

Le paradoxe de la chambre chinoise

Soumis par Philippe YONNET

18-11-2007

Dernière mise à jour : 18-11-2007

Les programmes de type Eliza ou Alice ne sont pas intelligents... Ils ont pour objectif d'imiter l'intelligence. Le problème est qu'il s'avère assez simple de tromper quelque temps un utilisateur peu averti, et de lui faire croire qu'il dialogue avec un robot à esprit avancé. Les philosophes du vingtième siècle ont étudié cette question : peut-on faire la différence entre un système qui donne des réponses préétablies à des questions attendues, et les réponses données par un humain ? La réponse est : non...

L'histoire de l'intelligence artificielle est parsemée de fraudes spectaculaires et de tromperies parfois géniales... La plus célèbre entre elles est celle du Baron Kempeln qui a parcouru les cours d'Europe au 18e siècle pour montrer son automate joueurs d'échec (en réalité piloté par un homme de petite taille caché dans son socle). On a souvent parlé de canular ou de tromperie à propos du programme Eliza... Sauf que là son créateur annonçait la couleur. A son grand étonnement, est le public qui avait du mal à croire qu'il dialoguait avec une suite de deux cent règles simples ! Il y a une maxime qui a cours chez les geeks : "on ne peut pas faire la différence entre une technologie réellement avancée et une démo astucieuse". Les programmes qui s'appuient sur un coup de pouce humain pour fonctionner correctement, ou un tour de magie, sont appelés "programmes pnambiques". Deux mots sur le mot "pnambique" Le terme pnambique, qui semble directement sorti d'une nouvelle de Lovecraft, est en fait bien tiré de la littérature, mais indirectement, en faisant un tour par le cinéma... Il s'agit d'un acronyme : P.N.A.M.B.I.C. tiré d'un dialogue du film "Le magicien d'Oz" : "Pay no attention to the man behind the curtain" (ne faites pas attention à l'homme derrière le rideau). PNAMBIC fut le premier nom du chatterbot Alice ALICE ne s'est pas toujours appelée Alice. Son premier nom était justement PNAMBIC, en hommage au rôle de la tromperie dans l'histoire de l'intelligence artificielle. Alice était le nom du premier serveur qui hébergeait Pnambic, ce qui a fait que tout le monde s'est mis à appeler Alice. Mais Alice est plus qu'un simple programme pnambique Le terme pnambique est-il adapté pour qualifier un programme de type Alice ? Oui et non, pour des raisons qui ont trait à l'ambiguïté profonde des techniques d'intelligence artificielle. D'un côté, les chatterbots de la famille d'Alice ont hérité d'Eliza des méthodes destinées à simuler une conversation humaine, ce qui finit par faire croire à l'utilisateur, soit qu'il a affaire à un programme très avancé, soit à une forme d'intelligence. Les "botmasters" de ce point de vue agissent comme le magicien d'Oz derrière son rideau. D'un autre côté, si Toto, le chien de Dorothy, vient à tirer le rideau derrière le chatterbot, il ne trouvera pas le "botmaster". Celui-ci est parti depuis longtemps, il a fait qu'écrire les règles qui dictent à l'avance le comportement du chatterbot... l'intelligence artificielle : une approche paradoxale Alan Turing, le célèbre mathématicien anglais père de l'informatique moderne, est à l'origine d'un célèbre test théorique, destiné à déterminer si une machine est ou n'est pas intelligente. Le test de Turing nous fait rester dans l'univers du magicien d'Oz. Il s'agit de faire interagir un humain avec une machine, en mode textuel. Si l'humain n'arrive pas à distinguer si les réponses sont celles d'un autre homme situé dans la pièce à côté, ou celles d'une machine, alors la machine est censée être dotée d'une "intelligence artificielle". Le problème du test de Turing, est qu'il part d'une hypothèse qui pose problème : on peut déduire qu'il y a une intelligence, mieux : une personne, derrière un programme en interagissant avec lui. Or, est faux ! Les déboires du père d'Eliza avec des utilisateurs du programme qui se comportaient avec lui comme avec leur psy (notamment sa propre assistante) ont provoqué une prise de conscience dans le monde de l'intelligence artificielle : il n'est nul besoin d'être intelligent pour bavarder avec un humain, et le fait que le système donne des réponses jugées correctes ne prouvent rien quant à l'intelligence du système. Pire : donner des réponses correctes ne signifie pas que l'on comprend la question, et pas plus que l'on comprend la réponse ! C'est le paradoxe décrit par Searle. Le paradoxe de la chambre chinoise de Searle Au cours des années 80 le philosophe John Searle a décrit une expérience théorique qu'il a dénommé "le paradoxe de la chambre chinoise". Le dispositif de l'expérience est simple : imaginez une personne enfermée dans une pièce. Il reçoit des messages en Chinois écrits sur des bouts de papier écrits et glissés sous la porte par une autre personne, un chinois, qui parle le mandarin. La personne enfermée dans la pièce ne parle pas le chinois, et encore moins ne sait le lire ou l'écrire. Par contre, l'occupant de la chambre chinoise dispose d'un manuel, qui contient des instructions : si vous voyez tel signe et tel signe, alors vous tracez tel trait comme ci et tel trait comme ça... Bref, il reçoit des textes en idéogrammes, les identifie avec le manuel, et trace d'autres idéogrammes en réponse, après les indications du manuel. Maintenant, si on se place du point de vue du chinois à l'extérieur de la chambre, tout se passe comme si l'occupant comprenait ses questions, puisque la réponse est parfaitement compréhensible et adaptée au contexte. Pourtant, l'occupant de la chambre chinoise ne comprend rien aux questions, ni aux réponses qu'il donne... Les programmes conversationnels de la famille d'Alice fonctionnent clairement de manière identique à cette expérience de la chambre chinoise. Les catégories d'Alice sont très semblables au livre de règles. Et l'intelligence là dedans ? Et bien, il y a bien des connaissances et une procédure établie (un programme et une base de données), mais le programme n'a pas besoin de comprendre pour sembler comprendre. C'est que Searle met en avant dans son livre. Cette expérience rappelle aussi un célèbre livre de science fiction de Verner Vinge, "True Names"... Dans ce classique, préfigurant le genre

