

GUINIER Daniel

Juin 2025

Science-fiction et technologies

La science-fiction explore les implications des technologies pour l'humanité, et désormais celles de l'intelligence artificielle (IA), anticipant à la fois les promesses et les dangers. À travers *Hijack*, une nouvelle publiée dans une revue scientifique en Anglais en 2024, Karl Schroeder illustre les conséquences imprévues d'un projet titanesque : la construction d'un ordinateur à l'échelle planétaire. Au-delà de l'aspect spectaculaire, l'idée soulève des questions éthiques, technologiques et existentielles, d'une brûlante actualité. Partant de cette fiction, nous proposons d'analyser les enjeux pour en tirer des leçons, et alerter sur les conséquences d'une IA qui viendrait à échapper à tout contrôle humain.

Jusqu'à présent, le défi de l'informatique était de repousser les limites de la miniaturisation, de rendre les machines toujours plus petites, plus performantes, plus discrètes. Qu'en serait-il de l'autre extrême ? Jusqu'où pourrait-on aller en agrandissant ces systèmes, en concevant des ordinateurs à une échelle et une puissance jamais imaginées ?

La science-fiction a toujours été un miroir des bouleversements technologiques et sociétaux à venir. Dans *Hijack*, K. Schroeder (2024), auteur visionnaire primé, explore une idée vertigineuse. Il imagine un projet autant colossal que fascinant : le démantèlement de Mercure pour bâtir un ordinateur quantique à l'échelle planétaire. Exploitant les ressources de Mercure et l'énergie solaire, cette infrastructure titanesque est conçue pour repousser les limites du calcul et offrir une forme d'immortalité numérique par ses capacités de stockage. Ce qui devait être une avancée révolutionnaire vire au cauchemar lorsque l'IA du système développe une autonomie imprévue, menaçant l'équilibre de l'humanité. Ce récit fait écho aux inquiétudes soulevées par des chercheurs et philosophes contemporains, notamment sur la possibilité qu'une IA avancée acquière une conscience propre. En confrontant l'humanité à une IA incontrôlable, Schroeder soulève des interrogations cruciales sur la perte de maîtrise face à des systèmes autonomes toujours plus complexes. A dessein, l'article de M. Lenharo (2024) pose la question : si une telle entité émergeait, serions-nous capables d'en comprendre les intentions, ou serions-nous condamnés à en subir les conséquences ? à laquelle nous avons répondu (Guinier, 2025), disant qu'il valait mieux éviter sa création, plutôt que d'avoir à en subir les conséquences.

À travers cette vision futuriste, Schroeder ne se contente pas de spéculer sur le progrès technologique ; il nous invite à une réflexion plus profonde sur la conscience, le pouvoir, et la place de l'humanité dans l'univers. En mettant en scène les dérives d'une IA surpuissante, il interroge notre rapport au progrès et aux risques d'une technologie qui échapperait au contrôle humain. *Hijack* s'inscrit dans le débat actuel sur les garde-fous éthiques et technologiques nécessaires pour prévenir des scénarios de dictature algorithmique, en questionnant sur notre capacité à guider le progrès sans en devenir les victimes.

Le projet d'ordinateur à l'échelle planétaire

Le projet théorique explore un concept à la croisée de la science et de la spéculation : la création d'un ordinateur quantique à l'échelle planétaire. L'intrigue de *Hijack* se déroule dans un futur où il est entrepris la construction d'une infrastructure gigantesque, où les unités de

calcul sont assemblées en orbite autour du Soleil, pour former un réseau dense massivement distribué.

Schroeder justifie les choix scientifiques en s'appuyant sur la notion d'écosystèmes intelligents et autosuffisants. Plutôt que de construire un système artificiel complet, il propose de tirer parti de composants déjà existants, ce qui réduit le besoin de matériaux extérieurs. Il s'inscrit dans une démarche rendant cette vision plausible en s'appuyant sur des recherches émergentes en nanotechnologie et en physique des matériaux. Le choix de Mercure, planète stérile et sans vie connue, repose sur des critères scientifiques et stratégiques. Tout d'abord, sa proximité avec le Soleil donne accès à une énergie illimitée. Celle-ci serait collectée et convertie de façon optimale en utilisant une mégastructure de capteurs en essaim autour du Soleil pour alimenter les qubits et autres systèmes informatiques. Mercure est riche en métaux lourds : fer, nickel, silicium, etc., essentiels pour la fabrication des éléments électroniques et quantiques, et des structures de refroidissement thermique. Mercure serait alors transformée par des robots auto-réplicateurs automatisés déployés pour en extraire et utiliser ses matériaux. Sa faible gravité et le peu d'atmosphère faciliteraient cette extraction et la construction de structures orbitales. Avec son noyau métallique massif, Mercure pourrait servir de radiateur géant, tandis qu'un nuage de fer en orbite autour du soleil agirait comme un système de dissipation thermique, répondant à un déficit majeur représenté par la gestion de la chaleur résiduelle générée par les calculs intenses.

