

# COMPRENDRE CHATGPT POUR EN FAIRE BON USAGE

Avec ChatGPT, la puissante efficacité de l'intelligence artificielle a fait irruption dans notre réalité, suscitant engouement, interrogations et craintes. Explications de **Patrick Flandrin** pour prendre la mesure de ce pas de géant, et des défis qu'il implique.

Dans le cadre de notre partenariat avec l'Académie des sciences, des académiciennes et académiciens analysent et apportent leur éclairage sur les grands enjeux du monde contemporain au travers de questions scientifiques qui font l'actualité.

**C**es dernières années, l'engouement pour l'intelligence artificielle (IA) a pris une ampleur sans précédent, avec des avancées remarquables dans les domaines de la reconnaissance vocale, de la vision par ordinateur et du traitement du langage naturel. Dans ce contexte, ChatGPT, un modèle de langage développé par OpenAI, a suscité un grand intérêt de la part des chercheurs, des entreprises et du grand public. Grâce à sa capacité à générer des textes en langage naturel qui peuvent être difficiles à distinguer de ceux écrits par des humains, ChatGPT est devenu un outil polyvalent et puissant pour les applications de l'IA telles que la réponse automatique aux clients, la génération de texte et la traduction automatique. Depuis sa mise à disposition en novembre 2022, pas un jour ne passe sans que ChatGPT ne s'invite dans les conversations et les médias, suscitant son lot de curiosités, d'interrogations et de craintes. L'IA, qui gardait à son évocation une part de science-fiction et de promesses pour un futur encore lointain, est devenue soudain une réalité tangible dont tout un chacun peut se convaincre de la redoutable efficacité. Preuve, s'il en fallait une : le premier paragraphe de cet article n'a pas été écrit par l'auteur de ces lignes ; c'est la transcription sans filtre de la réponse

que ChatGPT a donnée à la requête exprimée ainsi : « Écris-moi un paragraphe d'introduction à un article parlant de l'engouement pour ChatGPT » ! ChatGPT est ce que l'on appelle un « agent conversationnel » (ou chatbot), c'est-à-dire un logiciel d'IA destiné à interagir avec un utilisateur en simulant une conversation humaine. Sa force est de répondre en langage naturel aux questions qui lui sont posées. Le cœur de cette IA repose sur un modèle de langage baptisé GPT, pour Generative Pre-trained Transformer. Celui-ci procède par apprentissage : il s'appuie sur un très grand nombre d'exemples issus d'un corpus constitué de données glanées essentiellement sur Internet et les réseaux sociaux.

## EXPÉRIMENTATION ET RÉCOMPENSE

Fort de ces dernières, le modèle est entraîné pour en faire bon usage, et il y a pour cela deux phases : un mode supervisé et un autre par renforcement. Dans le cadre supervisé, la machine apprend à répondre correctement, sur la base d'exemples fournis par des opérateurs humains. L'apprentissage par renforcement est quant à lui basé sur l'expérimentation et la récompense, l'algorithme explorant le champ des possibles pour une tâche donnée et retenant les solutions lui valant les meilleures appréciations. Ce peut être, par exemple, la complétion d'une phrase dont les premiers mots sont donnés, la suite étant choisie parmi les plus attendus et les plus satisfaisants dans leur formulation. Des chatbots existaient déjà depuis quelques années pour des tâches spécifiques mais, avec l'arrivée de ChatGPT, un pas de géant a clairement été franchi. De la rédaction d'encarts publicitaires à la

**« CHATGPT N'EST PAS CONÇU POUR DONNER UNE RÉPONSE JUSTE, MAIS CHOISIE PARMIS LES PLUS PROBABLES APRÈS APPRENTISSAGE. »**



SIMON CASSANAS / ACADEMIE DES SCIENCES / ISP

**PROFIL**  
Physicien, directeur de recherche émérite au CNRS, laboratoire de physique de l'ENS de Lyon, **Patrick Flandrin** est membre de l'Académie des sciences (vice-président, 2019-2020, puis président, 2021-2022). Spécialiste du traitement du signal, ses travaux déterminants ont donné lieu à des applications en astrophysique, ingénierie biomédicale, modélisation du trafic Internet, entre autres.

