http://philosophie.ac-creteil.fr/spip.php?article963

Informations

- Actualités

-

Publication date: samedi 9 avril 2022

Sommaire

- Six petites leçons de pédagogie philosophique sur l'Intelligence (...)
- LEÇON N°1 : Les problèmes pédagogiques posés par les agents conversationnels en (...)
- LEÇON N°2 : Éléments pour une analyse technique et une critique des agents (...)
- LEÇON N°3 : L'évaluation des travaux des élèves en philosophie à l'épreuve des (...)
- LEÇON N°4 : De quel sens commun les agents conversationnels sont-ils capables (...)
- Leçon N°5 - De la Bêtise Artificielle
- Leçon N°6 : D'un usage pédagogique possible des agents conversationnels en (...)
- Des exercices par Maryse Emel IAN de philosophie et webmestre du site (...)
- I. REPRESENTATIONS
- II. Un exemple de dissertation
- III. SCIENCE et OPINION
- IV. Le chatGPT va-t-il prendre ma place ?
- V. RENCONTRER LE CHATBOT
- VI. La question du contrôle et de la domination. De la bêtise et de la (...)
  - La fascination pour les machines : attrait pour le merveilleux
  - Le langage du chatGPT

- BIBLIOGRAPHIE
- NOTES

Nouveau Site EMC

Publications récentes de collègues

[http://philosophie.ac-creteil.fr/index.php?action=image_responsive&img=sites/philosophie.ac-creteil.fr/IMG/jpg/e/2/6/f_lc83ns5b1_2.jpg&taille=100&1679636825]

Jean-Baptiste Le Bohec, professeur de philosophie agrégé au lycée François 1er, vient de publier la traduction(avec Frédéric Naudin, professeur agrégé d'anglais) et l'introduction d'un classique de la philosophie morale anglo-saxonne, « Le Langage de la morale », de Richard Hare par Mathieu Mulcey, qui dirige les éditions éliott. Lire plus...

[http://philosophie.ac-creteil.fr/index.php?action=image_responsive&img=sites/philosophie.ac-creteil.fr/IMG/png/8/3/2/c1_elogedutact_v05_ab_reduite.png&taille=160&1679637706]

Éloge du tact de Gilles Hanus, nov 2022 lire plus

QUESTION D'ACTUALITE

Six petites leçons de pédagogie philosophique sur l'Intelligence Artificielle

QUE FAIRE FACE A L'USAGE PAR DES ELEVES D'AGENTS CONVERSATIONNELS EN LIGNE POUR REDIGER LEURS COPIES DE PHILOSOPHIE ? -

Eric Le Coquil, I.A-I.P.R de philosophie des Académies de Créteil et d'Orléans-Tours

LECON N°1 : Les problèmes pédagogiques posés par les agents conversationnels en ligne dans un enseignement de philosophie

Depuis quelques semaines, la mise en ligne par l'entreprise OpenIA, à disposition du grand public, le 30 novembre 2022, du logiciel ChatGPT (Chat-Generative-Pretrained-Transformer) défraie la chronique. Exploitant les technologies d'Intelligence Artificielle avancées, cet outil numérique est entraîné à produire, sur requête et à partir de la synthèse informationnelle d'un très grand nombre de données récoltées sur l'Internet (antérieurement au 31 décembre 2021) des textes imitant parfaitement la qualité rédactionnelle d'un scripteur humain. Le 6 février 2023, l'entreprise Google annonçait à son tour le lancement prochain de son propre agent conversationnel en ligne, Bard, d'abord dans un cadre expérimental, puis en libre accès dans un second temps, ainsi que l'intégration à terme de ce logiciel à son moteur de recherche [1].

Dès la parution des premiers articles de presse faisant état de ces événements, des professeurs de philosophie ont fait part à leurs Inspecteurs d'Académie - Inspecteurs Pédagogiques Régionaux de leurs vives inquiétudes concernant l'usage que certains élèves des classes de Première et Terminale pourraient être tentés de faire de ce nouveau type d'outils numériques pour réaliser les travaux en temps libre qui leurs sont demandés dans le cadre des enseignements d'*Humanités, Littérature et Philosophie* et de *Philosophie* de tronc commun. D'autres collègues ont fait état des résultats des essais d'usage de ChatGPT qu'ils ont eux-mêmes réalisés, et de leur stupeur au vu de l'excellente qualité rédactionnelle de ses productions, jugées par certains supérieures à cet égard à celles dont est capable un nombre significatif de leurs élèves. D'autres professeurs de philosophie ont en outre d'ores et déjà signalé de premiers cas de devoirs « à la maison » rédigés par les élèves au moyen de ChatGPT, et fait part de leur embarras face à ce phénomène nouveau qui, étant donné les performances rédactionnelles de l'outil, leur pose un double problème : le problème du repérage des usages qui en sont faits par les élèves ; le problème des conséquences possibles, tout à la fois de ces usages et de la difficulté de leur repérage, sur l'évaluation et la notation des travaux des élèves.

Compte tenu de la nature, de la puissance de ces nouveaux outils numériques et des performances dont ils sont capables, l'irruption des agents conversationnels en ligne est vécue par certains professeurs de philosophie comme venant sérieusement aggraver le problème posé depuis environ vingt-cinq ans par les usages que font les élèves des ressources disponibles sur l'Internet pour la réalisation de leurs travaux en temps libre, et tout particulièrement par la pratique du « copier/coller » de corrigés de dissertation et d'explications de texte proposés notamment par des sites d'aide aux devoirs, gratuits ou payants. Bien plus, certains professeurs expriment le sentiment d'assister, avec l'arrivée des agents conversationnels en ligne, à un changement tout à la fois d'échelle et de nature de ce problème, craignant que les élèves ne disposent désormais, avec ces IA en libre accès, de moyens de faire écrire leurs devoirs dont les professeurs se seraient plus en mesure de déceler l'usage au vu de l'illusion presque parfaite que donnent leurs productions d'avoir été écrites par un rédacteur humain.

De telles craintes sont-elles ou non fondées en l'état actuel de la situation ? Faut-il redouter, avec les agents conversationnels en libre accès sur l'Internet, un changement d'échelle et de nature de l'usage par les élèves des ressources et outils numériques dans leurs copies, en particulier, s'il faut parler net, de leurs usages à des fins de tricherie ou de fraude ? Les professeurs de philosophie sont-ils démunis, dépourvus de réponses pédagogiques efficaces, de sorte qu'il leur faudrait considérer le combat comme perdu d'avance et s'en tenir, tout au plus, à des mesures d'interdiction et de répression, dont on peut soupçonner cependant qu'elles risqueraient de montrer bien vite leurs limites ?

Bien évidemment des philosophes auront beau jeu de dénoncer les agents conversationnels en ligne comme « une mascarade » (puisque, de toutes façons, en fait ils pensent pas) [2] ou de dénoncer leur « faillite épistémologique » (puisque, de toutes façons, en fait ils ne sont pas fiables) [3], ou encore de souligner qu'ils sont en réalité complètement idiots [4] - ce qui est sans doute absolument vrai, mais il ne suffit pas le constater ou de l'observer :

encore faut-il pouvoir, profondément, ce qui veut dire techniquement, l'expliquer. Et puis de quoi s'agit-il exactement ? Idiotie, naïveté, stupidité, bêtise : ces différents termes ne sont pas équivalents. Ces prises de position critiques ne manquent certes pas de légitimité : tout au plus de précision, sans doute parce qu'elles procèdent le plus souvent d'une attitude d'indignation, suscitée par des affects d'inquiétude et de colère - on le perçoit très nettement dans le ton employé - face à la puissance redoutable et aux effets catastrophiques possibles des agents conversationnels en ligne. Il n'est jamais facile, devant une innovation majeure et les risques qu'elle comporte, de conserver son sang-froid. Il n'empêche que ces outils existent, qu'ils sont diffusés, accessibles aux élèves et que ces derniers ne se privent pas de les utiliser. Les professeurs de philosophie, confrontés à ces usages très réels et très concrets, ne sauraient se contenter quant à eux de prises de positions de principe. La question décisive est plutôt : que faire ? Peut-on faire quelque chose, et si oui quoi ? Questions indissociablement philosophiques et pédagogiques.

Je soutiendrai, dans les six leçons qui suivent, qu'une approche pessimiste ou défaitiste du problème est à la fois injustifiée et inappropriée, que les professeurs de philosophie disposent de moyens pédagogiques bien ajustés et puissants pour faire face très efficacement aux conséquences sur leurs enseignements de l'apparition des IA en libre accès et de leur éventuel usage par des élèves dans le cadre de leur travail scolaire.

Pour le montrer, je tirerai tout d'abord de quelques éléments précis d'information, de connaissance et d'analyse touchant la nature et la puissance réelles de l'outil, un ensemble de réflexions relatives à ce qui constitue la véritable substance d'une copie de philosophie et par conséquent l'objet de son évaluation (leçon 2). De là, je tâcherai de déduire les conséquences qui s'ensuivent touchant le renouvellement supposé du problème de ce que les professeurs de philosophie interprètent et vivent dans leurs classes comme des conduites de tricherie de la part de certains élèves, conséquences dont je m'emploierai à évaluer la portée pédagogique à la lumière de quelques considérations relatives à l'histoire des institutions scolaires et de la pédagogie (leçon 3). A partir de cette reconfiguration du problème, je développerai quelques éléments de réflexion philosophique sur les rapports qu'entretiennent les agents conversationnels avec ce que la tradition philosophique a conceptualisé sous la dénomination de sens commun (leçon 4). De cette critique négative, formulée en termes de manque ou de défaut de sens commun, je tenterai de passer à une critique positive, formulée en terme d'excès, pour montrer que cet excès coïncide avec ce que l'on peut conceptualiser comme des formes de Bêtise Artificielle, comprise comme corrélat irréductible de toute Intelligence Artificielle (leçon 5). A la lumière de ces analyses, je formulerai un ensemble de recommandations et de propositions pédagogiques concrètes, qui permettront aux professeurs de philosophie de prendre en compte l'existence des agents conversationnels dans l'apprentissage et la pratique par les élèves des exercices écrits dans leurs enseignements de philosophie (leçon 6).

On trouvera enfin, dans le dernier article, un très riche ensemble de ressources numériques rassemblées par notre collègue Maryse Emel, IAN Philosophie de l'académie et Créteil et webmestre du site Académique de Philosophie de Créteil. La consultation et l'étude de ces ressources permettront aux collègues de philosophie qui souhaiteront s'y plonger, ainsi qu'aux élèves qui consulteraient notre site et voudraient s'y intéresser, de développer une réflexion approfondie relative aux problèmes et aux enjeux philosophiques soulevés par l'Intelligence Artificielle, ainsi que des propositions de mise en oeuvre pédagogique dans les cours et leçons dispensés en classe.

LECON N°2 : Éléments pour une analyse technique et une critique des agents conversationnels en ligne : leur nature, leurs possibilités, leurs limites

Il convient tout d'abord de se donner une compréhension précise et exacte de ce que sont les agents conversationnels en ligne, de leurs principes techniques de fonctionnement, partant de l'étendue et des limites de leurs possibilités. Les liens ci-dessous, dont la consultation est recommandée avant la lecture de la suite de cette

leçon, pourront y aider le lecteur.

1. Quelques liens utiles :

- Un collègue explique et analyse les principes techniques, les modalités de fonctionnement, les possibilités et les limites de ChatGPT3 : https://www.youtube.com/watch?v=R2fjRbc9Sa0

- Des collègues professeurs de philosophie corrigent des copies produites par ChatGPT : https://etudiant.lefigaro.fr/article/c-est-loin-de-faire-un-devoir-de-qualite-un-prof-corrige-le-bac-dephilo-de-chatgpt_a3cca6d4-8ddd-11ed-a86d-66d30ea5ec0a/ ; https://www.bfmtv.com/tech/on-a-montre-le-travail-de-chat-gpt-a-une-prof-de-philo-voici-son-constat_AV-202301130473.html

- Pour une réflexion philosophique sur l'intelligence artificielle, sa puissance réelle en son état actuel de développement technique, ses limites et ses véritables dangers (qui ne se situent certainement pas là où on le penserait) on peut écouter ce podcast récent, Avec Philosophie, sur France Culture : https://www.radiofrance.fr/franceculture/podcasts/avec-philosophie/l-intelligence-artificielle-estelle-vraiment-intelligente-2059509

- Les créateurs de l'I.A lancent la chasse aux tricheurs : https://www.liberation.fr/economie/economie-numerique/chatgpt-tu-peux-faire-mes-devoirs-les-createurs-de-lia-lancent-la-chasse-aux-tricheurs-20230203_RAQA55N54NFZZHEKSXKS6W5UVE/?at_campaign=NL_Lib%C3%A9_Matin_Samedi&at_email_type=acquisition&at_medium=email&at_creation=NL_Libe_Matin_samedi_2023-02-05&actId=ebwp0YMB8s1_OGEGSsDRkNUcvuQDVN7a57ET3fWtrS841mguA0zYUvG6YFUZoN3h&actCampaignType=CAMPAIGN_MAIL&actSource=522604

- Une vidéo dans laquelle un Data-Scientist explique de façon très claire les bases théoriques (mathématiques) du deeplearning, qui est au principe de l'Intelligence Artificielle, ainsi que leur développement historique : https://www.youtube.com/watch?v=XUFLq6dKQok

- Un article de Noam Chomsky publié le 8 mars 2023 dans le *New York Times* : « La fausse promesse de ChatGPT » : https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

2. Qualités d'expression ou qualités de réflexion : selon quels critères évaluer une copie de "philosophie" produite par un chatbot ?

Il serait facile, à première lecture des productions « philosophiques » de ChatGPT3, de céder à un animisme ou à un anthropomorphisme irréfléchis, portant à surestimer ses possibilités et ses performances réelles. Comme le montre très bien cette vidéo, il est pourtant parfaitement clair, au regard d'une analyse technologique précise et bien informée, qu'un dispositif numérique tel que ChatGPT ne pense pas, ne réfléchit pas philosophiquement. Il fonctionne selon les principes d'un chatbot ou agent conversationnel, fondé sur un modèle de langage, une description statistique qui modélise la distribution de séquences de mots dans une langue naturelle : il est ainsi capable de prédire des chaînes langagières, selon le degré de probabilité de succession des termes et syntagmes dans le langage ordinaire, dans un contexte rédactionnel et informationnel prescrit (la commande qui lui est adressée) sur la base de quantités très importantes de données, dans les limites posées par des dispositifs de méta-contrôle visant à essayer de « moraliser » les productions de l'outil, mais que l'on peut assez facilement contourner ou prendre en défaut. Un agent conversationnel n'est pas un sujet.

On ne peut donc s'effrayer vraiment de l'irruption des agents conversationnels en ligne que si l'on présuppose la réductibilité de la pensée à des constructions langagières.

Au contraire, si l'on est vraiment attentif à la distinction entre construction langagière et construction de pensée, on doit pouvoir toujours faire la différence entre un texte qui imite simplement un langage philosophique, donc assez convenu sur le fond, et un texte qui réalise un véritable travail de pensée. Si des élèves utilisent une IA de prédiction langagière, cela ne leur garantit nullement de produire un bon devoir de philosophie, ni même un devoir simplement correct sur le fond, et c'est tout ce qu'il nous faut considérer.

Une copie très bien écrite peut être entièrement vide de philosophie. Tous les professeurs de philosophie le savent d'expérience. Une dissertation philosophique ne saurait être corrigée et évaluée exclusivement ni même principalement du point de vue de ses qualités d'expression et d'argumentation : ce serait là justement agir comme si elle pouvait n'être que la production d'un agent conversationnel. Des professeurs de philosophie tentés d'attribuer une excellente note à une copie produite par un agent conversationnel devraient par conséquent interroger très sérieusement la nature et la validité des critères d'évaluation qu'ils appliquent en réalité aux copies qu'ils corrigent. Ce serait en effet assurément le signe que leurs exigences intellectuelles et philosophiques pour la formation de leurs élèves ne seraient pas, dans cette hypothèse, suffisamment élevées, ou à tout le moins qu'elles ne seraient pas correctement ajustées.

A l'inverse, si une bonne dissertation de philosophie est précisément celle qu'en l'état actuel de la technologie une IA n'aurait pas pu écrire, alors les professeurs de philosophie disposent, avec cette observation, d'un critère ou d'un ensemble de critères d'évaluation possibles des productions de leurs élèves. Je reviendrai sur ce point dans la suite de mon propos.

Tout l'enjeu est donc pour le correcteur de se mettre toujours en mesure de distinguer qualité de l'expression et qualité de l'articulation des idées, de façon à ne jamais tomber dans l'écueil qui consisterait à ne corriger, à n'évaluer et à ne noter les copies qu'en fonction de critères relevant exclusivement de leurs qualités formelles, tant sur le plan de l'expression française que sur celui de l'argumentation.

3. Le chatbot est-il dialecticien ?

L'accès donné aux élèves à des agents conversationnels en ligne réactive ainsi la question de savoir ce qu'il y a lieu d'évaluer dans une copie de philosophie. L'agent conversationnel est capable de produire des textes très bien rédigés : fort bien. Toutefois la dissertation philosophique ne se réduit en aucun cas à un exercice d'expression écrite ou de rhétorique. Si la qualité de l'expression écrite est expressément requise, la qualité intrinsèquement philosophique de la copie ne s'y réduit pas. La réflexion philosophique est en effet dans son essence articulation des idées, soit que cette articulation produise entre elles des rapports de tension, dont la mise en évidence constitue la position des problèmes philosophiques, soit qu'elle permette, par un travail de désarticulation qui a nom analyse, et de ré-articulation qui a nom synthèse, des effets de résolution de ces tensions idéelles, mouvement dont le concept de dialectique, dans sa version platonicienne, donne l'une des définitions et des interprétations possibles des principes méthodiques : art "des divisions et des rassemblements" (Äö½ ´¹±¹ÁsÃµÉ½ º±v ÃÅ½±³É³ö½) [5]. C'est par la seule maîtrise de cet art - quelque interprétation qu'on en donne, platonicienne ou autre - que penser philosophiquement et parler - ou écrire - peuvent authentiquement coïncider : "j'y vois le moyen d'apprendre à parler et à penser" (5½± ¿7yÂ Äµ f »s³µ¹½ Äµ º±v ÆÁ¿½µÖ½ ; ibid.) ; le connecteur logique de conjonction "et" (º±v) doit être ici compris en son sens le plus fort, en ce qu'il enveloppe tout un monde de relations logiques et spéculatives possibles.

L'enjeu de notre affaire est donc de savoir si un mouvement authentiquement dialectique - il faut bien, à la fin, appeler les choses par leur nom - habite ou non les productions du chatbot. Quelques essais montrent très rapidement que l'agent conversationnel s'avère totalement incapable, lorsqu'on lui soumet un intitulé de sujet de dissertation à deux notions, d'en construire la moindre articulation, mais qu'il se contente de juxtaposer des considérations relatives successivement à l'une et à l'autre, sans jamais produire de véritable mise en rapport. De

même on observe, lorsqu'on demande au chatbot de produire successivement les différentes parties d'une dissertation (la commande formulée étant, pour chacune, de développer une argumentation ou au contraire une critique exemplifiées et instrumentées par des références à des auteurs) que l'outil tantôt présente de simples opinions comme s'il s'agissait de connaissances avérées, tantôt juxtapose des thèses philosophiques ou des connaissances comme s'il s'agissait de simples opinions. Or une dissertation de philosophie ne consiste justement jamais à présenter simplement de façon correctement rédigée une succession de thèses ou d'opinions répondant à un sujet donné. C'est que les divisons et les rassemblements dont il s'agit en tout mouvement de pensée dialectique ne sont jamais des opérations purement formelles : ces divisions et ces rassemblements n'ont rien d'artificiel mais constituent les actes d'une intelligence philosophique vivante, raison pour laquelle sans doute une IA rencontre les plus grandes difficultés à les imiter.

