FUTURA

X. AI et TruthGPT : Elon Musk veut créer une IA qui ne dit que la vérité

Podcast écrit et lu par Emma Hollen

[*Générique d'intro, une musique énergique et vitaminée.*]

Elon Musk qui lance sa propre IA générative défenseuse de vérité, c'est l'actu de la semaine, dans Vitamine Tech.

[*Fin du générique.*]

L'IA avance à un rythme effréné. L'IA pourrait se révéler dangereuse pour l'humanité. L'IA a besoin d'un cadre éthique et de réglementations strictes pour se développer dans la bonne direction. Pour une fois, Elon Musk semblerait presque faire preuve de bon sens en se joignant à l'inquiétude légitime des experts affirmant que nous allons un peu trop vite en besogne avec les IA génératives. Dommage que dans la même interview, le patron de Tesla, SpaceX ou encore Twitter, annonce également avoir créé sa propre firme d'intelligence artificielle, grâce à laquelle il espère lancer prochainement son TruthGPT.

[*Une musique électronique calme.*]

If you can't beat them, join them, comme le dit le proverbe anglais. Si vous ne pouvez les vaincre, rejoignez-les. Après avoir cosigné une lettre demandant un moratoire de 6 mois sur le développement des intelligences artificielles supérieures à ChatGPT-4, Elon Musk retourne finalement sa veste et se lance lui-même dans l'aventure. Enfin, bien évidemment, ce n'est pas tout à fait comme ça qu'il présente sa version des faits. Lors d'une interview diffusée en début de semaine sur la chaîne d'informations *Fox News*, le multimilliardaire laisse entendre que c'est pour protéger l'humanité et rétablir la vérité qu'il souhaite à son tour entrer dans l'arène. Rien que ça. Dans la première partie de l'échange, il explique que l'IA a le potentiel de détruire les civilisations. Il souligne, avec justesse – reconnaissons-lui au moins ça –, que les régulations autour de l'IA sont très en retard par rapport à l'avancée supersonique que celle-ci a connu au cours des dernières années. Depuis quelques mois en particulier, des chatbots comme le célèbre ChatGPT, créé par OpenAI, ou Bard, lancé par Google, ont par maintes fois démontré la puissance de ces technologies génératives, mais également les risques qu'elles comportent, avec plusieurs dérapages dont nous avons déjà parlé dans Vitamine Tech. Ce qui pour l'instant se cantonne à des incidents isolés plus ou moins graves pourrait rapidement prendre des dimensions catastrophiques à mesure que les entreprises se précipitent pour intégrer l'intelligence artificielle à leurs solutions. Prenons un exemple concret : une entreprise souhaite utiliser une IA basée sur le machine learning pour trier les candidatures que son service de ressources humaines reçoit quotidiennement.

