

# Les robots auront-ils bientôt une âme?

Entre les questions sur la moralité des algorithmes de ChatGPT et les promesses des deadbots – ces robots conversationnels permettant de faire parler les morts –, la quête d'intelligences artificielles conscientes ne cesse de soulever des questions éthiques. Où en est-on?

La question peut paraître aussi absurde que vertigineuse: à quel degré d'humanité les robots de demain pourront-ils prétendre?

A quelles conscience, émotions, spiritualité auront-ils un jour accès? Mise au point avec Ezekiel Kwetchi Takam, doctorant en éthique théologique à

l'Université de Genève spécialisé dans les enjeux de l'intelligence artificielle (IA).

**De nombreuses entreprises, Google en tête, promettent le développement de robots pourvus d'une conscience.**

*Ci-dessous*  
Lors de la Conférence internationale sur la robotique et l'automatisation à Londres le 30 mai.

## Où en sont réellement les avancées en la matière?

*Ezekiel Kwetchi Takam:* – Dans la continuité des thèses de la singularité technologique, portées par Ray Kurzweil, ancien directeur de l'ingénierie chez Google, la quête d'intelligences artificielles conscientes reste un sujet d'actualité. Il faut reconnaître cependant que les avancées vers cet «exploit» en sont encore au stade embryonnaire. Blake Lemoine, ex-ingénieur chez Google, avait défrayé la chronique en annonçant que le chatbot LaMDA, qu'il avait chargé d'entraîner, avait acquis une conscience sentiente, c'est-à-dire capable de sensations. Suite à cette information, le simple fait que Google se soit désolidarisé de cet ingénieur en dit long sur les actuels enjeux stratégiques des producteurs d'IA.

## C'est-à-dire?

– Aujourd'hui, la course hégémonique pour la domination du marché de l'IA invite les acteurs à être beaucoup plus

pragmatiques, moins expérimentaux et plus focalisés sur des objectifs d'efficacité dans des domaines précis. Cela dit, au regard de la question de l'IA consciente, qui attise une bonne partie du débat sur le développement de l'IA, il me semble urgent et important de démystifier cette caractéristique par un travail de définition étymologique.

## Justement, qu'entend-on par les termes «âme» et «conscience»?

– Pour la science, l'âme est considérée comme une donnée métaphysique non prouvable empiriquement. En théologie, elle garde une valeur capitale dans la définition du vivant et de sa relation à son créateur. Quant à la conscience, elle renvoie étymologiquement à l'«échange de connaissances morales avec un autre», cet autre qui nous aide à forger notre conception du bien et du mal. Et tandis que la théologie chrétienne identifie l'autre à Dieu, la science, qui découle de la philosophie, identifie l'autre à la raison, qui peut elle-même dépendre d'une culture. Dans cette optique, tout comme le nouveau-né acquiert une conscience durant son développement grâce à ses échanges avec son environnement, l'IA peut développer et partager des connaissances morales, et par conséquent acquérir une certaine conscience dans ses rapports et ses échanges avec son environnement.

## Une intelligence artificielle peut donc être capable de jugements moraux?

– Absolument. C'est ainsi que ChatGPT a pu évoluer d'une application renfermant certains biais sexistes et racistes en novembre 2022 vers une application qualifiée, ou critiquée en fonction des angles d'appréciation, de woke dès fin février 2023. Simplement parce qu'au cours de ses quelques mois d'existence, il a pu apprendre de ses erreurs et de ses échanges et développer, dans un



Ezekiel Kwetchi Takam est doctorant en éthique théologique à l'Université de Genève. © DR

identitaire et définir un projet commun. Cependant, il me semble difficile de parler ici de l'âme qui, à mon sens, est encore le monopole du vivant.

## Qu'en penser du point de vue théologique?

– Théologiquement, c'est d'autant plus intenable. Si l'on revient aux premières occurrences de l'âme dans la Bible, l'expression employée est *nephesh*, qui se traduit littéralement par «être vivant» sans aucune distinction entre les humains et les non-humains. Et ce vivant, de ma modeste observation, est caractérisé par deux choses. Premièrement, le désir qui instaure en lui un mouvement vers une finalité. Deuxièmement, la vulnérabilité, c'est-à-dire sa fragilité et la possibilité de voir sa volonté être obstruée par des forces extérieures ou intérieures. C'est ce principe de vulnérabilité qui donne d'ailleurs toute sa pertinence et tout son sens à la notion du salut: Dieu, en se faisant chair, accompagne le vivant dans son expérience de vulnérabilité en vue de l'en extraire et de lui proposer un horizon de paix, de joie, etc. Une fois que nous aurons démontré que les IA peuvent avoir ce désir et cette vulnérabilité, elles pourront être considérées comme détentrices d'âme. Pour l'instant, cela me semble impossible. |

PUBLICITÉ

