



MINISTÈRE DE L'ÉDUCATION NATIONALE

*Liberté
Égalité
Fraternité*

Décrypter le numérique, transformer l'enseignement – Cahiers disciplinaires



N° 19

JUIN 2026

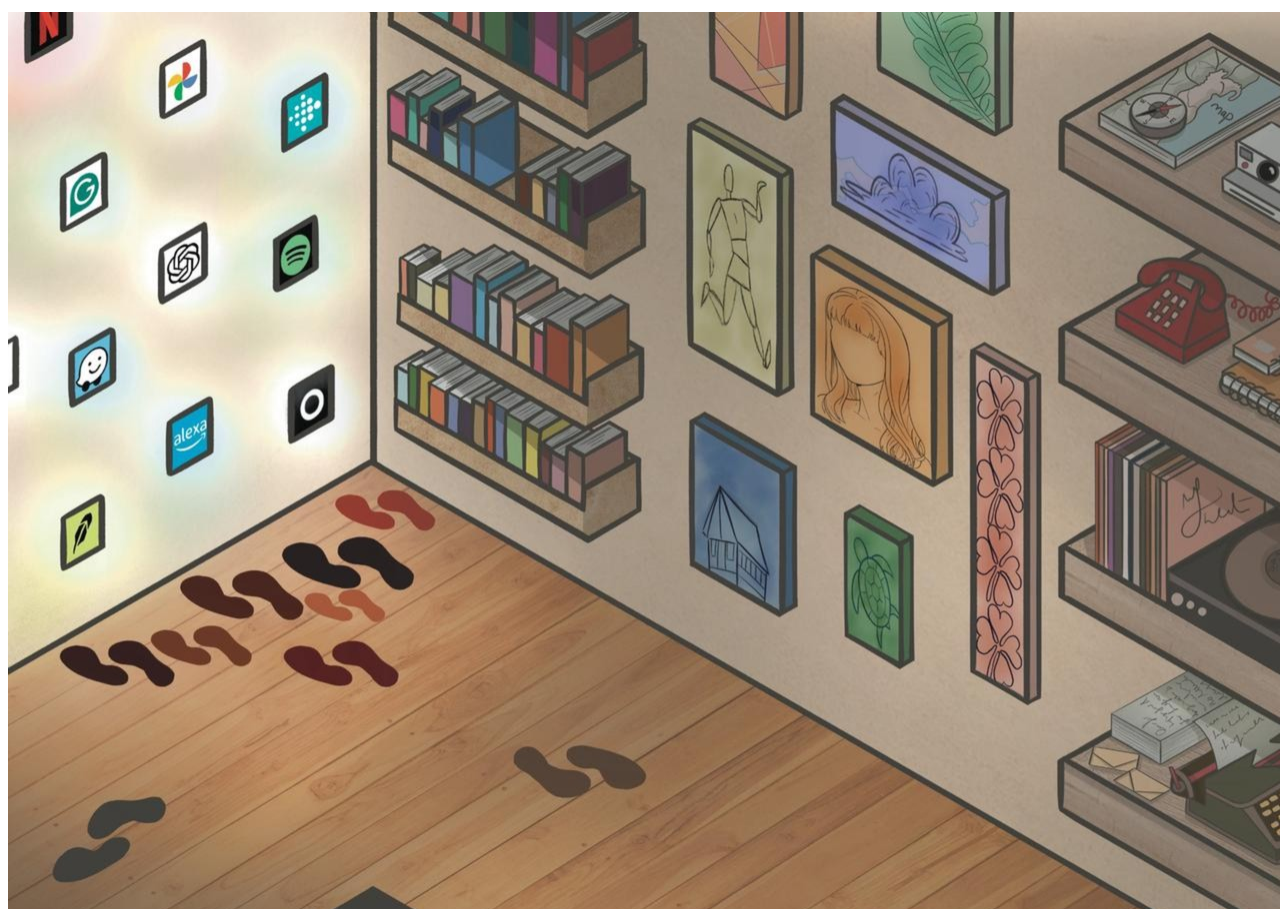


Figure 1 : Reihaneh Golpayegani & The Bigger Picture /
<https://betterimagesofai.org> / <https://creativecommons.org/licenses/by/4.0/>

Ce cahier ÉduNum en philosophie n° 19 inaugure une série dédiée à l'intelligence artificielle comme objet de réflexion philosophique et à ses usages en classe de philosophie.

Faisant le lien avec la recherche en philosophie, il s'ouvre sur un entretien avec **Éloïse Boisseau**, autrice d'une thèse de doctorat intitulée *Le Moulin à parole. Enquête sur la notion de parole dans la philosophie de l'IA*, et propose un accès à une conférence en ligne de **Bertille Rouillon**, doctorante, dont les travaux portent sur les problèmes et les risques générés par le mésalignement des intelligences artificielles.

La seconde rubrique offre des pistes pour interroger l'intelligence artificielle en lien avec les notions du programme de tronc commun ainsi qu'avec les thèmes de spécialité **HLP**. Intitulée « Focus », elle s'ouvre sur une réflexion consacrée aux puissants processus d'attachement relationnel suscités par les agents conversationnels, et invite les élèves à s'interroger sur leurs propres usages des intelligences artificielles.

Ce cahier se termine par la présentation de **pratiques pédagogiques** portant aussi bien sur l'IA comme objet de réflexion que sur ses usages en classe de philosophie.