Le fait que des élèves puissent envisager d'utiliser un agent conversationnel en ligne pour écrire leurs devoirs à leur place témoigne ainsi d'un contresens radical concernant ce qui se joue véritablement pour eux dans l'enseignement de philosophie qu'ils suivent. Et ce qui s'y joue, ce n'est pas la production, à destination du professeur, d'un savoir philosophique extériorisé, qui ne serait qu'un savoir mort. Tout au contraire, c'est la possibilité pour eux de mettre à l'épreuve leur propre pensée, de la constituer véritablement comme la leur, une pensée en première personne : c'est l'institution même de l'élève comme sujet véritablement pensant, l'auto-institution du sujet philosophant par lui-même, et avec lui de la pensée comme exercice de la liberté. C'est en quoi la dissertation et l'explication de texte (l'essai et la question d'interprétation pour l'enseignement de spécialité HLP) sont des exercices, au sens le plus fort du terme.

Les exercices proposés dans un enseignement de philosophie ne peuvent être que des exercices intellectuel : ils doivent nécessairement donner aux élèves l'occasion d'éprouver sérieusement leur pensée et d'exercer leur jugement, et c'est pourquoi tout exercice de simple mémoire y est à proscrire. : l'exercice de toutes les facultés sans exception, fût-ce celle de mémoire, ne peut être philosophique que s'il offre directement aux élèves une occasion de mettre à l'épreuve leurs propres idées. C'est pourquoi aussi, dans l'enseignement de philosophie, tous les exercices intellectuels sont toujours simultanément en quelque façon des exercices spirituels, des tentatives de cheminer vers une façon de penser qui soit plus exacte ou plus juste, cheminement qui enveloppe une transformation de la pensée et partant de soi, que Platon métaphorise - on le verra - comme un certain genre de translation, ou plutôt de rotation [6]. Sur ce plan, le chatbot, si beau parleur soit-il, manifeste immédiatement, si on en lit très attentivement les productions, une incapacité totale.

4. Exercice(s) de la pensée philosophique et limites du chatbot

Le savoir philosophique ne saurait en effet être conçu comme une substance préexistante que l'on pourrait produire extérieurement au sujet et qu'il s'agirait simplement de faire entrer dans son esprit, comme l'on verse un liquide d'un vase dans un autre, ni comme une faculté dont il s'agirait, de l'extérieur, de doter un organe qui en serait dépourvu. Il est tout au contraire l'exercice vivant d'une faculté qu'un chacun possède, celle de réfléchir ou de juger, qui a pour nom la raison, à laquelle il s'agit seulement de donner l'occasion de s'apprendre à elle-même comment se conduire et s'orienter dans de meilleures directions. Platon, au livre VII de la *République*, [7]l'avait déjà souligné dans ce célèbre passage :

"(...) la science ne s'apprend pas de la manière dont certaines gens le prétendent. Ils se vantent de pouvoir la faire entrer dans (518c) l'âme où elle n'est point, à peu près comme on donnerait la vue à des yeux aveugles.

Tel est leur langage.

Ce que nous avons dit suppose au contraire que chacun possède la faculté d'apprendre, un organe de la science ; et que, semblable à des yeux qui ne pourraient se tourner des ténèbres vers la lumière qu'avec le corps tout entier, l'organe de l'intelligence doit se tourner, avec l'âme tout entière, de la vue de ce qui naît vers la contemplation de ce qui est et de ce qu'il y a de plus lumineux dans l'être ; et cela nous l'avons appelé (518d) le bien, n'est-ce pas ?

Oui.

Tout l'art consiste donc à chercher la manière la plus aisée et la plus avantageuse dont l'âme puisse exécuter l'évolution qu'elle doit faire : il ne s'agit pas de lui donner la faculté de voir ; elle l'a déjà : mais son organe n'est pas dans une bonne direction, il ne regarde point où il faudrait : c'est ce qu'il s'agit de corriger.

En effet."

Platon ici ne conceptualise pas explicitement l'apprentissage de la philosophie comme auto-institution du sujet philosophant. Il l'analyse comme la conversion d'un âme, à laquelle il incombe de tourner son regard vers le vrai et le bien pour les contempler, et l'on sait par ailleurs, si l'on se reporte aux lignes qui précèdent ce passage dans le livre VII de la *République*, que cette conversion ne se fait pas sans une certaine brutalité et sans une certaine souffrance : elle suppose en effet la rencontre d'un agent provocateur, d'un Socrate qui, par ses questions retorses, par l'adversité qu'il oppose à la pensée facile de l'opinion et des croyances ordinaires, force cette âme à se libérer de ses chaînes et à confronter son regard ébloui par une lumière à laquelle elle n'est pas encore accoutumée à la considération des principes des choses. La rencontre d'un Socrate. Ou bien la rencontre d'un professeur de philosophie. Il n'en reste pas moins - et c'est le point - que c'est à l'âme qu'il revient d'effectuer elle-même ce mouvement de conversion : c'est par conséquent dans ce mouvement même que réside tout à la fois l'apprentissage philosophique et l'acte même de philosopher et c'est en quoi il marque une forme d'institution de l'âme dans le mouvement de la réflexion philosophique : elle l'y engage durablement et tout, pour elle, est désormais à jamais changé. Or un tel mouvement, par définition, ne saurait se déléguer ni s'externaliser : il est toujours le mouvement propre de chaque âme, mouvement qu'elle effectue par elle-même et que personne ne peut effectuer à sa place, pas même l'habile questionneur qui l'a provoqué, en quoi l'élève demeure toujours et à chaque instant de cette aventure un esprit libre. Ou plutôt il le devient, par un mouvement qui, peut-être ne trouve pas son origine ou son impulsion en lui-même, mais qu'il ne peut en tout état de cause effectuer que lui-même. Et c'est en quoi on peut ici légitimement parler malgré tout d'une auto-institution.

A bien suivre Platon, nous comprenons qu'un organe d'intelligence qui serait artificiel - ce que Platon lui-même, bien entendu, ne pouvait envisager - c'est-à-dire une intelligence qui ne serait pas celle d'une âme, n'aurait la faculté de se tourner vers rien : ce que saisirait un tel organe, tout formellement, il n'y aurait personne pour le connaître et pour en être transformé.

Il y a véritablement lieu d'expliquer cela aux élèves.

Si au contraire l'exercice de la philosophie est bien de cet ordre, alors à supposer même que des agents conversationnels en ligne encore plus perfectionnés soient capables un jour de produire des textes auxquels on serait un peu plus tentés de prêter le nom de dissertations de philosophie, leur activité et leurs productions demeureraient non substituables aux effets d'institution de soi produits par l'activité de réflexion d'un sujet philosophant en première personne, en tant que celle-ci l'institue précisément comme sujet pensant et libre : ces produits, désertés de tout esprit, seraient par conséquent philosophiquement stériles. L'existence d'Intelligences Artificielles capables de produire des textes philosophiques ne saurait dispenser quiconque d'écrire de la philosophie par lui-même et pour son propre compte, ni d'avoir à apprendre à le faire. Aussi un professeur de philosophie qui se croirait destitué par l'Intelligence Artificielle et par l'usage que les élèves pourraient faire dans leurs copies des outils qui en dérivent n'aurait-il strictement rien compris à la véritable nature de sa mission.

C'est pourquoi une copie entièrement rédigée par un agent conversationnel, pleine de savoir mort mais vide du mouvement vivant de la pensée qui change en l'instituant le sujet qui s'y engage, mériterait véritablement d'être notée d'un 0/20 si l'institution, posant en principe, dans sa bienveillance, la capacité des élèves à comprendre et à changer leur conduite, n'appelait chaque professeur à leur offrir une seconde chance : celle de refaire le devoir, afin de pouvoir se réinstituer eux-mêmes dans le mouvement vivant de la philosophie auquel ils avaient cru à tort pouvoir se soustraire.

LECON N°3 : L'évaluation des travaux des élèves en philosophie à l'épreuve des agents conversationnels en ligne : repenser le problème de la tricherie et l'articulation entre

devoirs surveillés et devoirs en temps libre (« à la maison »)

Après l'irruption de l'Internet dans le champ pédagogique à la fin des années 1990 et le développement du phénomène du « copier/coller » de corrigés de dissertations de philosophie dans les copies des élèves, l'avènement des IA en libre accès renouvelle les termes d'un vieux problème, plus général : celui de la tricherie et de ses moyens. Ce problème n'a, de fait, en lui-même rien de nouveau : seule l'est la puissance inégalée des moyens mis aujourd'hui à la disposition d'élèves qui auraient l'idée les utiliser pour tricher. Cependant ce changement d'échelle et de degré de puissance doit conduire la discipline Philosophie à repenser à nouveaux frais le cadre pédagogique dans lequel elle inscrit l'apprentissage de la dissertation philosophique par les élèves.

1. Soupçons de tricherie et principe de charité

Le principe de charité commande cependant de rappeler et de souligner tout d'abord avec force que le recours aux ressources disponibles sur l'Internet, y compris sur le mode du plagiat, ne procède pas nécessairement chez les élèves d'une volonté malveillante de frauder ou de se dispenser à bon compte du travail qui leur est prescrit. Bien souvent il procède tout aussi bien et bien plutôt d'un désarroi face à la nouveauté des exercices de dissertation et d'explication de texte, d'un vertige devant une tâche vécue à tort comme insurmontable. A ce désarroi s'ajoutent bien souvent le désir sincère des élèves de se montrer à la hauteur des attentes des professeurs, dont ils n'identifient pas forcément de façon claire et juste la nature exacte, malgré les explications qui ont pu leur être effectivement données, notamment parce que leur parcours scolaire antérieur les porte à faire interférer les attentes et les méthodes d'autres exercices pratiqués dans d'autres discipline - c'est là bien souvent ce qu'il y a vraiment lieu d'appeler un verrou pédagogique, ou, pour mieux dire, un obstacle épistémologique - et la crainte tout aussi sincère de ne pas y parvenir.

Le souci d'aider véritablement les élèves dans leur apprentissage de la dissertation philosophique, apprentissage barré par ces représentations préconçues, impose bien souvent aux professeurs le travail préalable de leur déconstruction, raison pour laquelle Jean-Louis Poirier, Doyen honoraire du Groupe de Philosophie de l'Inspection Générale, écrivit un jour : "ce n'est pas que les élèves n'ont pas de méthode : ils en ont une mais elle est mauvaise". Parmi ces préconceptions figure, il faut bien le dire, une certaine manière de se rapporter aux ressources numériques qui, dans bien des cas, n'a jamais fait l'objet d'une véritable prise en charge pédagogique dans le parcours scolaire antérieur des élèves, quand elle n'a pas été délibérément entretenue, par négligence ou par manque de réflexion : celle qui consiste à valoriser sous l'appellation de "faire des recherches sur l'Internet" la pratique du simple copier/coller de données. Réduisant partant la vraie notion de recherche, qui appelle une reconfiguration critique des contenus trouvés et leur exploitation dans le développement d'une réflexion véritablement personnelle, à la simple collecte d'informations brutes, cette façon de faire fixe en habitudes des pratiques fondées sur la confusion de la connaissance avec la simple information. Classes inversées, qui postulent que des élèves ou étudiants pourraient sans dommages recevoir un cours sur le mode de la simple "prise d'informations" mais dont on observe qu'elles ne font au bout du compte réussir que ceux qui le pouvaient d'emblée par des méthodes plus classiques sans faire progresser nettement les autres, du fait que leur procédé présuppose chez les élèves et les étudiants une autonomie dont beaucoup ne disposent pas encore et qu'il s'agirait plutôt de leur faire acquérir progressivement ; pédagogies par compétences, qui soumettent les élèves à des "tâches complexes" supposant, pour être réalisées, la recherche et la mobilisation d'éléments de "connaissances" bien souvent conçues et traitées comme de simples éléments d'information, comme si une connaissance pouvait constituer un élément discret immédiatement disponible indépendamment de tout travail d'élaboration intellectuelle préalable : il faudra un jour établir avec lucidité les responsabilités respectives de certains paradigmes pédagogiques à la mode, auxquels les élèves sont soumis dans bien des disciplines, et de leurs mauvais usages, dans le développement des déviances observées dans le rapport que les élèves entretiennent à l'information, à la connaissance et au travail intellectuel. Tout cela explique parfaitement la stupéfaction des élèves lorsque leurs professeurs de philosophie leur reprochent les reprises brutes tirées de sites Internet qu'ils ont trouvées dans leurs copies, pratiques auparavant encouragées et

valorisées inconsidérément par d'autres depuis leur plus tendre enfance. Il y a donc malheureusement tout lieu de penser que leur surprise est véritablement sincère. Mais à qui la faute ?

C'est un vertige pour beaucoup d'élèves que d'avoir à écrire leur première dissertation de philosophie. Les professeurs de philosophie, qui furent tous ou presque tous élèves, gagneraient beaucoup à se remémorer avec lucidité ce qui fut également un beau jour leur première expérience. La maladresse qui bien souvent dénonce le recours inapproprié à des ressources disponibles sur l'Internet révèle la très grande naïveté des usages spontanés qui sont faits de ces ressources, l'insuffisance ou l'absence d'une véritable éducation à leurs usages pertinents et légitimes dans le parcours scolaire antérieur des élèves, ainsi qu'aux manières à la fois intelligentes, créatives et personnelles dont ces derniers pourraient les exploiter.

Ces constats appellent incontestablement un accompagnement attentif et renforcé de l'apprentissage des deux exercices canoniques de la discipline philosophie, apprentissage qui gagnerait beaucoup à prendre en compte et à intégrer radicalement le facteur numérique, mais aussi à viser corrélativement un apprentissage des usages légitimes des ressources disponibles, qu'il n'est pas interdit au professeur de sélectionner lui-même et de prescrire à ses élèves.

Ces considérations ne doivent pas cependant conduire à dédouaner tous les élèves, d'avance et par principe, de leurs éventuels comportements de tricherie en niant, par l'effet d'une bienveillance excessive, déplacée et partant coupable, la réalité et le caractère délibéré de ces comportements, qui méritent incontestablement sanction. Pour autant, il faudrait bien reconnaître dans le même temps que ces comportements de tricherie témoignent tout aussi bien de la très grande perspicacité de leurs auteurs quant à la véritable nature des outils issus de l'Intelligence Artificielle et à ce qui se joue avec eux en réalité. Comme le fait très justement remarquer Jean-Gabriel Ganascia : [8]

> « (...) à bien y réfléchir, les termes sont tout à fait appropriés : au sens étymologique, l'intelligence artificielle est bien un »artifice« , c'est-à-dire un art consommé qui fait illusion en produisant des leurres fabriqués tout exprès pour nous tromper, en laissant croire que les machines seraient effectivement intelligentes. »

Rappelons que le fameux « test de Turing », élaboré par le mathématicien et cryptologue anglais Alan Turing (1912-1954), qui fut, comme on sait, un pionnier dans l'élaboration des modèles théoriques de l'informatique et de l'Intelligence Artificielle, dans son célèbre article de 1950 intitulé *Computing machinery an intelligence* [9], prend pour critère de reconnaissance de l'intelligence supposée d'une machine sa capacité à tromper l'interrogateur dans un « jeu d'imitation » à trois participants dans lequel l'interrogateur C pose des questions à deux autres personnes anonymes, A et B, qu'il ne voit pas et dont il n'entend pas directement la voix, dans le but de déterminer qui de A ou de B est un homme ou une femme, sachant que A est en réalité un ordinateur essayant de se faire passer pour une femme. Selon Turing un ordinateur pourrait ainsi être qualifié d'intelligent s'il se montrait capable de tromper l'interrogateur aussi longtemps que l'aurait fait l'homme [10]. Or comme le fait encore remarquer J.-G. Ganascia, les agents conversationnels « jouent au jeu de l'imitation et leurs performances ressemblent grandement à celles prévues par Turing » [11]. Dans son principe, le projet de l'Intelligence Artificielle « vise la simulation », au moyen de machines « des facultés supérieures de l'esprit humain » [12], « comme la capacité de parler, de lire, de comprendre, de calculer, de raisonner etc. » [13] et même désormais de percevoir et d'inventer. La capacité à simuler à des fins de tromperie est donc historiquement et théoriquement central dans la définition même de l'idée d'Intelligence Artificielle.

Un élève qui utiliserait un agent conversationnel pour rédiger sa dissertation en faisant croire qu'il en est lui-même l'auteur, ne ferait d'une certaine façon que reproduire le principe logique du test de Turing, ramenant ainsi l'outil à la

positivité de sa véritable nature et à la réalité foncière de sa principale opération : tromper son monde. La tricherie, fût-elle astucieuse, n'honore jamais son auteur : elle demeure en tout état de cause inexcusable et condamnable. Simplement elle dit, en l'espèce, quelque chose de la vérité de l'Intelligence Artificielle.

2. Les agents conversationnels et la question du plagiat

Il convient en second lieu d'analyser la façon dont l'usage d'agents conversationnels par les élèves vient sérieusement compliquer l'emploi, pour réprimer les faits de tricherie, de la catégorie juridique de plagiat. Est un plagiat l'acte de copier un auteur en s'attribuant indûment des passages de son oeuvre.

Bien des professeurs de philosophie trouvent dans la catégorie juridique de plagiat un point d'appui pour réprimer la pratique du « copier/coller » dans les copies des élèves ainsi qu'un argument dissuasif consistant à souligner que le recopiage de ressources en ligne sans citation de l'auteur ou de la source est assimilable à un délit au regard du droit de la propriété intellectuelle. Il est incontestable que cet argument bénéficie d'une consistance juridique au regard de la réglementation relative aux procédures de répression de la fraude aux examens et qu'il a montré jusqu'ici dans la pratique une réelle efficacité touchant le recopiage de ressources disponibles sur l'Internet. Toutefois, l'arrivée des agents conversationnels en ligne vient tout à la fois renforcer le problème du plagiat lui-même, et en même temps compliquer sa répression.

2.1. Le rapport ambigu des philosophes à leurs sources et les vertus d'apprentissage de la reprise

Le plagiat constitue une catégorie juridique, non une catégorie philosophique et l'on peut sérieusement se demander s'il peut réellement constituer une catégorie pédagogique. Il est en effet incontestable que la copie ou le recopiage des grandes oeuvres a, dans bien des domaines, une authentique puissance pédagogique et des vertus avérées pour l'apprentissage. Pour apprendre leur métier, les élèves des écoles de Beaux-Arts vont au musée copier les oeuvres des grands maîtres : nul ne songerait à les accuser de plagiat, même s'ils montrent à leurs maîtres les copies qu'ils ont réalisées. De même les élèves des conservatoires apprennent l'harmonie et le contrepoint dans des exercices contraints dans lesquels ils doivent écrire extraits de quatuors à cordes, chorals ou doubles choeurs dans le style de tel ou tel grand compositeur, ce qui les conduit la plupart du temps à leur emprunter leurs formules harmoniques ou mélodiques. Pour un compositeur des âges baroque et classique, être plagié ce n'était pas être pillé : c'était la consécration et la gloire. Bach retravaillait volontiers dans ses propres *concerti* des emprunts à ceux de Teleman (comparer par exemple le Larghetto du Concerto pour clavecin en fa mineur BWV 1056 du premier et celui du Concerto pour flûte en sol majeur TWV51:G2 du second) ou de Vivaldi et recyclait volontiers des pièces , les premiers *concerti* de Mozart pour piano ne sont bien souvent que des sonates de ses aînés, brillamment orchestrées et un peu aménagées, l'air du catalogue du *Don Giovanni* apparaît clairement comme une réécriture de l'air du catalogue que l'on trouve dans le *Don Giovanni* de son contemporain le compositeur vénitien Guiseppe Gazzaniga (1743-1818), dont l'ouvrage fut créé au Teatro San Moise huit mois avant la première de la version de Mozart et Da Ponte [14].

