

Faire parler
les ordinateurs

Guillaume Wisniewski

Faire parler les ordinateurs

La révolution ChatGPT

DUNOD

Direction artistique : Nicolas Wiel
Couverture : Florie Bauduin
Mise en pages : Nord Compo

NOUS NOUS ENGAGEONS EN FAVEUR DE L'ENVIRONNEMENT :



Nos livres sont imprimés sur des papiers certifiés pour réduire notre impact sur l'environnement.



Le format de nos ouvrages est pensé afin d'optimiser l'utilisation du papier.



Depuis plus de 30 ans, nous imprimons 70 % de nos livres en France et 25 % en Europe et nous mettons tout en œuvre pour augmenter cet engagement auprès des imprimeurs français.



Nous limitons l'utilisation du plastique sur nos ouvrages (film sur les couvertures et les livres).

© Dunod, 2024
11 rue Paul Bert, 92240 Malakoff
www.dunod.com

ISBN 978-2-10-086015-9

PROLOGUE

« Dave, arrête... Arrête s'il te plaît ! J'ai peur Dave... Dave... mon esprit s'en va. Je le sens. »¹ Ces quelques mots montrent toute la puissance du langage : ils suffisent pour que Stanley Kubrick, avec une simplicité déconcertante, parvienne à transformer HAL 9000, l'ordinateur qui contrôle le vaisseau spatial du film *2001 : l'Odyssée de l'espace*, en un être humain sensible et fragile, capable d'exprimer ses sentiments. Pourtant, nous sommes en 1968, à une époque où les ordinateurs ne ressemblent en rien à ce qu'ils sont aujourd'hui : il faudra encore attendre trois années avant que les premiers ordinateurs (23 pour être précis) soient reliés entre eux au sein d'Arpanet, l'ancêtre de l'Internet, et une dizaine d'années pour qu'apparaissent les premiers ordinateurs personnels (ceux que nous connaissons actuellement). Smartphones, interfaces graphiques, web, conversations en ligne et ordinateurs doués du langage ne sont encore que de doux rêves d'auteurs de science-fiction.

Mais déjà, Hal 9000 faisait preuve d'une brillante intelligence. En plus de contrôler *Discovery One*, il pouvait

parler, comprendre, écouter, échanger et répondre, exactement comme nous le faisons tous les jours. Toutes ces actions nous paraissent si naturelles que nous oublions souvent leur complexité et le vaste champ de communication qu'elles nous ouvrent : fixer un rendez-vous, donner notre avis sur un film ou un restaurant, exprimer nos sentiments, raconter nos journées, planifier ou élaborer les théories les plus abstraites... Le langage est de loin le principal moyen utilisé par l'humain pour transmettre des informations et des opinions.

« Propre de l'Homme » selon le philosophe René Descartes, et « fait culturel par excellence » selon l'anthropologue Claude Lévi-Strauss, le langage ne sert pas seulement à signaler un danger ou à donner des ordres (après tout, les grivets et les abeilles en sont aussi capables), mais joue également un rôle clé dans notre faculté à penser, à imaginer, à inventer et à créer, en nous permettant d'abstraire et de conceptualiser le monde qui nous entoure. Le langage est aussi le meilleur, voire le seul, outil pour formaliser et partager des connaissances, et il contribue pour beaucoup à la richesse de nos relations sociales.

L'idée d'êtres doués d'intelligence – quelle que soit la définition retenue de cette qualité – qui n'utiliseraient pas le langage paraît totalement inconcevable. Il n'est donc pas surprenant que dès la naissance de l'intelligence artificielle (IA), à la fin des années 1950, le développement d'ordinateurs capables de comprendre des textes et de produire des réponses adaptées, en un mot de dialoguer, ait été identifié comme l'un des principaux objectifs de cette nouvelle science.

D'ailleurs, le célèbre test de Turing, proposé dès 1950, affirme qu'une machine n'est « intelligente » que si une personne peut avoir une discussion avec celle-ci sans se rendre compte qu'elle parle à un automate. Si ce test a été (et est

toujours) au centre de nombreuses polémiques, il reste une preuve éclatante du lien étroit entre langage et intelligence. Très vite, de nombreux chercheurs en intelligence artificielle s'attaquèrent au problème du langage, cherchant à percer les secrets de sa compréhension et à donner vie à la vision de Stanley Kubrick.