Son but : améliorer et accélérer le processus de recrutement tout en introduisant un peu plus d'objectivité dans la façon dont les candidats sont sélectionnés. Le problème, c'est que pour apprendre à distinguer un bon d'un mauvais candidat, notre IA est bien obligée de s'appuyer sur quelque chose. On peut certes lui fournir des critères mesurables qui lui permettront de faire un premier tri, mais ce qui fait la puissance du machine learning, c'est sa capacité à ingérer une immense quantité de données pour en dériver des dimensions, des motifs récurrents. On va donc, pour expliquer les choses très schématiquement, fournir à l'IA des dizaines de milliers de CV, et lui indiquant que certains ont été retenus, et d'autres rejetés. L'IA va décortiquer ces CV pour identifier ce qui les distingue ou les rapproche et va ainsi pouvoir extrapoler l'ensemble des dimensions qui définissent un bon candidat ou un mauvais. Génial en pratique, mais il y a tout de même un petit problème. Bon, un gros problème. Car même si notre IA n'est rien de plus qu'une machine, les données qui ont servi à l'entraîner, elles, ont été produites par des humains. Ce sont des humains qui ont décidé si une candidature devait être retenue ou non et c'est sur la base de leurs choix que le machine learning a formé son algorithme. Or, je suis au regret de vous le dire, l'humain n'est pas exactement la plus objective des créatures. Même avec les meilleures intentions du monde, nous pouvons consciemment ou inconsciemment nous retrouver à ignorer plus facilement une candidature à cause de l'origine d'un candidat, de son choix d'établissements scolaires, de son nom ou de sa photo. En transmettant notre méthode à l'IA, nous lui transmettons donc aussi nos biais, nos préjugés et nos angles morts. Je vous en avais déjà parlé dans notre épisode sur le chatbot de Meta qui était devenu conspirationniste en un week-end. Ainsi, une entreprise de plusieurs milliers d'employés qui souhaiterait utiliser ce genre d'outil pour rendre ses recrutements plus objectifs risquerait au contraire de renforcer des biais déjà existants, nous entraînant dans une spirale sans fin. Elon Musk, pour en revenir à lui, affirme que les IA écrivent de manière de plus en plus convaincante et pourraient de ce fait manipuler l'opinion publique sur les réseaux sociaux sans que nous le sachions. Pour lui, il est donc essentiel d'établir un cadre réglementaire avant que l'IA n'aille trop loin. Il faudrait pour cela, créer une agence de régulation qui, je cite, « *chercherait d'abord à comprendre l'IA, puis solliciterait l'avis de l'industrie, et enfin proposerait des règles.* » Si certains jugent qu'il est encore trop tôt pour s'inquiéter du développement des IA génératives, de nombreux experts, qu'ils soient issus des milieux académiques ou industriels, n'ont pas hésité à tirer la sonnette d'alarme. Selon eux, ils est capital d'ouvrir dès à présent le débat, avant que la catastrophe ne survienne. Et pour une fois, Elon Musk semble délaisser ses bravades pour se joindre au rang des prudents. Ou peut-être que tout cela n'est que politique au final.

[*Virgule sonore, une cassette que l'on accélère puis rembobine*.]
[*Une musique de hip-hop expérimental calme*.]

Eh oui, car ne l'oublions pas, s'il affirme nous mettre en garde depuis des années contre l'IA, Elon Musk a tout de même grandement participé à son développement et à l'engouement qu'elle suscite. Il est après tout le cofondateur d'OpenAI, la firme à l'origine de ChatGPT. Il s'est certes retiré de l'entreprise en 2018, mais ce n'est pas pour autant qu'il a cessé d'investir dans l'intelligence artificielle. Neuralink, SpaceX et Tesla en particulier s'appuient tous, dans une mesure plus ou moins importante sur l'IA. Du côté de Twitter, Musk a déclaré dans un tweet le mois dernier qu'il prévoyait « *d'utiliser l'IA pour détecter et mettre en évidence la manipulation de l'opinion publique sur [la] plateforme.* » Et que dire d'Optimus, le robot intelligent présenté par Tesla pendant son annuelle journée de l'IA en 2022 ? Bref,