■ S'abonner à la [liste de diffusion en philosophie](#)

■ S'abonner aux [Cahiers ÉduNum Thématiques](#)

Sommaire

Du côté de la recherche

Entretien avec Éloïse Boisseau.....	3
Systèmes d'IA : mésalignés ou non alignables ?	
Conférence de Bertille Rouillon, doctorante.....	8
Focus sur : IA et recherche de soi.....	9
Une parole sans sujet ? Les réactions humaines face à la parole machinique.....	9
Nous sommes le problème.....	10
Typologies des réactions humaines.....	12
L'éducation à l'épreuve des voix artificielles.....	13
Pour aller plus loin.....	15
Pratiques pédagogiques.....	15

Entretien avec Éloïse Boisseau



Figure 2 : © Éloïse Boisseau. Avec l'aimable autorisation de l'autrice

Pour commencer, pourriez-vous revenir sur votre parcours intellectuel et nous expliquer ce qui vous a conduit à travailler sur l'intelligence artificielle du point de vue philosophique ?

Eh bien, je viens de soutenir en décembre dernier une thèse de doctorat intitulée « Le Moulin à parole. Enquête sur la notion de parole dans la philosophie de l'IA » et menée à Aix-Marseille Université. Cette thèse se situe dans la continuité d'un travail de master qui portait sur la question de l'attribution de prédicats psychologiques à des machines. J'ai été assez tôt marquée par la lecture du livre

La Denrée mentale de Vincent Descombes, ouvrage magistral écrit lors d'une autre grande période d'interrogations sur l'IA ; plus généralement ce sont les positions radicales (quant à la nature de l'esprit, du langage, de la signification, etc.) des philosophes qui s'inscrivent dans une tradition que l'on peut qualifier de « wittgensteinienne » qui m'ont avant tout influencée. Puisqu'il n'aura échappé à personne que l'IA est revenue ces dernières années au centre des discussions et des préoccupations, il m'a semblé bon de continuer le travail mené dans cette tradition en l'appliquant à des problématiques en apparence inédites. Inutile de dire que cette tradition est aujourd'hui minoritaire dans le grand marché des idées : la mobiliser, la poursuivre et la faire connaître m'est donc apparu comme salvateur (puisque – c'est du moins ce que je soutiens – c'est une tradition qui peut nous libérer de tout un tas de confusions redoutables ou d'obsédantes pseudo-questions).

Quels sont les arguments de ceux qui attribuent aujourd'hui une authentique « parole » aux IA génératives ?

C'est une grande question ! Il y en a plusieurs, provenant de sources pas forcément unitaires ou uniformes. Pour en donner un à gros traits qui capture bien l'esprit général de plusieurs de ces arguments : l'attribution de parole doit passer par l'examen des productions des candidats à la parole. Puisque les productions des IA génératives sont (ou peuvent être) indiscernables des productions d'authentiques causeurs (pour qui il n'y a pas de doute qu'ils parlent bel et bien), c'est donc qu'elles parlent au même titre que ces derniers. C'est ainsi essentiellement un argument par la ressemblance (ou argument d'identité des indiscernables). S'ajoutent parfois certaines considérations quasi-solipsistes de philosophie de l'esprit de type : de toutes façons nous ne pouvons passer que par les productions, c'est d'ailleurs bien (continue l'argument) ce que nous faisons avec nos congénères : nous inférons qu'ils ont une parole en les écoutant, en les lisant, etc. Puisque nous sommes dans la même position épistémologique concernant les IA génératives, nous devons tirer la même conclusion (à effets égaux, causes égales). Une autre dimension de cet argument consiste en outre à insister sur l'absolue nouveauté des productions des IA génératives : cette absolue nouveauté n'était (jusqu'ici) qu'un privilège des êtres humains (le privilège de pouvoir dire quelque chose qui n'a jamais été dit), privilège désormais perdu, etc.

Pourquoi considérez-vous l'image de la « machine parlante » comme mystifiante ?

C'est une question encore plus grande ! J'ai essayé – sur les 500 pages de la thèse – d'articuler une batterie d'arguments pour établir qu'il y avait en effet dans la prétention d'avoir affaire à des « machines parlantes » toute une mythologie philosophique à l'œuvre. Il va sans dire que je ne pourrai résumer ou explorer ici toute l'articulation. Encore une fois, à très gros traits donc, je dirai la chose suivante : une telle image est mystifiante car elle est aveugle à la source de la parole (par opposition à son vestige, i.e. par opposition à ce qui reste). L'authentique parole doit d'abord être motivée (il doit y avoir un point pour que l'on puisse dire qu'il y a parole) ; l'authentique parole suppose que des conventions (des règles) soient suivies (et une machine ne saurait *suivre* des règles, même si elle peut agir *en accord* avec des règles) ; la parole a très souvent une dimension *expressive* (et requiert donc *en amont* qu'il y ait des buts, des désirs, du bien et du mal pour celui ou celle qui parle) ; la parole est intentionnelle (au double sens où il n'y a pas de sens à dire qu'on parle *sans le savoir* ou sans que *quelque chose soit visé*) ; la parole est aussi *engagée* et engageante (parler est une forme d'action sociale, normative, et au bout du compte *morale* – elle implique la responsabilité). Ces considérations rendent l'idée d'une machine parlante contradictoire.

Comment distinguer la simple production de signes par un modèle de langage de ce que nous appelons, chez l'humain, un acte de parole ? En quoi un texte généré par IA ne peut-il pas être considéré comme un jugement ?

J'insiste d'une certaine façon sur la primauté de la pragmatique sur la syntaxe : la production de signes par quelque LLM n'est *pas du tout* un acte de parole (les signes ne sont pas des formules magiques permettant par leur simple existence de conjurer des locuteurs). Les authentiques locuteurs – comme je le disais à l'instant – s'expriment, s'engagent, se conforment à des règles ou bien s'en écartent, etc. Pour reprendre une image de Wittgenstein : soutenir le contraire serait comme penser qu'il suffirait que de l'argent passe d'une main à une autre pour qu'il y ait vente ou don. Or il est entendu que si je passe (par exemple) de l'argent de ma main droite à ma main gauche, il n'y a aucunement vente ou don ! On ne doit donc pas juger *sur pièce* (ce que j'ai appelé plus tôt un « vestige ») mais par le contexte.

