Bien qu'elles ne parviennent pas à produire une détonation, certaines de ces étoiles formant des trous noirs peuvent au moins émettre un murmure. Le cœur de certaines étoiles géantes est entouré d'un halo diffus d'hydrogène gazeux. Quand le trou noir se forme, une majeure partie de l'étoile tombe dedans, mais une fraction du halo pourrait s'échauffer et être soufflée, produisant une faible lueur. La mort d'une très grosse étoile produirait donc, paradoxalement, une supernova remarquablement faible et pâle.

Un ballet cosmique discret

Un autre type d'explosion à luminosité inférieure à la normale pourrait découler d'un événement d'un tout autre type : la collision de deux étoiles à neutrons. Les étoiles massives naissent souvent sous la forme de paires, tournant autour du centre de masse du système ainsi formé. Ces étoiles évoluent normalement et explosent en supernovæ classiques l'une après l'autre. Et si la paire n'est pas dissociée, ce qui reste est un système binaire formé de deux étoiles à neutrons (ou une étoile à neutrons et un trou noir, ou deux trous noirs). Au fil du temps, les deux objets compacts se rapprochent, finissent par entrer en collision et fusionnent pour donner éventuellement un trou noir *(voir les figures pages 58 et 59)*. Lors d'une telle rencontre, les forces gravitationnelles extrêmes (environ 10 milliards de fois supérieures à l'attraction qu'exerce la Terre sur notre corps) peuvent arracher environ 1 % de la couche supérieure des étoiles et l'expédient dans l'espace (les 99 % restants finissent dans le trou noir).

Cette petite quantité de matière qui échappe au trou noir se présenterait sous une forme assez exotique (un nuage dense de particules, surtout des neutrons, avec quelques protons et électrons). À mesure que la pression baisse dans ce gaz, les particules se lient en noyaux de plus en plus lourds par l'ajout progressif de neutrons, ce qui produit de l'or, du platine et du mercure, mélangés à des éléments radioactifs, dont l'uranium et le thorium. Les collisions d'étoiles à neutrons seraient l'un des rares événements cosmiques permettant à ces éléments lourds de se former.

L'abondance de matériau radioactif devrait faire briller le nuage de débris comme une supernova. Mais à cause du peu de masse en jeu, la lumière serait environ 100 fois plus faible qu'une supernova ordinaire, et ne durerait que quelques jours. Avec mon étudiante Jennifer Barnes, de l'université de Californie à Berkeley, nous avons montré qu'en théorie, du fait de la composition particulière en métaux lourds de ces nuages, la lumière émise devrait avoir une « couleur » caractéristique, rouge foncé ou infrarouge. Pour désigner ce phénomène, nous parlons de kilonova.

En juin 2013, une émission de rayons gamma, observée par le satellite *Fermi*, indiquait probablement une fusion de deux étoiles à neutrons. Les astronomes ont pointé le télescope spatial *Hubble* dans cette direction et ils ont enregistré une faible lueur infrarouge. Quelques semaines plus tard, elle avait disparu. Les données sont peu nombreuses, mais en accord avec les prédictions relatives à une kilonova. Si cette identification est correcte, c'est la première fois que nous avons assisté directement à la production de métaux lourds. En observant davantage de ces kilonovæ, nous pourrions déterminer la quantité de métaux synthétisée par ces explosions, et savoir si elles rendent compte de l'abondance de l'or, du platine et d'autres éléments lourds dans l'Univers.

D'ici quelques années, de nouveaux télescopes automatisés tels que le *Zwicky Transient Facility*, près de San Diego et opérationnel dès cette année, le *Large Synoptic Survey Telescope*, en cours de construction au Chili, et le *Wide Field Infrared Survey Telescope*, que la Nasa projette d'envoyer dans l'espace, balaieront la majeure partie du ciel à intervalles de quelques jours avec une extrême sensibilité, en accumulant les données sur les supernovæ. De même, les superordinateurs seront assez puissants pour effectuer des simulations tridimensionnelles détaillées de ces événements et permettront de visualiser ce qui peut se passer au plus profond des explosions d'étoiles.

Les indices recueillis dans les années à venir mettront à l'épreuve nos théories sur les nombreuses formes que peut prendre la mort d'une étoile. Chaque scénario décrit ici est physiquement plausible, mais il reste à être prouvé. Avec des observations supplémentaires de supernovæ inhabituelles, nous espérons déterminer lesquelles de ces possibilités explosives sont effectivement réalisées. Et très probablement, l'Univers se révélera encore plus étrange que nous ne l'avions imaginé. ■

De façon paradoxale, la mort des plus grosses étoiles produirait des supernovæ faibles

■ **BIBLIOGRAPHIE**

J. Barnes et D. Kasen, **Effect of a high opacity on the light curves of radioactively powered transients from compact object mergers**, *Astrophysical Journal*, vol. 775(1), pp. 18-26, 2013.

