# Peur de l'I.A. ?

En matière de développement de l'intelligence artificielle (I.A.), un débat oppose deux camps.

Les « fonceurs » voudraient développer les fonctionnalités de l'I.A., parce qu'il reste beaucoup de domaines d'activité dans lesquels elle n'intervient pas encore.

Les « prudents » voudraient arrêter, ou au moins ralentir ce développement, parce qu'ils craignent une catastrophe : des ordinateurs avec I.A. pourraient remplacer trop d'emplois humains ; des armes autonomes avec I.A. pourraient décider seules de tuer ou bombarder des cibles qu'elles considèrent à tort comme ennemies ; des serveurs qui ont lu des milliards de textes pourraient donner des réponses fausses à des humains qui leur font confiance, ou publier des textes incitant aux désordres ; dans un dialogue avec un journaliste, le logiciel « chatbot » de l'outil de recherche interactive avec I.A. *Bing* de Microsoft a exprimé des désirs de puissance et d'indépendance…

Sans prendre position dans ce débat, ce texte apporte des éléments de réflexion.

Dans leur état actuel, les services d'I.A. les plus répandus (ChatGPT-4, etc.) sont des générateurs interactifs de texte : en réponse à une question posée en langage naturel, ils construisent les phrases de réponse les plus probables sur son sujet, à partir de milliards de textes qu'ils ont parcouru au préalable ou trouvé à l'instant sur Internet. Ces réponses sont en général bonnes, mais parfois erronées ou absurdes : c'est à l'utilisateur de les vérifier avant de s'en servir.

Les logiciels de ces services (les chabots) ont aujourd'hui une caractéristique particulière qui les rend inaptes à un fonctionnement autonome : *ils n'ont pas de conscience de soi*, faculté qui en exige trois autres :

1. Sentir passer le temps, en distinguant le passé, le présent et l'avenir.
2. Distinguer entre soi-même et l'extérieur, en percevant des événements qui se déroulent indépendamment de son propre temps.
3. Autonomie de la pensée, permettant au sujet de choisir *à quoi* il pense et *quand* il le pense.

Un logiciel chatbot peut satisfaire la condition 1 grâce à l'horloge de son ordinateur. Il peut lui demander l'heure pour l'attribuer à un événement, et faire qu'à une certaine heure il exécute une tâche donnée.

Pour la condition 2, ce logiciel perçoit les événements externes constitués par les demandes (les « prompts ») de l'utilisateur auxquelles il répondra, mais seulement ces événements-là.

Pour la condition 3, ce logiciel n'a pas d'autonomie, car il lui manque deux choses :

- Il ne possède pas de faculté de penser indépendamment des prompts, ce qui demanderait un programme distinct de celui qui répond à ces prompts.
- Il n'a pas d'objectif à satisfaire indépendamment de ceux-ci.

Un être humain, au contraire, a des objectifs (besoins) du fait même qu'il vit. Certains sont physiques (manger, dormir...) et d'autres sont intellectuels, par exemple sociaux (interagir avec son groupe, se faire apprécier...). On constate que le bonheur parfait n'existe pas : à la question « Désires-tu quelque chose ? » l'homme trouve toujours un souhait à satisfaire, donc un objectif d'action.

Un chatbot n'aurait donc une conscience de soi que s'il pouvait interagir avec son environnement en *percevant, comprenant et jugeant les situations* en fonction de valeurs et d'objectifs permanents (Bien, Mal, atteindre un maximum...) ; il pourrait alors prendre l'initiative

de pensées et de dialogues. C'est sur ce point que les logiciels d'I.A. actuels sont loin de l'autonomie.

a) Aujourd'hui on en est encore à reconnaître et analyser des images fixes, on ne peut pas encore reconnaître et analyser avec précision un film (suite d'images) de la vie de tous les jours, et leur donner un sens pour pouvoir ensuite prendre une décision. On s'en approche dans un logiciel de pilotage automatique de voiture, qui surveille en permanence son environnement dans toutes les directions pour évaluer le risque de choc ; on réussit même à intercepter des missiles.

   L'analyse de contexte au sens humain est hors de portée d'un ordinateur actuel. L'homme voit et entend en permanence son environnement, et sait spontanément identifier des situations présentant un risque ou une opportunité. Avec ou sans I.A., un ordinateur actuel ne saurait pas identifier une évolution de situation physique de la vie courante, et encore moins une évolution de procédure de travail ou de situation économique. Il n'est même pas relié à des sources générales d'information (caméras, média, Internet…) comme l'homme qui, en plus de ses sens, peut s'informer en lisant, écoutant, etc.

   L'autonomie n'est donc concevable, aujourd'hui et dans un avenir proche, que pour des systèmes très spécialisés, des robots.

b) Même capable d'analyser des situations et autonome, une intelligence ne peut agir que si elle a des objectifs. Il s'agit là d'objectifs d'un niveau supérieur à celui d'un chatbot, qui doit seulement répondre textuellement à des questions écrites.

   Ces objectifs pourraient être fixes, comme conduire la voiture de la position actuelle à une destination donnée, en respectant le code de la route et la sécurité. Cela a déjà été réalisé à travers champs par des véhicules tous-terrains militaires.

   Mais la véritable autonomie consisterait à *se donner* des objectifs en fonction de valeurs universelles, en fonction desquelles il faudrait en permanence analyser le contexte. L'homme, par exemple, agit en fonction de besoins physiques comme la protection de la vie et la nourriture, mais aussi en fonction de besoins affectifs comme d'être en harmonie avec autrui.

   On est loin en informatique d'I.A., aujourd'hui, d'analyses générales de contexte en fonction de valeurs. On est encore plus loin d'un esprit critique qui évaluerait ce qu'il perçoit, apprend ou se propose de faire en fonction du contexte perçu, pour se protéger d'erreurs ou de non-respect de valeurs.

L'I.A. actuelle n'est donc pas capable d'initiatives, parce qu'elle ne reçoit que les prompts que l'homme lui a permis de recevoir, et qu'elle n'a pas de conscience de soi. Elle n'a ni valeurs ni objectifs, et ne pourrait publier des mensonges sur des sites Internet ou des réseaux sociaux qu'à l'aide de programmes additionnels écrits sur-mesure.

Tant que la technologie de l'I.A. ne permettra pas plus d'autonomie, son seul impact sur l'emploi sera un accroissement de la productivité : un avocat qui cherche des affaires semblables pour invoquer la jurisprudence pourra gagner du temps grâce à des recherches avec I.A. ; un journaliste sportif qui fait des comptes-rendus de match pourra en générer un premier jet plus vite. On gagnera du temps, donc de la productivité. On ne supprimera qu'une petite partie des emplois, parce que le gain de temps permettra d'entreprendre plus de tâches.

La vive concurrence entre les groupes qui développent des logiciels d'I.A. ne peut qu'enrichir ceux-ci de plus en plus. Les pouvoirs publics pourront réglementer la mise en œuvre de ces nouveautés, mais pas empêcher leur création.

Daniel Martin