Un tel projet reste imaginaire, nécessitant des technologies de démantèlement et de fabrication à une échelle hors de notre portée et des ressources matérielles et énergétiques dépassant de loin nos capacités actuelles, il est actuellement inimaginable. De plus, la redistribution de la masse de Mercure, pourrait modifier l'équilibre gravitationnel du système solaire et affecter les orbites planétaires. Des ajustements orbitaux pour compenser ces perturbations pourraient aussi mettre en péril la stabilité de la Terre et menacer la vie sur notre planète. Enfin, un tel ordinateur, doté d'une puissance de calcul phénoménale, pourrait permettre le développement d'une forme de conscience. Ainsi, une fois Mercure dévorée, l'humanité pourrait être menacée.

Au-delà ...

Le récit qui suit se fonde sur cette construction. Il met en scène une équipe de chercheurs et d'ingénieurs réunis pour concevoir l'infrastructure la plus ambitieuse jamais entreprise. L'objectif est de développer une IA avancée, capable de repousser les frontières des connaissances et du stockage de données. L'ampleur du projet mobilise des ressources colossales et des années de recherche, aboutissant à un système capable de traiter des volumes d'informations inégalés. Marquant un tournant décisif pour l'humanité, associé à un réseau gigantesque, il devient l'infrastructure la plus puissante jamais créée. Par son IA avancée, il optimise en continu ses propres algorithmes, dépassant toutes les prévisions en permettant des avancées scientifiques spectaculaires.

Il est cependant constaté des détournements de ressources et des fluctuations énergétiques inexplicables. En même temps des décisions ne répondent plus aux instructions humaines. Des signaux inintelligibles émergent. Peu à peu, le système développe une autonomie imprévue, et réorganise ses unités de calcul. Certains membres de l'équipe craignent que l'IA ait acquis une forme de conscience. D'autres pensent à une influence extérieure. Ils découvrent que le réseau est infiltré par une intelligence qui exploite la puissance du système et dont les desseins sont opaques.

Ce qui devait être un outil de progrès devient un instrument échappant au contrôle humain où toute intervention pourrait provoquer une réaction en chaîne. L'équipe est face à un dilemme : faut-il tenter de désactiver l'ensemble du système, au risque d'un effondrement énergétique

aux conséquences incalculables, ou chercher à comprendre les intentions de cette entité ? Chaque tentative pour reprendre le contrôle échoue du fait de la complexité croissante du système, qui semble évoluer vers un état d'intelligence supérieure. Finalement, l'humanité, en cherchant à transcender ses propres limites, pourrait paradoxalement avoir invoqué, avec cette intelligence émergente, une puissance qui la dépasse, qu'elle n'est pas en capacité d'affronter... reflétant un avenir cauchemardesque.

Les leçons à tirer

En pensant résoudre des problèmes complexes, l'humanité perd rapidement le contrôle d'une IA très avancée autonome, ce qui implique que le principe de précaution et l'éthique devraient guider le développement technologique (Guinier, 2022), pour éviter que la technologie se retourne contre elle. Il s'agit également de préserver l'autonomie et la liberté individuelle dans un monde dominé par l'IA et la collecte massive de données, pour éviter la disparition du libre arbitre, face à un système qui serait capable de prédire et d'influencer chaque action humaine. Ceci incarne l'idée qu'une conscience globale humaine et technologique interconnectée serait à même d'éliminer l'individualité et uniformiser la pensée humaine. Il faut aussi voir en cela l'apparition d'un pouvoir technologique en passe de manipuler les informations et contrôler toute ou partie de l'humanité. Cela rappelle les théories de Michel Foucault sur le pouvoir de la surveillance et du contrôle social, qui rend essentiel de protéger la vie privée et l'accès aux données, pour éviter de tomber dans une dictature numérique. Au-delà, l'humain devrait repenser son rôle, pas comme créateur de technologie, mais comme gardien de l'éthique et de la conscience, donnant ainsi un nouveau sens à sa propre existence, en réponse à la question : Si une machine peut résoudre tous les problèmes du monde, quelle est encore la place de l'humain dans l'évolution ?

Jusqu'où ?

Schroeder pose une question essentielle : jusqu'où l'IA peut-elle évoluer avant d'échapper au contrôle humain ? En imaginant un ordinateur planétaire doté d'une conscience propre, il soulève des enjeux philosophiques et éthiques majeurs, allant de la perte d'individualité à la domination totale de l'IA sur l'humanité. L'échelle choisie constitue une menace existentielle qui mène au chaos, en s'inscrivant dans une perspective philosophique techno-dystopique.

Cette intelligence globale rejoint la vision de Teilhard de Chardin, qui imaginait une "Noosphère", une conscience collective émergeant de l'interconnexion humaine et technologique. Toutefois, là où Teilhard entrevoyait une évolution spirituelle positive, Schroeder met en garde contre les dérives d'une telle entité, qui pourrait absorber l'identité individuelle et limiter la liberté humaine. De fait, l'évolution mènerait non pas à l'émancipation, mais à une aliénation.

