, plutôt que de traiter de la météo ou des accidents de la route.

Int. : ChatGPT peut raconter n'importe quoi et même inventer des citations. Jusqu'où peut-il garantir que ses réponses sont fondées? Peut-on le solliciter pour comprendre ses couches de neurones? En d'autres termes, quid de la rétro-ingénierie?

C. de L. : Quand je joue avec ChatGPT, je me retrouve assez bien dans le personnage de Susan Calvin, la psychologue pour robots inventée par l'écrivain Isaac Asimov et dont la mission est de comprendre les actions de ces derniers. Pour autant, il ne s'agit pas de rétro-ingénierie à proprement parler, car nous ne pouvons pas étudier ce qui se passe dans ces matrices que sont les réseaux de neurones. Il s'agit de véritables boîtes noires que nous ne savons pas interpréter.

Syllabs vs ChatGPT

Int. : Comment affrontez-vous l'arrivée de ChatGPT? Comment réagissent vos clients?

C. de L. : L'arrivée de ChatGPT ne m'a pas effrayé, car nos clients ont besoin de contenus immédiatement publiables, sans relecture. Notre système est fiable. Les LLM que nous utilisons sont contrôlés par notre propre outil. En outre, même si cette technologie génère automatiquement des *prompts* très précis – grâce à des instructions très strictes – et systématiquement contrôlés, nous faisons relire nos contenus pour tenir notre engagement contractuel de production de contenus sans erreur. La situation est un peu plus compliquée avec nos prospects. Toutefois, si l'arrivée de ChatGPT complexifie les ventes, elle nous fait aussi de la publicité. Nous bénéficions d'une certaine aura et nous sommes écoutés.

Par ailleurs, l'utilisation de ChatGPT peut être très ludique et toucher le grand public ou les influenceurs, ce qui n'est pas le cas de notre technologie. En revanche, il est quasiment impossible de le faire écrire en respectant strictement la véracité et une ligne éditoriale – ce que nous proposons.

Int. : Certes votre système ne fait pas d'erreurs et vos contenus ne demandent pas de relecture, contrairement à ChatGPT, mais ne craignez-vous pas que votre bonne monnaie soit chassée par la mauvaise, compte tenu de la puissance médiatique et commerciale de ChatGPT?

C. de L. : Non, d'autant que je ne qualifierais pas ChatGPT de mauvaise monnaie. Il fait autre chose que notre outil, et il nous arrive d'ailleurs de l'utiliser pour gagner du temps. Certaines "mauvaises graines" ne manqueront pas de l'utiliser pour rédiger leurs articles sans vérification et publieront alors des contenus erronés, mais il est possible d'en faire bon usage. Le risque majeur est celui de la pollution informationnelle. Certes, il est rassurant de constater que les informations, y compris les photographies, sont de plus en plus vérifiées. Toutefois, alors

que l'écosystème informationnel est régulièrement attaqué par les "fermes à contenus", trouver la bonne information est un défi de taille.

Int. : *Alors que la production de ChatGPT a été instantanément acceptée par tout le monde, avez-vous des regrets de ne pas avoir opéré tel ou tel développement technologique ?*

C. de L. : Non ! La différence avec notre technologie est que ChatGPT touche le grand public et que chacun peut se l'approprier – ce qui explique que tout le monde en parle. Notre système est très spécialisé et personne n'a envie de "s'amuser" avec lui, tandis que ChatGPT est ludique et prenant.

Par ailleurs, plusieurs entreprises ont proposé d'acheter notre technologie pour pouvoir l'utiliser pour elles. Nous avons aussi été régulièrement approchés pour produire de faux avis. De même, un média chinois souhaitait que nous diffusions sa propagande. Nous avons évidemment refusé toutes ces demandes. Nous entendons conserver la maîtrise de notre outil, pour éviter qu'il alimente la désinformation.