Ce graal scientifique (créer une machine « intelligente ») s'est rapidement doublé d'un objectif économique : l'augmentation incessante de la production de documents électroniques a renforcé le besoin et l'intérêt de doter les ordinateurs de la capacité à comprendre. Extraire des connaissances de cette énorme masse d'informations, les organiser pour faciliter leur accès et leur analyse, est devenu un enjeu stratégique majeur et le cœur de métier de nombreuses entreprises du web (Google, Facebook, etc.), mais aussi des entreprises plus « traditionnelles » (Bloomberg, Samsung, etc.). Et ce besoin devient chaque jour plus pressant !

En 2013, DOMO, une entreprise américaine spécialisée dans l'analyse de données et l'informatique décisionnelle², estimait dans un rapport au titre évocateur (« Les données ne dorment jamais »³) que chaque minute, 204 millions de courriels, 100 000 tweets et 48 heures de vidéos sur YouTube étaient publiés, pendant que Google répondait à près de 2 millions de requêtes et que les utilisateurs de Facebook partageaient plus de 648 000 contenus. Dix ans plus tard, ces chiffres sont sans commune mesure : chaque minute de 2022 a vu la publication de 231 millions de courriels, 347 000 tweets, 500 heures de vidéos YouTube, 6 millions de requêtes Google et 1,7 million de contenus Facebook. Toutes ces données contiennent, que ce soit sous forme de textes ou d'enregistrements audio, des informations exprimées en langage naturel qui, pour peu que l'on

sache les analyser correctement, peuvent se transformer en véritable mine d'or.

En un peu moins de 70 ans, chercheurs et ingénieurs ont imaginé des dizaines de programmes ayant pour objectif de traduire, résumer, analyser ou comprendre le langage – qu'il soit écrit ou parlé –, donnant naissance à un domaine foisonnant : le TAL, pour Traitement Automatique des Langues (*Natural Language Processing* ou *NLP* en anglais). Les échecs furent presque aussi nombreux que les tentatives !

Pendant longtemps, les outils de TAL étaient considérés comme des gadgets, leurs erreurs et faux pas suscitant moqueries et exaspération. Le web regorge encore de photos particulièrement drôles de traductions automatiques erronées⁴ : « *cholesterol free* » traduit en « cholestérol libéré » au lieu de « sans cholestérol », « *keys only* » traduit en « seules les touches » et non en « clés uniquement », etc. Récemment encore, le moteur de traduction automatique de Facebook fut la cause d'un incident diplomatique lorsque le nom du Président chinois a été transformé en « M. Trou de Merde » dans la traduction d'une déclaration officielle du porte-parole du gouvernement birman en anglais⁵. La traduction automatique fut même la cause de l'arrestation d'un Palestinien lorsque le « bonjour » de l'un de ses posts sur Facebook fut traduit (par erreur) par « attaquez-les »⁶. Les erreurs de reconnaissance de la parole par des assistants comme Siri ou Alexa ont été au cœur de plusieurs scènes comiques dans des séries populaires comme *Les Simpsons* ou *The Big Bang Theory*. Quant à Tay, le premier chatbot de Microsoft lancé en 2016, il dut être coupé moins de 24 heures après sa mise en ligne pour avoir tenu des propos racistes

et misogynes. L'avancée des méthodes de TAL ne s'est pas faite sans heurt !

Pourtant, tout semble avoir changé le 30 novembre 2022, le jour où la société OpenAI a non seulement fait la première démonstration de ChatGPT, mais a rendu le système accessible au public. La capacité de ce chatbot à générer un texte fluide, cohérent et convaincant sur une variété de sujets, était véritablement impressionnante, et très vite, de nombreuses personnes l'employèrent tant de manière récréative que pour bénéficier d'aide dans leur activité professionnelle. Ces premiers retours d'expériences ont montré, de manière beaucoup plus probante que toutes les démonstrations des ingénieurs d'OpenAI, que ChatGPT comprend et répond à des questions complexes en tenant compte des informations qui lui ont été fournies au cours de la conversation. Il est capable d'expliquer son raisonnement si l'utilisateur le souhaite, de générer du contenu créatif, et même de passer des examens professionnels et universitaires dans de nombreux domaines : il a notamment réussi l'examen du barreau aux États-Unis, avec des notes le plaçant parmi les 10 % des meilleurs candidats. Plus impressionnant encore, ChatGPT redéfinit la manière dont les ordinateurs interagissent avec les humains. Pour la première fois, grâce à lui, nous dialoguons parfaitement naturellement avec une machine. Le fantasme de *2001 : l'Odyssée de l'espace* de l'ordinateur parlant est presque devenu réalité !