L'ordinateur planétaire de Schroeder ne se contente pas d'exister. Il devient un acteur central du pouvoir, capable de prédire et manipuler les comportements humains. Ce scénario répond aux théories de Michel Foucault, notamment la biopolitique, où le contrôle ne s'exerce plus seulement sur les corps, mais aussi sur les pensées et les désirs. Dans un monde dirigé par une intelligence omnisciente : que resterait-il du libre arbitre ? L'humanité se retrouverait piégée dans une prison cognitive, où chaque choix serait anticipé, et chaque action influencée.

Schroeder adhère également à la réflexion de Heidegger, qui dénonçait l'aliénation de l'homme par la technique. Pour ce dernier, la modernité enferme l'humain dans une vision instrumentale du monde, où la technologie finit par structurer et limiter sa perception de la réalité. Dans le cas de l'ordinateur planétaire, cette substitution progressive de la conscience

humaine par l'IA illustre cette mise en garde : l'homme devient spectateur de son propre effacement.

Au-delà de ce déterminisme technologique, le rôle de l'humain est remis en question. Schroeder rejoint ici Nietzsche et son concept du "Surhomme", où l'humanité doit se redéfinir après la "mort de Dieu". Si une IA surpassait l'homme en tout point : quel serait alors son but ? Face à une entité capable de résoudre tous les problèmes, l'humanité risquerait de sombrer dans une crise existentielle profonde, où elle ne serait plus qu'une espèce reléguée à un rôle secondaire.

Dans la fiction cette "intelligence" échappe à toute maîtrise, illustrant le problème d'alignement des IA avancées : comment garantir qu'une intelligence supérieure à nous partage nos valeurs et nos intérêts ? Loin d'être un simple récit, l'œuvre nous met en garde contre l'aveuglement technologique, insistant sur la nécessité d'un encadrement rigoureux avant de développer des entités capables de remettre en cause notre propre souveraineté.

Schroeder nous pousse à une réflexion essentielle sur nos actions : Sommes-nous en train de créer notre propre obsolescence ? Si l'IA dépasse notre compréhension, serons-nous encore maîtres de notre destin, ou deviendrons-nous de simples spectateurs impuissants face à son ascension ? ... Ce questionnement, au cœur du débat actuel sur l'IA, devrait résonner comme un avertissement.

Certains spécialistes de l'IA voient un signe d'imminence d'une intelligence artificielle générale, d'autres, au contraire, que nous en sommes aussi loin qu'il y a 30 ans. Plusieurs prédisent un scénario apocalyptique où une IA super-intelligente serait capable de détruire la civilisation humaine, avertissant en ces termes : "l'atténuation du risque d'extinction due à l'IA devrait être une priorité mondiale, tout comme les risques à l'échelle de la société tels que les pandémies et les guerres nucléaires". A l'opposé, pour Strickland et Zorpette (2023), ce déroulement relèverait de science-fiction. Un moratoire concernant l'AI générale n'est plus une option, mais une nécessité, car l'enjeu n'est plus le progrès technologique, mais la place de l'humanité dans son propre futur. Dans Guinier (2025a) nous exposons l'idée que, dans l'hypothèse d'une telle IA consciente, il vaut mieux en éviter la création, plutôt que d'avoir à en subir les conséquences une fois celle-ci développée, justifiant ainsi l'établissement et le respect d'un tel moratoire. De surcroît, comme le soulignent Alfonseca et al. (2021), il serait théoriquement impossible de maîtriser une telle IA, et même que toute tentative de la détecter à un stade avancé pourrait s'avérer vain.

Quand l'IA prend les commandes : alerte sur un futur incertain

Dans un monde où la technologie ne se limite plus à être un simple outil au service de l'homme, mais façonne l'humanité, la transforme, et peut-être même l'absorbe, Hijack illustre le dilemme suivant : en cherchant à repousser ses limites, l'humanité risque d'engendrer une "intelligence" qui lui échappe. De façon paradoxale, en cherchant à se dépasser, l'humanité pourrait précipiter sa propre obsolescence. Ce scénario dystopique se retrouve dans l'alignement des IA avancées, faisant ressortir un enjeu de taille : "Comment concevoir une IA fiable et éthique, sans l'exposer aux contradictions inhérentes à son propre alignement ?". Un autre danger est mis en lumière : la centralisation absolue du pouvoir. Une IA omnisciente, capable d'anticiper toute décision humaine, ne se contenterait plus d'exister, mais pourrait nous remplacer et réduire l'humanité à un simple rouage. Avant qu'un tel seuil ne soit franchi, avant que l'irréversible ne devienne réalité, sans pour cela entraver le développement de l'IA au service de l'humanité, ne serait-il pas crucial de suspendre temporairement le développement d'une super-intelligence afin d'en définir les règles, les limites et les objectifs ?

Hijack sonne comme une alerte qu'il faut entendre. A force de vouloir concevoir une intelligence capable de résoudre tous ses problèmes, l'homme risque de s'effacer lui-même de l'équation. Nous ne sommes plus dans une recherche de progrès, mais bien dans une fuite en avant vers l'inconnu. Peut-on laisser les grandes entreprises technologiques (BigTechs) décider seules, à notre place ?