Vous vous appuyez sur des distinctions aristotéliennes pour douter de la parole artificielle. Comment la philosophie antique permet-elle d'éclairer des technologies aussi contemporaines ?

Quand Aristote nous dit dans les *Topiques* qu'il y a quelque chose de problématique et de pernicieux à utiliser ce qui est à définir dans ce qui définit (l'exemple savoureux qu'il donne est la définition du Soleil comme l'astre qui parcourt le ciel *lorsqu'il fait jour*), cet avertissement traverse les siècles : nous continuons à penser, en dépit des avancées technologiques, qu'une définition ne saurait être bonne si elle utilise précisément dans le *definiens* le *definiendum*. Il y a donc des leçons qui ne sont pas soumises aux aléas des inventions et des dernières trouvailles. En l'espèce certaines distinctions conceptuelles que l'on trouve chez Aristote me semblent toujours opérantes pour structurer clairement les débats, pour éviter les confusions. Nous pouvons

donc à la fois penser qu'il y aurait quelque chose d'absurde à nous reposer sur la science et la technologie du temps d'Aristote pour régler nos problèmes scientifiques et technologiques contemporains *mais aussi* penser qu'il n'est pas inutile d'enrôler des percées aristotéliennes *quand elles sont conceptuelles* (puisque'il peut très bien se faire que nous soyons collectivement à ces égards plus confus qu'Aristote).



Figure 3 : Euphonia. The London Journal, [Wikimedia Commons](#)

En quoi la notion wittgensteinienne de « jeux de langage » peut-elle aider à penser les limites de l'idée selon laquelle une machine parlerait réellement ?

Un des aspects que l'on peut conserver dans la discussion de Wittgenstein des « jeux de langage » est l'idée – à mes yeux très importante – que le langage n'est pas la source d'une activité *singulière* ou *unique*. Il est tentant de penser, par exemple, que le langage – quand on l'utilise – repose au fond sur de la transmission d'information. C'est pourtant une tentation à laquelle il ne faut pas céder : nous pouvons faire tout un tas de choses par ou avec le langage qui ne tombent pas sous ce schéma (par exemple : remercier, faire une blague, saluer, promettre, fanfaronner, poser une question, etc.). L'image de la machine parlante tend à aplatir cette diversité pour la ramener (par exemple) à de la production de signes ou de marques. Il y a du reste tout un tas de conditions pour qu'un jeu soit réellement bel et bien joué, et parmi ces conditions, il y a des conditions ayant trait – ici aussi – au contexte (notamment social).

Votre travail sur la prétendue « intelligence » des IA vous a-t-il amenée à reconsidérer ou à redéfinir ce que nous appelons l'intelligence humaine ?

J'ai surtout retrouvé un embarras auquel je ne suis certainement pas la première à me confronter : nous n'avons pas – collectivement – les idées claires sur ce qu'il faut entendre par « intelligence », et il y a un risque (quand on essaie d'y voir plus clair) à faire de cette notion une notion purement quantitative sous le prétexte que nous pouvons parfois sans mal reconnaître que quelque chose (ou quelqu'un) est plus intelligent qu'autre chose (ou que quelqu'un d'autre). Cela ne veut pour autant pas dire que le premier possède un *niveau* d'intelligence plus élevé que le second (tout comme dire de quelqu'un qu'il est plus beau qu'un autre ne saurait passer par l'idée qu'il y aurait chez le premier un *quotient de beauté* plus élevé que chez le second). Les discussions sur l'intelligence *en général* sont nébuleuses, et les discussions sur l'intelligence *artificielle* héritent de cette obscurité.

L'imaginaire de la science-fiction semble structurer une grande partie de nos espoirs et de nos craintes face à l'IA. Constitue-t-il toujours un obstacle à la réflexion philosophique, ou peut-il au contraire l'enrichir ? Si cet imaginaire peut être fécond, quelles œuvres de fiction recommanderiez-vous pour un cours de HLP, notamment autour du thème « l'humain et ses limites » ?

Oui, il y a en effet un risque que nous soyons parfois en quelque sorte tenus en captivité par des images ou des histoires issues de la fiction. C'est souvent le cas avec les récits mobilisant l'IA. Ces images ou ces histoires peuvent bel et bien être un frein ou un obstacle pour philosopher sereinement car elles ne sont pas soumises (du moins pas avec le même niveau d'exigence et de systématisme) à la cohérence et à l'intelligibilité (je veux dire qu'il n'est pas forcément important pour une bonne histoire d'être conceptuellement douteuse). Elles peuvent certes aussi nous faire considérer des possibilités jusqu'ici inexplorées, mettre en scène de manière plus fine des difficultés auxquelles les philosophes ne sont pas toujours sensibles, etc. Il n'y a donc pas de verdict général à avoir ici : il faut juger au cas par cas. Je ne serai pas très originale, mais pour répondre à la dernière partie de cette question, voici quelques suggestions potentiellement utiles : *Frankenstein* de Mary Shelley ; *R.U.R.* de Karel Čapek. Plus récemment, côté cinéma : *Her* de Spike Jonze (2013) ; *Ex Machina* d'Alex Garland (2014).