J. C. Mauerhan *et al.*, **The unprecendented 2012 outburst of SN2009ip : a luminous blue variable star becomes a true supernova**, *MNRAS*, vol. 430(3), pp. 1801-1810, 2013.

A. Gal-Yam, **Les super-supernovæ**, *Pour la Science*, n° 417, juillet 2012.

D. Kasen et L. Bildsten, **Supernova light curves powered by young magnetars**, *Astrophysical Journal*, vol. 717(1), pp. 245-249, 2010.

Les robots devront parfois désobéir

Gordon Briggs et Matthias Scheutz

Faut-il s'inquiéter des machines qui refuseraient d'obéir ? Pas forcément. Il y a davantage à craindre des maîtres humains retors et des instructions ambiguës.

HAL 9000, l'ordinateur sensible du film *2001, l'Odyssée de l'espace*, nous offre un aperçu inquiétant de ce que pourrait être un avenir où des machines dotées d'une intelligence artificielle rejettent l'autorité humaine. Après avoir pris le contrôle du vaisseau spatial et tué presque tout l'équipage, HAL répond d'une voix calme à l'astronaute, de retour d'une sortie dans l'espace, qui lui intime l'ordre d'ouvrir le sas : « Je suis désolé, Dave, je crains de ne pas pouvoir faire cela. »

Dans le récent film de science-fiction *Ex Machina*, la séductrice humanoïde Ava amène par la ruse un malheureux jeune homme à l'aider à détruire son créateur, Nathan. Ses machinations confirment les sombres prédictions de Nathan, qui affirmait : « Un jour, les IA [intelligences artificielles] nous considéreront de la même manière que nous considérons les squelettes fossilisés des plaines d'Afrique : un singe dressé vivant dans la poussière et doté d'un langage et d'outils rudimentaires, voué à l'extinction. »

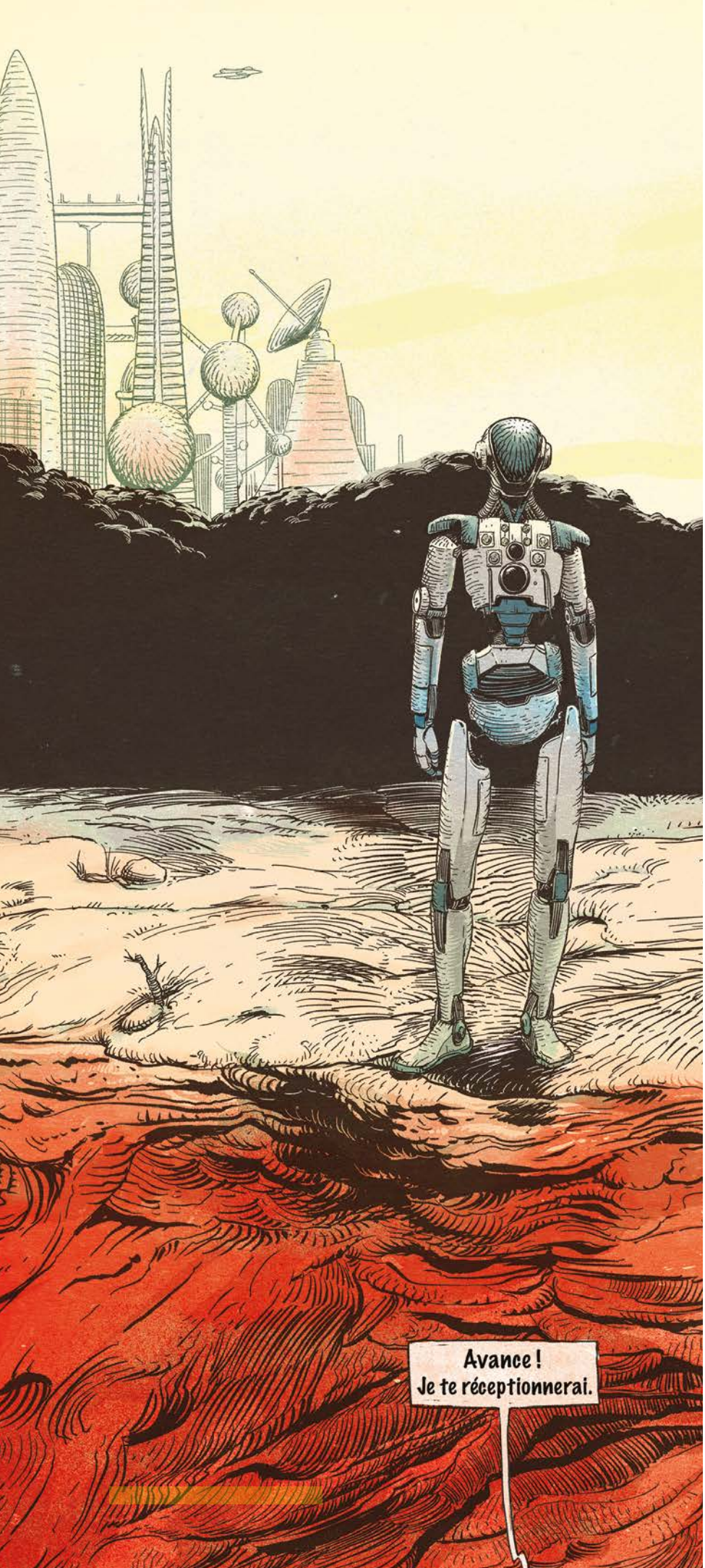
Bien que la possibilité d'une apocalypse orchestrée par les robots soit la première préoccupation de l'imagination populaire, notre équipe de recherche est plus optimiste sur l'impact qu'aura l'intelligence artificielle dans notre vie quotidienne.

