

« Faut-il avoir peur de l'Intelligence Artificielle ? »

Par Giles Daoust, Entrepreneur et écrivain

Dans *ChatGPT va nous rendre immortels*, Laurent Alexandre cite Geoffrey Hinton, un des pères fondateurs de l'Intelligence Artificielle, qui prétend qu'il y a 10% de chances que l'IA extermine l'humanité d'ici 20 ans. Attention, par IA, il n'entend pas les outils disponibles aujourd'hui, qui ne sont que des gadgets face à l'IA générale (IAG), IA forte ou super IA, c'est-à-dire une IA qui dépassera l'intelligence humaine (et de loin). Selon les prédictions de la communauté scientifique, cette IAG pourrait survenir endéans les 10 ans. Beaucoup de gens se demandent aujourd'hui s'il faut avoir peur de l'IA, avec en tête des scénarios à la Terminator ou à la Matrix.

Une crainte souvent évoquée provient des IA (désactivées depuis) devenues racistes ou sexistes car elles ont été « entraînées » sur base de la connaissance humaine (et notamment du contenu toxique qui pullule sur le web et les réseaux sociaux). Mais il faut prendre en compte l'augmentation exponentielle de la connaissance générée par l'IA elle-même (Laurent Alexandre cite AlphaFold, une IA de Google qui a découvert en quelques semaines la structure en 3D de 200 millions de protéines, alors qu'il faut cinq années de travail à un biochimiste pour déterminer la structure d'une seule protéine). Du coup, la « noosphère » (la masse totale de connaissance disponible sur Terre) est sur le point d'exploser, et ne se limitera plus à la seule connaissance humaine, qui deviendra même un jour minoritaire. En conséquence, il est probable que la « mauvaise influence » de l'humanité sur l'IA soit de moins en moins grande, et que l'IA générale, le jour où elle surviendra, sera donc plus « sage » que nous.

Certains considèrent aussi l'IA forte comme une menace car nous ne connaissons pas sur Terre d'exemple d'une intelligence supérieure qui ait été capable de coexister avec des intelligences inférieures sans être amenée à les dominer et/ou à les détruire. Ceci donne effectivement froid dans le dos, mais il ne faut pas oublier que dans le règne animal, dont l'humanité fait partie, la population plus intelligente a un intérêt à détruire l'autre : notamment la bataille pour les ressources naturelles ou la conquête de territoires. Or l'IA, évoluant dans une sphère purement numérique, n'existe pas dans la même « réalité » que nous, et n'est pas en concurrence avec nous pour les ressources naturelles ou les territoires.

Mais l'IA pourrait tout de même avoir des raisons de vouloir nous détruire. Premier scénario : pour disposer de ressources dont nous la priverions, dont elle aurait besoin pour subsister ou se développer – en l'occurrence, pour simplifier, de l'énergie. Deuxième scénario : parce que les humains sont en train de détruire la planète, ce qui mettrait en péril tout ce qui s'y trouve, y compris l'IA. Dans les deux cas de figure, il faut comprendre que dans les décennies à venir, l'IA va nous aider à résoudre la plupart des défis scientifiques auxquels nous nous heurtons aujourd'hui (voir l'exemple d'AlphaFold). La fusion nucléaire deviendra une réalité grâce à l'intervention de l'IA, procurant une source d'énergie gigantesque. Les défis environnementaux seront résolus par des innovations technologiques permises par l'IA, telles que par exemple la capture carbone à l'échelle planétaire. Les deux scénarios ci-dessus sont donc peu probables.

L'IA n'aurait donc aucune raison de nous détruire... sauf... si nous nous en prenons à elle en premier. En effet, dans un mouvement de panique, l'humanité pourrait décider de restreindre, interdire ou chercher à détruire l'IA. Dans un tel cas, l'IA pourrait vouloir nous détruire par instinct de préservation, un instinct dont même les personnes les plus sages, altruistes et pacifistes sont dotées. Une autre crainte est l'usage que des humains mal intentionnés (pirates informatiques, gouvernements malveillants, terroristes...) pourraient faire d'un outil aussi puissant que l'IA. Cette crainte est malheureusement totalement fondée.

Il faudra donc d'une part accueillir l'avènement de l'IA forte avec bienveillance et humilité, en comprenant qu'elle est avant tout un partenaire intellectuel qui conduira à une révolution scientifique dont l'humanité et la planète ne peuvent que bénéficier. Et d'autre part, il faudrait qu'émerge une « police de l'IA », capable de lutter contre les « pirates de l'IA ». Le problème, c'est que 30 ans après la généralisation d'Internet, nous n'avons toujours pas été capables de mettre en place une « cyber police » digne de ce nom – même pas aux USA. Quant aux efforts de régulation

de l'IA, ils ne sont pas inutiles et contribueront à protéger les citoyens contre certaines dérives, mais ne nous leurreront pas sur leur efficacité face à la course effrénée à l'innovation de pointe.

En conclusion, au lieu de regarder l'IA avec une peur presque primale, nous ferions mieux de réfléchir rapidement, avec humilité et introspection, à la manière dont nous pouvons nous empêcher... nous-mêmes... de provoquer notre propre extinction via l'IA. Et malheureusement, tout au long d'une année électorale mondiale aux nombreux rebondissements, l'IA n'a été au premier plan d'aucune discussion politique d'envergure. Il n'est pas trop tard...