traduction automatique, de la vérification de codes informatiques à des conseils de cuisine, il semble qu'on puisse lui demander à peu près tout (et n'importe quoi). Affable et pédagogue dans ses réponses, prompt à reconnaître ses erreurs et à présenter ses excuses, ChatGPT entretient avec aisance le fil d'une conversation, y compris en sautant d'une langue à une autre. L'illusion est le plus souvent parfaite d'une interaction avec un interlocuteur humain, ne serait son insistance à rappeler sa nature d'IA dès lors qu'on sollicite des avis personnels.

## L'AISANCE D'UN EXPERT HUMAIN

Les résultats impressionnants qu'un chatbot comme ChatGPT fournit lorsqu'on l'interroge incitent naturellement à en faire une alternative aux moteurs de recherche, ou à tout le moins un complément. Les grandes plateformes ne s'y sont pas trompées, certaines proposant d'ores et déjà d'interfacer l'IA de ChatGPT avec leur moteur (Bing pour Microsoft).

Cette approche appelle cependant un certain nombre de réserves. La première est purement technique et repose sur le fait que, par construction, ChatGPT n'est pas conçu pour donner une réponse qui soit nécessairement juste, mais une réponse choisie, avec une part de hasard, parmi les plus probables après apprentissage. Il peut en résulter – et c'est souvent le cas – des erreurs factuelles telles que des éléments incongrus ou inexacts dans une biographie, des dates erronées, des citations plausibles mais inventées, etc. Erreurs qui plus est fluctuantes au gré d'interrogations répétées en posant la même question ! La deuxième réserve provient de ce que l'utilisateur prend d'autant plus pour argent comptant une réponse qu'elle est faite avec l'aisance attendue d'un expert humain. Plus qu'un travers de l'IA, on pourrait dire paradoxalement que c'est au contraire un succès à mettre à son actif dans la mesure où l'on se trouve alors dans une situation semblable à celle d'une conversation avec un interlocuteur humain »

» à qui il nous appartient de faire confiance ou pas. Dans le conte de Charles Perrault, le roi croit naïvement le Chat botté qui lui affirme avec aplomb l'existence et les qualités de son maître, le supposé marquis de Carabas. Les réponses qu'un chatbot ou un moteur de recherche « chatbotté » nous propose sont en fait autant de marquis de Carabas potentiels. À charge pour nous d'aller au-delà de ce qui nous est dit, de croiser les informations et de remonter à leurs sources.

Si les potentialités d'un outil comme ChatGPT sont gigantesques, leur utilisation soulève de nombreuses questions. En dehors du rapport incertain que cette IA entretient parfois avec la vérité, c'est aussi, implicitement, une vision du monde qu'elle porte en filigrane dans ses réponses. Comme dans tout processus d'apprentissage, cette vision s'enracine à la fois dans la nature et le contenu du corpus utilisé dans la phase de supervision, qui est essentiellement anglo-saxon, et dans la sélection ou le rejet de certaines réponses dans la phase de renforcement. Si des biais existent dans l'apprentissage, les réponses en porteront la marque puisque l'algorithme a précisément été entraîné pour être en cohérence avec les données qui lui sont fournies. La modération mise en œuvre dans la phase de renforcement pour éviter des contenus inappropriés (haineux, injurieux ou discriminatoires) pose néanmoins la question de savoir ce qui est jugé offensant ou pas, et à l'aune de quelle autorité. Il en résulte dans les conversations un ton caractéristique, neutre et lisse, opérant une forme de thermalisation de la pensée, certes politiquement correcte mais sans aspérité ni éclat.