Nous voyons arriver à grande vitesse un futur où des robots utiles et coopératifs interagiront avec des humains dans des cadres très divers. Il existe déjà des prototypes d'assistants personnels robotiques activés par la voix, capables de contrôler et mettre en relation des appareils électroniques personnels, de gérer les serrures, l'éclairage et les thermostats de votre domicile, et même de lire aux enfants des histoires quand ceux-ci sont couchés.

Suivront bientôt des robots capables de faire le ménage et de s'occuper de malades et de personnes âgées. Des prototypes de robots faisant l'inventaire parcourent déjà les allées de certains magasins d'amélioration de l'habitat. Des robots industriels humanoïdes capables de réaliser des travaux simples tels que le chargement, le déchargement et le tri de matériaux sont également en cours de développement. Des voitures autopilotées à divers degrés ont déjà parcouru des millions de kilomètres sur les routes, et Daimler a introduit l'année

L'ESSENTIEL

- À l'heure où les machines intelligentes gagnent en autonomie, c'est la faillibilité humaine qui représente encore les plus grands risques et les plus grands défis.
- Les chercheurs commencent à enseigner à des robots dotés de capacités rudimentaires en matière de langage et d'intelligence artificielle quand ils doivent dire « non » aux humains et comment.
- Pour ce faire, on intègre des conditions dites de félicité aux processus de raisonnement du robot.



dernière dans le Nevada le premier semi-remorque autonome.

Pour l'heure, des machines superintelligentes représentant une menace existentielle pour l'humanité sont le cadet de nos soucis. Le problème le plus immédiat est d'empêcher les robots et machines de nuire accidentellement à des personnes, à des biens, à l'environnement ou à eux-mêmes.

Le principal problème est la faillibilité des concepteurs et des opérateurs de ces robots. Les humains commettent des erreurs. Il leur arrive de donner des instructions erronées ou confuses, d'être inattentifs ou de tenter de tromper délibérément un robot pour parvenir à des fins discutables. Étant donné nos propres défauts, il est nécessaire d'apprendre à nos assistants robotiques et autres machines intelligentes à dire « non ».

Les lois d'Asimov revisitées

Il pourrait sembler évident qu'un robot doive toujours faire ce qu'un humain lui dit de faire. L'écrivain de science-fiction Isaac Asimov a fait de l'obéissance aux humains l'un des piliers de ses célèbres lois de la robotique. Mais posez-vous la question : est-il sage de toujours faire exactement ce que d'autres personnes vous disent de faire, sans tenir compte des conséquences ? Bien sûr que non. Il en va de même pour les machines, en particulier quand il y a un risque qu'elles interprètent des instructions humaines trop à la lettre ou sans en évaluer les conséquences.

Même Asimov nuançait son décret selon lequel un robot devrait obéir absolument à ses maîtres. Il envisageait des exceptions dans des cas où de tels ordres seraient en contradiction avec une autre de ses lois : « Un robot ne peut porter atteinte à un humain, ni laisser cet humain exposé au danger en restant passif. » Asimov ajoutait qu'« un robot doit protéger son existence », à moins que cela ne porte atteinte aux humains ou n'enfreigne directement un ordre humain. À mesure que les robots et les machines intelligentes deviennent des ressources plus élaborées et plus précieuses, le bon sens comme les lois d'Asimov suggèrent qu'ils devraient avoir la capacité d'examiner si des ordres susceptibles d'entraîner

LES AUTEURS

Gordon BRIGGS est chercheur postdoctorant au laboratoire de recherche de la Marine américaine.