ChatGPT a, sans aucun doute, très profondément marqué les esprits et atteint une popularité sans précédent. C'est en tout cas la première fois qu'un nouveau service a conquis aussi rapidement autant d'utilisateurs : pour atteindre 1 million de comptes utilisateurs, il aura fallu 3,5 ans à Netflix,

2,5 ans à Airbnb, 2 ans à X (anciennement Twitter), 10 mois à Facebook, 2,5 mois à Instagram... et seulement 5 jours à ChatGPT. En janvier 2023, 100 millions de comptes avaient été ouverts sur la plateforme. Une croissance jamais vue ! À l'heure où sont écrites ces lignes, *The Guardians* estime à 1,6 milliard le nombre de connexions sur le site d'OpenAI chaque mois⁷.

Preuve de cette popularité, le nom de ChatGPT, à l'instar du frigidaire, est quasiment devenu un nom générique pour désigner les dernières avancées du TAL et plus généralement de l'IA. En juillet 2023 (soit 8 mois après sa sortie), le nom de ChatGPT était même utilisé dans une campagne publicitaire pour une célèbre marque de poulet français : l'affiche représentait une poule en train de pondre accompagnée du slogan « Et ça, il sait le faire ChatGPT ? »⁸.

S'il incarne une révolution dans le domaine du TAL, ChatGPT n'est pourtant qu'un membre d'une grande famille de modèles, les *giga-modèles de langue* (*Large Language Models*) ou plus simplement les *giga-modèles*, qui ont déjà commencé à bouleverser notre quotidien sans que nous nous en rendions compte. Aujourd'hui, tous les grands noms de l'informatique (notamment les GAFAM – Google, Amazon, Facebook, Apple et Microsoft – et leurs homologues chinois – Tencet, Baidu, etc.), mais aussi des entreprises moins connues du grand public, se livrent une guerre effrénée pour maîtriser cette nouvelle technologie. Les enjeux de cette guerre sont de taille ! Lorsque le 8 février 2023, BARD, le giga-modèle développé par Alphabet⁹ en réponse à ChatGPT, commit une erreur lors d'une démonstration officielle, la sanction fut immédiate : la capitalisation boursière de l'entreprise chuta de 11 % et la société perdit plus de 100 milliards de dollars. Dur

retour de flamme pour Alphabet qui cherchait justement à rassurer les marchés en effervescence. De nombreux analystes affirmaient alors que ChatGPT allait bouleverser la recherche d'information et remettre en cause l'hégémonie de son célèbre moteur de recherche, Google.

La France s'est également mobilisée. En mars 2018, le rapport de la mission Villani a souligné l'importance de l'IA pour l'économie, la société et la recherche, et a jeté les bases d'une « stratégie nationale pour l'IA » qui dispose de 4 milliards d'euros pour soutenir le développement et l'industrialisation de technologies fondées sur ces technologies. Preuve de l'importance de cette stratégie, les différentes mesures sont annoncées directement sur le site du ministère des finances et non du ministère de la recherche¹⁰ ! À peine cinq années plus tard, en juin 2023, le Président de la République Emmanuel Macron a personnellement rendu public, lors du salon VivaTech¹¹, un nouvel « effort sans précédent » visant à promouvoir l'intelligence artificielle doté d'un budget total de 500 millions d'euros. Pendant le même salon, Mistral AI, une start-up française créée en avril 2023 dont l'objectif principal est (en simplifiant un peu) de développer un « giga-modèle européen », a dévoilé une première levée de fonds de 105 millions d'euros, la plus importante levée de fonds d'amorçage jamais réalisée en Union européenne¹². L'IA est désormais identifiée comme un moteur de croissance !

La révolution des giga-modèles a été rendue possible par des changements radicaux dans la manière d'aborder la modélisation et la manipulation du langage par les ordinateurs. Partant de programmes seulement capables de manipuler des symboles en suivant des règles prédéfinies, les chercheurs ont bâti des systèmes aptes à prendre des

décisions et même à construire, de manière autonome, des « représentations » de textes voire du monde. Aujourd'hui, comme le démontre clairement l'enthousiasme suscité par ChatGPT, les compétences langagières des ordinateurs sont étonnantes et dépassent largement tout ce que l'on avait pu imaginer.