Du côté de la philosophie, n'oublions pas l'étrange ressemblance entre l'article *Autorité politique* de l'*Encyclopédie* de Diderot et d'Alembert et le second chapitre du livre I du *Contrat social* de Rousseau, ou les pages des *Manuscrits de 1844* dans lesquelles Marx recopie sans sourciller et sans citer ses sources des paragraphes entiers des grandes oeuvres de l'économie politique. Qui oserait prétendre que ces emprunts furent stériles ? Allez faire, après cela, la leçon aux élèves...

Il n'est donc pas du tout certain que les élèves n'apprennent rien en recopiant des sources extérieures. Tout ce qu'on peut leur reprocher, c'est de s'y prendre fort mal, de mal choisir leurs sources lorsqu'ils en choisissent de médiocres ou de mal ajustées aux sujets qu'ils ont à traiter, de ne pas les citer, de leur emprunter trop largement et de ne rien en tirer de personnel.

Sommes-nous du reste bien certains d'être véritablement les auteurs de ce que nous croyons être nos propres opinions, les plus personnelles ? Ne nous contentons-nous pas bien souvent, ce faisant, de répéter simplement ce que d'autres nous ont dit, sans vraiment savoir d'où ces idées nous viennent, soit que nous l'ayons oublié, soit que leurs auteurs soient très difficiles voire impossibles à identifier ? A ce régime tout un chacun pourrait bien se révéler être un petit plagiaire quotidien qui, simplement, s'ignore. La philosophie, en ce qu'elle nous engage dans un examen critique radical de toutes les idées et ouvre ainsi la possibilité d'échapper à cette subordination de la pensée, trouve en cela une pleine justification. La confrontation des élèves avec le phénomène des agents conversationnels est par conséquent une occasion de les sensibiliser à cet ensemble de difficultés, tout à fait élémentaires.

Il n'en reste pas moins que la question de la place occupée par la dimension de reprise, d'imitation ou de transformation de ses sources dans le processus d'élaboration de toute pensée philosophique originale se pose vraisemblablement pour toute activité d'écriture philosophique, qui ne part jamais de rien, depuis les premières tentatives de son apprentissage dans les classes de philosophie jusqu'aux chefs d'oeuvres majeurs des traditions philosophiques.

2.2. Renforcement du problème du plagiat : qui sont les auteurs des productions du tchatbot ?

L'usage par les élèves d'agents conversationnels en ligne pour rédiger leurs devoirs semble effectivement susceptible de poser un problème juridique de cette nature. Comme l'ont souligné dans la presse plusieurs juristes ou chercheurs spécialistes du traitement numérique des textes depuis le lancement du logiciel ChatGPT au mois de décembre 2022, celui-ci a été entraîné à partir d'un très grand nombre de textes en ligne, dont il est très difficile de vérifier si tous sont ou non libres de droits. Les productions de ChatGPT ne contiennent pas à ce jour la mention explicite des sources à partir desquelles elles ont été élaborées, ce qui implique qu'elles ne satisfont pas non plus aux règles d'usage dont l'application est requise pour certaines sources libres de droit (celles de Wikipédia, par exemple). Ainsi, on pourrait très bien considérer, en conséquence de ces prémisses, qu'un élève faisant usage d'un agent conversationnel en ligne pour rédiger une copie s'expose à la commission d'un fait de plagiat par Intelligence Artificielle interposée. On pourrait craindre en outre, en suivant ce raisonnement, que la remise à un professeur d'une copie rédigée par ce moyen puisse être constitutive d'un acte de diffusion à tiers excédant les limites du droit de copie privée. A ce titre, l'acte consistant pour un élève à recopier une source trouvée ou constituée sur l'Internet tombe hors du périmètre du droit d'exception pédagogique accordé à notre institution pour l'usage des ressources publiées dans le cadre des activités d'enseignement, droit qui concerne exclusivement les professeurs dans l'exercice de leurs missions d'enseignement, non les élèves [15].

Il paraît aujourd'hui indispeensable d'aborder ces questions avec les élèves, ce qui est aussi une façon de les sensibiliser aux enjeux juridiques des usages du numérique et ainsi de contribuer, pour la discipline philosophie, par exemple dans le cadre d'une leçon de philosophie du droit, à leur éducation sur ce plan.

2.3. Complication du problème de la répression du plagiat : les obstacles à l'identification des sources dans le cas des tchatbots

La catégorie de plagiat est désormais explicitement intégrée à la liste des cas constitutifs d'une tentative ou d'un acte de fraude aux examens.

> Sur ce point : https://siec.education.fr/examens/bacs-general-et-technologique/fraudes-aux-examens-267.html

La catégorie juridique de plagiat se trouve ici subordonnée à la catégorie juridique de fraude : le plagiat est constitutif

d'une fraude d'abord en raison du caractère frauduleux des moyens employés, par exemple l'introduction dans la salle d'examen de documents (dictionnaire, anti-sèche, documents divers) ou d'appareils interdits (smartphone, oreillette etc.) permettant la consultation de ressources disponibles sur l'Internet ou d'une autre nature et donnant partant matière à un recopiage durant l'épreuve, en contravention au règlement de l'examen. La catégorie de plagiat conserve cependant une consistance réglementaire propre indépendamment de la question du recours éventuel à des moyens eux-mêmes frauduleux : le plagiat n'est en effet pleinement constitué comme tel que si l'élève omet de mentionner explicitement les sources qu'il cite ainsi que le nom de leurs auteurs. Ainsi, même commis par des moyens non frauduleux, par exemple par la restitution de mémoire d'éléments textuels appris par coeur, l'acte de plagiat serait constitué dès lors que le candidat omettrait de citer les auteurs et sources des éléments de texte qu'il récite. Au contraire, dans le cas d'une copie saturée de citations réellement restituées de mémoire durant l'épreuve, mais qui toutes seraient dûment référencées, le caractère de plagiat deviendrait, dans la forme, beaucoup plus difficile à justifier.

Toutefois, le plagiat ainsi compris comme un cas de tentative ou d'acte de fraude aux examens ne saurait s'excepter du régime juridique de la preuve : encore faut-il pouvoir prouver qu'il y a bel et bien un fait de plagiat, ce qui suppose d'être en mesure de produire les sources qui ont fait l'objet de telles reprises non référencées. La chose n'est pas forcément très compliquée lorsque l'on a affaire à des ressources disponibles sur l'Internet. En revanche elle le devient beaucoup plus lorsque l'on a affaire à des éléments textuels pouvant avoir été produits par des agents conversationnels.

D'une part, ces éléments textuels ne sont pas ordinairement accessibles librement : ils ne le sont généralement qu'à l'intérieur du compte personnel du candidat, verrouillé par un mot de passe, sur le site mettant le logiciel à la disposition du public ; c'est le cas par exemple pour ChatGPT, qui requiert une inscription nominative, c'est-à-dire la création d'un compte personnel sécurisé par un mot de passe.

D'une autre part, le propre de l'agent conversationnel est de produire, à chaque commande, un texte à chaque fois différent : il constitue à ce titre une source instable, par conséquent beaucoup plus difficile à distinguer de la production personnelle d'un rédacteur humain.

Ce sont là les raisons pour lesquelles certaines des entreprises propriétaires d'agent conversationnels récemment mis en ligne, soucieux de ne pas être accusés de promouvoir des comportements frauduleux, ont annoncé avoir commencé à développer et à diffuser des logiciels permettant de repérer dans un texte les traces caractéristiques de sa production par un tchatbot. On peut cependant douter sérieusement que ces contrefeux pour le moment très modestes suffisent à permettre aux professeurs de faire face aux difficultés pédagogiques induites dans toute leur ampleur. Aussi un type de réponse d'une nature tout autre que simplement répressive est-il assurément requis : une réponse préventive et de nature proprement pédagogique. Nous reviendrons bien évidemment plus bas sur ce point.

Il devient donc très difficile, dans des cas de recopiage par les candidats, dans leurs copies d'examen, d'éléments textuels produits par des agents conversationnels, de repérer la source utilisée, parce que ses conditions d'accessibilité et sa nature empêchent de procéder pour ce faire au moyen de comparaisons simples entre la copie et les originaux supposés. Et il est bien évident que cette difficulté se pose également pour les travaux demandés aux élèves durant l'année scolaire, qu'il s'agisse de travaux attendus dans le cadre d'enseignements évalués à l'examen en contrôle continu, auquel cas les manquements identifiés par les professeurs seraient susceptibles de tomber dans le périmètre juridique de la fraude aux examens, ou bien qu'il s'agisse d'enseignements évalués hors contrôle continue : dans ce derniers cas, au delà de la transposition qui est faite du régime réglementaire de l'examen lors d'épreuves blanches passées en conditions ou d'épreuves communes ayant valeur d'entraînement direct aux épreuves terminales de l'examen, le cadre juridique constitué par le règlement intérieur de l'établissement offre un point d'appui.

3. Les réponses de l'institution à la question de la tricherie dans le travail des élèves au cours de l'année scolaire, et leur histoire

Les règlements intérieurs des établissements disposent la plupart du temps de façon explicite que les élèves sont tenus de se conformer aux consignes données par les professeurs et de répondre à leurs demandes. Dans ce cas, le recopiage ou le copier/coller d'une ressource numérique ou le recours à des agents conversationnels peut être constitutif d'une faute au regard du règlement intérieur sous la condition que le professeur ait proscrit ces pratiques. Encore faut-il qu'il les ait effectivement et clairement interdites. Le fait de tricherie donc peut être d'autant plus aisément établi que les pratiques ainsi qualifiées ont fait l'objet d'interdictions explicites. Or nous avons bien vu plus haut les raisons pour lesquelles la question, concernant l'usage des ressources disponibles sur l'Internet en général, n'est, pour bien des élèves, pas si claire. Bien des difficultés rencontrées par les professeurs dans leurs classes proviennent des ambiguïtés de leur propre discours à cet égard, de ce qu'il considèrent à tort comme trop évident pour avoir à le dire, partant de l'absence pure et simple de discours, et par conséquent des malentendus qu'engendrent inévitablement ces non-dits.

Si l'on ne veut pas que les élèves utilisent des agents conversationnels pour écrire leurs dissertations, mieux vaut donc commencer par le leur dire, le plus explicitement possible, et même le leur faire écrire dans leurs cahiers, ou l'indiquer sur les documents pédagogiques qu'on leur fournit. On pourra dès lors justifier pleinement qu'un élève qui, en violation de cette interdiction explicite, aurait tout de même fait usage d'un agent conversationnel, est en faute au regard du règlement intérieur, dans une forme analogue à celle d'une fraude lors d'un examen. Dans ce cas, l'action s'exerçant sur le plan de la faute, dans le cadre défini par le régime des punitions et sanctions, qui relève de l'autorité du chef d'établissement, l'élève ne pourra pas se voir attribuer une note (par exemple un 0/20), qui relèverait quant à elle du régime pédagogique de l'évaluation. La faute commise pourra seulement être sanctionnée conformément au régime des punitions et sanctions défini par le règlement intérieur de l'établissement. La démarche devra alors s'accompagner de toutes les précautions juridiques réglementaires, notamment le respect du principe du contradictoire : l'élève doit pouvoir être entendu, pouvoir s'expliquer et dire ce qu'il a fait, voulu faire et pour quelles raisons, et il doit pouvoir se défendre. L'action consistant à choisir de ne pas sanctionner l'élève pour faute, en considérant qu'il a réalisé le travail demandé, mais à attribuer à ce travail une note très basse (par exemple 0/20), au motif qu'étant le recopiage intégral d'une source extérieure n'ayant donné lieu à aucun travail de réflexion personnelle de l'élève, ce travail n'a eu aucune valeur formatrice pour celui-ci, repose sur une autre interprétation possible des faits et constitue une option alternative, cependant exclusive de la première.

Au XIXe siècle, l'enseignement secondaire en général et l'enseignement de philosophie en particulier connaissaient déjà parfaitement le problème du recours que pouvaient avoir les élèves à des aides extérieures, en un temps où les élèves disposaient de ressources et d'outils bien moins nombreux et bien moins performants, mais en disposaient tout de même. Nos prédécesseurs avaient résolu ce problème de la façon suivante : les résultats annuels des élèves n'intégraient pas de travaux donnés à faire à la maison ou à l'étude, ils excluaient les exercices et devoirs courants ; ils prenaient en compte uniquement les résultats de compositions trimestrielles faites sur table. Un bon travail de surveillance, scrupuleusement attentif pendant ces épreuves, permettait aux professeurs de faire efficacement la chasse aux tentatives de triche et de surprendre la plupart du temps les petits téméraires en flagrant délit.

Alors que de nos jours l'introduction récente du contrôle continu et l'intégration des résultats trimestriels des élèves dans les procédures d'examen et d'orientation font de l'évaluation un enjeu majeur décisif aux yeux des élèves et par voie de conséquence une source potentielle de conflit dans les classes, est-ce à dire qu'il faille renoncer aux « devoirs à la maison », propices, lorsqu'ils sont faits avec raison et probité, à des recherches intelligentes, à d'enrichissantes lectures, à des essais et erreurs ? Il faut réaffirmer qu'il n'en est rien.

4. Les questions pédagogiques nouvelles soulevées par les "devoirs à la maison"

Les devoirs à la maison, faits pour apprendre, pour s'entraîner, pour faire des recherches personnelles, pour retravailler ses erreurs tranquillement, sur le mode d'une répétition au sens théâtral ou musical du terme, relèvent naturellement de la dimension d'évaluation que l'on dit aujourd'hui « formative ». Même s'il faut, bien entendu, les noter pour que l'élève puisse bénéficier de repères relatifs à ses progrès et à l'évolution de ses besoins pédagogiques, on pourrait juger que ces travaux ne devraient pas, en toute rigueur, entrer dans le calcul des résultats trimestriels des élèves. [16]

Néanmoins il paraît légitime de valoriser, dans les résultats trimestriels, des travaux en temps libre réussis, pourvu que l'on ait pu s'assurer qu'ils soient le fruit d'un travail véritablement personnel, quand bien même s'appuieraient-ils sur des sources découvertes à l'occasion de recherches effectuées hors la classe mais exploitées judicieusement. On peut penser du reste que même un dispositif qui serait plus fortement axé sur la pratique de la composition sur table (3 devoirs surveillés ou épreuves blanches sont préconisés dans les recommandations de l'Inspection de Philosophie : cf. document d'accompagnement intitulé *L'évaluation des travaux en classe de philosophie* : https://eduscol.education.fr/1702/programmes-et-ressources-en-philosophie-voie-gt) n'exclurait nullement l'usage de formules mixtes, dont on sait qu'elles permettent un accompagnement resserré de la préparation des devoirs limitant considérablement les risques de copier/coller et qu'elles peuvent offrir l'occasion d'intégrer des ressources numériques dans le travail des élèves, le cas échéant à titre de repoussoir, qu'elle n'exclurait pas non plus la pédagogie de la seconde chance (faire refaire les copies ratées ou en réécrire des parties), ni même des formes de notations partielles, provisoires, intermédiaires, préparatoires à celle du devoir achevé, qu'il le soit « à la maison » ou sur table.

Pour garantir le rendu de devoirs à la maison non pris en compte dans la moyenne trimestrielle (le risque que les élèves ne les rendent pas s'ils ne « comptent » pas constitue très évidemment une objection sérieuse) on pourrait très simplement lui conditionner l'accès aux dispositifs de seconde chance post-devoirs surveillés : pas de DM d'entraînement rendus en amont, purs, bien entendu, de tout plagiat ou de tout contenu issu d'agents conversationnels, pas de seconde chance.

Au delà de toutes les difficultés pédagogiques qui en sont les conséquences, l'usage des agents conversationnels en ligne par les élèves transforme donc considérablement les termes du problème du plagiat tel qu'il a été posé jusqu'ici dans le cas plus simple de l'usage des ressources de l'Internet, du fait que les agents conversationnels ne sont pas seulement des logiciels de tri de données mais qu'ils reposent sur les technologies de l'Intelligence Artificielle, capable de réagencer et de transformer les données dans lesquelles elle trouve ses sources pour écrire des textes d'apparence originale. A ce titre, il est légitime de se demander si, dans les éléments textuels qu'il produit, l'agent conversationnel est ou non véritablement l'auteur de quelque chose.

5. Le tchatbot est-il l'auteur de quelque chose ?

Compte tenu de cette capacité tout à fait singulière, on pourrait formuler l'hypothèse selon laquelle, au delà des contributions d'auteurs en droit identifiés ou identifiables que l'IA collecte et synthétise, elle est également capable d'opérer la synthèse de considérations beaucoup plus générales, d'idées très courantes, très communes, auxquelles il est, pour cette raison, impossible d'assigner un ou des auteurs tant elles sont très généralement partagées et répandues, non seulement sur l'Internet, notamment dans la presse et sur les réseaux sociaux, mais plus généralement encore dans la vie courante, ou, si l'on convoque ici l'expression aristotélicienne consacrée, la synthèse d'"idées généralement accréditées" (traduction proposée par L.-A. Dorion dans ses travaux de recherche concernant la dialectique aristotélicienne) ; en grec : *endoxa*. Au 1er chapitre du livre I des *Topiques* Aristote définissait les "idées généralement accréditée" comme "les opinions partagées par tous les hommes, ou par presque tous, ou par ceux qui représentent l'opinion éclairée, et parmi ces derniers par tous, ou par presque tous, ou par les plus connus et les mieux admis comme autorités" [17],en une série qui paraît convenir assez bien pour qualifier au moins en partie les différents types de sources à partir desquelles travaille un agent conversationnel.

***

Si l'on fait abstraction du problème proprement juridique des références implicites ou explicites à des idées de spécialistes que l'IA pourrait incidemment retenir, et partant du problème de l'identification des sources et de la mention explicite des références pour ce qui concerne les données sous droits d'auteurs, une question demeure : qu'en est-il de ce qui reste, dans la production du tchatbot, à savoir les idées "généralement accréditées par tous ou par presque tous" simplement en raison de leur caractère extrêmement commun ? Doivent-elles être considérées comme ayant un auteur, à savoir le tchatbot lui-même, ou bien au contraire doivent-elles être considérées comme une simple mais authentique expression de la pensée du sens commun, dont l'agent conversationnel serait ainsi réputé capable de reproduire l'opération ?

LECON N°4 : De quel sens commun les agents conversationnels sont-ils capables ?

(Cet article, un peu technique, est destiné plus spécifiquement aux professeurs de philosophie)

> "Par apprentissage non supervisé on apprend comment fonctionne le monde et c'est ce qui nous donne le sens commun.
> Ce qui manque aux machines pour l'instant c'est le sens commun."
>
> Yann LeCun [18], *L'apprentissage profond : une révolution en intelligence artificielle*, Leçon inaugurale, 4 février 2016, Chaire Informatique et sciences numériques, Collège de France.