Perspectives

Int. : *Comment interprétez-vous les récents propos d'Elon Musk selon lesquels il faut tout arrêter ?*

C. de L. : Il n'y croit pas lui-même ! C'est sans doute pour faire parler de lui, mais aussi pour se légitimer en tant que visionnaire. Yuval Noah Harari, qui a cosigné cette pétition, est lui-même brillant, mais c'est un conteur qui ne cite pas ses sources, qui plagie et dont certaines affirmations ne sont pas prouvées. En somme, je n'ai pas une totale confiance en ces meneurs ! En revanche, d'autres signataires, acteurs éminents du secteur, méritent d'être écoutés lorsqu'ils sonnent l'alarme. Ces outils présentent à coup sûr un risque et méritent d'être surveillés et contrôlés. Même si certaines menaces sont agitées à tort, il est bel et bien question d'outils surpuissants utilisés par des millions de personnes, dont des criminels. Des garde-fous, y compris législatifs, sont inévitables.

Int. : *ChatGPT pourra-t-il apprendre à se contenir et à se montrer prudent dans ses réponses ? Est-il possible de lui apprendre l'éthique ?*

C. de L. : Les probabilistes estiment que les humains n'apprennent pas grâce à des règles. Par exemple, ils savent reconnaître un chat parce qu'on leur en a montré initialement. Néanmoins, si les règles ne servaient à rien, personne n'irait à l'école ! Des pages entières de règles ont été rédigées manuellement pour apprendre à GPT-4 à bien se comporter, en ne tenant pas de propos racistes, en n'affichant pas d'opinion politique, en disant qu'il ne sait pas, en s'excusant de s'être trompé, etc. En outre, pour que la machine apprenne à reconnaître un chat, il faut lui montrer plusieurs milliers d'images de chats – même si les performances en la matière s'améliorent grandement.

L'outil de veille Flint axe sa communication sous l'angle de l'école des robots. De fait, alimenter les machines avec une quantité phénoménale de contenus est un véritable apprentissage scolaire. Quant au travail sur l'éthique, il est nécessairement humain et difficile à normer, car ce concept varie selon les cultures. Cela dit, j'ignore si l'on peut enseigner l'éthique, même s'il est possible d'apprendre à la machine à ne pas faire ou dire certaines choses. Certains se sont d'ailleurs déjà amusés à hacker ChatGPT pour le pousser à faire des choses qu'il n'est pas censé faire.

Int. : *Peut-on doter ChatGPT de la faculté de se bloquer lorsqu'un utilisateur se comporte mal ?*

C. de L. : C'est envisageable, mais cela pourra sans doute être contourné.

Int. : *Pour contrer la méfiance grandissante envers les boîtes noires que sont les algorithmes, faut-il généraliser l'open source ? Ne faut-il pas rendre l'explicabilité obligatoire ?*

C. de L. : Non. En revanche, il serait utile de connaître le corpus d'entraînement de GPT-4, ce qui impose d'obliger OpenAI à révéler les bases sur lesquelles le système a appris.

Ensuite, mettre les algorithmes en open source ne servirait pas à grand-chose, car tout dépend de la façon dont les paramètres s'autoconfigurent. Ainsi, l'open source ne permettrait pas de savoir ce que la boîte noire

a appris, la façon dont elle l'a appris, comment elle interprète les questions et comment les *prompts* sont mis en relation avec ce qu'elle a appris. Même en ouvrant la matrice, il n'est pas possible de savoir ce qu'il y a derrière un réseau de neurones. Je suis plutôt favorable à l'open source, mais l'accès à tous inclus aussi les mafias et les personnes mal intentionnées. Doit-on rendre les fusils mitrailleurs accessibles à tous, comme aux États-Unis ?

L'explicabilité est donc la seule clé pour maîtriser ce qui se passe, d'un point de vue informationnel.

Int. : *Mettriez-vous la technologie de Syllabs en open source ?*

C. de L. : Non, pour diverses raisons d'ordre économique. Pour autant, tout est explicable dans la mesure où notre système est fondamentalement symbolique. Pour chaque question, les équipes sont capables de remonter à la donnée et au graphe des chemins possibles, afin de savoir pourquoi tel élément a été produit.

■ Présentation de l'orateur ■

Claude de Loupy : Ingénieur et docteur en informatique, cofondateur et président de Syllabs, start-up spécialisée en génération automatique de texte, il a travaillé au CNRS, à Bertin Technologies, à l'université d'Avignon, à l'université Paris-Nanterre et à Sinequa, une entreprise spécialisée dans les moteurs de recherche.

Diffusion en septembre 2023

**Retrouvez les prochaines séances et dernières parutions
du séminaire Management de l'innovation sur notre site www.ecole.org.**