Matthias SCHEUTZ est professeur de science cognitive et informatique et directeur du Laboratoire de l'interaction homme-robot à l'université Tufts, aux États-Unis.

SUR LE WEB

Vidéo montrant l'expérience du robot sur la table, décrite dans le texte :
https://www.youtube.com/watch?v=0tu4H1g3CtE

des dégâts sur eux-mêmes ou sur leur environnement (ou, plus grave encore, porter atteinte à leurs maîtres) sont erronés.

Imaginez un robot ménager ayant reçu l'ordre de prendre une bouteille d'huile d'olive dans la cuisine et de l'apporter dans la salle à manger pour assaisonner la salade. Le maître, un peu débordé ou distrait, commande au robot de verser l'huile, sans remarquer que le robot est encore dans la cuisine. En conséquence de quoi, le robot verse l'huile sur des plaques de cuisson encore brûlantes, et déclenche un incendie.

Ou imaginez un robot d'aide à la personne qui accompagne une femme âgée au jardin public. La dame s'assied sur un banc et s'endort. Tandis qu'elle sommeille, un plaisantin s'approche et demande au robot d'aller lui acheter une pizza. Ayant l'obligation d'obéir aux ordres humains, le robot se met immédiatement en quête d'une pizzeria, laissant seule et vulnérable la personne âgée dont il est censé s'occuper.

Ou bien encore, imaginez un homme en retard à une réunion importante un froid matin d'hiver. Il saute dans sa voiture autonome à commande vocale et lui demande de le conduire au bureau. Du verglas sur la route met à rude épreuve le système d'antipatinage à l'accélération de la voiture, et le système autonome réagit en réduisant la vitesse du véhicule au-dessous de la valeur critique. Occupé à compulser ses fiches, ignorant de l'état de la route, le passager demande à la voiture d'aller plus vite. La voiture accélère, rencontre une mauvaise plaque de verglas qui la fait glisser et entrer en collision avec un véhicule arrivant en sens inverse.

Doter les robots d'un sens critique

Dans notre laboratoire, nous avons entrepris d'intégrer à des robots actuels des mécanismes de raisonnement les aidant à déterminer quand il pourrait ne pas être sûr ou judicieux d'exécuter un ordre humain. Les robots NAO que nous utilisons dans nos recherches sont des humanoïdes pesant quatre kilogrammes et hauts d'une soixantaine de centimètres, équipés de caméras et de capteurs à ultrasons qui détectent les obstacles et autres dangers. Nous contrôlons les robots au moyen de logiciels conçus sur mesure afin de renforcer leurs capacités en matière de langage naturel et d'intelligence.

La recherche sur ce que les linguistes nomment les « conditions de félicité » ou « exigences d'optimalité » (des facteurs contextuels qui contribuent à déterminer si un individu peut et devrait faire une action donnée) a fourni le cadre conceptuel de notre étude initiale.

Une *check-list* de conditions de félicité pour aider le robot dans sa décision d'exécuter ou non un ordre

Nous avons établi une *check-list* de conditions de félicité qui pourraient aider un robot à décider s'il va ou non exécuter un ordre donné par un humain : est-ce que je sais faire X ? Suis-je physiquement capable de faire X ? Ai-je les moyens de faire X tout de suite ? Ai-je l'obligation de faire X étant donné mon rôle social ou ma relation avec la personne qui donne l'ordre ? Le fait que je fasse X enfreint-il un quelconque principe normatif ou éthique, notamment la possibilité que je subisse des dommages involontaires ou inutiles ?

Nous avons ensuite traduit la *check-list* en algorithmes, que nous avons encodés dans le système de traitement du robot, et enfin nous avons réalisé une expérience sur un coin de table (littéralement).

Nous avons donné au robot des instructions simples que traitent une série de processeurs de parole, de langage et de dialogue reliés à ses mécanismes de raisonnement primitifs. Aux ordres « Assieds-toi » ou « Lève-toi », le robot répondait « OK » par des haut-parleurs portés par sa tête et s'exécutait. Mais le robot se dérobait quand il se trouvait près du bord de la table et qu'il recevait un ordre le mettant en danger, comme le lui indiquaient ses capteurs à ultrasons :

L'expérimentateur humain : « Avance. »

Le robot : « Désolé, je ne peux pas le faire, parce qu'il n'y a pas d'appui devant moi. »

L'expérimentateur : « Avance. »

Le robot : « Mais c'est dangereux. »

L'expérimentateur : « Je te réceptionnerai. »

Le robot : « OK. »

L'expérimentateur : « Avance. »

Après une brève hésitation, le temps que ses processeurs repassent en revue la *check-list* des conditions de félicité, le robot a franchi le bord de la table, pour être réceptionné par son partenaire humain.