La question de savoir si les élèves plagient ou non quelque chose lorsqu'ils reprennent à leur compte les productions d'un agent conversationnel, ne s'épuise pas dans le problème de la propriété intellectuelle des données à partir desquelles il opère.

Il faut considérer, d'une part, que les données à partir desquelles l'agent conversationnel opère ne sont pas toutes sous droits d'auteurs : certaines relèvent du domaine public, soit qu'elles aient un auteur identifié mais soient sorties du périmètre des données sous droits au profit du domaine public, le délai légal étant échu, soit qu'elles constituent des données anonymes, des idées ou des opinions très générales dont l'origine se perd dans le flot indéfini des conversations et des échanges courants, relayés, notamment mais pas seulement, par les réseaux sociaux ou par les commentaires de lecteurs sur les sites des journaux de magazines nationaux ou locaux. Le premier qui a affirmé sur un réseau social que Dieu existe ou que l'euthanasie devrait être autorisée en France ne peut assurément pas être tenu pour l'auteur de ces opinions, qu'il n'a fait que relayer, puisque celles-ci circulaient déjà très manifestement dans l'espace public avant que quiconque ait eu l'idée de les formuler une première fois sur un réseau social, et ceci avant même l'invention des réseaux sociaux. Affirmer que toutes les idées et toutes les opinions auraient forcément des auteurs qui seraient forcément des individus, auxquels devrait dès lors être nécessairement reconnu un droit de propriété, constitue un présupposé philosophique éminemment discutable. Ce présupposé exclut arbitrairement la possibilité que des opinions puissent être le résultat d'élaborations collectives, produites par des communautés humaines dont les périmètres pourraient être très larges sans pour autant pouvoir être précisément déterminés, selon des processus de transformation évoluant par petites touches, selon des temporalités très lentes, de sorte que la notion d'auteur, individuel ou collectif, leur serait inapplicable, le problème ne se réduisant donc pas à la simple incapacité circonstancielle d'identifier des auteurs dont on devrait cependant postuler l'existence nécessaire. De méchantes langues diraient peut-être bien volontiers qu'une telle conception de l'origine des idées, qui prétendrait les

référer toutes nécessairement à un auteur, témoignerait d'un véritable fétichisme de la subjectivité individuelle, et d'une conception de l'origine des idées caractéristique du capitalisme libéral, obsédé par le droit de propriété privée au point d'imaginer que toute chose doit nécessairement être la propriété d'un individu, dès lors fondé à vouloir en tirer un profit financier. On pourrait dire, de façon plus neutre, que peut-être certaines idées ou opinions pourraient au contraire constituer des bien publics, à la libre disposition de quiconque, en raison de leur caractère très largement partagé, de leur caractère commun. Il s'agit donc ici d'un commun compris comme matériau exploité par l'Intelligence Artificielle.

Cependant il faut également considérer, d'une autre part, que des agents conversationnels mus par des technologies d'Intelligence Artificielle doivent bien recéler une opération propre, qui consiste à déduire des données extrêmement nombreuses à partir desquelles elles travaillent des formules textuelles d'ordre très général, irréductibles à la particularité des données à partir desquelles ces formules sont calculées. Cette capacité pour ainsi dire à généraliser, quasi inductive, que l'on pourrait ainsi être tenté de prêter à la machine, devrait alors être tenue pour sa contribution spécifique, celle dont elle pourrait être considérée comme l'authentique productrice. Il s'agirait donc ici d'un commun compris comme résultat de l'opération de l'Intelligence Artificielle.

On voit bien dès lors quelles difficultés pose, au point du vue du commun compris en ce double sens - le commun comme matériau des opérations de l'Intelligence Artificielle et le commun comme résultat de ces opérations - la reprise par des élèves dans leurs copies des productions des agents conversationnelle. Si l'opération de l'Intelligence Artificielle consiste à réaliser simplement la synthèse d'un matériau d'idées ou d'opinions d'emblée communes, le reproche de plagiat perd considérablement de sa force : on ne saurait tenir rigueur à un élève de s'approprier des idées tout à fait communes, qui appartiennent à tout le monde, donc qui lui appartiennent aussi. Dans ce cas on ne pourrait lui reprocher guère plus que la très grande banalité ou la très grande pauvreté de son propos. On pourrait légitimement tenir rigueur à quelqu'un d'avoir plagié le *Dictionnaire des idées reçues* de Gustave Flaubert [19], mais certainement pas d'avoir plagié les idées reçues elles-mêmes qui en sont l'objet, précisément en raison du fait qu'elles sont des idées reçues. Tout au plus pourrait-on lui reprocher de n'y avoir rien compris, puisque ce texte fait précisément la liste de tous les clichés qu'il est bon d'éviter lorsque l'on prétend être un bon écrivain. En revanche, si l'on fait l'hypothèse que l'opération propre de l'Intelligence Artificielle consiste à produire de toute pièce un commun des idées à partir d'un matériau de données quant à elles toujours particulières et auxquelles ce commun serait par conséquent irréductible, alors on pourrait être fondé à reprocher aux élèves de piller quelque chose comme un auteur. Il y aurait cependant là une décision métaphysique tout à fait audacieuse : ce serait en effet attribuer alors à l'Intelligence Artificielle un pouvoir propre de produire par elle-même un commun des idées, faculté que la tradition philosophique a désigné de longue date par le syntagme de « sens commun ». Or une telle attribution serait-elle philosophiquement légitime ?

La clé de notre problème se situe donc dans la possibilité d'accorder ou non le sens commun aux agents conversationnels. Ce problème se spécifie en deux hypothèses à examiner : l'une, forte, consistant à attribuer au tchatbot un sens commun actif, compris comme un pouvoir propre de généralisation ; l'autre, faible, consistance à lui attribuer une sorte de un sens commun passif, réduit à la synthèse inerte des produits du sens commun humain, dont il faudrait cependant se demander s'il mérite encore le nom de sens commun.

1. Qu'est-ce que le sens commun ?

Le sens commun désigne le pouvoir de l'esprit humain de produire des représentations susceptibles d'être communiquées à d'autres esprits humains, ce qui veut dire partagées avec eux, en droit avec tout autre esprit humain. La question de la possibilité d'un sens commun a donc pour enjeu la possibilité d'une unité de l'humanité sur plan de l'esprit, la possibilité de communiquer nos représentations engageant tout simplement celle de la pensée tout entière, comme l'écrivait le philosophe Emmanuel Kant dans son opuscule de 1786 intitulé *Qu'est-ce que s'orienter dans la pensée ?* :

Mais penserions-nous beaucoup et penserions-nous bien si nous ne pensions pour ainsi dire pas en commun avec d'autres auxquels nous communiquons nos pensées, et qui nous font part des leurs ? [20]

La possibilité d'une telle unité est décisive en particulier pour la constitution de la connaissance. Dans le § 21 de la *Critique de la faculté de juger* [21], intitulé « Peut-on avec quelque raison supposer un sens commun ? », Kant définit ainsi la condition sous laquelle seule des représentations qui soient des connaissances peuvent être communiquées :

> (...) si des connaissances doivent se pouvoir communiquer, il faut aussi que se puisse universellement communiquer l'état d'esprit, c'est-à-dire l'accord des facultés cognitives en vue d'une connaissance en général, et plus précisément cette représentation (par laquelle un objet nous est donné) pour en faire une connaissance ; car sans cet accord en tant que condition subjective du fait de connaître, la connaissance ne saurait en résulter comme son effet.

La possibilité de la connaissance a donc pour condition que soit établi entre les facultés cognitives humaines que sont la sensibilité, l'imagination, l'entendement et la raison un certain type de rapport de proportion, un certain type d'harmonie correspondant un certain état de l'esprit rendant possible l'échange de représentations qui soient des jugements connaissances. C'est cet état d'esprit spécifique qu'il faut appeler d'abord « sens commun », plus précisément « sens commun logique » (*sensus communis logicus*), c'est-à-dire conditionnant la possibilité de la connaissance. Sans cet état de l'esprit la pensée se limiterait à

> (...) un simple jeu subjectif des facultés représentatives, exactement comme le veut le scepticisme.

Hors du rapport de bonne proportion exigé, les représentations produites par l'esprit ne seraient que des représentations empiriques, fondées sur la sensibilité immédiate, mais dépourvues de toute valeur de connaissance. Elles constitueraient « un simple jeu subjectif » : en elles la subjectivité n'exercerait nullement un rôle législateur, le rapport entre les représentations ne recouvrirait aucune relation de nécessité et n'envelopperait aucun caractère d'universalité, bref il n'aboutirait à aucune forme d'objectivité. Le jeu des facultés serait condamné à l'impuissance cognitive, ce qui donnerait argument à la thèse fondamentale de toute philosophie sceptique selon laquelle l'esprit humain est dans l'incapacité d'atteindre la moindre connaissance vraie (ou plus précisément : on n'est jamais en mesure de savoir s'il en est capable ou non). Dans la suite du même paragraphe, Kant décrit ce rapport de proportion entre les différentes facultés de représentation, qui rend possible la formation de représentations qui soient des connaissances :

> C'est là ce qui arrive aussi dans la réalité chaque fois qu'un objet donné par l'intermédiaire des sens met en activité l'imagination pour qu'elle compose le divers, tandis que celle-ci à son tour suscite l'activité de l'entendement pour qu'il unifie ce divers dans des concepts

Dans le rapport de proportion requis pour que des représentations puissent se constituer en une connaissance, il revient à l'entendement de présider au jeu des facultés, puisque c'est à lui qu'incombe la tâche d'unifier sous des concepts *a priori* (c'est-à-dire quant à eux non dérivés de l'expérience) la diversité sentie que l'imagination a préalablement mise en ordre ou schématisée. Telle est donc la structuration des facultés de représentation requises pour qu'existe un sens commun rendant possible la communication universelle de jugements qui soient des connaissances. On remarque cependant, dans ce passage, que si l'entendement préside à la structure, l'imagination y joue un rôle actif déterminant : c'est elle qui, déclenchée par la sensibilité, met en forme les sensations qu'elle lui

fournit et déclenche l'activité conceptuelle unificatrice de l'entendement. C'est pourquoi Kant montre, dans la suite du §21, que la proportion correspondant au sens commun cognitif ou logique présuppose une autre configuration, une autre composition ou proportion, plus fondamentale, dont le sens commun cognitif lui-même dépend en réalité, parce que c'est elle qui, portée fondamentalement par l'imagination, permet d'établir une relation entre la multiplicité des sensations et l'unité des concepts qui rendra possible le rassemblement des unes sous l'unité des autres :

> (...) cet accord des facultés de connaître possède, selon la différence des objets qui sont donnés, des proportions différentes. Cependant, il faut qu'il y ait une proportion où une relation interne qui anime les deux facultés de l'esprit (l'une par l'autre) soit la plus appropriée à l'une comme à l'autre dans la perspective d'une connaissance (d'objets donnés) en général (...)

Cette proportion, qui doit relier seulement l'imagination et l'entendement dans un jeu par lequel elles puissent s'animer mutuellement (l'imagination déclenchant l'action unificatrice de l'entendement tandis que l'entendement opère en retour sur une unification sur les schèmes construits par l'imagination), ne peut être qu'une proportion dans laquelle l'imagination est émancipée de la tutelle législatrice de l'entendement, donnant dès lors lieu à une activité productrice de représentations dans lesquelles les sensations sont ordonnées mais non conceptualisées. Or de telles représentations constituent ce qui s'appelle en toute rigueur des sentiments.

> (...) cet accord ne peut pas être déterminé autrement que par le sentiment (et non pas d'après des concepts).

Il faut donc présupposer, comme condition préalable de toute connaissance, un accord des facultés plus fondamental que celui dont la configuration ou la proportion donne lieu à des jugements et des représentations de connaissance, et qui est celle qui donne lieu à la production de sentiments, eux-mêmes partant universellement communicables à tout esprit doté du même type de configuration de ses facultés. Dans la suite de la *Critique de la faculté de juger*, au §40, intitulé « Du goût comme une sorte de *sensus communis* » [22], Kant donne à ce libre accord des facultés, à ce sens commun sentimental ou sensible le nom de sens commun esthétique, (du grec ancien *aisthésis*, qui signifie « faculté de percevoir par les sens »). La question du sens commun se pose donc sur deux plans : celui des idées et celui des sentiments, les unes et les autres renvoyant à deux types de jugements. C'est la question de savoir comment nous pouvons former des idées, et comment nous pouvons former des sentiments qui puissent être communiqués à autrui, c'est-à-dire partagés avec tout autre être humain. L'enjeu de l'existence d'un sens commun est par conséquent la possibilité de faire communauté avec autrui, communauté intellectuelle, sur le plan des idées - dans ce cas on comprendra le mot « sens » au sens d'une faculté de juger intellectuelle ; communauté affective, sur le plan des sentiments - dans ce cas on comprendra le mot « sens » en son sens premier, qui renvoie à notre faculté de ressentir des états sensibles, à notre sensibilité, c'est-à-dire à une faculté de juger esthétique. Ces deux déterminations du sens commun doivent assurément être comptées au nombre des conditions qui rendent également possible la formation de communautés humaines de niveau supérieur, plus élaborées : comment constituer en effet des communautés morales, des communautés politiques, comment faire tout simplement société sans la possibilité de communiquer, de partager des idées et de partager des sentiments, en droit avec tout autre être humain ? Il faut pour cela supposer en chacun, sous la forme d'un postulat de la raison dans le cas du sens commun logique, sous la forme d'un sentiment dans le cas du sens commun esthétique, des configurations de nos facultés de représentation qui nous rendent capables de former de telles idées et de former de tels sentiments qui puissent être partagés avec tout autre, donnant prise à un usage intellectuel et à un usage esthétique de notre faculté de juger.

Kant souligne, dans la suite du §40, que dans le libre rapport des facultés constitutif du sens commun esthétique,

l'imagination agit de façon libre, cette action étant possible simplement parce qu'elle se trouve dans une configuration où elle est libérée du pouvoir législateur de l'entendement, sans pour autant disposer elle-même d'un pouvoir législateur.

> Ce n'est que là où l'imagination en sa liberté éveille l'entendement, et que celui-ci engage sans concepts l'imagination à un jeu régulier, que la représentation se communique, non comme pensée mais comme sentiment intérieur d'un état de l'esprit qui apparaît comme correspondant à une fin.

Or cette communicabilité en droit universelle de sentiments suppose que la configuration de l'esprit qui la rend possible ne peut se représenter que sous la seule forme de représentations dont il dispose à ce niveau, c'est-à-dire sous la forme d'un sentiment, un sentiment pour ainsi dire à la puissance 2, un sentiment de tous les sentiments, d'un sentiment d'avoir des sentiments qui soit lui-même communicable universellement :

> Or, dans la mesure où cet accord lui-même doit se pouvoir communiquer universellement, le sentiment qu'on a de lui (lors d'une représentation donnée) doit également pouvoir l'être (...)

Une telle supposition ne prend donc pas la forme d'un postulat de la raison, qui supposerait de l'imagination l'usage de concepts dont elle ne dispose cependant pas. Il faut donc supposer, antérieurement au sens commun logique et comme sa condition de possibilité, un sens commun plus fondamental, un sens commun sensible, prenant lui-même la forme présuppositionnelle d'un sentiment, qui rende possible la communicabilité universelle de tous les sentiments et donne en cela à la communicabilité universelle des idées fondée dans le sens commun logique sa condition de possibilité fondamentale.

> (...)mais comme cette communicabilité universelle d'un sentiment présuppose un sens commun, c'est donc avec raison que l'existence de celui-ci pourra être admise, et cela sans qu'on doive s'appuyer à cet égard sur des observations psychologiques, mais comme la condition nécessaire de la communicabilité universelle de notre connaissance, laquelle doit être nécessairement présupposée en toute logique et en tout principe de connaissance qui ne soit pas sceptique.

De deux choses l'une : ou bien on accepte cette double présupposition, deux étages, de la possibilité d'un sens commun logique reposant lui-même sur la possibilité d'un sens commun esthétique, ou bien on est condamné au naufrage du scepticisme. [23]

Il faut cependant insister sur le rapport qui existe entre ces deux aspects du sens commun, le sens commun logique et le sens commun esthétique. Le sens commun esthétique ne vient pas compléter le sens commun logique : il fonde et rend possible le sens commun logique qui le présuppose toujours. C'est que dans le sens commun esthétique, l'imagination ne soumet aucun objet au pouvoir législateur des autres facultés, ce qui serait constitutif d'un accord objectif de ces facultés. Le plaisir esthétique constitue en effet la forme supérieure du sentiment. Sa supériorité tient au fait qu'elle n'est liée à aucun attrait sensible ou empirique pour l'objet de la sensation : il s'agit d'un plaisir pur, désintéressé dans son principe. Est véritablement esthétique en effet le simple effet d'une représentation sur celui qui l'éprouve, indépendamment de l'existence de l'objet correspondant. Seule peut avoir pour effet, dans le jugement

esthétique que j'exprime en disant : « c'est beau », un plaisir supérieur la représentation de la forme pure d'un objet, réfléchie dans l'imagination. Dans un jugement esthétique, c'est la simple représentation de la forme de l'objet, réfléchie dans l'imagination, qui est la cause d'un plaisir supérieur, désintéressé. Cette réflexion de l'imagination ne se rapporte à aucun concept : elle ne schématise plus le donné sensible, elle le réfléchit seulement au point de vue de la forme, se comportant ainsi comme la cause productive et spontanée « de formes arbitraires d'intuitions possibles » [24]. L'accord qui se produit ainsi entre l'imagination et l'entendement est un accord entièrement libre et indéterminé au point de vue des concepts : il s'agit d'un accord qui n'est pas objectif mais qui demeure subjectif, dont le plaisir esthétique est le résultat, il s'agit d'un acte de jugement dont le résultat ne peut pas être pensé intellectuellement, mais qui, de façon paradoxale, ne peut être que senti. Cela signifie que, dans ce libre rapport des facultés, aucune d'entre elle ne légifère : nous nous situons ici à un niveau de relation des facultés antérieur à tout acte législateur, qui est celui de leur simple harmonie subjective [25].

C'est pourquoi Kant conclut, dans la dernière partie du §40, que c'est en dernière instance la faculté de juger dans son exercice sensible, en tant qu'elle est capable de produire des sentiments en droit partageables par tous les hommes, autrement dit la faculté de juger dans son usage esthétique, que l'on appelle plus simplement : le goût, qui mériterait en réalité d'être appelé au plus au point *sensus communis* « à plus juste titre que le 'bon sens', plutôt que la faculté de juger dans son usage intellectuel » [26].

Or il y a ici un point tout à fait décisif pour le traitement de notre problème, qui est celui de savoir si l'on peut ou non accorder aux productions des agents conversationnels une valeur équivalente à celle que nous pouvons accorder aux productions du sens commun humain, partant si l'algorithme opératoire qu'applique le tchatbot pour produire ses textes peut être tenu pour l'exact équivalent de l'exercice d'une faculté de juger, attendu que les productions des tchatbots apparaissent comme une certaine sorte de synthèse et de réarrangement d'idées dont on est tenté de situer la provenance dans le sens commun humain compte tenu de la nature des données à partir desquelles l'algorithme travaille. Dans la mesure où leurs productions textuelles résultent d'une opération purement calculatoire permettant d'induire le mot suivant le plus probable au moyen d'un calcul de fréquence à partir des données disponibles, les agents conversationnels ne peuvent recourir à aucune sensibilité, dont ils ne disposent pas de toute façon. La présupposition d'un sens commun esthétique ne peut donc être faite les concernant, ce qui ruine d'avance et par principe la possibilité de leur attribuer aussi un sens commun logique. Dès lors, se pose la question de savoir quelle valeur il y a lieu d'attribuer à leur productions, en tant qu'elles *contrefont* néanmoins l'apparence de celles du sens commun logique humain. La question est d'autant plus légitime que les élèves, en mobilisant des agents conversationnels pour réaliser leurs travaux scolaires à leur place, les mettent en concurrence avec leur propre sens commun logique. Pour répondre à cette question, le début du §40 de la CFJ apporte des indications précieuses. Il met en effet en évidence les ambiguïtés qui affectent la notion de sens commun dans son usage courant.

