



## **Non, une Intelligence Artificielle n'est pas une personne !**

### **Partie 2**

La tentative d'humaniser les machines à travers un assistant vocal comme Google Duplex cacherait-elle des motivations autres que commerciales ?

#### ***Un choix idéologique et un agenda posthumaniste ?***

Raymond Kurzweil, directeur de l'ingénierie chez Google et prophète du posthumanisme, nous invite à nous préparer à la singularité technologique. Il annonce pour 2045 ce basculement, où une Intelligence artificielle, dite forte, aura conscience d'elle-même. Kurzweil milite pour lui donner le droit de vote. Dès lors, d'aucuns pourraient voir dans ces choix éditoriaux **une volonté délibérée** de pousser des millions d'utilisateurs à considérer les IA comme nos égales.

Une autre entreprise, Boston Dynamics, a basculé dans la *vallée de l'étrange* depuis plusieurs années. Son robot bipède Atlas marche dans la neige, court, tombe puis se relève, exécute des sauts périlleux sans tomber, avec une facilité déconcertante compte tenu de son stature et de son poids<sup>i</sup>. Son robot quadrupède Spot ressemble à un chien, pourvu d'un bras articulé sur le dos pour ouvrir et fermer les portes<sup>ii</sup>. Les vidéos de ces machines ont fait le tour de la planète. Boston Dynamics, qui travaille pour l'armée américaine, a été rachetée en 2017 par la société Softbank, elle-même propriété du milliardaire Masayoshi Son, connu pour son militantisme transhumaniste.

#### ***Ne jamais cacher la nature non-humaine de la machine***

Oren Etzioni, du Allen Institute for Artificial Intelligence, a édicté trois lois sur les systèmes à base d'Intelligence Artificielle<sup>iii</sup>, dont **l'obligation pour la machine de déclarer ouvertement sa nature non-humaine**. Le chercheur note qu'en ces temps de fausses citations ou de vidéos falsifiées circulant sur la Toile, de chatbots diffusant de fausses nouvelles (*fake news*), nos sociétés doivent avoir l'assurance que les IA se présentent comme telles. La vérité et la transparence sont devenues des exigences partagées par tous.

Dans le domaine commercial, le recours lors d'une campagne de télémarketing, à des chatbots plutôt qu'à des humains sans en informer le client, pourrait être « de nature à altérer de manière substantielle le comportement économique du consommateur ». Ce qui constitue une pratique commerciale déloyale ou trompeuse, d'après la loi française.

La connaissance de l'origine humaine ou informatique d'un incident peut revêtir une importance vitale. Dans la nuit du 25 au 26 septembre 1983, en pleine période de tension entre les grandes puissances, les satellites de surveillance soviétiques ont détecté une attaque nucléaire américaine. Selon le protocole en vigueur, l'officier de garde sur une base d'alerte stratégique au sud de Moscou aurait dû déclencher une riposte et, par voie de conséquence, un conflit nucléaire mondial. Heureusement, il a estimé qu'une attaque américaine devait impliquer une centaine de missiles et pas cinq ou six. Il a conclu à une erreur des programmes



et il a annoncé à ses supérieurs une fausse alerte. Aurait-il réagi de la même manière si des humains avaient déclenché l'alarme ? Personne ne le saura<sup>iv</sup>.

### ***Faire cesser la confusion***

La confusion entre l'inerte et le vivant doit cesser. **Tout un chacun, y compris les personnes malentendantes ou malvoyantes, doit distinguer sans équivoque la machine de l'homme et, plus généralement, du vivant.** Concrètement, je propose que **toute Intelligence Artificielle échoue sans ambiguïté à un test de Turing étendu aux êtres vivants, à travers la vue, l'ouïe et le toucher.**

Le mathématicien Alan Turing a anticipé dans les années cinquante, le jour où un programme se ferait passer pour un humain. Il a conçu un test, qui porte son nom, pour l'évaluer. Le test de Turing consiste à confronter un humain à l'aveugle à un programme informatique, via des messages textuels. Si, après plusieurs échanges, la personne n'arrive pas à dire lequel des deux est un ordinateur, le test est considéré comme réussi par la machine. Jusqu'à présent, aucun logiciel, pourvu des dernières évolutions de l'Intelligence Artificielle, n'a réussi le test.