"cyberpunk", un certain nombre de renégats se connectent régulièrement à un réseau global, en masquant leur identité réelle. Un jour, un nouveau personnage apparaît : mais on se rend compte plus tard que l'homme derrière le pseudo n'existe pas, c'est en fait un programme, une intelligence artificielle, créée par le gouvernement ! Une Intelligence Artificielle qui passe visiblement le test de Turing, puisqu'elle réussit à tromper les vrais membres du réseau. Le rôle de la psychologie dans l'interaction homme - machine Le chatterbot Eliza a mis en lumière le comportement particulier des utilisateurs face à ce type de programmes. Il s'avère que la fascination véritable des gens pour ce type d'applications provient d'un mélange détonant entre trois éléments de la psychologie humaine. Le besoin de croire au merveilleux d'abord. Tout le monde sait qu'il y a un truc derrière les exploits d'un David Copperfield : et pourtant, on se laisse facilement aller à se dire "il a réellement fait disparaître la Statue de la Liberté, mince alors !". Il suffit qu'un chatterbot réponde à trois questions de suite de manière jugée pertinente, pour que l'utilisateur se dise "mince, alors ! Cette chose comprend mes questions ! Elle est intelligente". Même les esprits les plus rationnels réagissent en disant "Wow ! Je sais qu'il y a un truc, mais wow !". Par ailleurs, un élément renforce cette impression : les ordinateurs sont les premières entités non humaines de la Création qui peuvent utiliser le langage humain. Cela fait trois générations seulement que nous avons l'expérience d'interaction avec ce type de machines, et seulement une génération qui en a une expérience quotidienne. Pour nos contemporains, certaines prouesses de ces machines relèvent toujours de la sorcellerie, certains pensent donc que derrière le rideau, il y a un homme qui tire les ficelles, d'autres que par une diablerie, ces machines pensent vraiment... Ensuite, il y a le côté gratifiant de l'échange et de l'interaction avec autrui. Les humains aiment bavarder, et peu importe si c'est de sujets sans intérêt... On peut donc réellement prendre plaisir à échanger avec une machine, même idiote, et finir par croire que l'on échange avec un nouvel ami, comme on le fait sur un chat sur internet (d'ailleurs, on peut connecter des chatterbots sur les programmes de chats ou d'IRC, et dans ce contexte, ils font souvent longtemps illusion). D'ailleurs, souvent, le besoin d'écoute est supérieur à la volonté d'avoir des réponses... Le plus important est donc que la machine semble comprendre les questions, pas qu'elle sache y répondre de manière exacte, et même appropriée... Enfin, les humains aiment jouer à trouver les failles dans les programmes d'intelligence artificielle. Alan Turing l'avait déjà analysé, en déclarant : "A chaque fois qu'un programme d'intelligence artificielle se voit poser une question délicate, [...] et qu'il répond à côté, cela nous donne un sentiment de supériorité." Le résultat, c'est que nous éprouvons du plaisir à jouer avec des machines imparfaites, et qu'à contrario, les machines parfaites peuvent nous faire peur au lieu de nous amuser et de nous ennuyer ! Turing, dans le même texte, annonçait toutefois que la supériorité de l'homme sur la machine était relative et sans doute provisoire... "Ce sentiment [de supériorité] est-il illusoire ? Il est sans doute assez naturel, mais je pense qu'il ne faut pas lui accorder une grande importance. Nous aussi nous donnons souvent des réponses erronées, pour pouvoir justifier une telle satisfaction devant les défauts de ces machines." De plus, cette supériorité ne peut s'éprouver que dans le contexte de la machine aux dépens de laquelle nous avons obtenu notre petit triomphe mesquin. Il ne saurait être question de triompher de toutes les machines. En peu de temps, une machine peut surclasser toutes les autres par son intelligence, et d'autres machines peuvent encore derrière devenir plus intelligentes" Eliza faisait illusion de temps à autre. Alice y parvient la plupart du temps, sans passer le test de Turing... Mais demain ? Le paradoxe de la chambre chinoise annonce qu'on ne saura plus s'il y a un magicien derrière le rideau du magicien Oz, bien avant de savoir si l'entité qui répond est intelligente, ou même, comprend les questions qu'on lui pose... Nota Bene : cet article est librement inspiré par un article du Pr Wallace sur le même sujet.