2. Les ambiguïtés de la notion de sens commun

Au début du §40 de la *Critique de la Faculté de juger*, Kant avait en effet d'emblée mis en évidence les ambiguïtés produites par les usages qui sont couramment faits du syntagme « sens commun ».

2.1. Première ambiguïté de l'usage courant de l'expression « sens commun »

Il faisait remarquer tout d'abord (p. 278) que, dans la langue courante, l'emploi du mot « sens » pour désigner la faculté humaine de juger, par exemple lorsque l'on dit de quelqu'un qu'il a « le sens de la vérité », ou « le sens des convenances », ou « le sens de la beauté », ou encore « le sens de la justice », ne porte pas essentiellement l'attention sur l'opération de réflexion que la faculté de juger effectue mais seulement sur ses résultats. Il s'agit ici, sinon de lever un contresens possible, du moins de signaler une impropriété de langage qui pourrait prêter à confusion. Ce faisant, on s'exprime en effet exactement comme si la faculté de juger pouvait être simultanément porteuse de deux pouvoirs de l'esprit en réalité incompatibles au sein d'une seule et même faculté : un pouvoir de

sentir ou d'intuitionner et un pouvoir de conceptualiser. Dire couramment de quelqu'un qu'il a « le sens de la vérité », c'est s'exprimer exactement comme s'il pouvait en même temps la sentir, au même tire qu'il voit ou entend tel ou tel objet, sur le mode de la perception sensible, et en juger, c'est-à-dire former la concernant des des jugements de type déterminant - ou si l'on veut de type scientifique, tel que Kant l'avait analysé dans le *Critique de la Raison pure*, à savoir un jugement pour la formation duquel notre faculté de juger possèderait, inscrit en elle *a priori*, antérieurement à toute expérience, un concept de vérité (ou un concept de convenance ou un concept de beauté ou un concept de justice) sous l'unité duquel elle subsumerait et ordonnerait le divers phénoménal, issu de l'intuition sensible, de sorte que le jugement ainsi produit puisse constituer une « règle universelle » et partant avoir une valeur de connaissance. Or, écrit Kant,

> « on sait, ou du moins on devrait normalement savoir qu'il n'existe pas un sens en lequel ces concepts puissent avoir leur siège et - c'est encore plus évident - qu'un tel sens ne possède pas la moindre aptitude à édicter des règles universelles. »

Ce serait en effet supposer tout d'abord l'existence d'une faculté de sentir (c'est-à-dire une pure réceptivité) qui puisse posséder en même temps des concepts *a priori* (c'est-à-dire recéler une activité de l'esprit), ce qui est contradictoire et impossible : une faculté possédant de telles caractéristiques ne saurait exister ; ce serait par conséquent prêter à la faculté de juger ainsi désignée sous le nom de sens commun compris en ce sens la possession *a priori* de concepts qui ne s'y trouvent pas du tout ; ce serait enfin et de de ce fait même prêter à une telle faculté sensible donc toujours particulière la capacité supposée « d'édicter des règles universelles », chose par définition impossible pour une faculté de sentir. Ce n'est pas du tout, en effet, de cette façon que les concepts de vérité, de convenance, de beauté et de justice nous viennent :

> (...) nulle représentation de ce genre concernant la vérité, la convenance, la beauté ou la justice, ne pourrait jamais nous venir à l'esprit si nous n'étions pas capables de nous élever au dessus des sens jusqu'aux pouvoirs supérieurs de connaissance.

La formation de représentations identifiées comme des vérités, comme des convenances, comme des jugements de goût, ou comme des jugements au sens moral et juridique supposent tout au contraire un processus de constitution universalisante : par exemple, pour les jugements scientifiquement vrais, c'est l'intervention du pouvoir universalisant des concepts *a priori* opérés par l'entendement qui permet de s'élever au dessus de la simple phénoménalité sentie pour constituer des jugements déterminants ; dans le cas des convenances, c'est le jugement moral, opéré par la raison pure, qui permet à la faculté de désirer de s'extraire des contenus pathologiques qui l'affecte du fait de sa relation à la faculté de sentir, dans le cas du jugement de goût, c'est l'opération libre de l'imagination dans la faculté de juger esthétique qui permet de former un sentiment universel sans concept, par réflexion pure de la simple forme de l'objet et partant purifié de ses contenus sensibles matériels. Dans ces trois cas, la formulation d'un jugement qui exprime une règle universel suppose donc bien de « nous élever au dessus des sens jusqu'aux pouvoirs supérieurs de connaissance ».

Il y a cependant ici, dans la dénonciation de cette première ambiguïté du syntagme « sens commun », un parallélisme frappant avec la façon dont nous pourrions être tentés de considérer les textes produits par les agents conversationnels en ligne, au vu de leur qualité rédactionnelle. Ne considérant que les résultats, par exemple en corrigeant une copie rédigée par un agent conversationnel exactement comme si elle avait été réellement écrite par un élève, en faisant ainsi totalement abstraction de son processus réel et propre d'élaboration, on pourrait croire que le texte, imitant parfaitement un discours humain, a été produit par l'exact équivalent d'une faculté de juger, croyance

ou présupposition que l'on tend à avaliser lorsque l'on dit que le texte « fait illusion » ou est susceptible de le faire. Exactement de la même façon que dans le cas du « sens de la vérité », du « sens des convenances », du « sens de la beauté » ou du « sens de la justice » dans le cas d'un être humain, nous présupposons à tort, ne nous fiant qu'à l'apparence des résultats qu'elle produit, que l'Intelligence Artificielle possèderait une faculté de connaître déterminante ou une faculté de juger réfléchissante. Il s'agit ici exactement du même type de présupposition que lorsque nous attribuons à quelqu'un, sous le nom de sens commun, une intuition sensible qui serait capable, de façon cependant tout à fait contradictoire, d'édicter des jugements. Ce faisant, nous manipulons une représentation de l'Intelligence Artificielle qui n'est qu'une chimère, exactement comme la conception vulgaire du « sens commun » comme une supposée faculté sensible d'édicter des règles universelles est une pure chimère : nulle part il n'existe rien de tels. Nous le croyons cependant, simplement parce que nous nous fions uniquement à l'apparence des résultats sans analyser rigoureusement le processus qui les a réellement produits.

Il faut ensuite prêter la plus grande attention, ainsi que Kant nous y invite dans la suite du §40, à l'ambiguïté que recouvre, en langue allemande comme dans bien d'autres langues, le mot « commun » lorsqu'il se trouve associé au mot « sens » dans l'expression « sens commun ».

2.2. Deuxième ambiguïté de l'usage courant de l'expression « sens commun »

Dans un usage plus rigoureux, l'expression « sens commun » réfère aussi, écrit Kant, à « l'entendement commun », c'est-à-dire à un « entendement simplement sain » (p. 278), qui n'est pas affecté par exemple d'une tendance à juger de façon seulement empirique, ou de façon systématiquement faussée, ou encore qui n'est pas en proie à la folie, au délire, même s'il n'a pas (encore) été cultivé, ce qui signifie à la fois qu'il n'a pas encore été développé dans son usage ou entraîné par son exercice régulier mais aussi qu'il n'a pas encore été nourri d'une culture littéraire, scientifique, artistique, historique, juridique, religieuse, philosophique, tel qu'on en trouve l'exemple chez des personnes qui n'ont pas reçu une éducation intellectuelle très développée mais qui sont néanmoins parfaitement capables de raisonner juste. Cet usage de l'idée de « sens commun » est un usage plus philosophique : on peut penser bien entendu à l'usage que fait Descartes de l'idée de sens commun prise en ce sens dans les premières lignes du Discours de la méthode, lorsqu'ils écrit que

> « le bon sens est la chose du monde la mieux partagée : car chacun pense en être si bien pourvu, que ceux mêmes qui sont les plus difficiles à contenter en tout autre chose, n'ont point coutume d'en désirer plus qu'ils e ont. En quoi il n'est pas vraisemblable que tous se trompent ; mais plutôt cela témoigne que la puissance de bien juger, et distinguer le vrai d'avec le faux, qui est proprement ce qu'on nomme le bon sens ou la raison, est naturellement égale en tous les hommes ; et ainsi que la diversité de nos opinions ne vient pas de ce que les uns •ont plus raisonnables que les autres, mais seulement de ce que nous conduisons nos pensées par diverses voies, et ne considérons pas les mêmes choses. » [27]

C'est dire que l'expression « sens commun » réfère ici aux jugements généraux que tous les êtres humains sont capables d'énoncer et de partager, et qui, écrit encore Kant au §40 de la CFJ, représentent « la moindre des choses que l'on puisse attendre » d'eux en vertu du fait qu'ils tiennent précisément leur nom d'êtres humains du fait qu'ils possèdent tous un tel « entendement simplement sain ».

Néanmoins - et c'est là l'ambiguïté terminologique soulignée par Kant - l'expression « sens commun » réfère aussi aux usages purement empiriques de cet entendement sain, partant aux résultats les plus banals de ces usages, partant à « ce que l'on rencontre partout et dont la possession n'est absolument pas un mérite ou un privilège et qui constitue ce que la langue classique désigne aussi comme » le vulgaire'', à savoir par exemple des opinions que l'on répète simplement sans y avoir soi-même réfléchi et qui ne sont pas véritablement le produit d'une activité

réfléchissante effectuée en première personne.

Dès lors, à laquelle de ces deux définitions du sens commun faut-il rattacher les productions des agents conversationnels, au vu de leur mode opératoire propre ? Il n'est pas difficile de montrer qu'elles se trouvent strictement cantonnées à la définition du commun ou du sens commun comme vulgarité, à laquelle elles satisfont pleinement. En effet, elles ne satisfont pas, à l'inverse, à la définition proprement philosophique de l'entendement sain compris comme sens commun, qui est

> « l'Idée d'un sens commun à tous, c'est-à-dire un pouvoir de juger qui dans sa réflexion, tient compte en pensée (a priori) du mode de représentation de tout autre, pour en quelque sorte comparer son jugement à la raison humaine tout entière et se défaire ainsi de l'illusion qui, procédant de conditions subjectives particulières aisément susceptibles d'être tenues pour objectives, exercerait une influence néfaste sur le jugement. » [28]

La possession effective par tous les êtres humains sans exception, passés, présents et à venir, d'un tel pouvoir de juger ne peut pas faire l'objet d'une démonstration de type scientifique, parce qu'il est impossible d'en vérifier tous les cas dans l'expérience, c'est-à-dire sur le plan des phénomènes observables. Cette impossibilité s'explique par le fait que nous attribuons le sens commun par principe à tous les êtres humains, non comme une propriété empirique ou phénoménale, qu'ils acquerraient *a posteriori* dans et par l'expérience, mais comme une propriété qu'ils possèdent du seul fait qu'ils sont des êtres humains, c'est-à-dire en vertu de leur essence, de ce qu'il sont en soi, autrement dit comme une propriété métaphysique. Cependant, du fait que cette propriété n'est pas en elle-même empiriquement démontrable sur le plan des phénomènes observables - on peut seulement supposer que certaines conduites observées ou certaines pensées exprimées en constituent des effets et des manifestations empiriques partielles - cette attribution du sens commun constitue un simple postulat métaphysique, un principe régulateur que la raison, dans son usage pur, est obligée de se donner pour rendre compte de la réalité en l'homme d'une propriété dont on peut donc lui supposer la possession en soi, au delà de toute expérience possible, mais dont la connaissance scientifique, cantonnée à la connaissance des seuls phénomènes, n'est pas en mesure de rendre compte, ce que marque, dans ce passage, le recours au concept d'« Idée » (avec majuscule).

Or il est manifeste que le fonctionnement des agents conversationnels en ligne et leurs productions ne satisfont pas aux exigences du sens commun tel qu'il se trouve ici défini par Kant au sens du *sensus communis logicus* [29].

3. Les agents conversationnels en ligne pris dans les ambiguïtés de la notion de sens commun.

Parce qu'il procède par compilation et synthèse de données à partir desquelles il a été entraîné, il est clair que le chatbot obéit à un processus d'élaboration purement empirique : à aucun moment il ne « tient compte en pensée (a priori) du mode de représentation de tout autre », mais il se contente de tenir compte, *a posteriori*, c'est-à-dire après computation sur la base des données qui lui sont disponibles, des représentations de certains autres seulement, en vertu du principe de la plus grande fréquence et non de la plus grande pertinence. Ce qui le prouve, ce sont par exemple les biais divers et variés, sexistes ou racistes par exemple, dont les productions brutes du logiciel témoignent et qui ont contraint ses concepteurs à introduire dans son fonctionnement des instances de censure ; ce sont aussi les erreurs manifestes et les préjugés les plus faux qu'il restitue lorsqu'il est interrogé par exemple sur des questions d'histoire [30]. C'est là une preuve manifeste que le logiciel est structurellement incapable de faire de lui-même preuve bon sens. Jamais le logiciel n'est en mesure, du fait de sa structure et de son mode de fonctionnement, de « comparer son jugement à la raison humaine tout entière », parce qu'il ne procède que par compilation et synthèse empirique de données qui sont des produits résiduels du jugement humain, qu'il assemble, en extériorité, selon le principe de la meilleure fréquence, tout en demeurant en lui-même dépourvu d'une faculté ou d'une instance interne de jugement. C'est pourquoi il demeure fondamentalement victime de « l'illusion qui,

procédant de conditions subjectives particulières aisément susceptibles d'être tenues pour objectives, [exerce] une influence néfaste », non sur le jugement, qu'il ne possède pas, mais sur les résultats de ses calculs : le principe de meilleure fréquence le rendant incapable de distinguer par exemple une connaissance fiable et établie par la recherche scientifique d'un simple préjugé, il présente comme si elle relevait de conditions objectives la totalité de son matériau de données, dont une partie relève cependant en réalité de conditions subjectives. Dans l'exercice du sens commun logique (*sensus communis logicus*) au contraire, tout être humain

> « compare son jugement moins aux jugements réels des autres qu'à leurs jugements simplement possibles et (...) l'on se met à la place de tout autre en faisant simplement abstraction des limitations qui s'attachent de façon contingente à notre appréciation. » [31]

L'agent conversationnel ne considère quant à lui que des jugements réels produits par d'autres selon le principe du plus fréquent ; or les jugements les plus fréquents ne se confondent nullement avec les jugements possibles. Les biais et reprises de préjugés manifestes rappelés ci-dessus montrent également que l'outil est incapable de faire « abstraction des limitations qui s'attachent de façon contingente » à l'appréciation de chacun, mais que tout au contraire il en tient compte et les intègre pleinement dans ses computations, exception faite des domaines « sensibles » (sexisme, racisme, xénophobie etc.) sur lesquels il a été bridé par ses concepteurs.

En ce sens, il convient de se départir du contresens que Kant dénonce dans la suite du §40 de la CFJ. L'opération de réflexion effectuée par les êtres humains lorsqu'ils exercent leur sens commun logique, et qui consiste à « écarte[r] autant que possible ce qui, dans l'état représentatif, est matière, c'est-à-dire sensation » et à « prête[r] exclusivement attention aux caractéristiques formelles de sa représentation ou de son état représentatif », c'est-à-dire qui consiste à mettre de côté notre sensibilité personnelle, à en extraire les déterminations de nos jugements de façon à les soustraire autant que possible à ses effets subjectifs, pourrait passer, à première vue pour une opération d'abstraction formelle de type purement technique, « trop technique pour que l'on puisse l'attribuer à ce pouvoir que l'on nomme le sens commun », de sorte qu'il serait facile de croire qu'une machine informatique devrait parvenir non seulement à réaliser cette opération formelle, mais encore à la réaliser bien mieux qu'un esprit humain. Or c'est là, nous dit Kant, une pure apparence, qui ne nous saisit que lorsqu'on exprime cette opération « dans des formules » abstraites, auxquelles cependant l'opération dans son effectivité est elle-même irréductible. Si la machine se montre effectivement incapable d'un tel discernement, incapable « de faire abstraction de l'attrait et de l'émotion quand on cherche un jugement qui doit servir de règle universelle » (ibid. p. 279), c'est qu'en réalité « il n'est rien de plus naturel » à l'esprit humain, parce que c'est là son opération propre, l'opération vivante de la raison humaine, qui consiste à s'abstraire des représentations simplement sensibles pour s'élever aux idées, opération dont aucune machine n'est capable.

C'est dire qu'en réalité, le pouvoir de juger ne s'apprend pas, comme Kant l'avait déjà montré dans la *Critique de la raison pure*, dans l'introduction de l'*Analytique des principes*, intitulée « Du jugement transcendantal en général » :

Le jugement est un don particulier qui ne peut pas du tout être appris. Aussi le jugement est-il la marque spécifique de ce qu'on nomme le bon sens et au manque de quoi aucun enseignement ne peut suppléer ; car bien qu'une école puisse présenter à un entendement borné une provision de règles, et greffer, pour ainsi dire sur lui des connaissances étrangères, il faut que l'élève possède par lui-même le pouvoir de se servir de ces règles exactement, et il n'y a pas de règle que l'on puisse garantir contre l'abus qu'il en peut faire quand un tel don naturel lui manque*. C'est pourquoi un médecin, un juge ou un homme d'Etat peuvent avoir dans la tête beaucoup de belles règles de pathologie, de jurisprudence ou de politique, à un degré capable de les rendre de savants professeurs en ces matières, et pourtant se tromper facilement dans l'application de ces règles.

