Sciences et Golems

Jean-Gabriel Ganascia

Jean-Gabriel Ganascia est professeur d'informatique à l'Université Pierre et Marie Curie et chercheur au LIP6. Outre ses travaux sur l'intelligence artificielle, la modélisation cognitive et la philosophie computationnelle, il s'interroge sur les conséquences éthiques et politiques du développement des technologies de l'information.

Ambivalence du « Golem »

La tradition cabalistique juive rapporte l'existence d'une statue d'argile, le « Golem »[1], fabriquée, au milieu du XVI[e] siècle par le rabbin Loew, plus connu sous le nom de « Maharal de Prague ». À l'instar des ordinateurs contemporains, cette machine s'animait lorsqu'on passait un message derrière ses dents. Usuellement, elle vaquait aux occupations domestiques quotidiennes comme un serviteur zélé et assidu. Beaucoup de légendes ont couru autour de cette statue extraordinaire. Selon l'une d'entre elles, le rabbin Loew aurait oublié, un samedi, jour de prière, d'enlever le message de derrière les dents du Golem et ce dernier aurait commencé à s'agiter, à crier et à effrayer tous les voisins pendant que son maître remplissait ses devoirs saints à la synagogue. De retour chez lui, le rabbin Loew aurait détruit son œuvre de peur qu'elle ne recommence à prendre de fâcheuses initiatives. Selon une autre légende, sur le front du Golem était écrit le mot EMETH qui signifie vérité en hébreu ; un jour celui-ci aurait effacé au couteau la première lettre du mot, faisant apparaître le mot METH *(mort)*. Il résulte de toutes ces mythologies une ambivalence du Golem qui annonce celle des sciences et des techniques contemporaines. D'un côté, le rabbin Loew, qui était capable par son savoir de fabriquer un objet si perfectionné, fut grandement loué, et même vénéré, au point que son fauteuil est toujours visible dans la vieille synagogue de Prague. D'un autre côté, un tel Golem risque d'échapper à ses maîtres et créateurs qui doivent toujours se garder d'une telle éventualité. Notre responsabilité devant les œuvres techniques et, en particulier, devant la machine, se trouve là si clairement posée que Norbert Wiener s'y réfère explicitement dans *God and Golem*[2], ouvrage qu'il consacre tout entier aux enjeux éthiques de la cybernétique et des machines qu'il qualifie de téléologiques – adjectif dérivé du grec *téléo-*, qui signifie fin, et de *logos*, discours rationnel, science – parce que leur mouvement n'épouse pas simple-

[1] Idel, M., Atlan, M. et Aslanof, C. (1992), *Le Golem*, Éditions le Cerf, Paris.

[2] Wiener N., (2001), *God & Golem inc. : Sur quelques points de collision entre cybernétique et religion*, Éditions de l'Éclat, collection Premiers secours.

L'origine du terme « robot »

Le terme « robot » provient du tchèque *robota* qui signifie « travail pénible, corvée » ; il a été inventé par un écrivain Tchèque, Karel Čapek, dans une pièce de théâtre intitulée « Rossum's Universal Robots ». Il désigne des travailleurs artificiels corvéables à merci. Ceux-ci souffrent de notre indifférence à leur égard. D'après Karel Čapek, ces êtres que nous avions fabriqués pour nous servir méritent notre attention dès lors qu'ils possèdent une conscience. Grâce à une jeune femme sensible et intelligente qui est la propre fille de leur inventeur, ces robots humanoïdes se révoltent contre un ordre social qu'ils jugent injuste et obtiennent la reconnaissance humaine avant de détruire cette humanité qui leur avait donné naissance. Cette pièce, écrite en 1920, suscita beaucoup d'intérêt ; elle faisait écho à des préoccupations sociales brûlantes. Venu au pouvoir, Hitler s'en inquiéta. Il semble même que Karel Čapek fût pressenti pour le prix Nobel de littérature et que seule la crainte de froisser le dictateur retint l'académie Suédoise de lui décerner cette distinction.