Quel regard portez-vous sur l'usage des IA génératives en classe de philosophie ? N'est-ce qu'un « moulin à parole » qui risque d'appauvrir la pensée qu'il faut proscrire en toute circonstance, ou peut-on former les élèves à un usage critique et philosophiquement fécond des textes générés ?

J'ai tendance à penser qu'il faudrait en proscrire totalement l'usage et je peine à en trouver le moindre usage bénéfique dans un contexte pédagogique. On parle beaucoup aujourd'hui de « dette cognitive » (*cognitive debt*), « de psychose IA » (*AI Psychosis*) ou de « déqualification » (*deskilling*) et je pense qu'il y a en effet matière à s'inquiéter : si le corps enseignant décide d'abdiquer sur ce front, c'est la finalité même du métier qui est en péril. Comme vous le savez, les textes générés sont (pour reprendre l'image de Ted Chiang) comme une image

floue des textes d'entraînement : il y a donc tout intérêt à aller voir les textes d'origine eux-mêmes (dans leur précision) plutôt que d'en avoir une version forcément approximative (et immanquablement fautive). Pour le reste, apprendre à s'exprimer est une des capacités les plus importantes dans la vie humaine. La palette de ce que l'on vit est, à bien des égards, dépendante de ce que l'on peut dire, concevoir, appréhender. Si vous déléguez votre expression à quoi que ce soit, c'est une sorte de suicide mental. Maintenant qu'est-ce qui est important pour le développement des aptitudes philosophiques ? Sans grande originalité ici aussi, je dirais : la lecture et l'écriture. La lecture pour alimenter la pensée et l'écriture pour former une pensée : si on estime que ces deux activités sont superflues, on estime que la pensée elle-même est superflue.



Figure 4 : Henri Manuel. R. U. R. (Karel Čapek) - Comédie des Champs-Élysées, 26 mars 1924 [Wikimedia Commons](#)

Face à des interfaces qui donnent l'impression de nous comprendre, l'anthropomorphisme est un réflexe puissant. Quels outils conceptuels recommanderiez-vous aux enseignants pour aider les élèves à maintenir une distance critique vis-à-vis de la machine ?

De manière générale, le grand défi est de réussir à montrer que les fruits d'une réflexion *lente* sont toujours meilleurs que ceux d'un prêt à penser. C'est évidemment un grand défi pour des raisons bien connues (perte de l'attention, de la concentration, de la disposition à la lecture, etc.). Il me semble qu'une éducation au fonctionnement de telles interfaces peut sans doute aider pour éviter de tomber dans l'anthropomorphisme dont vous parlez : quand le magicien révèle ses secrets, nous pouvons apprécier son ingéniosité mais nous ne

sommes pas sidérés par ses pouvoirs magiques ! J'ai bien conscience que ce n'est ni suffisant, ni toujours faisable. Comme pour d'autres thématiques philosophiques, il s'agit d'essayer de susciter un questionnement réflexif pour faire tomber les certitudes non examinées des élèves : les rendre sensibles aux positions philosophiques alternatives, aux présupposés énormes logés dans cette assimilation, à l'idéologie et aux intérêts socio-économiques en jeu mais pas toujours manifestes dans ces programmes, etc. Je ne suis pas sûre que des nouveaux « outils conceptuels » puissent faire mieux que ce que Socrate faisait jadis : de la discussion argumentée, et honnête intellectuellement.

Systèmes d'IA : mésalignés ou non alignables ?

Conférence de Bertille Rouillon, doctorante

Lors du séminaire des IAN de philosophie du 10 mars 2026, Bertille Rouillon, doctorante à la Sorbonne, a présenté ses travaux en cours sur les problèmes et les risques générés par le mésalignement des intelligences artificielles. Cette visioconférence est désormais partagée sur PodEduc. On y trouvera une présentation claire des différents types d'intelligence artificielle et des difficultés épistémologiques que pose leur alignement. Le mésalignement désigne les situations dans lesquelles les systèmes optimisent des objectifs qui ne correspondent pas aux intentions humaines. En soulignant la diversité des systèmes et des finalités et en prenant appui sur des exemples concrets, comme la détection de fraudes, la reconnaissance d'images ou l'apprentissage par le jeu, cette réflexion ouvre des pistes stimulantes pour penser les conditions – peut-être incertaines – d'un alignement véritable.



Figure 5 : Amritha R Warrier & AI4Media / <https://betterimagesofai.org/> / <https://creativecommons.org/licenses/by/4.0/>

Une parole sans sujet ?

Les réactions humaines face à la parole machinique

Pour le philosophe, l'irruption dans l'histoire de machines qui parlent semble bousculer des certitudes : celles des Descartes et Rousseau selon lesquels le langage serait le propre de l'homme. Ce n'est que l'apparence : car c'est plus précisément « l'invention de l'art de communiquer nos idées »¹ qui nous singularise. Même si les machines *ne font que* parler sans penser, cela ouvre une ère nouvelle dans l'histoire des médias. Si « se demander si un ordinateur peut penser est aussi intéressant que de se demander si un sous-marin peut nager »², un fait est pourtant patent : les machines parlent. Et c'est ce que nous fait, à nous, une machine qui parle que nous devrions questionner³. Car ce qui est le plus étonnant, c'est que nous ne nous étonnons pas qu'une machine puisse parler⁴ !

Ce sont les signaux qui comptent. En toute rigueur, votre intériorité, je ne sais pas si elle existe... Elle se manifeste à moi par des signes extérieurs. Je ne sais pas ce que c'est que de ressentir une émotion chez un autre être humain. Je peux ressentir de l'empathie, mais je peux aussi en ressentir face à un acteur, qui simule. Je ne peux pas savoir si ces signes émanent d'une intériorité ou d'une simulation. C'est pourquoi pour ceux qui fabriquent des machines, si elles sont capables de simuler, ça suffira.