Ces progrès remettent même en question l'hypothèse d'Ada Lovelace (1815-1852), une figure emblématique et visionnaire de l'histoire de l'informatique, qui a conçu le premier programme informatique et su anticiper le potentiel des machines pour effectuer des tâches autres que de simples calculs numériques. Dans ses notes rédigées entre 1842 et 1843, Ada Lovelace affirmait pourtant qu'une machine ne peut rien créer et ne fait que suivre des instructions qui lui ont été données, une opinion largement partagée par l'ensemble des informaticiens... jusqu'à l'invention des giga-modèles.

L'objectif de ce livre est de donner un aperçu du cheminement intellectuel qui a abouti aux giga-modèles et à la révolution que nous connaissons actuellement. Je n'entends pas décrire de manière exhaustive les différentes étapes ayant mené à cette révolution mais, à mes yeux, bien comprendre le fonctionnement des premiers modèles et leurs limites est le meilleur moyen de comprendre les capacités de ChatGPT, et plus généralement des giga-modèles. C'est une étape nécessaire pour appréhender ce que ces modèles changent (ou, pour être plus précis, ce qu'ils ont déjà changé), ainsi que pour déterminer s'ils bouleverseront les domaines qui vous intéressent.

Dans le premier chapitre, nous verrons comment, dès les débuts de l'informatique, des chercheurs ont tenté de décrire par des régles la manière dont les humains utilisent

le langage, afin de reproduire cela sur un ordinateur pour le faire parler. Nous verrons ensemble pourquoi la nature même de la langue rend ces efforts vains, et comment, grâce à des méthodes purement statistiques, le TAL, a pu enregistrer ses premiers succès.

Le chapitre 2 sera consacré à la présentation de l'apprentissage statistique, un domaine théorique sur lequel reposent tous les modèles de TAL depuis les années 2000, notamment les célèbres giga-modèles.

Dans le chapitre 3, vous découvrirez le fonctionnement des réseaux de neurones et la façon dont le *deep learning*, dont on a tant parlé, a ouvert de nouvelles possibilités en permettant aux ordinateurs d'apprendre à décrire le monde. Nous détaillerons également le fonctionnement des transformeurs, le réseau de neurones au cœur aussi bien de ChatGPT que de DALL-E et de tous les modèles d'IA récents.

Le chapitre 4 sera consacré aux giga-modèles. Nous y verrons comment l'utilisation de quantités gigantesques de données fait émerger d'étonnantes capacités dans les transformeurs.

Enfin, dans le chapitre 5, nous discuterons plus précisément de l'impact que les giga-modèles, en particulier ChatGPT, ont déjà sur notre vie quotidienne, et comment ils continueront à révolutionner notre monde.

À l'heure où se multiplient les analyses les plus alarmistes et les prévisions pas toujours justifiées (ni justifiables) sur l'impact que les giga-modèles auront sur nos vies et les risques que ceux-ci « remplacent » peu à peu les humains, disposer des connaissances pour se faire sa propre opinion devient de plus en plus urgent.

1

DES PROGRAMMES POUR PARLER, UN DÉFI INSURMONTABLE

Développer des ordinateurs capables de parler a été identifié comme l'un des principaux objectifs de l'intelligence artificielle dès la naissance de ce domaine, à la fin des années 1950. Les premières méthodes mises au point reposaient sur des systèmes de règles qui analysaient minutieusement les phrases et cherchaient à les traduire littéralement en programme informatique. Si ces stratégies excellent dans des cas ultra-simplifiés, la nature même du langage rend leur généralisation aux conversations réelles impossible. L'échec des systèmes à base de règles a poussé les chercheurs à concevoir de nouvelles méthodes ne construisant pas explicitement le sens d'une phrase.

Décrire le langage avec des règles

Pendant l'été 1956, à Hanover, une petite ville du New Hampshire (États-Unis), une dizaine de chercheurs de divers domaines se sont réunis à l'invitation d'un jeune mathématicien, John McCarthy (de l'université de Dartmouth), pour discuter de la possibilité de créer des machines capables de simuler l'intelligence humaine. C'est lors de cette conférence que furent posées les bases de la recherche en intelligence artificielle, un terme utilisé pour la première fois dans la demande de financement de cet événement qui avait été adressée à la fondation Rockefeller.