bien que la défiance d'Elon Musk envers l'intelligence artificielle soit tout à fait légitime et paraisse même sincère, le dirigeant ne semble pas pour autant prêt à prendre du retard sur ses concurrents. Mais ça ne s'arrête pas là puisque dans un nouveau retournement de situation, nous avons appris ce week-end qu'il avait fondé sa propre firme d'IA générative, baptisée X.AI, en mars dernier. Il aurait également fait l'acquisition de 10 000 processeurs graphiques, des blocs fondamentaux dans la conception, vous l'avez deviné, d'intelligences artificielles. À peine avons-nous eu le temps de nous remettre de nos émotions que *Fox News* diffusait une interview de Musk par Tucker Carlson, un journaliste connu pour ses aspirations républicaines… et conspirationnistes. Nous en avons évoqué la première partie, alors venons-en au morceau où l'entrepreneur annonce le lancement de son IA, TruthGPT. Se voulant une alternative à ChatGPT, celle-ci aurait pour but, à en croire Musk, de maximiser la recherche de la vérité. Pourquoi ? Eh bien parce que selon lui, les chatbots actuels auraient été bridés pour être politiquement corrects, ou woke en anglais, ce qui en reviendrait à leur apprendre à mentir. Et ça, c'est dangereux, mortellement dangereux. En tout cas, c'est ce qu'il affirme sans vraiment nous donner plus de précisions. Pour lui, avoir appris à une IA à ne pas tenir de propos racistes, misogynes ou transphobes, menacerait la vérité et possiblement l'humanité. C'est peut-être pour ça que mardi, Twitter a discrètement retiré une ligne de sa charte de conduite qui protégeait jusqu'alors les personnes transsexuelles et transgenres contre certaines attaques. Les utilisateurs mal intentionnés retrouveront désormais le droit de mégenrer leurs victimes ou d'utiliser leur prénom de naissance. Triste programme en perspective, que le patron du réseau social défend comme une volonté de maintenir un espace de libre expression. Son IA, dénuée de tout filtre contre les propos discriminatoires, aura d'après lui pour mission, je cite « *de comprendre la nature de l'univers* » et de ne partager rien d'autre que les faits. Inutile de dire qu'en face de lui, le journaliste de la *Fox* jubile. Musk semble également persuadé que TruthGPT offrira une alternative sûre pour l'humanité, car l'IA, en étudiant l'univers, s'apercevra forcément que nous sommes une partie intéressante du monde et que nous méritons d'être préservés. Il explique ainsi que nous aurions pu décimer les chimpanzés mais que nous avons préféré les protéger, eux et leur habitat. L'IA en fera donc probablement de même pour nous. Drôle d'exemple étant donné que l'espèce, en danger d'extinction, est sur le déclin, et que la déforestation poursuit son cours à un rythme effréné dans plusieurs pays d'Afrique. C'est peut-être pour cette même raison qu'une IA véritablement dotée d'objectivité choisirait plutôt d'éliminer l'être humain, afin de préserver ce qu'il n'a pas encore détruit. Au final, l'entrepreneur n'est pas clair sur ses intentions. Souhaite-t-il vraiment créer une IA en quête de vérité, et donc d'une objectivité qui devra l'amener à considérer notre espèce sous un angle purement pratique plutôt que romantique comme l'imagine Musk ? Ou veut-il introduire une IA dotée d'une forme de vérité, comme celle que Trump défendait sur les canaux de *Fox News* au cours de son mandat ? Le nom de TruthGPT n'est pas sans rappeler celui de Truth Social, le réseau social à majorité républicaine et conspirationniste créé par l'ancien président des États-Unis. Alors que les campagnes électorales se mettent déjà en marche dans le pays, Elon Musk affirme que Twitter devrait jouer un grand rôle dans les débats à venir, à domicile mais aussi à l'international. Son IA influencera-t-elle la direction vers laquelle penchera la balance ? Espérons que non et que nous garderons la liberté de pensée et d'expression que Musk semble si férocement défendre, de manière de moins en moins convaincante.

[*Virgule sonore, un grésillement électronique*.]

C’est tout pour cet épisode de Vitamine Tech. Si le podcast vous plaît, pensez à vous y abonner et à nous laisser un commentaire sur votre app de prédilection : Apple Podcast, Spotify, Podcast Addict ou encore Podchaser offrent tous cette option. Pour en découvrir plus sur le monde des inventions et sur la façon dont elles peuvent aussi œuvrer à créer un monde meilleur, je vous invite à écouter notre podcast Jeunes Pousses, où Thibault Caudron interviewe toutes les deux semaines un acteur de l’innovation positive. Pour le reste, je vous souhaite à toutes et tous une excellente journée ou une très bonne soirée et je vous dis à la semaine prochaine dans Vitamine Tech.

[*Un glitch électronique ferme l’épisode*.]