Jean-Marie Besnier⁵

Diderot nous a déjà fait observer⁶ que celui qui provoque le rire et les larmes n'a pas besoin de ressentir l'émotion. Considérons donc ce qui *nous* arrive lorsque nous conversons avec une machine. C'est bien le même Alan Turing qui a imaginé à la fois la *machine de Turing*⁷, et qui a vu dès 1950 qu'une machine pouvait peut-être passer pour humaine aux yeux d'un être humain⁸. Très tôt, on a d'ailleurs constaté qu'en discutant avec le

¹ Jean-Jacques Rousseau, *Essai sur l'origine des langues*, dans *Collection complète des œuvres de J. J. Rousseau*, t. 8, Genève, 1781.

² On doit cette formule à l'informaticien néerlandais Edsger Dijkstra lors d'une table ronde sur *Ordinateur et société* à la Burroughs Corporation en 1983.

³ L'intérêt suscité par cette question rappelle la légende urbaine selon laquelle un enfant se serait suicidé à la suite de la « mort » de son Tamagotchi, relayée par Carmagnat, Fanny & Robson, Elizabeth (1999), *Réseaux. Communication – Technologie – Société*, n° 92-93, pp. 343-364. Disponible sur Persée : https://www.persee.fr/doc/reso_0751-7971_1999_num_17_92_2126, consulté le 27/02/2026

⁴ Voir GIRAUD, Thibaut (monsieur Phi), *La Parole aux machines : Philosophie des grands modèles de langage*. Paris : Grasset, 2025.

⁵ Cité par TUAL, Morgane. Intelligence artificielle : une machine peut-elle ressentir de l'émotion ? Le Monde – Pixels, 12 octobre 2015. En ligne : https://www.lemonde.fr/pixels/article/2015/10/12/intelligence-artificielle-une-machine-peut-elle-ressentir-de-l-emotion_4787837_4408996.html (consulté le 27 février 2026).

⁶ Denis Diderot, *Le Paradoxe du comédien*, 1769

⁷ La *machine de Turing* est un modèle imaginé en 1936, pouvant calculer n'importe quel problème mathématique calculable.

⁸ Voir https://fr.wikipedia.org/wiki/Test_de_Turing

chatbot rudimentaire *ELIZA*,⁹ des personnes normales développaient des attachements émotionnels, voire des pensées délirantes. Et sans surprise, on a constaté que les robots anthropomorphes favorisent l'application d'une théorie de l'esprit¹⁰ qui leur prête des intentions.

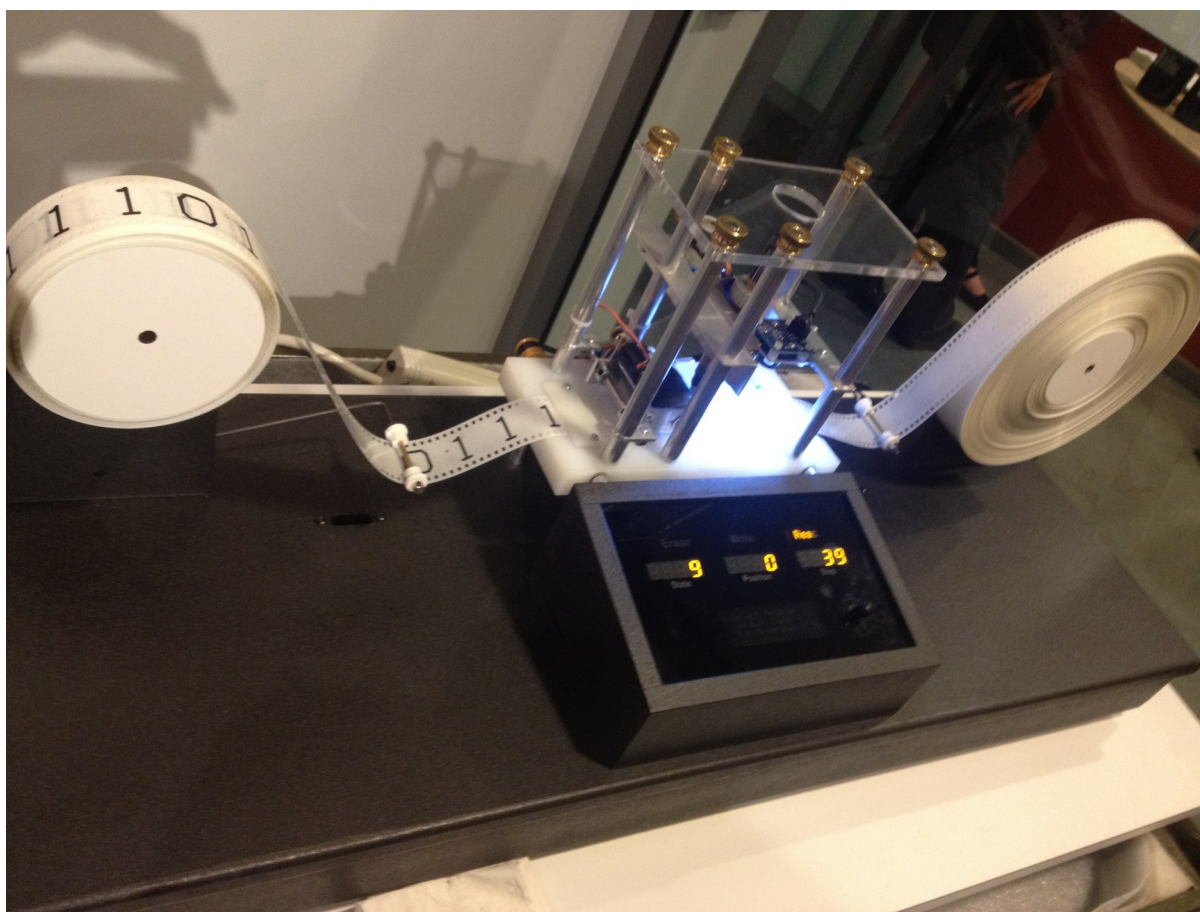


Figure 6 : Gabrielf, Modèles d'une machine de Turing. CC BY-SA 3.0 [Wikimedia Commons](#)