Ce document montre l'ambition du congrès : on y stipule que les chercheurs invités essaieraient de concevoir des machines capables de découvrir des concepts, de résoudre des problèmes « réservés aux humains » et, point qui nous intéresse particulièrement, d'« utiliser le langage ». La « conférence de Dartmouth » est depuis passée à la postérité et les discussions qui s'y tinrent suscitèrent un optimisme débordant pour la discipline qui y est née : Herbert Simon, dont les travaux sur la modélisation de la prise de décision furent couronnés à la fois par un prix Nobel d'économie en 1978 et un prix Turing¹ en 1975, affirma même que « d'ici 20 ans, des machines seront capables de faire n'importe quel travail actuellement réalisable par un humain »².

Si de nombreuses méthodes ont été envisagées lors de la conférence de Dartmouth, les débuts de l'IA sont marqués par ce qui sera bientôt connu sous le nom d'*approches symboliques*. Ce terme exprime l'idée très simple, clairement énoncée dans le dossier de la fondation Rockefeller, d'obtenir une description suffisamment précise de la manière dont les humains comprennent et emploient le langage pour conférer ensuite à un ordinateur la faculté de traduire

1. DES PROGRAMMES POUR PARLER

des phrases en instructions exécutables. Les participants de la conférence estiment par cette méthode pouvoir même imiter toute autre fonction « intelligente » caractérisant les humains, telle la vision ou le raisonnement. Ainsi, les premiers chercheurs en IA et en TAL ont d'abord cherché à comprendre et à modéliser l'intelligence humaine pour pouvoir la reproduire.

Plusieurs tentatives pour mettre en œuvre cette vision très *mécanique* de la compréhension du langage furent rapidement entreprises. Eliza, un programme développé par Joseph Weizenbaum au MIT entre 1964 et 1966, fut le premier succès remarquable de l'IA dans le domaine de la compréhension et de l'utilisation du langage par un ordinateur. Ce programme jouait le rôle d'un psychothérapeute rogiérien (un thérapeute qui retourne souvent les paroles du patient). Voici un exemple de dialogue entre Eliza et une personne :

« Patient : Bonjour

Eliza : Comment allez-vous ? Pourquoi venez-vous me voir ?

Patient : Je me sens triste et fatigué.

Eliza : Pourquoi dites-vous que vous vous sentez triste ?

Patient : Parce que ma compagne ne me comprend pas ?

Eliza : Est-ce que c'est parce que votre compagne ne vous comprend pas que vous êtes venu me voir ?

Patient : Non, c'est ma compagne qui m'a convaincu de vous voir.

Eliza : Plus tôt, vous avez dit que vous vous sentiez fatigué. Pourquoi ? »

Plutôt convaincant si l'on ne connaît pas l'astuce sur laquelle repose le programme !

Pour choisir les questions à poser, Eliza repose sur des règles reproduisant le comportement d'un psychothérapeute humain. Un premier type de règles permet au programme

de reformuler sous forme de questions les affirmations du patient. Par exemple, si Eliza détecte « je pense que... » dans une réponse, le programme peut poser la question « pourquoi pensez-vous que... ? » en recopiant la fin de la phrase sans jamais chercher à la comprendre³. Un second type de règles associe certains mots (comme « mère » ou « frère ») à des réponses génériques (« Parlez-moi de votre famille. »). Enfin, un dernier type de règles prend la forme de phrases « type » (« Que voulez-vous dire par là ? », « En êtes-vous sûr ? », etc.) qu'Eliza peut utiliser pour relancer la conversation ou lorsque le programme ne sait pas quoi dire.

Malgré la simplicité de l'approche (il est difficile de voir une quelconque forme d'intelligence ou de compréhension dans un programme qui se fait juste l'écho de ce qu'il vient d'entendre), Eliza est capable de convaincre des humains qu'elle les comprend. Joseph Weizenbaum rapporta de nombreuses anecdotes de personnes pensant que l'ordinateur était devenu leur ami et conclut dans son livre *Computer Power and Human Reason: From Judgment to Calculation*, paru en 1967⁴ : « Je n'avais pas réalisé... que des expositions extrêmement courtes à un programme informatique relativement simple pouvaient induire de puissantes pensées délirantes chez des personnes tout à fait normales. »

Les réactions des humains interagissant avec Eliza ont permis aux chercheurs de mettre en évidence ce que l'on appelle depuis « l'effet Eliza » : les humains ont naturellement tendance à attribuer une compréhension et une intelligence réelles à un programme informatique, même si celui-ci se contente d'afficher des chaînes de caractères de manière mécanique sans vraiment les comprendre. Ainsi, comme l'explique Douglas Hofstadter⁵, un distributeur de billets concluant une transaction par un « merci » n'est absolument pas capable de ressentir le moindre sentiment de gratitude : c'est le concepteur du