Pour la vue (cas des malentendants), je propose que **chaque robot présente une plaque transparente, bien visible, rappelant sa nature mécanique interne**, comme le recommande l'Institut pour l'étude des relations homme-robots (IERHR)<sup>v</sup>. Elle mesurera un minimum de **20 centimètres de côté**, soit l'empan d'une main humaine.

Pour le toucher (cas des personnes malvoyantes), **le corps du robot comportera une partie lisse (par exemple plastique ou métallique), accessible aisément, de la taille minimale d'un empan.** Cette partie de la machine pourra correspondre à la plaque transparente, requise pour la vue.

Pour l'ouïe (personnes malvoyantes ou assistant vocal), **le logiciel s'identifiera explicitement dès le début de la conversation, comme une « machine », un « programme informatique » ou un « robot ».** Les termes d'« intelligence artificielle », d'« assistant vocal » ou d'« assistant conversationnel », qui prêtent à la machine des capacités humaines, seront proscrits. **Une personne qui se joindra à une conversation en cours devra reconnaître la machine.** Il existe de nombreuses solutions techniques pour donner à une voix une tonalité artificielle. Il appartiendra aux fabricants de choisir celle qui leur conviendra le mieux, de sorte qu'on puisse interdire les appels en provenance de machines.

En cas d'échange par messages (*chat* ou *mail*), **le programme pourra aussi s'identifier par le préfixe « chatbot ».**

Les nouveaux modèles de voitures passent aux « mines ». La chercheuse Laurence Devillers propose **un dispositif similaire d'homologation pour les robots<sup>vi</sup>.** Le test de Turing ci-dessus serait réalisé à cette occasion. Ces systèmes à base d'Intelligence Artificielle s'adaptent constamment. Elle recommande de **tracer leur comportement, au quotidien. Un dispositif équivalent aux « boîtes noires »** des avions ou des poids-lourds serait implanté dans chaque machine.



Le législateur ne peut pas déléguer notre avenir à des intérêts qui se retrouvent à la fois juge et partie. Si l'éthique a un sens, il ne consiste pas à cautionner la confusion entre l'homme et la machine, entre le vivant et l'inerte. Dans ce contexte, l'initiative européenne en cours visant à définir des lignes directrices en matière d'éthique de développement de l'Intelligence Artificielle est une excellente nouvelle<sup>vii</sup>.

Emmanuel Bertrand-Egrefeuil.

## NOTES

---

<sup>i</sup> <https://www.youtube.com/watch?v=rVlhMGQgDkY>

<sup>ii</sup> <https://www.youtube.com/watch?v=fUyU3IKzoio>

<sup>iii</sup> Une IA doit se conformer aux lois existantes ; une IA ne doit pas cacher sa nature de machine ; une IA ne doit pas conserver ou diffuser une information sans l'accord explicite de sa source. Voir Oren Etzioni, *How To Regulate Artificial Intelligence*, New York Times, 1er septembre 2017.

<sup>iv</sup> Stanislav Srevov a reçu le Dresden Peace Prize en 2013. Il est décédé le 19 mai 2017.

<sup>v</sup> <https://www.ierhr.org/charte-ethique/>. Consulté le 21/03/2019. Elle demande « *qu'une partie de l'intérieur des robots soit toujours visible (protection transparente) pour rappeler leur caractère de machine.* »

<sup>vi</sup> *Ibid.* p. 144.

<sup>vii</sup> Un groupe de 52 experts internationaux, *GEHNIA ou Groupe d'Experts de Haut Niveau sur l'Intelligence Artificielle*, a été mis en place par l'UE en 2018 pour réfléchir sur les impacts éthiques et sociétaux de l'Intelligence Artificielle.