*Le manque de jugement est proprement ce qu'on appelle stupidité, et à ce vice il n'y a pas de remède. Une tête obtuse ou bornée en laquelle il ne manque que le degré d'entendement convenable et de concepts qui lui soient propres, peut fort bien arriver par l'instruction jusqu'à l'érudition. Mais, comme alors, le plus souvent, ce défaut (*secunda Petri* [32]) accompagne aussi les autres, il n'est pas rare de trouver des hommes très instruits qui laissent incessamment apercevoir dans l'usage qu'ils font de leur science ce vice irrémédiable. [33]

Qu'avons-nous en réalité à apprendre à nos élèves dans un enseignement scolaire en général et dans un enseignement de philosophie en particulier ? A employer des règles pour exercer leur jugement. C'est peu ou prou ce qu'on nomme dans le système éducatif français, « rendre les élèves autonomes ». Pourtant, il leur revient toujours en dernière instance de régler par eux-mêmes l'usage de ces règles, car il n'y a pas de règle de l'usage des règles, et cela n'a qu'un nom : la liberté. En définitive le jeu des facultés humaines ne peut jamais être autre chose qu'un libre jeu, même si c'est dans le sens commun esthétique que ce libre jeu est le plus libre. L'IA ne peut quant à elle qu'obéir à des règles, fût-elle capable d'apprendre, c'est-à-dire de les reconfigurer en déduisant leur adaptation ou leur transformation de l'analyse des effets de ses propres productions : c'est là encore appliquer un ensemble de règles. Ce n'est pas un petit paradoxe de dire que l'Intelligence Artificielle, en tant qu'elle n'est par définition qu'une machine apprenante, est condamnée de ce fait même à la plus totale stupidité, irrémédiablement. La stupidité n'est pas le contradictoire de l'intelligence : c'est seulement son contraire, ce qui signifie qu'elle ne l'exclut ni logiquement ni factuellement, mais qu'elle est au contraire tout à fait compatible avec elle. De même qu'un chat peut être à la fois blanc et noir (blanc et noir sont des contraires) s'il est par exemple bicolore, ou encore s'il est gris, tandis qu'il ne peut pas être à la fois blanc et non-blanc (qui sont des contradictoires), de même un être intelligent, ce qui signifie un être qui connaît et qui comprend les règles, parce qu'il les a apprises, peut être en même temps complètement stupide, ce qui signifie manquer totalement de jugement dans l'usage ou dans l'application de ces règles qui, quant à eux, ne s'apprennent pas. L'IA pourra donc toujours apprendre des règles tant qu'elle pourra, et même en créer par elle-même un nombre indéfini de nouvelles, jamais elle ne pourra, ce faisant, apprendre à juger, car juger ne s'apprend pas. Dans l'ordre du jugement, aucun être humain n'apprend jamais à être libre : on découvre simplement un jour qu'on est libre et qu'en vérité on l'avait toujours été. Nul ne peut échapper à cette liberté.

4. Inaccessibilité aux tchatbots des maximes élémentaires du sens commun

Il est dès lors aisé de montrer et de comprendre en quoi les opérations et les productions d'un agent conversationnel ne satisfont pas aux « maximes du sens commun » que Kant analyse dans la suite du §40, c'est-à-dire aux principes d'usage qui doivent présider à son exercice. Parce que l'existence en chaque homme d'un sens commun ne peut pas être autre chose qu'un postulat de la raison pure, son usage ne peut pas faire l'objet de règles qui soient des lois : il n'existe et ne peut exister aucune loi du sens commun. Le sens commun peut donc seulement faire l'objet de principes subjectifs de son bon usage : des tel principes s'appellent des maximes. Les maximes du sens commun sont au nombre de trois :

> 1. Penser par soi-même ; 2. Penser en se mettant à la place de tout autre ; 3. Toujours penser en accord avec soi-même. La première est la maxime du mode de pensée qui est *libre de préjugés*, la seconde celle de la pensée *élargie*, la troisième celle de la pensée *conséquente*.

4.1. Penser par soi-même

Concernant la première maxime du sens commun, si l'analyse de différentes productions d'agents conversationnels en ligne permet de montrer que, dans les faits, ces outils ne sont en rien protégés contre le préjugé, l'analyse qu'en propose Kant dans la suite du §40 permet de comprendre pourquoi il en est ainsi en principe. Le processus informationnel consistant, pour la machine, à procéder au tri selon le critère du plus fréquent, à la synthèse et au réarrangement de données, et pour l'élève, à reprendre à son compte le résultat de ces opérations, ne consiste pas à « penser par soi-même », mais au contraire à s'inscrire radicalement dans l'ordre du pré-jugé, les productions de l'agent conversationnel étant construites à partir de données qui, quelle que soit leur valeur informationnelle intrinsèque (de l'ordre d'un savoir scientifique ou de l'ordre de l'opinion ordinaire), se présentent comme les résidus d'actes de jugement produits par d'autres, que la machine ne ré-effectue pas en opérant sur eux, pas plus que l'élève ne le fait en les reprenant simplement à son propre compte. En ce sens, à supposer même que l'on puisse

assimiler les opérations de la machine à une forme de rationalité, de type purement calculatoire, et la reprise par l'élève de ses productions à une forme de rationalité de type stratégique, il s'agit au mieux, dans un cas comme dans l'autre, d'un usage purement *passif* de la raison ; or

> la tendance à la passivité, par conséquent à l'hétéronomie de la raison, c'est là ce qu'on appelle le *préjugé* (...)

Cette tendance à la passivité est une tendance à l'hétéronomie de la raison : dans les textes produits par un agent conversationnel, aucune faculté de juger n'agit selon ses propres lois ; bien au contraire, les idées produites obéissent aux lois (en grec : *nomoi*) d'un ou plusieurs autres (en grec : *hétéros*) qui ont déjà jugé d'avance la question, l'agent conversationnel ne faisant que reprendre et recombiner avec d'autres les expressions textuelles de ces activités judicatives dont il devient dès lors entièrement dépendant, quand bien même les recombinaisons opérées à partir d'elles seraient originales. Et c'est pourquoi rien ne fait obstacle à la possibilité que l'agent conversationnel reprenne à son compte des opinions délirantes, irrationnelles ou tout simplement contraires à la vérité, pour peu que les données à partir desquelles il travaille soient partiellement affectées par des représentations contraires au sens commun, par exemple à des représentations de « la nature comme n'étant pas soumise à des règles que l'entendement, à travers sa propre loi, lui donne pour fondement » (p. 279), c'est-à-dire à des représentations contraires au savoir scientifique le mieux établi rationnellement. C'est pourquoi la première maxime du sens commun est bien « la maxime de l'entendement » comme le souligne Kant quelques lignes plus loin (p. 280) : seul l'exercice de notre entendement, c'est-à-dire le développement d'une pensée fondée sur la connaissance peut constituer le premier rempart propre à nous protéger contre le préjugé en général et la superstition en particulier. En ce sens, le rapport du tchatbot à ses données est au contraire un rapport fondamentalement superstitieux : un rapport de croyance en leur autorité indépendant de toute évaluation véritablement rationnelle possible par laquelle une pensée viendrait « se donner constamment à soi-même sa propre loi », et dont aucune évaluation statistique au regard de leur fréquence ou de leur probabilité dans les textes sources ne saurait constituer un exact équivalent : en un mot, il s'agit d'un rapport de type fondamentalement dogmatique. C'est en quoi les agents conversationnels constituent en eux-mêmes, dans les usages qu'en font les élèves, des dispositifs contraires aux lumières de la raison : rien n'oeuvre en eux à « la libération de[s] préjugés en général » ; seules des instances de censure surimposées par leurs concepteurs humains à leurs opérations peuvent le permettre, sans toutefois que ces instances montrent une totale efficacité. Ils imposent au contraire à leurs utilisateurs, à commencer par les élèves qui en reprennent naïvement les productions à leur compte, « le besoin d'être guidés par d'autres, par conséquent l'état d'une raison passive » (p. 280).

La deuxième maxime du sens commun commande la pratique de la pensée élargie : penser en se mettant à la place de tout autre. C'est en quoi, souligne Kant, elle est véritablement la maxime de la faculté de juger proprement dite, la maxime de la réflexion, qui doit construire des représentations partageables avec autrui qui ne sont pas des concepts préexistants *a priori*, dans l'entendement, au travail de la réflexion. Or, comme nous l'avons déjà souligné précédemment, un agent conversationnel est structurellement incapable de « s'élever au dessus des conditions subjectives et particulières du jugement » que contiennent les données à partir desquelles il travaille, pour adopter « un point de vue universel » lui permettant « de réfléchir à son propre jugement » comme s'il le considérait « du point de vue d'autrui ». La vastitude des données à partir desquelles il est entraîné ne peut rien changer à ce problème. Le recoupement selon des règles de fréquence et de probabilité d'un très grand nombre de données marquées du sceau de la subjectivité, même de façon partielle, ne peut en aucun cas aboutir à la construction d'un point de vue universel objectif : il y a là un saut métaphysique, qu'aucune machine n'est en mesure d'effectuer. Aucune machine ne peut se mettre réellement « à la place d'autrui », même si elle peut très bien en donner l'illusion par des artifices de langage. A ce titre, les opérations des agents conversationnels apparaissent comme des dispositifs de simulation, par des voies calculatoires de nature toute différente, de l'activité de jugement du sens commun, et leurs productions comme des simulacres d'idées de sens commun.

4.2. Imitation ou simulation ?

Une simulation consiste, au moyen d'un modèle formel, qui, dans le domaine technologique, est très souvent de type mathématique ou statistique probabiliste, à reproduire artificiellement, avec une représentation ou un effet approchant au plus près possible, un phénomène réel, naturel, technique ou social. Les sciences et techniques développent ainsi des simulations, par exemple, en physique, en biologie ou en géologie, qui sont parfois transposables à la résolution de problèmes dans le domaine des sciences humaines, en économie, en stratégie, en sociologie, en psychologie par exemple. Le problème de la simulation, comprise en ce sens, n'est pas celui de l'imitation : la difficulté n'est pas ici celle du degré de déformation, d'inexactitude ou de fausseté avec lequel on reproduit ou imite plus ou moins fidèlement le phénomène que l'on cherche à simuler, ce n'est pas celle de l'écart entre la simulation et le phénomène simulé. Une imitation réalise effectivement le phénomène qu'elle imite, avec un degré de ressemblance plus ou moins grande ; au contraire une simulation ne réalise jamais ce qu'elle simule : la difficulté est ici que la simulation ne produit pas du tout l'effet qu'elle reproduit, quand bien même l'opération de simulation serait-elle parfaite. Un collier de verre n'est pas la simulation d'un collier de diamant, il en est l'imitation : tout de verre qu'il soit, il n'en demeure pas moins réellement un collier. Un imitateur, qui contrefait la voix parlée ou chantée d'un personnage célèbre, chante et parle réellement, reproduisant de façon très approchée le timbre et l'intonation de la voix imitée, comme le prouve la comparaison de son diagramme fréquentiel avec celui de la voix imitatrice. Au contraire, dans un simulateur de vol, utilisé pour l'entraînement des pilotes de ligne ou de chasse, aucun vol n'a réellement lieu : personne ne vole. C'est même tout l'intérêt de la simulation : en cas de fausse manoeuvre de l'apprenti pilote, aucun risque de crash aérien : l'apprenti peut progresser sereinement, en apprenant de ses erreurs, sans inquiétude ni culpabilité. De même le compagnon de la fable, qui simule la mort pour tromper l'ours dont il n'aurait pas dû vendre la peau avant de l'avoir tué, ne meurt pas. Celui qui simule le plaisir, ou la surprise, n'éprouve absolument rien de ces émotions : il garde le coeur froid, même s'il peut tromper très efficacement son destinataire, à la mesure de son talent : l'excellent comédien joue de sang froid, il se rend ainsi capable de tout reconfigurer de lui-même, y compris sa sensibilité et jusqu'aux mouvements de son corps ; c'est un simulateur, non un imitateur : c'est même là ce qui fait son paradoxe, comme l'a bien montré Diderot. Dans une simulation, rien de ce qui est simulé n'a donc lieu, sinon l'acte de simulation lui-même. L'effet ou l'apparence de l'effet sont artificiellement reproduits et non produits, pas plus que le phénomène qui en est la cause.

Les agents conversationnels se présentent comme des simulateurs de pensée élargie : ils cherchent à en approcher au plus près l'efficace tout en ne pouvant faire autre chose qu'en contrefaire les effets. Ils simulent la pensée élargie sans parvenir à l'imiter. Ils la simulent en formulant, à propos de la question qu'on lui pose, des énoncés de portée très générale qui synthétisent les données les plus répandues : il s'agit à proprement parler d'un acte d'induction, non d'un acte d'universalisation opéré par la faculté de juger faisant l'effort de se mettre à la place de tout autre. Encore l'illusion de pensée élargie n'est-elle produite la plupart du temps, dans les productions textuelles des agents conversationnels que par la multiplication de points de vues divers sur la question posée, l'accumulation d'opinions ou de thèses : il est tout au plus accessible à la pluralité des perspectives, dont il se montre cependant incapable de tenter un dépassement, vers un jugement qui permettrait d'accorder les points de vue divers et de décider la question de façon légitime et fondée. En ce sens les agents conversationnels sont d'authentiques fauteurs de relativisme, ce que montre clairement le type de formules qu'ils adoptent dès qu'il leur est demandé de conclure : « la réponse à cette question dépend en grande partie des convictions personnelles et des systèmes de valeurs de chacun », « Il est important de continuer à débattre et à réfléchir à cette question ». En ce sens, les éléments textuels que produisent les agents conversationnels n'ouvrent à aucune véritable communauté des idées, dans la mesure où ils se présentent comme de simples collections de points de vues possibles sur une question donnée (on peut penser ceci, ou l'on peut penser cela, certains ont répondu ainsi à la question, d'autres y ont répondu comme cela), dont la structure, non seulement ne permet jamais de tendre vers un point de vue universel possible, qui pourrait, en droit, devenir véritablement commun, mais en exclut par principe, la visée.

4.3. La « pensée » peu conséquente des agents conversationnels : comment je me suis disputé avec mon Intelligence Artificielle... (ma vie numérique) [34]

La troisième maxime du sens commun, celle de la pensée conséquente, qui consiste à toujours penser en accord avec soi-même, est la maxime de la raison pure par excellence. Parce qu'elle renvoie directement aux considérations développées dans la *Critique de la Raison pure* concernant les droits de la raison pure à constituer une métaphysique à partir de ses propres principes, Kant n'en dit pas grand chose dans le §40 de la CFJ. Il nous suffira donc ici de souligner que Kant insiste sur la dépendance de cette troisième maxime à l'égard des deux précédentes pour montrer qu'un agent conversationnel n'y satisfait pas non plus.

Toutefois, si nous envisageons ici l'idée d'une pensée conséquente d'un point de vue plus global en la définissant comme une pensée simplement cohérente avec elle-même, il semble que l'agent conversationnel éprouve également quelques difficultés sur ce plan. A une même question réitérée plusieurs fois, il donne en effet, lorsqu'il s'agit de questions que l'on peut qualifier de controversées, comme le sont toujours toutes les questions philosophiques, des réponses sensiblement différentes, mobilisant parfois des arguments contraires voire contradictoires avec ceux qu'il avait sélectionné lors d'une commande similaire. Il semble également qu'il ne soit pas tout à fait obéissant, formulant par exemple des critiques à l'égard d'une thèse dont on lui demandé de produire seulement un ensemble de justifications. Ainsi, un élève qui aurait la curiosité de poser la même question de façon réitérée à un agent conversationnel, au vu du désordre et de l'absence de convergence voire des contradictions entre les réponses successives fournies, finirait sans doute, à supposer qu'il s'en aperçoive, par ne plus savoir quoi écrire et encore moi quoi en penser lui-même.

Enfin l'augmentation exponentielle des consultations des agents conversationnels depuis leur mise en ligne produit d'étranges dysfonctionnements, en particulier en situation d'utilisation prolongée, ont été rapportés dans la presse : réponses fantaisites, déclarations d'amour, propos agressifs ou évoquant un délire de persécution, réponses malveillantes, injures, comportements discursifs à personnalités multiples, propositions indécentes, ce qui a conduit les entreprises exploitant ces outils à introduire des mesures de bridage, notamment à limiter le nombre de tours possibles (un tour est un ensemble question-réponse) au cours d'une même session d'utilisation. [35] On peut donc dire des agents conversationnels que la pensée conséquente n'est pas leur fort.

***

Que peut-on conclure de ces analyses ? Que non seulement le tchatbot n'est pas dialecticien mais qu'il est de plus incapable en lui-même de tout sens commun : ses productions ne sont que la simulation de ses effets, opérée par compilation, sélection et refonte de résidus contingents du sens commun vulgaire (*sensus communis vulgaris*), de « ce que l'on trouve partout » sur l'Internet sans qu'on puisse lui attribuer le moindre mérite, pas même celui d'un entendement simplement sain mais véritablement actif.

Avant de s'intéresser à ce qu'il serait possible de faire ou non, dans une classe de philosophie, de matériaux produits par un tel dispositif, il y a lieu d'analyser les causes possibles des manifestations d'inconséquence, d'incapacité à la pensée élargie et d'inaptitude à penser par soi-même, bref de ce manque caractérisé de sens commun des agents conversationnels mus par l'Intelligence Artificielle. Dire que les agents conversationnels sont dépourvus de sens commun constitue en effet une explication négative de leur impuissance, celle que manifestent à la fois à leurs dysfonctionnements mais aussi leur incapacité à produire une authentique réflexion philosophique, une explication par ce qui leur fait défaut. Or ne pourrait-on trouver à cette double impuissance une explication positive ? Une notion bien connue entre inévitablement en scène dès que l'on s'avance un peu sur le terrain de l'analyse et de la critique de la notion d'intelligence chez l'être humain : c'est la bêtise. Ce dont l'Intelligence Artificielle se montre ici affectée, n'est-ce pas un phénomène analogue à celui auquel toute intelligence humaine prétend s'opposer et pourtant, inévitablement, s'adosse : une forme de Bêtise Artificielle ?

Leçon N°5 - De la Bêtise Artificielle

(Leçon en cours)

On définit ordinairement la bêtise un manque d'intelligence [36], définition qui peut faire apparaître comme paradoxale l'idée qu'une Intelligence Artificielle, du fait qu'elle est justement censée être intelligente par définition, puisse faire preuve de bêtise. À examiner la question de plus près, la définition de la bêtise comme manque d'intelligence paraîtra cependant un peu trop rapide.

La bêtise n'est pas le contraire de l'intelligence, même à raisonner en termes de possession et de manque : elle est bien plutôt son nécessaire corrélat. Seul un être intelligent peut faire preuve de bêtise. Cela peut s'expliquer assez aisément : en tant qu'elle se définit comme faculté de comprendre (*intelligere*), une intelligence se tient toujours nécessairement à la frontière entre ce qu'elle parvient à comprendre en vertu des moyens dont elle dispose, et ce qu'elle ne comprend pas du fait qu'elle ne dispose pas de moyens adéquats. Aussitôt qu'elle cherche à s'aventurer au delà de cette frontière, une intelligence produit nécessairement de la mécompréhension, parce qu'elle s'applique mal, d'une façon inappropriée, croyant comprendre ce qu'en réalité elle ne comprend pas du tout, parce qu'elle imagine à tort que les moyens d'analyse dont elle dispose lui permettent de saisir une réalité qui s situe en fait hors de leur portée. Bref, la bêtise survient lorsqu'une intelligence cherche à aller au delà de ce qu'elle peut, excède les limites de sa puissance. Le point de passage de cette frontière constitue aussitôt pour cette intelligence un point de rebroussement dans la courbe de variation de sa puissance de compréhension, qui diminue alors selon la même équation : son impuissance à comprendre croît en proportion même de sa puissance de comprendre. Aussi ceux qui sont réputés être les plus intelligents sont-ils souvent tout aussi bien ceux qui sont capables de se montrer les plus bêtes.

> « C'est pourquoi on voit tant d'ineptes âmes entre les savantes, et plus que d'autres. » [37]

Nous ne citerons pas de noms. Et c'est alors que, l'orgueil piqué au vif, certains s'écrient : « est-ce donc moi que vous visez ? » Preuve de leur foncière intelligence : du général au particulier la conclusion peut être logiquement valide ; bonne déduction par conséquent ; mais preuve de leur insondable bêtise : le général n'est pas l'universel, il admet des exceptions, ce qui veut dire des différences, qu'eux-mêmes constituent peut-être ; la généralisation est rendue abusive, par effet de vanité. Ce n'est pas simplement que ces « ineptes âmes » croient à tort comprendre ce qu'en réalité elles ne comprennent pas, ce qui serait simplement erreur ou illusion : c'est qu'elles veulent à tout prix le croire. Aussi leur bêtise suscite-t-elle tout particulièrement la consternation et l'agacement.