Référence

Čapek, K. (1921), *RUR - Rossum's Universal Robots*, pièce traduite du tchèque et présentée par Jan Rubeš, Éditions de L'Aube, « Regards croisés », 1997.

ment le jeu des forces qui les meuvent, mais qu'elles asservissent leurs actions à la réalisation d'un but.

Beaucoup d'inventions contemporaines suscitent en nous un sentiment d'ambivalence analogue à celui que suscitait le Golem de Prague dans la tradition cabalistique : d'un côté, elles rendent des services considérables et attestent de la grande pénétration de leur concepteur, d'un autre côté elles font craindre le pire, car du fait même de leur perfection elles acquièrent une autonomie qui les rend imprévisibles et potentiellement dangereuses. Songeons par exemple aux robots autonomes : certains aideront à explorer la Lune et des planètes comme Mars ou Vénus, ce qui explique l'intérêt qu'on leur porte ; d'autres promettent de faire une guerre sans pertes inutiles pour les assaillants, ce qui peut faire trembler... Or, à la différence du Golem de Prague qui demeurait métaphorique, les « Golems modernes » acquièrent, du fait des progrès des sciences contemporaines, une existence tangible dans le monde actuel. Ils ne sauraient donc nous laisser indifférents. Puisqu'ils reproduisent réellement un certain nombre de facultés humaines, tout en étant autonomes et en s'affranchissant de notre tutelle, beaucoup de gens reprochent aux scientifiques qui les conçoivent de jouer aux apprentis sorciers. Est-ce exact ? Ces réalisations sont-elles dénuées de dangers ? Ces dangers correspondent-ils aux craintes que l'on nourrit usuellement ? Telles sont les questions dont nous allons essayer de traiter dans la suite.

Le monde des Androïdes Autonomes

La crainte des robots est ancienne ; la fascination qu'ils exercent aussi. Ces sentiments préexistent aux robots, non seulement à leur réalité matérielle, mais à leur nom aussi.

Très tôt, on s'inquiéta de la responsabilité de l'homme devant les automates. Soulignons qu'en 1921, date à laquelle la pièce de Karel Capek est publiée, les robots ont une existence essentiellement fantasmagorique.

Certes, on fabriqua beaucoup d'automates au XVIII^e^ et au XIX^e^ siècle. Mais, ces répliques mécaniques de nos semblables restaient très maladroites. Il n'en demeure pas moins que l'on aspirait déjà, et depuis très longtemps, à fabriquer des travailleurs artificiels. Ainsi, on trouve dans le chant XVIII de l'Iliade[3] un étrange passage où le dieu forgeron, Héphaïstos, se fait servir par des robots :

« Il dépose à l'écart, loin du feu, ses soufflets. Dans un coffre d'argent il met tous ses outils. Puis avec une éponge il se lave le front, les bras, le cou puissant, la poitrine velue. Il revêt sa tunique et prend un gros bâton, puis il sort en boitant. Deux servantes en or viennent alors le soutenir. Bien qu'elles soient en or, on les prendrait vraiment pour des filles vivantes. La raison les habite ; elles ont voix et force ; les Immortels leur ont appris à travailler. Pour soutenir leur maître elles vont et s'affairent. »

