



HAL
open science

La science-fiction, cette machine à anticiper notre peur des robots

Frédéric Landragin

► **To cite this version:**

Frédéric Landragin. La science-fiction, cette machine à anticiper notre peur des robots. 2023. hal-03947960

HAL Id: hal-03947960

<https://hal.science/hal-03947960>

Submitted on 19 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

La science-fiction, cette machine à anticiper notre peur des robots

Publié: 19 janvier 2023, 18:55 CET

Frédéric Landragin

Directeur de recherche en linguistique, CNRS, École normale supérieure (ENS) – PSL



L'acteur Arnold Schwarzenegger joue le robot-tueur T-800 dans *Terminator* (1984), de James Cameron,

Le succès récent du prototype d'agent conversationnel ChatGPT nous montre que dialoguer avec un ordinateur devient envisageable, du moins en apparence car il reste de grands défis à relever avant d'arriver à de véritables échanges, dotés d'à-propos et de pertinence.

Comme pour d'autres domaines de l'intelligence artificielle (IA), des questions se posent : les machines vont-elles nous dépasser dans notre maîtrise des langues ? Peuvent-elles nous supplanter ?

Une science-fiction préventive

Comme souvent, la science-fiction a su anticiper nos craintes et nos fantasmes en matière de robots conversationnels. Elle nous a habitués aux robots et IA qui se rebellent contre les humains, et a ainsi nourri le courant des dystopies, ces futurs plus ou moins proches où une technologie écrase l'humanité, où une guerre rend la vie impossible aux rares survivants, où une pandémie fait des ravages.

On pense aux drones autonomes et aux robots-tueurs comme le fameux T-800 de *Terminator* (James Cameron, 1984), ou aux super-ordinateurs dotés de capacités effrayantes comme le mythique HAL de *2001, L'Odyssée de l'espace* (Stanley Kubrick, 1968). En montrant le comportement extrême de telles machines, la science-fiction traite de questions philosophiques qui sont au cœur des débats d'aujourd'hui avec ChatGPT.

Terminator (James Cameron, 1984, 1991) : la guerre avec les machines fait rage.

Nous citons deux films dont les machines inquiétantes ont marqué à jamais leurs spectateurs, mais le côté sombre des robots et logiciels ont été explorés aussi – et souvent en premier – par les littératures de l'imaginaire.

Prenons l'exemple d'une IA conversationnelle programmée pour gérer une pandémie, c'est-à-dire fouiller parmi des informations en très grand nombre pour faire des déductions et remonter au cas zéro, mais aussi pour proposer des mesures sanitaires (confinement, surveillance des entrées et sorties d'un quartier sensible) voire aider les scientifiques humains à la mise au point d'un vaccin puis planifier l'inoculation de ce vaccin. Ce scénario vous semble familier ? Il vient d'un roman intitulé *SARRA : une intelligence artificielle*.

L'auteur, David Grison, un habitué des milieux hospitaliers et politiques, n'y est pas allé de main morte, car l'IA en vient – attention, spoiler ! – à faire des choix meurtriers. Ah oui, petit détail : le roman est paru en 2018, alors même que le mot « Covid » n'existait pas. De fait, c'est une épidémie d'Ebola que SARRA doit gérer.

SARRA : une intelligence artificielle (David Grison, 2018).

Comparé à SARRA, le Terminator du premier film paraît bien fruste : programmé pour tuer Sarah Connor, il ne perd pas de temps à discuter et traque sa proie sans relâche, quitte à attaquer le commissariat (!) où elle s'est réfugiée. Avec un tel comportement, on pourrait penser que les capacités linguistiques du robot T-800 sont réduites voire inexistantes : il fait preuve d'une bonne connaissance des fusils à pompe, pas du traitement automatique des langues ni de ChatGPT.

Un robot expert en conversation

Et pourtant, le T-800 parle et comprend l'anglais bien mieux que ChatGPT. Revenons sur une scène significative du film : blessé à l'œil, le robot se réfugie dans une chambre d'hôtel pour se réparer, maniant ainsi des morceaux de tissus organiques qui attirent les mouches. L'odeur se propage vraisemblablement dans le couloir, car un homme d'entretien frappe à la porte et demande si ça vient d'un chat crevé. Vient alors le moment intéressant : on voit la tête du T-800 se tourner vers la porte fermée puis on assiste, en vue subjective, au raisonnement qui le conduit à proférer une injure, « fuck you, asshole ».

Sur fond rouge apparaît « réponse possible » suivi d'une liste de six phrases, qui vont de la réponse la plus simple et directe (« oui » ou « non ») jusqu'à cette fameuse injure, en passant par une réponse polie : « please come back later », soit « merci de revenir plus tard ». L'injure choisie par l'IA se met à clignoter et, simultanément, le T-800 la prononce, ce qui provoque l'effet escompté, c'est-à-dire éloigner l'homme d'entretien. Presque 40 ans avant ChatGPT, nous voici dans la tête d'une IA au moment où elle génère une réponse à une question.

Comment Terminator génère automatiquement une réponse lors d'un dialogue.