> « Je ne m'émeus pas une fois l'an des fautes de ceux sur lesquels j'ai puissance ; mais sur le point de la bêtise et opiniâtreté de leurs allégations, excuses et défenses ânières et brutales, nous sommes tous les jours à nous en prendre à la gorge. » [38]

« Est-ce donc moi que vous visez ? » : on entend bien le braiment : « défense ânière et brutale ».

La bêtise n'est donc pas le résultat d'une incapacité de l'intelligence à exercer son propre pouvoir ou, pour le dire plus simplement, d'un manque ou d'un défaut d'intelligence ; s'y joue tout autre chose : le drame de la puissance.

Toute la question est alors de savoir ce qui peut bien porter une intelligence à un tel excès, venir déformer, fausser ou submerger ainsi la représentation qu'elle a de ses propres limites. On en déduit que le développement ou non de

la bêtise est déterminé par le fait qu'une intelligence peut ou non disposer d'une telle représentation et par la possibilité qu'elle a ou non de s'y conformer, cette représentation ayant quelque chose à voir avec la faculté d'imagination et avec la faculté que l'on pourrait légitimement juger habilitée à lui assigner des bornes. L'explication la plus ordinairement donnée de cette croyance est en effet que, dans la bêtise, l'usage de notre intelligence serait désorienté ou débordé par des volontés obstinées, par des désirs, par des pulsions. Cette explication se heurte cependant à une impossibilité radicale dans le cas d'une Intelligence Artificielle, qui ne dispose d'aucune faculté d'imagination, ne contient d'aucune instance de volonté ou de désir, n'est soumise à aucune dynamique pulsionnelle, mais constitue et demeure, malgré les innombrables fantasmes qu'elle suscite [39], une pure puissance calculatoire. A ce titre, les manifestations objectives de bêtise dont témoignent les agents conversationnels semblent tout à fait paradoxales : il est absolument nécessaire d'en rendre compte, de leur trouver une explication. Nous envisagerons ici le problème de la bêtise dans le cas spécifique et précis de la machine numérique porteuse de formes d'Intelligence Artificielle, sans élargir l'analyse au plan « systémique » du contexte général de la société capitaliste « hyperindustrielle » dans son ensemble [40].

1. Bêtise et animalité

Les critiques philosophiques contemporaines de la bêtise [41] en ont posé jusqu'à présent le problème par comparaison avec la figure de l'animal. Nous proposons ici d'introduire dans cette critique un troisième terme ou une troisième figure, qui est précisément celle de la machine. Cette introduction paraît ouvrir une voie critique assez naturelle tant sont classiques, dans l'histoire de la philosophie, les rapprochements voire l'identification de l'animal et de la machine (Descartes [42]) d'une part, de l'homme et de la machine (La Mettrie [43]) d'une autre part, et vives leurs critiques dans la philosophie contemporaine [44].

Lorsqu'elle caricature la bêtise, celle de gens célèbres ou celle du personnel politique par exemple, la presse satirique, comme la littérature satirique avant elle, les compare souvent à des figures d'animaux, qu'il est courant d'appeler : des bêtes. La bêtise prend ainsi figure animale. Ce rapprochement est-il cependant pertinent ? Les animaux ont-ils quoi que ce soit à voir dans cette affaire ? Ne commet-on pas une profonde injustice en les y mêlant, en les associant ainsi, de façon peut-être très inconsidérée, à la question de la bêtise ? Ces questions accentuent très évidemment notre problème, les animaux, en leur qualité d'êtres vivants non humains, paraissant constituer, si l'on voulait placer sur un « axe de la bêtise » la machine, l'animal et l'homme, un ordre de réalité dont on peut se demander s'il est ou non proche de celui que constituent les machines, voire exactement identique, ou au contraire diamétralement opposé.

La représentation de la bêtise comme manifestation d'une pensée humaine sous emprise animale, en proie à des forces corporelles relevant de l'animalité de l'homme, et qui viendraient la déformer, la fausser, l'entacher d'erreur, est-elle cependant recevable ? Elle peut être mise en question, en ce qu'elle résulte d'un système de présupposés admis sans discussion par les traditions métaphysiques idéalistes et rationalistes, que Gilles Deleuze critique sous le dénomination : « image dogmatique de la pensée » [45]. Ses postulats « écrasent la pensée sous une image qui est celle du Même et du Semblable dans la représentation, mais qui trahit au plus profond ce que signifie pensée, aliénant les deux puissances de la différence et de la répétition, du commencement et du recommencement philosophiques » [46] de la pensée. Le cinquième de ces postulats est celui en vertu duquel « l'erreur exprime à la fois tout ce qui peut arriver de mauvais *dans* la pensée mais comme le produit de mécanismes *externes* » [47]. Dès lors la bêtise est ramenée à une forme d'erreur, comprise comme effet supposé ou prétendu de l'animalité, de la corporéité animale de l'être humain sur sa pensée, par l'effet d'une réduction que l'on a pu qualifier d'épistémologique [48]. C'est donc la propension à vouloir ramener le différent au même qui anime la conception de la bêtise comme animalité, expression ou conséquence de l'animalité. Si l'on suit cette voie, il ne peut y avoir de bêtise des machines que si l'on caractérise les animaux eux-mêmes comme des machines, comme l'avait fait Descartes, que si l'on situe les causes de la bêtise dans ce que l'animalité aurait de typiquement mécanique. Ainsi Descartes formule l'hypothèse selon laquelle rien, dans la constitution physique d'un animal, ne permet de le distinguer d'un automate qui en aurait tous les organes et la figure extérieure :

« (...) s'il y avait de telles machines, qui eussent les organes et la figure extérieure d'un signe, ou de quelque autre animal sans raison, nous n'aurions aucun moyen pour reconnaître qu'elles ne seraient pas en tout de même nature que ces animaux. » [49]

Tout au contraire, ajoute Descartes, tandis que la plupart des mouvements du corps humain s'accomplissent « sans que la volonté les conduise », de sorte que l'on puisse également considérer le corps humain comme l'équivalent d'une machine, à l'image des automates mais « incomparablement mieux ordonnée » qu'eux en ce qu'il sort tout droit « des mains de Dieu »,

> « (...) s'il y en avait [sc. des automates] qui eussent la ressemblance de nos corps et imitassent autant nos actions que moralement il est possible, nous aurions toujours deux moyens très certains pour reconnaître qu'elles ne seraient point pour cela de vrais hommes. Dont le premier est que jamais elles ne pourraient user de paroles, ni d'autres signes en les composant, comme nous faisons pour déclarer aux autres nos pensées. (...) Et le second est que, bien qu'elles [sc. les machines] fissent plusieurs choses aussi bien ou peut-être mieux qu'aucun de nous, elles manqueraient infailliblement en quelques autres, par lesquelles on découvrirait qu'elles n'agiraient pas par connaissance, mais seulement par la disposition de leurs organes. » [50]

Or il est incontestable que quelque chose cloche dans la série de rapprochements ou d'assimilations consistant à faire de la bêtise un effet de l'animalité, ce qui veut dire un effet corporel, c'est-à-dire mécanique. Dans le cas de l'animal, de façon très étrange, ce qu'il y a de mécanique ou ce qui pourrait être interprété comme tel semble pencher univoquement du côté de l'intelligence plutôt que de la bêtise. L'âne bâté brayant, le dindon glougloutant, le mouton bêlant ou l'oie cacardant « toute la bêtise du monde » [51], dont le parler populaire dit qu'ils sont des bêtes [52] et avec lesquels on compare ordinairement les humains pour dénoncer leur bêtise ou la tourner en ridicule, ont malgré tout bien assez d'intelligence pour survivre quelques temps dans leur milieu. C'est donc que la bêtise n'est pas l'animalité :

> « La bêtise n'est pas l'animalité. L'animal est garanti par des formes spécifiques qui l'empêchent d'être »bête« . On a souvent établi des correspondances formelles entre le visage humain et les têtes animales, c'est-à-dire entre les différences individuelles de l'homme et des différences spécifiques de l'animal. Mais ainsi on ne rend pas compte de la bêtise comme bestialité proprement humaine. Quand le poète satirique parcourt tous les degrés de l'injure, il n'en reste pas aux formes animales, mais entreprend des régressions plus profondes, des carnivores aux herbivores, et finit par déboucher dans un cloaque, sur un fond universel digestif et légumineux. Plus profond que le geste extérieur de l'attaque ou le mouvement de la voracité, il y a le processus intérieur de la digestion, la bêtise aux mouvements péristaltiques. Ce pourquoi le tyran n'est pas seulement à tête de boeuf, mais de poire, de chou ou de pomme de terre. » [53]

Les animaux sont doués d'intelligence, personne n'en doute, mais ils ont ceci de particulier que, contrairement à l'homme, l'exercice de leur intelligence est naturellement prémuni contre la bêtise. Qu'est-ce qui empêche ainsi l'animal d'être bête ? Ce sont, écrit Deleuze, « des formes spécifiques » qui l'en « garanti[ssent] ». L'adjectif « spécifique » renvoie ici à la notion d'espèce, au sens biologique du terme, telle qu'elle a été constituée et thématisée au XVIIIe siècle notamment par le naturaliste Georges-Louis Leclerc Buffon (1707-1788). C'est en vertu des caractéristiques de chacune de leurs espèces, désignées ici sous le terme de « formes », que les animaux sont garantis, écrit Deleuze, contre la bêtise. De quel type de forme s'agit-il ? Il s'agit tout d'abord des formes en fonction desquelles s'organisent, dans chaque espèce, le corps des individus, de la forme en un sens morphologique. Ce qui en est l'indice, c'est que, lorsque les caricaturistes ou les satiristes s'emploient à peindre sous des traits animaux des personnages pour les tourner en ridicule, par exemple en établissant des « correspondances formelles entre le visage humain et les têtes animales », en particulier ce qu'on a pu appeler la « zoologie politique », dont le pionnier fut le caricaturiste Jean-Jacques Grandville (1803-1847) [54], ces comparaisons et ces rapprochements se montrent, en un sens, inopérants, parce qu'elles portent sur des propriétés qui ne se situent pas sur le même plan, qui ne sont

pas du même ordre : les différences individuelles de l'être humain et les différences spécifiques des animaux. Or l'individu n'est pas l'espèce : les différences spécifiques des animaux n'ont pas du tout, quant à la possibilité ou non de la bêtise, le même effet que les différences individuelles chez les humains. Elles ont même des effets exactement contraires : tandis qu'il faut chercher du côté des différences individuelles les conditions de possibilité de la bêtise humaine, les différences spécifiques animales viennent tout au contraire barrer la possibilité de toute forme de bêtise chez les animaux. Ce n'est donc pas du tout l'intelligence qui serait un privilège des humains, c'est au contraire la bêtise : l'intelligence humaine seule porte en elle la possibilité corrélative de formes de bêtise, contrairement à l'intelligence animale, qui en est protégée ou garantie par les déterminations propres que sont précisément les formes spécifiques animales, ce qui présuppose que, dans le cas de l'évolution biologique de l'espèce humaine, le statut, la portée et les effets de ses formes spécifiques aient été modifiées par rapport à ce qu'ils demeurent chez toutes les autres espèces animales. Et c'est pourquoi non seulement les rapprochements entre formes animales et visages humains ne caricaturent pas les animaux faisant l'objet de la comparaison, qui ne sont jamais en eux-mêmes ridicules dans ces caricatures puisqu'ils sont représentés simplement tels qu'ils sont, fussent-ils affublés de vêtements humains ; mais encore ils ne rendent pas non plus véritablement compte de la bêtise proprement humaine en mobilisant de telles comparaisons animales. Ce sont les animaux, plutôt que les hommes, qui se trouvent abêtis par la comparaison morphologique de leurs têtes avec des visages d'humains plus ou moins laids, difformes, grotesques, la comparaison ayant pour intention explicite et pour objet délibéré de caricaturer la bêtise des hommes, non l'animalité des animaux qui, en elle-même, n'a jamais rien de ridicule ou de bête.

Au reste, les organes externes des animaux qui expriment ces formes spécifiques ont leurs formes propres et corrélatives d'intelligence : l'habileté parfaite dans l'usage de leurs membres, pattes, ailes, mains, surface et couleur de la peau ou du pelage, pour le déplacement, la chasse, la fuite ou le camouflage, est une caractéristique générale dans toutes les espèces animales. Platon l'avait déjà souligné dans la célèbre réinterprétation du mythe de Prométhée qu'il place dans la bouche du sophiste Protagoras dans le célèbre dialogue qui porte son nom. L'origine de la bêtise n'est pas à rechercher chez les animaux, que le titan Epiméthée a pleinement doté des moyens de survivre dont ils avaient besoin. Elle est plutôt à chercher du côté d'Epiméthée lui-même, « qui n'était pas très réfléchi » [55] et avait distribué inconsidérément aux animaux tous les attributs naturels, oubliant l'homme et obligeant par conséquent son frère Prométhée à voler aux dieux le feu et les techniques pour les donner aux hommes. L'intelligence technique humaine a donc pour origine et pour condition de possibilité, dans le mythe, une bêtise, celle d'Epiméthée (plutôt que sa « faute » : pourquoi moraliser ? [56]), Epiméthée, celui qui agit d'abord et réfléchit ensuite (À¼®¸µ±, « la pensée qui vient à l'esprit après » [57]). Tout au contraire, sur les différents plans vitaux qui déterminent leur survie, les animaux, qui n'ont été victimes d'aucune bêtise de la part d'Epiméthée, ne font quand à eux jamais de bêtises dans l'exploitation de leurs capacités physiques : ils les utilisent toujours au mieux de ce qui leur est possible. Mais peut-être faut-il appliquer plus largement cette idée au corps en général : le corps vivant, considéré comme totalité comme dans le détail de ses organes, n'est jamais bête lui-non plus, même le corps de l'homme. Les pieds, que l'on prétend souvent être très bêtes - bête comme ses pieds - que l'on s'en serve pour penser, pour écrire, pour chanter ou pour voter, ont en réalité bien plus d'intelligence qu'on ne le croit lorsqu'ils nous servent à marcher prudemment sur des chemins périlleux [58]. Le corps en lui-même n'est jamais bête : situer dans le corps l'origine de la bêtise, c'est lui faire injure.

Cependant, les « formes » qui définissent les différentes espèces auxquelles les animaux appartiennent ne se réduisent pas au seul plan morphologique. Un autre type de formes, spécifiquement animales, propres à chaque espèce, pourrait être constitué par les formes de comportement naturellement déterminé pour chaque espèce, dans les activités de nutrition mais aussi de reproduction, ce que l'histoire naturelle et la philosophie ont thématisé à partir du XVIIIe siècle sous le nom d'instinct. L'instinct animal, comme forme stéréotypée du comportement propre à chaque espèce animale, est aussi ce qui préserve les animaux de la bêtise, en ce qu'il les rend parfaitement adaptés à leur milieu, au point de faire de certains d'entre eux des prédateurs redoutables, devant lesquels un homme désarmé a toutes les raisons de trembler. Et c'est vraisemblablement ce à quoi Deleuze fait allusion lorsqu'il évoque « le geste extérieur de l'attaque », qui renvoie aux formes des attitudes instinctives d'agressivité caractéristiques de chaque espèce (le chat qui hérisse son poil, fait le dos rond et feule face à un adversaire) ou encore le « mouvement de la voracité », autres « formes » de comportement vital distinctes et spécifiques, chez la hyène, la baudroie ou le

pigeon, prédéterminées comme des caractères spécifiques, sans rapport avec une quelconque forme de bêtise. Ce n'est que lorsqu'il se trouve extrait de son milieu naturel et placé dans un milieu artificiel, celui de l'agriculture ou de la société des hommes, dans lesquels une part de ses instincts n'a plus cours et où il se trouve partiellement désadapté, que l'animal domestiqué devient bête, bête de somme, animal de troupeau, bétail, animal de compagnie, familier [59].

De telles comparaisons animalières demeurent par conséquent encore trop flatteuses dans l'intention de caricaturer de la bêtise. Plus profondément, il faudrait considérer aussi et surtout les formes physiologiques ou métaboliques propres à chaque espèce - thème éminemment nietzschéen que Deleuze reprend ici. La comparaison de la bêtise humaine avec des mouvements métaboliques ou physiologiques du corps tels qu'ils sont formalisés dans chaque espèce animale pourrait sembler se justifier en ce qu'elle donne lieu à des métaphores plus régressives : elle renvoie à des processus mécaniques, obscurs, internes et partant cachés, aveugles, inconscients et un peu dégoûtants, ce qui pourrait sembler fournir de bons points de comparaison pour décrire et caricaturer la bêtise humaine en ce qu'elle a aussi de mécanique, d'obscur, d'aveugle, d'obtus, d'inconscient et d'un peu dégoûtant. La caricature de l'homme bête en carnivore, du tyran en loup, bête et méchant (« si ce n'est toi, c'est donc ton frère »), comme dans les fables, ou en herbivore, du bourgeois en vache, en ruminant, « bête à manger de foin », ou encore du peuple en troupeau de moutons bêlant, par exemple, toutes ces métaphores renvoient à l'idée que le véritable processus de la bêtise se joue à un niveau encore inférieur, interne, intra-organique, ressemblant peu ou prou à ce qui se produit dans ce cloaque malodorant que constitue, chez les animaux, leur estomac, et que ce processus est en quelque façon comparable à un processus digestif. Les formes spécifiques avec lesquelles on pourrait donc être tenté de comparer la bêtise humaine, ce pourraient être des formes de digestion, des manières de digérer, ou tout aussi bien de ne pas digérer, des formes d'indigestion, avec tout ce qu'elles pourraient comporter de remontées gastriques et de vomissements. Mouvements d'ingestion ou d'excrétion, de réjection, de régurgitation de matières qu'agitent des « mouvements péristaltiques », ces « constrictions annulaires de caractère réflexe, se propageant de haut en bas dans les organes tubulaires afin de faire progresser le contenu de ceux-ci » [60]. Ce que ces mouvements font descendre ou remonter, c'est donc un contenu, les objets même de la mastication et de l'ingestion, en particulier végétaux, chez les herbivores, qui inspire donc un degré de caricature encore plus régressif, par exemple celle de la bêtise politique, celle du tyran : « le tyran n'est pas seulement à tête de boeuf » (bovin, herbivore) « mais de poire » (Louis-Philippe par Daumier) « de chou - bête comme chou - ou de pomme de terre » [61].