### DES IA CRÉATIVES ?

Parce qu'elle a « fait ses classes » en identifiant lors de l'apprentissage des régularités, des proximités, des liens dans ses données d'entraînement, ce sont ces régularités qu'une IA comme ChatGPT utilisera et reproduira lorsqu'elle sera amenée, face à des situations nouvelles, à généraliser ce qu'elle a appris. Il ne s'agit donc pas de « créations » à proprement parler, mais de propositions plausibles fondées sur le mimétisme et des variations dans la reproduction, l'irruption du jamais-vu restant improbable. Le résultat est néanmoins saisissant lorsqu'il s'agit de traduire un texte, d'en écrire (ou réécrire) un « à la manière de », pourquoi pas de résumer « la Recherche » de Proust en un haïku (« L'auteur se souvient / Le temps passé revient vivant / La recherche est finie »). Il l'est peut-être encore davantage dans le cas de génération d'images à partir de descriptions textuelles (ainsi que le permettent des IA comme Dall-E, Midjourney, etc.) qui synthétisent des scènes

## « L'ALGORITHME A PRÉCISÉMENT ÉTÉ ENTRAÎNÉ POUR ÊTRE EN COHÉRENCE AVEC LES DONNÉES QUI LUI SONT FOURNIES. »

visuelles originales au réalisme impressionnant... alors même qu'aucune réalité n'y est présente. Des textes apocryphes aux photomontages, les manipulations n'ont pas attendu l'IA pour prospérer et brouiller les lignes de partage entre réalité et artifice. Néanmoins, avec les avancées spectaculaires actuelles, on fait face à un changement d'échelle, tant quantitatif que qualitatif. De plus, chacun peut désormais faire usage de façon très simple d'outils extrêmement puissants. Le fond de la prouesse en revient néanmoins toujours à l'apprentissage et aux données d'entraînement, les objets virtuels « créés » portant en creux la marque de données réelles.

### ENTRE CONFIANCE AVEUGLE ET INTERDICTION

Cela ne va pas sans poser des problèmes de droit, d'appropriation éventuelle d'œuvres protégées et, en retour, des questions de propriété des productions qui en résultent. Le monde de l'art s'en émeut (pour paraphraser Walter Benjamin, quid de « L'œuvre d'art à l'heure de l'IA » ?) tout autant que celui de l'éducation (quel recours admissible à ChatGPT à l'école ou à l'université ?) et du débat public (quelle porte ouverte à la désinformation ou à la manipulation des opinions ?).

Sans doute des réponses techniques peuvent-elles être apportées, comme le marquage de productions chatbottées pour en certifier l'origine (quasi impossible aujourd'hui) ou le bridage de versions trop puissantes pour être mises en libre circulation. Il n'en reste pas moins que la technologie évolue sans cesse (GPT4, payant, est d'ores et déjà beaucoup plus performant que la version gratuite de ChatGPT) et à des échelles de temps beaucoup plus courtes que celles qui sont nécessaires aux réflexions pouvant amener à légiférer pour en contrôler l'usage ou conduire des réflexions, par exemple éthiques, relatives à des questions déjà dépassées lorsqu'on leur apporte des réponses.

Une autre question à ne pas négliger est le coût énergétique et environnemental très élevé d'une IA comme ChatGPT, qu'il s'agisse de l'entraînement ou de l'exécution, avec une amplification naturelle de l'impact liée à une adoption de plus en plus grande. Entre confiance aveugle et interdiction, la clé d'un bon usage est certainement à chercher dans une appropriation raisonnée d'outils qui, s'ils n'ont pas à proprement parler une intelligence, sont à même de compléter la nôtre. ●

### EN SAVOIR PLUS

Le site de l'Académie des sciences : [www.academie-sciences.fr](http://www.academie-sciences.fr)

« Intelligence artificielle, intelligence humaine. La double énigme », de Daniel Andler, Gallimard, coll. « NRF Essais », 2023.

« Comment fonctionne ChatGPT ? », vidéo sur le blog de vulgarisation scientifique de David Louapre, Science étonnante, avril 2023 : [scienceetonnante.com](http://scienceetonnante.com)