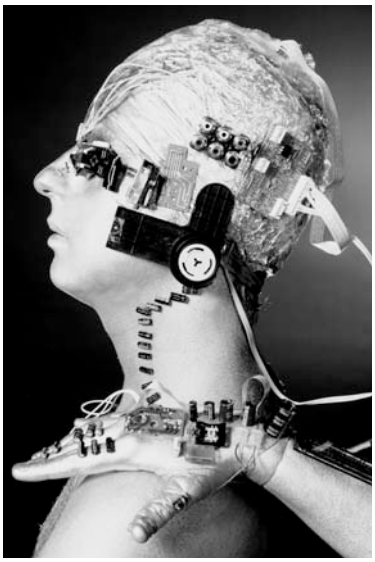
En 1938, lassé de lire tant d'histoires montrant des robots envahisseurs et agressifs, le biologiste d'origine russe Isaac Asimov développe une série de nouvelles et romans[4] qu'il organise autour de trois lois immuables de la robotique auxquelles il adjoindra des compléments nécessaires au fur et à mesure de l'évolution de son œuvre. Ces lois sous-jacentes à la création des androïdes sont censées les empêcher de nuire aux humains :

- *Un robot ne peut porter atteinte à un être humain ni, en demeurant passif, laisser cet être humain exposé au danger.*
- *Un robot doit obéir aux ordres donnés par les êtres humains, sauf si de tels ordres sont en contradiction avec la 1^ère^ loi.*
- *Un robot doit protéger son existence dans la mesure où cette protection n'est pas en contradiction avec la 1^ère^ ou la 2^e^ loi.*

Cette incursion hâtive dans la (pré)histoire de la robotique nous montre l'ancienneté des robots, ou plus exactement de leur existence imaginaire, qui remonte, au moins, à l'Antiquité grecque, bien avant l'existence de robots matériels. Conjointement à cette présence fantasmatique de Pygmalion, de Pinocchio et de toutes sortes de statues animées, naquirent des préoccupations éthiques : comment veiller à ce que ce qui s'anime tout seul et se meut de soi-même ne prenne son essor et n'acquiert la liberté d'agir pour son propre compte ? Quelles limites doit-on s'imposer pour que nos œuvres ne nous ensevelissent pas ? Parfois, il faut avoir le courage de détruire ce que l'on a fait : telle est la leçon que nous enseigne la légende du Golem. Dans tous les cas, le robot acquiert un statut analogue au nôtre. Il se présente comme notre double. Nous sommes, vis-à-vis de lui, dans la même situation qu'un Dieu créateur – ou que la Nature dans une perspective athée – vis-à-vis de nous. Cela rend l'entreprise d'autant plus audacieuse et risquée. Cela explique aussi que l'on songe à imposer à ces êtres artificiels des interdits analogues à ceux que nous nous imposons face à Dieu. Mentionnons enfin que, symétriquement à cette exigence de soumission du robot à l'homme, on évoque parfois une responsabilité de l'homme

[3] Homère, *Iliade*, traduction Paul Mazon, Gallimard, collection Folio, série Classique, 1975

[4] Asimov, I. (2004) I, *Robot*, Spectra, New York, NY.



vis-à-vis des robots qui serait analogue à celle de Dieu vis-à-vis de l'homme. À cet égard, l'énigme de la pièce de Karel Capek, est éloquente. D'autres récits de la littérature de science-fiction, par exemple la nouvelle de Brian Aldiss, intitulée « Supertoys Last All Summer Long » et dont on a tiré le film « AI, Artificial Intelligence », reprennent à leur compte cette vue.

La réalité des robots et les agents autonomes

Nous assistons aujourd'hui à une prolifération de robots et d'agents autonomes de toutes sortes qui reproduisent des apparences et des facultés humaines, mais qui, à bien des égards, se distinguent fortement de nous. Songeons aux robots aspirateurs « Roomba »[5] que l'on vend dans le commerce ou aux tondeuses à gazon de la même marque : ce sont des vrais robots qui se meuvent de façon autonome et qui appréhendent tant leur environnement que leur état interne, par exemple, l'état de leur batterie. Ils agissent en fonction de la représentation qu'ils se font de leur situation et de leurs objectifs. Ce sont des travailleurs artificiels utiles. Pour autant on ne saurait les qualifier d'androïdes : même s'ils accomplissent quelques fonctions qui requièrent des facultés cognitives qui étaient jusqu'ici propres aux humains, et en dépit de tous nos efforts d'imagination, ces cylindres colorés et plats demeurent fort dissemblables de nos congénères. Sans compter que, parmi les agents autonomes, certains apparaissent uniquement dans les mondes qualifiés de virtuels, sous la forme de programmes informatiques.