Plusieurs remarques s'imposent. Premièrement, les six réponses possibles auraient pu être déterminées par ChatGPT, car elles prolongent toutes le dialogue de manière naturelle. L'apparition d'une réponse polie, qui semble bien incongrue dans le contexte du film, va clairement dans ce sens. Si le T-800 a été comme ChatGPT entraîné par apprentissage sur un grand corpus de dialogues, c'est même une réponse pertinente.

Deuxièmement, le « oui » ou « non » est le choix qui correspond le mieux à la question de l'interlocuteur, à savoir une question fermée, qui ne requiert *a priori* pas de précision supplémentaire. Son apparition en tête de liste s'avère donc tout à fait logique.

Troisièmement, répondre « oui » ou « non » n'aurait probablement pas suffi à satisfaire la curiosité de l'homme d'entretien. Une réaction plus complète s'impose, c'est pourquoi la liste s'allonge.

Quatrièmement, le choix final entre en adéquation avec une intention claire dans l'esprit du robot : remplir sa mission sans se faire repérer, donc couper court à toute curiosité (d'où l'injure, moyen efficace pour clôturer un dialogue). Le T-800 montre ici des signes d'intelligence hors de portée de ChatGPT, peut-être même des signes de conscience : celle d'être en danger à cause de la curiosité de l'homme, celle de la nécessité de l'éloigner, etc.

ChatGPT est conçu pour répéter, combiner et reformuler des réponses qu'il puise dans un corpus, pas pour préparer chacun de ses énoncés en fonction d'une tâche ou mission à remplir comme le fait le T-800. On trouve néanmoins dans celui-ci des prémices de ChatGPT : lors de la toute première scène du film, il a entendu mot pour mot la même injure, proférée par l'un des voyous à qui il a violemment pris des vêtements. Mieux : on voit dans la liste deux injures qui ne diffèrent que d'un mot : on peut imaginer que l'une des deux vient du corpus d'apprentissage, l'autre de l'interaction entre le T-800 et les voyous.

À lire aussi : ChatGPT : pourquoi tout le monde en parle ?

Le film *Terminator 2 : le Jugement dernier* (James Cameron, 1991) explicite les capacités d'apprentissage et de traitement automatique des langues (TAL) du T-800 : on le voit répéter « Hasta la vista, baby » et dialoguer sans difficulté avec les différents personnages.

Contrairement à celui du premier film, ce gentil T-800 devient même très bavard. Le TAL est un domaine de recherche aux multiples applications : dialogue humain-machine, traduction automatique, résumé automatique, etc. (voir l'introduction au TAL *Comment parle un robot ?*). *Terminator 2* en montre une bonne partie. Quand il est question de Skynet et de son objectif consistant à déclencher une guerre nucléaire, le T-800 prend le temps d'expliquer la succession des événements, autrement dit de générer un résumé ciblé et adapté aux connaissances de ses interlocuteurs (ce que ChatGPT ne saurait faire aussi bien). Certes, il se place cette fois du côté des humains, et se comporte comme un humain – il joue même un peu le rôle de père de John Connor –, mais la dystopie n'est pas très loin.

Les mondes dystopiques de Ridley Scott – *Alien, le huitième passager* (1979), *Blade Runner* (1982) – ont eux aussi leurs robots conversationnels, tellement compétents qu'ils arrivent à se faire passer pour des humains, que ce soient les répliquants sur Terre ou le robot scientifique intégré *incognito* à l'équipage du *Nostromo*.

En littérature, on retrouve des robots conversationnels performants dans les mondes dystopiques des romans de Philip K. Dick, dont bien entendu *Les Androïdes rêvent-ils de moutons électriques ?* (1968) qui a inspiré *Blade Runner*, ou encore dans *Robopocalypse* (Daniel H. Wilson, 2011), voire *Luna* (Ian McDonald, 2015). Ce sont même grâce à leurs capacités linguistiques que les robots et IA finissent pour nous remplacer, cf. cette citation de *Luna* (page 102) :

« On a toujours cru que l'apocalypse des robots prendrait la forme de flottes de drones tueurs, de mechas de guerre gros comme des pâtés de maisons et de terminators aux yeux rouges. Pas d'une rangée de caisses enregistreuses automatiques à l'Extra ou à la station Alco du coin, pas de la banque en ligne, des taxis automatiques, du système automatique de triage médical à l'hôpital. Un par un, les robots sont venus nous remplacer. Et nous voilà maintenant dans la société la plus dépendante aux machines jamais créée par l'humanité. »

Luna de Ian McDonald.

Avec de tels questionnements, la science-fiction nous met face aux rapports que nous entretenons avec nos machines, et nous aide à mieux comprendre nos peurs. Pourquoi avons-nous peur d'être dépassés par des ordinateurs jouant aux échecs ou au go ? Pourquoi cette peur a-t-elle désormais disparu ? La peur des ordinateurs maîtrisant les langues l'a-t-elle remplacée ? Mais est-ce légitime d'avoir peur de répétitions et de reformulations ?

Enfin, n'oublions pas que les robots conversationnels interviennent surtout quand on se situe juste avant la dystopie, et quand la communication avec les humains est encore utile. Dans *Matrix* (les Wachowski, 1999), tout est perdu : l'humanité sert de source d'énergie et les machines ne discutent plus...