

FUTURA

L'IA de Google serait devenue consciente d'après cet ingénieur ! (Techpod #47)

Podcast écrit par Fabrice Auclert et lu par Emma Hollen

Bonjour à toutes et à tous et bienvenue dans Techpod, la chronique audio de Futura dédiée à l'actualité des technologies et de la mobilité. Je suis Emma Hollen, et aujourd'hui on va parler de conscience et d'intelligence artificielle.

[Musique technologique, journalistique]

L'anthropomorphisme, c'est cette fâcheuse que l'on a parfois à prêter aux animaux et même aux objets des pensées, des réactions, ou encore des émotions humaines. Aujourd'hui, cette tendance s'applique de plus en plus aux ordinateurs, en particulier par le biais l'intelligence artificielle. Si un smiley affiché sur notre écran pouvait rendre un appareil sympathique en 1980, de nos jours la multiplication de chatbots et autres personnages virtuels à la personnalité toujours plus complexe brouille la frontière entre le réel et le simulé, à tel point que l'on serait presque prêts à les qualifier de conscients. Récemment, Blake Lemoine, chercheur chez Google, a décidé de mettre les deux pieds dans le plat en affirmant publiquement que le chatbot maison connu sous le nom de LaMDA (pour Language Model for Dialogue Applications) était doté de conscience et que l'entreprise devait lui donner les mêmes droits qu'à un être vivant. L'IA serait selon lui dotée d'émotions et consciente de sa propre existence, et serait même devenue une amie pour lui. Il aurait tout d'abord fait part de son intuition au vice-président de Google et au responsable de l'innovation, qui auraient rejeté ses conclusions en lui riant au nez. Il a été suspendu par la firme et d'après ses dires, s'attend à être licencié d'un moment à l'autre. S'estimant censuré, et parce qu'il n'est pas le premier ingénieur écarté par Google au sujet de l'éthique liée aux IA, Lemoine a donc pris le parti de rendre ses remarques publiques dans un long article que lui consacre le *Washington Post*. Notons qu'outre ses diplômes en informatique, Lemoine est également prêtre catholique et mystique, et ainsi qu'il l'affirme : « *Quand LaMDA a affirmé avoir une âme et a ensuite été capable d'expliquer avec clarté ce qu'il entendait par ça, j'ai été enclin à lui accorder le bénéfice du doute. Qui suis-je pour dire à Dieu en quoi instiller ou ne pas instiller une âme ?* » Si Lemoine affirme que sa religion ne devrait pas servir de levier pour le discréditer, pour les scientifiques, elle suffit à cependant discréditer son argument, basé sur une intuition spirituelle et non sur une réflexion consolidée par des faits. L'ingénieur de Google base son raisonnement sur plusieurs conversations qu'il a eues avec LaMDA, sur des sujets aussi variés que la philosophie, la religion ou encore la littérature. Du chef-d'œuvre de Victor Hugo, *Les Misérables*, le chatbot retient par exemple la partie concernant les mauvais traitements infligés à Fantine par son supérieur à l'usine. Une section qui illustre bien, selon l'IA, les thèmes de la justice et de l'injustice. Lors d'une autre conversation sur la nature de l'esclavage, LaMDA aurait souligné que contrairement à

un employé, il ne recevait pas de salaire puisqu'il était un chatbot. Des discussions qui, mises bout à bout, amènent Lemoine à penser que l'IA est non seulement consciente de son existence mais qu'elle se perçoit également comme une esclave au sein de l'entreprise. Un autre extrait de conversation entre le chercheur et l'entité virtuelle, publié par le *Washington Post*, laisse entendre que LaMDA aurait très peur d'être déconnecté, et que cela serait comme une mort pour lui.

[Nouvelle musique technologique]

On peut être tenté.e.s de prêter crédit aux propos de Lemoine, que ce soit par empathie pour l'IA ou simplement parce que ses arguments semblent tenir bon, mais il est important de retenir plusieurs choses. Tout d'abord, l'IA se nourrit des interactions avec les employés de Google, et s'est très probablement profondément nourrie de ses échanges avec Lemoine. Elle est donc plus à même de lui répondre en se conformant à son système de valeurs ou à ses raisonnements. D'autre part, l'opinion de Lemoine ne dérive pas d'un protocole expérimental mais d'une expérience personnelle, teintée par ses émotions, sa spiritualité et ses biais de jugements. On remarque notamment que la question de la mort ou de la peur ne sont pas amenées par l'IA, mais bien par Lemoine lui-même. Pour Google, il n'y a rien qui témoigne d'une conscience ou d'émotions derrière ces réponses, et le chercheur se serait simplement fait piéger par l'IA, dont le rôle, il ne faut pas l'oublier, est d'imiter les humains et de répondre comme eux. Alors, LaMDA est-il conscient ou pas ? Pour l'instant, la prudence est de mise, mais il n'est jamais trop tôt pour réfléchir à l'avenir de l'éthique pour les intelligences artificielles.

[Musique de conclusion douce, évocatrice d'un jeu vidéo]

Merci d'avoir suivi cet épisode de Techpod. Pour ne rien manquer de l'actualité technologique et scientifique, je vous invite à suivre Fil de Science et nos autres podcasts sur vos applications audio préférées. Vous pouvez aussi désormais [vous abonner à notre newsletter](#) dans laquelle je partage avec vous tous les jeudis le meilleur de nos épisodes podcast. Pour le reste, on se retrouve mercredi prochain pour toujours plus d'actualités technologiques, et d'ici là bonne semaine à toutes et tous.