Robots, agents autonomes, webbots, etc. tous ces êtres artificiels ont fait progressivement leur apparition autour de nous. Désormais, ils peuplent notre univers, à la fois dans le monde matériel et dans l'infosphère ; nous les côtoyons journellement, le plus souvent à notre insu. Pour donner une idée de leur importance, prenons des indicateurs numériques : la population des robots en activité dans les industries manufacturières s'élève aujourd'hui à environ un million d'« individus » et à cela s'ajoutent les robots domestiques, les robots de compagnie, les robots spatiaux, les robots médicaux, les drones, les robots soldats etc., sans compter les innombrables robots virtuels qui parcourent la toile en tous sens, et se démultiplient à loisir, en fonction des besoins. Il faut savoir, à cet égard, que les moteurs de recherche comme Google font appel à des armées de robots qui aspirent régulièrement le contenu de sites, les indexent et les stockent. Bref, nous vivons aujourd'hui dans un monde nouveau où les hommes coexistent, de

[5] Le lecteur intéressé peut visiter un site où sont décrites les performances de ces robots aspirateurs : http://www.idealrobot.com/

plus en plus, avec les robots. De nombreuses questions en résultent : comment assumer la nouvelle condition d'une humanité vivant en symbiose avec les robots ? Ne risquons-nous pas de devenir les esclaves des machines que nous avions initialement conçues pour nous servir ?

Certes, nous concevons aisément qu'un robot domestique prenne des « initiatives » malheureuses ou qu'il fasse des bêtises, brûlant la chemise qu'il repasse ou avalant le fil d'une lampe électrique lorsqu'il passe l'aspirateur. Nous pardonnons aux automates tous ces défauts de jeunesse qui nous font parfois sourire. En même temps, nous appréhendons tous de voir une machine qui, au chevet d'un malade, déciderait, sans consulter d'avis médical, d'augmenter démesurément la dose d'un médicament. Nous tremblons plus encore à l'idée d'un robot soldat qui effectuerait sa « mission » à la perfection, en supprimant systématiquement tous ceux qui font obstacle à sa progression. Ce sont pourtant là des situations tout à fait banales auxquelles nous devrons faire face d'ici peu. Les études effectuées dans les laboratoires et les projets des militaires doivent nous en convaincre : la réalité du monde contemporain créé par les techno-sciences se trouve là. Pire encore, il arrive que, dans des situations imprévues, les robots prennent des décisions déroutantes, voire choquantes. Lorsque des risques sont connus, nous devrions être en mesure de nous en protéger. Mais comment se prémunir contre un danger que l'on ne connaît pas ?[6]

En cela, les mises en garde et les critiques formulées par certains de nos congénères à l'encontre des conséquences du développement des sciences et des techniques apparaissent tout à fait fondées. Sans doute ne faut-il pas laisser aveuglément libre cours à toutes les applications techniques que les sciences rendent possibles. Pour autant, cela ne signifie pas que la leçon du Golem s'applique *aujourd'hui* aux scientifiques, du moins à ceux qui œuvrent dans le domaine de la robotique.

Autonomie des robots et des agents autonomes

En effet, en dépit des risques que les robots nous font encourir et de leur présence dans le champ social, le danger lié à leur prolifération ne se trouve pas exactement là où on l'imaginait auparavant : malgré leur ressemblance avec le Golem, les robots ne se substituent jamais vraiment aux hommes. Non seulement, ils ne nous ressemblent pas physiquement, mais de plus, ils ne se présentent pas comme nos *alter ego*, autrement dit comme d'autres nous-mêmes à qui nous ferions toute confiance. *A fortiori*, ils ne s'arrogent pas, tous seuls, des qualités que nous ne leur avons pas attribuées. Ils ne se substituent aux hommes que pour effectuer à leur place des tâches qualifiées d'ingrates, parce que pénibles et délicates, et uniquement lorsque nous leur avons explicitement commandé de le faire.