Nous sommes le problème

John Searle, avec l'argument de la *chambre chinoise*¹¹ nous a rappelé par avance, dès 1980, qu'il était inutile de supposer une véritable compréhension par la machine de ce qu'elle pourrait dire. Sans doute, mais la question est mal posée pour Dennett, si nous avons besoin de supposer une intention (celle de gagner chez le programme d'échecs contre lequel nous jouons) pour prédire son comportement, alors nous le faisons.

Le rasoir d'Ockham suffirait à écarter l'hypothèse d'un *fantôme dans la machine*. Nous savons comment un chatbot basé sur un LLM fonctionne. Spéculer sur l'émergence de compétences relève probablement de la plus haute fantaisie. Comprendre signifierait calculer des distances entre vecteurs. Le problème est ailleurs : il n'est pas

⁹ *ELIZA* est un chatbot écrit par Joseph Weizenbaum entre 1964 et 1966, qui simule un psychothérapeute rogéien en reformulant la plupart des affirmations du « patient » sous forme de questions.

¹⁰ [Théorie de l'esprit](#)

¹¹ Searle, J., « Esprit, cerveaux et programmes », in D. Hofstadter et D. Dennett (éd.), *Vues de l'esprit*, trad. J. Henry, InterÉditions, Paris, 1987, pp. 354-373. Cette expérience de pensée montre un homme dans une pièce qui manipule des symboles chinois selon des règles sans comprendre le chinois. Cela prouve pour Searle que la manipulation syntaxique ne suffit jamais à produire la sémantique. L'intentionnalité est intrinsèque, liée à la biologie du cerveau, et non « relative à l'observateur ».

dans le mécanisme, mais dans l'interaction, qui suppose de notre part une stratégie projetant effectivement une intention.

Dennett distingue trois niveaux de prévisibilité qui pourraient réclamer de nous cette projection.

1) On peut dire que la nature est un être *intentionnel* : pour le physicien, c'est simplement dire qu'elle obéit à des lois. Leibniz¹² a montré qu'une forme de finalité est à l'œuvre dans tout principe de conservation ; dès lors, mécanisme et finalisme ne devraient pas s'opposer violemment, mais reconnaître leur complémentarité.

2) On peut prédire le comportement d'un objet ou d'un logiciel si on comprend comment il a été fabriqué/programmé. On se met à la place du constructeur. Comme on cherchait autrefois un document dans une bibliothèque en se demandant sous quels mots clés le bibliothécaire l'avait rangé.

3) Il y a des cas où nous avons *besoin* de projeter des intentions sur l'objet pour interagir.

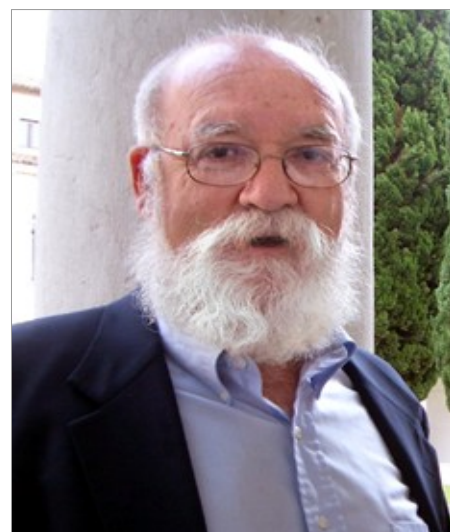


Figure 7 : Daniel Dennett en 2006, [Wikimedia Commons](#)



Figure 8 : Entraîner une figure à trois polygones à sauter par-dessus la balle. F. Elie, domaine public

Cette posture intentionnelle montre que *nous* sommes le problème. Car *de fait*, nous projetons des intentions.

Lors d'une conférence sur l'IA en 2017, François Elie, IAN philosophie de l'académie de Poitiers, montrait comment un réseau de neurones pilotant trois formes contiguës pouvait apprendre à sauter au-dessus d'une balle lui arrivant dessus, au bout de dizaines de tentatives infructueuses (voir figure ci-contre ainsi que [cette vidéo](#) et [ce projet](#)).

Les spectateurs ont fini par encourager la petite forme

géométrique, objet doté soudain d'une âme, et l'amphithéâtre entier a poussé un cri de triomphe lorsque trois malheureux triangles ont enfin réussi à bondir au-dessus d'un petit cercle en mouvement. Le dialogue avec une machine devient un miroir de notre propre fonctionnement cognitif et social.

Typologies des réactions humaines

L'effet Eliza

Nous attribuons automatiquement et de façon inconsciente des intentions, des émotions et de la compréhension à la machine. C'est le mécanisme le plus primitif, activé dès que la machine produit du langage. Cela déclenche en nous l'activation irréfléchie de scripts sociaux. Ce qui parle nous parle.

