

Les thèmes du post-humanisme, de la singularité, les utopies et les dystopies de l'IA

Cette proposition consiste à exposer les différents scénarios exposés par Max Tegmark quant à l'avenir de notre société sous l'influence d'intelligence artificielle avancées, voir dans des sociétés post-singularité, et à les discuter. Max Tegmark est un cosmologiste d'origine suédoise et actuellement en poste au Massachusetts Institute of Technology. Il est également l'un des fondateurs du *Future of Life Institute*, une association de volontaire cherchant à diminuer les risques existentiels, notamment ceux provenant de l'intelligence artificielle. L'institut a reçu dix millions de dollars de soutiens de la part d'Elon Musk. Il s'agira de discuter les différents scénarios mis en avant dans son livre *La vie 3.0, Être humain à l'ère de l'intelligence artificielle*. Il est également à noter que Max Tegmark commente lui-même les possibilités, les avantages et les inconvénients de chaque scénario :

- L'utopie libertarienne : il s'agit d'un monde où les humains cohabitent pacifiquement avec la technologie. La vie sur Terre se subdivise en trois catégories, à savoir des zones réservées pour les machines, d'autres pour les humains, et enfin des zones mixtes (auquel on pourrait rajouter des réserves naturelles pour le reste du vivant). Le monde des machines, semblant morne à l'extérieur, serait pourtant extraordinairement animé à l'« intérieur », créant des mondes d'une rare richesse. Dans les zones mixtes, les robots, les IA, les transhumains, posthumains et tout être intermédiaires, joueraient avec les catégories au gré de leur convenance. Dans le monde des humains, les intelligences artificielles générales ou supérieures, autrement dit égales ou supérieures à l'intelligence humaine, seraient bannis. Ce monde ressemblerait au nôtre, à ceci près que tout les problèmes économique, sociaux, et environnementaux y seraient résolus.
- Le dictateur bienveillant : dans ce scénario, la souffrance a disparu, ou du moins elle est réduite au strict minimum, « grâce » à l'hégémonie absolue d'une intelligence artificielle prenant soin de tous les êtres vivants, humains, et entités artificiellement intelligentes. La plupart des humains se sentent satisfaits de ce mode de vie, soit par conditionnement, soit par modification génétique. L'IA, omnisciente voir omnipotente avec les moyens mis à sa disposition, contrôle ou supervise chaque aspect de l'existence.
- L'utopie égalitarienne : il s'agit en quelque sorte de l'inverse de l'utopie libertarienne, au sens où la cohabitation entre humain, cyborgs, IA et robots n'est pas assuré par la protection absolue du droit de propriété, mais par l'abolition de celle-ci.
- Le gardien : Le « gardien » est une IA avancée chargée d'éviter que d'autres IA avancées n'apparaissent, au moyen d'une surveillance de masse et de l'arrêt de ses propres progrès. Le sort de l'humanité lui est indifférent tant que son objectif de gardien est assuré.
- Le dieu protecteur : il s'agit d'une IA avancée chargée de veiller sur le destin de l'humanité pour son plus grand « bien », sans que celle-ci n'en soit nécessairement au courant. Ainsi, l'IA peut agencer les événements de manière à ce que ceux-ci paraissent « naturels », ou influencer via des nanomachines les idées des humains, de manière à ce qu'ils croient en être les véritables auteurs.
- Le dieu enchaîné : la superintelligence artificielle est au service absolu de l'humanité, avec toutes les conséquences qui pourraient en résulter.
- Le conquérant : une superintelligence artificielle cherche à conquérir l'ensemble de l'univers qui lui est physiquement accessible, ce qui inclue l'élimination des humains, d'une manière sans doute inimaginable actuellement, en vue d'objectifs dépassant vraisemblablement notre compréhension.
- Les descendants : les IA avancées sont considérées ou se considèrent comme les descendants des humains. Ces IA pouvant même encouragés leur disparition, éventuellement avec la collaboration de ceux-ci, par exemple en encourageant l'absence de natalité et en distrayant les humains.
- Le gardien de zoo : dans ce scénario, la plupart des humains peuvent avoir disparu, mais certains sont gardés en vie par des IA, un peu à la manière d'animaux dans un zoo ou dans des réserves.
- 1984 : les progrès de l'IA sont stoppés par une dictature technophobe cherchant à préserver l'humanité.
- Réversion : une dictature menant un retour en arrière technologique pour diminuer les chances de super-IA.
- Autodestruction : l'humanité a disparu avant l'apparition d'une super-IA.

Biographie

Après un Mémoire sur « L'implémentation d'éthique en intelligence artificielle » lors de son Master en Histoire et Philosophie de sciences, cohabilitées par l'Université de Paris et l'École Normale Supérieure de Paris, Alexandre Bretel a commencé une thèse à l'Université Grenoble Alpes en philosophie sur « L'apport de la philosophie de la technique à l'éthique de l'intelligence artificielle. Étude de la notion de responsabilité appliquée à la civilisation technologique et à la société de l'innovation ». Il est sous la supervision de Thierry Ménissier, professeur de science politique à l'Institut de Philosophie de Grenoble (IPhiG) et porteur de la chaire « Éthique & IA », faisant parti des projets 3IA, et sous la supervision de Jean-Gabriel Ganascia, professeur d'informatique à Sorbonne Université, chercheur au Laboratoire d'informatique de Paris 6 et spécialiste d'éthique de l'IA. Alexandre Bretel sera coordinateur de projet pour les États généraux de l'éthique de l'IA, qui auront lieu à Grenoble en septembre 2021.

Bibliographie

- *La vie 3.0 : Être humain à l'ère de l'intelligence artificielle*, Max Tegmark, Dunod, 2018.
- *Une vie après la vie, les robots avenir de l'intelligence*, Hans Moravec, Émile Jacob, 1982.
- *"New Heaven", in A History of the World in 10½*, Julian Barnes, Jonathan Cape, 1989.
- *Manna: Two Visions of Humanity's Future*, Marshall Brain, BYG Publishing, Inc, 2003.
- *On Intelligence*, Jeff Hawkins et Sandra Blakeslee, Time Books, 2003.
- *The Dreaded Comparison: Human and Animal Slavery*, Marjorie Spiegel, Broché, 1996.
- *Superintelligence: Paths, Dangers, Strategies*, Nick Bostrom, Dunod, 2014.
- *1984*, George Orwell, Poche, 1972.
- *Humanité 2.0 (The Singularity Is Near: When Humans Transcend Biology)*, Ray Kurzweil, CBS, 2005.
- *Cycle de Fondation*, Isaac Asimov, Folio, 1942–1993.

Filmographie

- Ex Machina (2014)
- Transcendance (2014)
- Her (2004)
- Docteur Folamour (1964)