---

[6] Et, dans l'éventualité d'un robot coupable d'actes répréhensibles, qui doit être tenu pour responsable ? Doit-on incriminer le concepteur du robot, son fabriquant ou le propriétaire ? Pour certains, il faudrait faire évoluer la loi afin de définir le statut de dispositifs matériels complexes dont les comportements échappent aussi bien à leurs concepteurs qu'à leurs propriétaires.

Certes, les robots et les « agents autonomes » disposent d'une *autonomie de mouvement* fascinante. Toutefois, elle ne doit pas faire illusion : elle ne permet pas aux robots contemporains d'accéder à une *autonomie de la volonté* ; en cela, ce ne sont pas des Golems. Ce point est tout à fait central, car de la confusion entre ces deux formes d'autonomie naissent bien des malentendus et des craintes sur la réalité des sciences contemporaines et sur les moyens requis pour parer aux effets néfastes de leurs applications.

Pour préciser les choses, rappelons que l'autonomie, au sens étymologique, désigne la capacité à se donner soi-même ses propres lois. À strictement parler, est autonome un sujet disposant de libre arbitre qui décide des règles qu'il impose à son comportement. Par extension, on qualifie d'autonome des machines qui déterminent *par elles-mêmes* les mouvements qu'elles doivent accomplir, en fonction d'objectifs prédéterminés. Or, dans cette dernière acception de l'autonomie, les mouvements sont asservis à la réalisation de buts fixés *a priori* et de façon extérieure. Ainsi, si les machines autonomes agissent *par elles-mêmes,* elles n'agissent pas *pour elles-mêmes*, afin de réaliser des buts qu'elles se seraient fixées seules. En cela, et à strictement parler, les machines qualifiées d'autonomes que l'on fabrique aujourd'hui sont hétéronomes, car elles obéissent à des objectifs qui leur sont donnés et qui leur sont extérieurs ; leur volonté n'est pas la leur ; ce n'est que dans le feu de l'action, et du fait de leur complexité, que ces robots nous apparaissent doués d'une volonté propre, mais cela n'est qu'une illusion.

Ce ne sont pas des sujets, au sens philosophique du terme. D'ailleurs, l'expérience montre que ce n'est pas l'excès d'intelligence qui rend les robots dangereux, mais leur stupidité ! Et, nous ne sommes pas vraiment victime de l'autonomie de robots qui prendraient leur ascendant sur nous, mais bien plutôt de la paresse qui nous a conduits à déléguer de plus en plus de tâches à certains robots, sans nous soucier des limites de leurs compétences.

En dépit de ces différences de nature entre les robots contemporains – fussent-ils doués d'une autonomie mécanique – d'un côté, et le Golem de la tradition de l'autre, il se pourrait qu'un jour, les progrès de la science, en particulier de l'informatique, des sciences cognitives, des nanotechnologies et de la biologie, autrement dit de ce que l'on appelle parfois la convergence NBIC, permettent de réaliser des automates doués de conscience et de volonté. Dans cette éventualité, les leçons de la légende du Golem s'appliqueraient fort bien. Et, il faudrait alors avoir le courage de détruire ces œuvres, aussi parfaites fussent-elles. Mais, pour l'heure, il n'en va pas ainsi : les robots et les agents autonomes contemporains ne disposent pas d'autonomie au sens propre.

Le risque pour l'humanité de devenir leur esclave n'est pas d'actualité. Pour l'instant, il convient seulement d'examiner les conséquences sociales de leur déploiement massif et de parer à leurs effets néfastes. ■