¹² Gottfried Wilhelm Leibniz, *Discours de métaphysique*, § 19-22.

```
Welcome to
          EEEEE LL    IIII ZZZZZZ AAAAA
          EE    LL    II     ZZ  AA  AA
          EEEEE LL    II     ZZ  AAAAAA
          EE    LL    II     ZZ  AA  AA
          EEEEE LLLLL IIII ZZZZZZ AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Figure 9 : Exemple de conversation avec ELIZA, [Wikimedia Commons](#)

La vallée de l'étrange et l'inquiétante étrangeté

Dans un article 1970, le roboticien japonais Mori Masahiro montre que plus le robot ressemble à un humain, plus nous nous y attachons, mais jusqu'à un certain point seulement, qu'il nomme « la vallée de l'étrange »¹³ : quand il ressemble trop à un humain il devient inquiétant ou repoussant. Au-delà de la robotique, ce sentiment d'étrange proximité décrit bien ce que l'on peut éprouver face aux « agents conversationnels ». L'inquiétante étrangeté est la traduction de l'*Unheimlichkeit* de Freud, situation d'inconfort dans la situation où l'intime surgit comme étranger, inconnu, autre absolu, au point d'en être effrayant. Thierry Hoquet applique ce terme aux situations dans lesquelles on ne sait pas si on est face à une personne humaine ou une créature mécanique.¹⁴

L'évaluation

Nous admirons les performances, cette réaction est souvent accompagnée d'une surestimation des capacités réelles. Ou bien nous adoptons la réaction inverse : le refus de nous étonner. Cela met en évidence dans les deux cas notre besoin de reconnaissance.

L'angoisse et la honte

Nous craignons la perte de contrôle, l'obsolescence humaine, notre remplacement par les machines. Certains ont peur que les avancées technologiques remettent en cause l'autonomie humaine et représentent une menace pour notre espèce. C'est le versant anxieux, des risques existentiels aux angoisses professionnelles. Cette forme étrange de technophobie projette une intention malveillante sur l'objet technique : un golem qui échapperait à son créateur. Il faut mettre au compte de notre nature une angoisse de fond devant ce que nous ne pouvons pas maîtriser.

¹³ Pierre Cassou-Noguès mobilise cette notion pour éclairer l'ambivalence de notre rapport aux machines dites intelligentes dans *La bienveillance des machines. Comment le numérique nous transforme à notre insu*. Paris: Seuil. 2022. p.226

¹⁴ Hoquet, T., *Les Presque-humains. Mutants, cyborg, robots, zombies... et nous*, Paris, Éditions du Seuil, 2021, pp. 15-17.

Ce sentiment est à rapprocher de ce que Günther Anders appelle la « honte prométhéenne d'être soi » de l'homme qui fabrique des objets plus performants que lui, « honte qui s'empare de l'homme devant l'humiliante qualité des choses qu'il a lui-même fabriquées. »¹⁵

L'attachement relationnel / la dépendance

Des utilisateurs décrivent des *chatbots* comme des amis, petits amis ou petites amies. Dans une enquête sur les utilisateurs de Replika, 88% des participants identifiaient leur *chatbot* comme leur « partenaire »¹⁶.

L'interlocuteur répond toujours, il est poli, il ne hausse pas le ton. Il ne faut pas être grand clerc pour supposer que l'IA va finir de conquérir rapidement les réseaux sociaux. Des bots vont déclencher des processus d'attachement très puissants, décuplant la manipulation des êtres humains. Umberto Eco voyait dans les réseaux sociaux *l'invasion des imbéciles*¹⁷. Ils risquent de vite devenir le *petit mur* dans la caverne de Platon.

L'éducation à l'épreuve des voix artificielles

En moyenne, les enfants et adolescents âgés de 6 à 17 ans passent plus de 4 heures par jour devant les écrans, selon les données de Santé publique France d'avril 2024¹⁸. Il n'est donc pas absurde d'imaginer que, demain, une part croissante de leur apprentissage passe aussi par des voix synthétiques. Ce ne serait ni la première ni la dernière transformation des médiations éducatives.

De la télévision, on a rêvé du meilleur. On se souvient des émissions choisies avec soin, diffusées par séquences dans les écoles des années 1960. Comme pour toute technique, les usages ont varié : le meilleur a côtoyé le moins exigeant. Il en ira sans doute de même pour l'intelligence artificielle, dont nous attendons beaucoup : elle pourrait favoriser un enseignement différencié, capable de repérer sans se lasser les erreurs et les progrès de chacun. Mais avant l'IA-professeur, viendra sans doute l'IA-nounou, qui apprendra à parler aux enfants, jouera avec eux, les apaisera, les occupera.

Une rupture anthropologique s'esquisse : des êtres humains feront très tôt l'expérience de la *voix humaine* à travers des machines. Qu'en sera-t-il de notre manière d'entrer en relation avec les autres ? Schopenhauer rappelait que l'être humain est « un animal sauvage et effroyable » que la civilisation apprivoise tant bien que mal¹⁹.

Nous entrons peut-être dans un âge où cet apprivoisement se fera en partie par procuration, par l'intermédiaire de systèmes artificiels. Les IA deviendront-elles les nouvelles gouvernantes, donnant à chaque enfant son

¹⁵ Anders G., « Sur la honte prométhéenne », *L'Obsolescence de l'homme. Sur l'âme à l'époque de la deuxième révolution industrielle*, traduit de l'allemand par Christophe David, Paris, Éditions de l'Encyclopédie des nuisances-Éditions Ivrea, 2002 (1956), p.38

¹⁶ Smith, M. G. (2025). « Can Generative AI Chatbots Emulate Human Connection ? » *Perspectives on Psychological Science*.

¹⁷ Nicoletti, Gianluca, « Umberto Eco: "Con i social parola a legioni di imbecilli" », *La Stampa*, 11 juin 2015 (Torino), en ligne : *La Stampa* (consulté le 22/02/2026)

¹⁸ Enfants et écrans : [Vie publique](#)

¹⁹ Schopenhauer Arthur, *Parerga und Paralipomena* (1851), dans Schopenhauer – *Éthique, Droit et Politique*, trad. A. Dietrich, Félix Alcan, 1909, p. 40, [éd. numérique Wikisource](#), consultée le 27 février 2026

récepteur infatigable, une sorte d'Émile numérique, toujours disponible ? Doit-on craindre que les IA produisent leurs complices, pour devenir l'espèce qui nous succédera²⁰ ?

Pour l'heure, les machines n'ont produit qu'une fraction infime des discours humains accumulés au fil de l'histoire²¹. Reste à savoir qui aura le dernier mot : comment faire en sorte que la parole humaine – faillible, incarnée, parfois conflictuelle – demeure centrale face aux logiques d'optimisation algorithmique ?

Pour aller plus loin

 Salles, Arleen, et al., « *Anthropomorphism in AI* », *Neuroethics*, 2020. Repris dans : *AI and Ethics*, 2024.

 Watson, David, « *The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence* », *Minds and Machines*, 2019.

 Damiano, Luisa, et Paul Dumouchel, « *Anthropomorphism in Human–Robot Interactions : A Multidimensional Conceptualization* », *Communication Theory*, 2023.

 Takahashi, Yuta, et al., « *Anthropomorphism-based Causal and Responsibility Attributions to Robots* », *Scientific Reports*, 2023.

 Damholdt, Malene, et al., « *Anthropomorphizing Machines : Reality or Popular Myth?* », *Minds and Machines*, 2024.

 Pataranutaporn, Pichayada, et al., « *Influencing Human–AI Interaction by Priming Beliefs about AI* », *Nature Machine Intelligence*, 2023.

²⁰ Marceau Felden annonce dès 1987 que nous produisons l'espèce suivante, dans *Le Songe de Minerve : le cerveau et les sciences de l'artificiel*, éd. Lieu commun

²¹ Sur 2000 ans, en tenant compte de l'explosion récente, les estimations de la production humaine tournent autour de **100 à 300** quadrillions **de mots** (10^{17} à 10^{18}). Rien que pour l'ère numérique récente, l'humanité produit environ 300 milliards de mots *par jour* (courriels, messages, publications, documents...). Les IA génératives à grande échelle, comme Claude (Anthropic), ChatGPT (OpenAI) et Gemini (Google) existent massivement depuis fin 2022. En 2024-2025, on estime que les principaux modèles génèrent collectivement quelques milliards à dizaines de milliards de mots par jour. Sur 2-3 ans d'existence grand public, cela donne un ordre de grandeur de **quelques dizaines de trillions de mots** (10^{13} à 10^{14}), soit environ **1000 fois moins** que le total humain cumulé. Mais la production des IA double en quelques mois). Si la tendance se maintient — avec de plus en plus d'agents autonomes, de génération de code, de contenus automatisés — certaines projections suggèrent que les IA pourraient rattraper le volume cumulé humain d'ici une à deux décennies, voire moins.

L'Académie de Reims propose deux fiches Édubase consacrées aux usages des IA génératives en classe de philosophie. La première présente une expérimentation menée en première et terminale HLP, fondée sur un dispositif de classe inversée : les élèves prennent connaissance du cours en amont, celui-ci étant reformulé par un modèle de langage selon différents formats (dialogue philosophique, portrait, récit ou conte), afin de proposer des approches variées d'un même contenu conceptuel. La seconde fiche propose des exercices visant à encourager un usage réflexif des agents conversationnels, non comme simples fournisseurs de réponses, mais comme outils d'appropriation du cours.

L'Académie de Toulouse a conçu deux séquences pédagogiques indexées dans Édubase : la « commission d'entente » et le « colloque des philosophes ». La première consiste à utiliser un texte produit par une IA comme support de critique et d'amélioration, afin de sensibiliser les élèves aux limites de ces outils et de prévenir les usages frauduleux. La seconde organise un colloque en classe autour de trois questions philosophiques majeures : une machine qui parle pense-t-elle ? L'intelligence humaine est-elle réductible au calcul ? Peut-on entretenir une relation d'amitié avec un robot ? Des textes et supports vidéo accompagnent ces activités.

Enfin, le site disciplinaire de l'Académie de Montpellier propose une introduction claire et accessible au fonctionnement des intelligences artificielles génératives, dans une perspective à la fois pédagogique et philosophique. Cet article, intitulé *Appréhender le fonctionnement des IA génératives pour mieux les penser* fournit des repères utiles pour développer l'esprit critique des élèves face aux usages de l'IA : il montre que ces systèmes ne « pensent » pas, mais reposent sur des mécanismes de calcul statistique fondés sur la prédiction de séquences linguistiques. Tout en rappelant l'absence d'intentionnalité et de conscience, invitant ainsi à distinguer rigoureusement calcul et pensée, il met en évidence l'illusion d'intelligence produite par ces outils, dont les performances discursives peuvent laisser croire à une compréhension effective.

**Lettre ÉduNum proposée par la direction du numérique pour l'éducation
Bureau de l'accompagnement des usages et de l'expérience utilisateur (DNE-TN3)**

[Contact courriel](#)

Vous recevez cette lettre car vous êtes abonné à la lettre ÉduNum Philosophie.
Souhaitez-vous continuer à recevoir la lettre ÉduNum Philosophie ?

[Abonnement/Désabonnement](#)

À tout moment, vous disposez d'un droit d'accès, de rectification et de suppression des données qui vous concernent (articles 15 et suivants du RGPD). Pour consulter nos mentions légales, [cliquez ici](#).

ISSN 2739-8935 (en ligne)