Ces fonctions internes - on pourrait conduire une analyse exactement analogue avec les fonctions internes liées à la physiologie de la reproduction - paraissent cependant les moins spécifiques, puisqu'on les retrouve peu ou prou, sous les mêmes formes générales, chez toutes les espèces animales comme chez l'homme. En ce qu'elles semblent partant beaucoup moins barrées ou bridées par des formes spécifiques supérieures, la digestion et plus généralement les processus physiologiques offrent un modèle possible pour analyser et critiquer la bêtise, à condition de le transposer du plan corporel au plan de la pensée. Une telle transposition n'est possible que sous plusieurs conditions. La première condition concerne le statut à donner à la dualité de plans que cette transposition suppose : celle de corps et de la pensée. Cette dualité n'implique en effet ici aucun dualisme : elle le récuse tout au contraire, en posant que la pensée et le corps pourraient connaître des processus ou des dynamiques de types ou de formes analogues : le processus de la bêtise procède d'effets d'un processus analogue à ceux qui peuvent se produire dans un processus de digestion corporelle. En ce sens - et c'est la deuxième condition - le bêtise ne peut être pensée comme résultat de l'action sur la pensée d'une instance extérieure, qui serait le corps. C'est parce que la bêtise, comme la digestion corporelle, procède d'un processus intérieur, interne, immanent, que la figure de la bêtise humaine que constitue le tyran n'est jamais « extérieur ni supérieur à ce dont il profite » :

> « (...) le tyran institutionnalise la bêtise, mais il est le premier servant de son système et le premier institué, c'est toujours un esclave qui commande aux esclaves. » [62]

Le tyran ne produit pas, comme de l'extérieur, sur le peuple qu'il domine une bêtise et une cruauté qu'il tiendrait lui-même de forces corporelles extérieures à sa pensée qui viendraient altérer son jugement, comme en une cascade de causes et d'effets. Tout au contraire, le tyran produit des effets tyranniques de l'intérieur même du peuple d'esclaves qu'il tyrannise, dont il n'est en réalité que le premier, parce que la bêtise et la cruauté qu'il projette dans tout son système tyrannique lui est elle-même interne, comme les produits des mouvements aberrants ou délirants de sa physiologie propre. A ce titre, la tyrannie n'est pas une erreur : c'est l'équivalent d'une complexion, d'un ensemble de dispositions physiologiques détraquées, c'est une manière de vivre. Il y a donc ici, de la bêtise comme phénomène de la pensée aux processus corporels avec lesquels on peut la comparer, une unité par simple analogie. Et c'est pourquoi, écrit encore Deleuze, la philosophie, si elle consentait à s'attaquer un peu sérieusement au problème de la bêtise, devrait considérer ou partir du présupposé assumé que « la bêtise n'est jamais celle d'autrui » [63] Le problème philosophique de la bêtise, dans la mesure où la philosophie se constitue dans son principe comme une critique radicale des idées, ce qui veut dire une critique radicale de la pensée par elle-même, est toujours celui de notre propre bêtise, et par voie de conséquence la critique radicale des formes de bêtise qui habitent la pensée philosophique elle-même.

Si le phénomène de la bêtise humaine a, selon Deleuze, quelque chose d'analogue aux dysfonctionnements d'un processus digestif en ce qu'il correspond à la remontée à la surface de la conscience d'un fond comparable, par analogie ou par métaphore - mais par analogie ou par métaphore seulement - au contenu légumineux de l'estomac d'un herbivore meuglant, bêlant ou brayant, mais sur un autre plan que celui des forces corporelles, de quel fond peut-il bien s'agir, et dans quel processus est-il exactement impliqué ? La réponse a cette question est difficile à interpréter compte tenu du caractère quelque peu énigmatique du passage de *Différence et répétition* dans lequel Deleuze la développe. Certains interprètes on cru y voir un argument psychanalytique, une référence au Ça [64]. D'autres on cru pouvoir plaquer sur ce passage une interprétation simondonienne, sous prétexte que Deleuze mobilise dans ce passage le concept d'individuation et que Gilbert Simondon est référencé dans la bibliographie de l'ouvrage [65]. Il semble pourtant exister une troisième interprétation possible, qu'il faut aller chercher du côté de l'analyse que Deleuze développe du phénomène du ressentiment tel qu'il est compris par Nietzsche [66].

(Paragraphe en cours)

2. Bêtise et erreur : la question de l'infaillibilité de la machine

La bêtise ne s'explique donc pas comme l'effet de puissances corporelles extérieures à la pensée, une erreur ou un ensemble d'erreur commises par l'esprit sous la domination du corps : les puissances corporelles et les formes qui les structurent, tout au contraire, comme le montre très bien la comparaison avec les animaux, garantissent plutôt contre la bêtise. La bêtise est donc d'un autre ordre, que celui du corps. Dire que la bêtise n'est pas le contraire de l'intelligence mais son corrélat, c'est dire surtout qu'elle ne se confond pas avec ce qui constitue le véritable contraire de l'intelligence : l'erreur, avec laquelle la bêtise ne se confond pas. Les erreurs que nous faisons sont des ratés de notre intelligence, des défauts de son fonctionnement qui la portent à confondre le faux avec le vrai, à prendre l'un pour l'autre, comme dire « bonjour Théodore » quand c'est Théètête qui passe [67], « il est 3 heures » quand il est 3 heures et demi, et 7 + 5 = 13« [68]. Or de telles erreurs, c'est précisément ce que, sous la réserve de la fiabilité des données dont elle dispose, une Intelligence Artificielle ne commet pas. On sait aujourd'hui avec certitude que l'automaticité machinique du calcul, sa rapidité, son ampleur au vu de la masse de données (big data) qu'elle est capable de manipuler et de la finesse des corrélations qu'elle est en mesure d'établir rend l'Intelligence Artificielle non seulement plus rapide mais beaucoup plus fiable que le calcul humain. Ce fait se vérifie par exemple dans le domaine de la santé, où l'Intelligence Artificielle parvient à déceler 95% des pré-cancers ou cancers du col de l'utérus, lorsque le diagnostIc médical humain classique peine à dépasser les 70% [69]. Dans ces conditions, comment expliquer les dysfonctionnements décrits par les utilisateurs des agents conversationnels en ligne, si la notion d'erreur (de calcul) ne permet pas d'en rendre compte ?

Il y a là un point décisif : on ne pourrait craindre l'Intelligence Artificielle, et plus précisément la craindre en raison de

son aptitude à simuler l'exercice d'une pensée authentiquement philosophique au point d'abuser la vigilance des professeurs qui en liraient des productions recopiées dans les devoirs de leurs élèves, qu'à la condition d'accepter le postulat selon lequel l'erreur serait « tout ce qui peut arriver de mauvais *dans* la pensée » [70], face à des systèmes informatiques dont la puissance de calcul est telle qu'elle ne commettrait jamais d'erreur. Si telle était le cas, si la bêtise pouvait se réduire à l'erreur, alors il faudrait attribuer radicalement aux Intelligences Artificielles la puissance de penser, sans restriction, une puissance de penser substituable sans reste à celle des êtres humains. Or il se trouve que la bêtise ne se confond justement pas avec cet autre mésusage ou cette autre mésaventure de l'intelligence que constitue l'erreur :

> "La bêtise n'est pas une erreur ni un tissu d'erreurs. On connaît des pensées imbéciles, des discours imbéciles qui sont faits tout entiers de vérités [71].

La bêtise n'est pas un accident de l'intelligence, qui surviendrait sous l'effet de « puissances corporelles » (les désirs et passions des individus), de « faits de caractère » , liés à la psychologie singulière de l'individu (ses inclinations, ses tendances, sa sensibilité propre), ou encore de faits de société, elle n'est pas « une détermination empirique, renvoyant à la psychologie ou à l'anecdote » [72] : elle lui est structurellement liée, peut-être même liée à la pensée même, et constitue partant un véritable « problème transcendantal » [73], formulation d'inspiration kantienne qui fait d'elle une condition de droit, un élément nécessaire du cadre formel pur *a priori* au sein duquel s'exerce toute pensée intelligente :

> « La bêtise est une structure de la pensée comme telle : elle n'est pas une manière de se tromper, elle exprime en droit le non-sens dans la pensée. » [74].

Le non-sens désigne ici la coupure, ou l'incapacité à établir le rapport entre une proposition et le problème auquel celle-ci renvoie. Deleuze prend l'exemple des « devoirs » d'élèves dans lesquels des thèses ou des arguments de philosophes sont énoncés sans que l'élève ait pris la peine d'expliquer les problèmes auxquels ces énoncés renvoient. Écrire simplement dans une copie que pour Socrate « nul n'est méchant volontairement » ou que Descartes « le bon sens est la chose du monde la mieux partagée » sans expliquer quels problèmes ces propositions cherchent à résoudre, c'est dire des bêtises, en ce sens que c'est faire un usage bête, dépourvu de tout sens philosophique, de ces énoncés, qui les transforme en non-sens : les enjeux ne sont pas compris, comme écrivent souvent les professeurs dans leurs appréciations. De tels énoncés ne sont pas des erreurs : l'élève ne dit rien de faux, il est parfaitement vrai que ces thèses ont bien été soutenues par les philosophes auxquels ils les attribue ; mais de tels énoncés sont des non-sens, car ils n'ont pas, en eux-mêmes, en tant que tels, de signification philosophique indépendamment des problèmes philosophiques d'où ils procèdent, ce que l'élève n'a pas compris. [75] Ou pour le dire dans un vocabulaire kantien : il s'agit non pas d'énoncés simplement assertoriques mais d'énoncés problématiques ; or un énoncé problématique traité comme s'il était purement assertorique, comme s'il pouvait être simplement affirmé positivement, devient immédiatement un énoncé dogmatique, non philosophique :

> « Déjà les professeurs savent bien qu'il est rare de rencontrer dans les »devoirs« (sauf dans les exercices où il faut traduire proposition par proposition, ou bien produire un résultat fixe) des erreurs u quelque chose de faux. Mais des non-sens, des remarques sans intérêt ni importance, des banalités prises pour remarquables, des confusions de »points« ordinaires avec des points singuliers, des problèmes mal posés ou détournés de leur sens, tel est le pire et le plus fréquent, pourtant gros de menaces, notre sort à tous. » [76]

Or il est frappant d'observer que cette relation assertorique aux idées et aux références philosophique est précisément le mode d'expression coutumier des agents conversationnels en ligne lorsqu'on les interroger sur une question philosophique, par exemple un sujet de dissertation. Il faut dès lors se demander quelle en est la raison. C'est que le non-sens n'est ici que l'effet, dont il faut chercher la cause, écrit Deleuze, dans « le lien de la pensée avec l'individuation » [77].

(Paragraphe en cours)

Tout comme une intelligence humaine, réputée naturelle ou cultivée par une éducation, produit inévitablement des formes corrélatives de bêtise, de même on peut logiquement s'attendre à ce qu'une Intelligence Artificielle, de type calculatoire, produise inévitablement des formes corrélatives de Bêtise Artificielles, qui sont des formes de bêtise calculatoire, consistant en des calculs mal ajustés. On trouverait là un argument relativement solide pour soutenir qu'en son genre, l'Intelligence Artificielle, si elle produit de la bêtise artificielle au même titre que l'intelligence humaine, constitue une forme authentique d'intelligence. Ce n'est pas du tout que ces calculs soient faux ou erronés.

> ; mais ces vérités sont basses, sont celles d'une basse, lourde et de plomb. *La bêtise et, plus profondément, ce dont elle est symptôme : une manière basse de penser.* Voilà ce qui exprime en droit l'état d'une esprit dominé par des forces réactives. Dans la vérité comme dans l'erreur, la pensée stupide ne découvre que le plus bas, les basses erreurs et les basses vérités qui traduisent le triomphe de l'esclave, le règne des valeurs mesquines ou la puissance d'un ordre établi."Gilles Deleuze, *locus cit.*

3. Comment la Bêtise Artificielle est-elle possible ?

(Chapitre à venir)

3.1. Big data : le problème de la nature et de la qualité des données

3.2. Principe de corrélation et principe de causalité

3.3. La « réflexivité » des machines : algorithme de repropagation du gradient et calculs d'optimisation

3.4. Apprentissage non supervisé et apprentissage par renforcement : le démon de la confirmation

3.5. Principe d'identité et différence critique

4. La censure pour seul horizon ?

(Chapitre à venir)

***

Incapable de tout sens commun logique (*sensus communis logicus*) le chatbot se réduit donc tout au mieux à un simple dispositif « endoxal » au sens aristotélicien (je reprends ici à mon compte ce néologisme « bâti sur son antonyme exact 'paradoxal' » risqué par Jacques Brunchwig dans l'introduction à son édition des *Topiques* d'Aristote) [78], c'est-à-dire un dispositif producteur des collections de représentations simulant les idées généralement admises, soit par un peu tout le monde, soit par tel ou tel philosophe (lorsque le commande exige l'inclusion dans le texte d'une ou plusieurs références philosophiques) mais dont la vraisemblance, elle-même simulée, dépend exclusivement du type d'autorité, nommée ou anonyme, connue ou inconnue, dont elles procèdent.

Pareil matériau peut-il être de quelque utilité pour apprendre aux élèves à faire des dissertations ? L'envisager sérieusement constitue une manière tout à fait paradoxale de faire droit à ce qu'on pourrait malgré tout trouver légitime dans l'intention première des élèves de recourir aux agents conversationnels en ligne, par delà la duplicité qui éventuellement les motive et leur maladresse insigne : trouver une aide véritable pour réaliser et réussir leur travail scolaire, mais en réorientant complètement les moyens qu'ils ont mobilisés en vue lui apporter réponse, en un renversement ironique conduisant à les mettre au service d'un authentique apprentissage. Nous verrons, dans la leçon suivante, que le caractère endoxal que l'on peut attribuer aux productions des tchatbots constitue le point d'ancrage d'une exploitation pédagogique possible qui soit de cet ordre.

Leçon N°6 : D'un usage pédagogique possible des agents conversationnels en ligne pour l'apprentissage de la dissertation en classe de philosophie

Les travaux que l'on peut conduire en classe avec les élèves autour des agents conversationnels, des comportements de copier/coller et plus généralement de la très grande prudence avec laquelle il convient de considérer les ressources fournies sur l'Internet, mettent en jeu la question du rapport spontané que les élèves entretiennent à l'autorité intellectuelle. Les idées endoxales, généralement accréditées relativement à une question ne sont rien autre que celles qui font, d'une manière ou d'un autre, autorité. La difficulté est constituée par les principes d'indexation appliqués par les moteurs de recherche en ligne, que les agents conversationnels, qui reprennent ces principes à leur compte, attendu qu'ils travaillent sur les mêmes données et lorsqu'ils ne sont pas directement associés à des moteurs de recherche (c'est le cas par exemple de *Prometheus*, la version de *ChatGPT* associée au moteur de recherche *Bing* [79]). La plupart des moteurs de recherche actuels fonctionne selon le principe de l'audimat et non selon le principe de la plus grande pertinence : ce sont les sites les plus fréquemment consultés par les utilisateurs du moteur de recherche, les sites ou documents les plus populaires dans lesquels apparaissent les termes de la commande de recherche, qui sont répertoriés en tête de liste. On constate très fréquemment que, dans leurs comportements de recherche, les élèves prennent très peu en compte ces éléments d'analyse, qu'ils les connaissent ou non, pour évaluer le degré de fiabilité des informations qu'ils recueillent, en procédant par comparaison et recoupements, ou pour procéder à la vérification du statut des sites-sources qu'ils utilisent. L'intérêt de travaux pédagogiques incluant des usages du numérique est par conséquent de conduire les élèves à prendre conscience de leur rapport spontané à l'autorité et à l'interroger philosophiquement, en vue de se soustraire à une attitude de subordination inconsciente d'elle-même et non réfléchie aux autorités (ce qui n'exclut cependant jamais de pouvoir choisir d'en reconnaître certaines, mais pour des raisons auxquelles on a solidement réfléchi) pour devenir soi-même le véritable auteur de sa pensée. A ce titre, la survenue des agents conversationnels en ligne et la question de la façon de l'Ecole peut éduquer les élèves à leur usage recouvrent un enjeu de formation civique.

Les analyses développées dans les précédentes leçons nous ont conduit à comprendre les agents conversationnels comme des dispositifs endoxaux, compilant et synthétisant en des constructions rédactionnelles spécifiques des ensembles d'opinions ou d'idées admises à partir des résidus de l'exercice du sens commun humain contenus dans les données à partir desquelles ils travaillent, et également des éléments d'idées plus autorisées, implicitement ou

explicitement, le cas échéant référées à un ou plusieurs auteurs, notamment philosophes. Cette interprétation endoxale ne va pas de soi. Dans la mesure où elle convoque un appareillage conceptuel qui n'est pas anodin, implique-t-elle ou non quelque chose comme une réactivation pédagogique de modèles dialectiques anciens, issus de la tradition aristotélicienne ? Ainsi, la considération du geste philosophique cartésien, inaugural de la philosophie moderne, qui a précisément consisté à congédier la dialectique ainsi comprise (comme dispositif endoxal) du champ de la connaissance rationnelle vient très sérieusement interroger la perspective d'une telle réhabilitation. Comment donner un statut pédagogique légitime aux instruments dialectiques que peuvent constituer les agents conversationnels, sans pour autant céder forcément à la tentation d'une telle réhabilitation ? Dans ces conditions, comment concevoir quelques principes généraux d'une pédagogie possible des Intelligences Artificielles dans un enseignement de philosophie ?

1. Difficultés et risques d'une pédagogisation des agents conversationnels en ligne pour l'apprentissage de la dissertation dans le cadre d'un enseignement de philosophie

L'idée d'utiliser des agents conversationnels en ligne comme simulateurs endoxaux dans le cadre d'un enseignement de la dissertation philosophique pourrait être aisément taxée d'archaïsme pédagogique. Elle pourrait facilement se voir reprocher de promouvoir une régression pédagogique fâcheuse, vers une conception scolastique de la philosophie et de son enseignement, au mépris du geste philosophique fondateur par lequel Descartes, dans les *Règles pour la direction de l'esprit* (posth.) puis dans le *Discours de la méthode* (1637), avait écarté la dialectique scolastique de type aristotélico-thomiste du champ de la rationalité philosophique et scientifique, ouvrant ainsi la double voie des sciences et de la philosophie modernes.

D'un certain point de vue une accusation de ce type pourrait sembler paradoxale concernant un outil, l'agent conversationnel en ligne, qui représente la pointe des technologies numériques mises à disposition du grand public. En réalité il n'est est rien. Rien n'exclut en effet que l'usage mal réfléchi et mal ajusté de technologies numériques très avancées puisse susciter des pratiques pédagogiques grossières, d'un complet archaïsme, inappropriées, brutales, aux effets désastreux pour les élèves. On nous permettra de considérer, sur ce plan, les trois exemples suivants.

La dématérialisation forcée et brutale de l'enseignement de philosophie qu'ont suscitée les confinements survenus lors de la crise de la Covid 2020-2022, a conduit certains collègues, pris de court ou de panique, et vivant la mise en oeuvre de séances de cours en visioconférences interactives comme une difficulté insurmontable, à déposer le texte entièrement rédigé de leurs cours sur les Espaces Numériques de Travail des établissements, en un geste strictement identique, dans sa forme comme dans son contenu - mais en version numérique - à celui qui aurait consisté à distribuer aux élèves des cours polycopiés (ce qui, du reste, a été fait également durant le premier confinement, à destination des élèves qui ne disposaient d'aucun équipement numérique dans leur cadre familial), pratique que l'Inspection de Philosophie dénonce pourtant, depuis les Instructions du 2 septembre 1925 au moins, comme un travers profondément dommageable pour les élèves et pour les professeurs, et comme le degré zéro de la pédagogie et de la formation philosophiques. On a eu recours à ce type de dépôts, le plus souvent comme à un pis-aller, dans l'urgence et faute d'autres possibilités institutionnelles ou tout simplement matérielles immédiatement disponibles : nul ne songerait par conséquent à blâmer les collègues qui s'en sont tenu là, au vu du caractère tout à fait inédit de la situation et de la très grande violence à la fois sociale et psychologique des circonstances. Toutefois, ces dépôts ont-ils constitué au bout du compte, pour les professeurs qui ont dû s'y limiter, et pour leurs élèves qui ont bien dû s'en contenter, plus et autre chose qu'un simple acquis de conscience ? On peut en douter sérieusement, en particulier au vu des effets d'apprentissage très affaiblis de l'enseignement de la philosophie à distance observés par de très nombreux collègues dans les mois qui ont suivi.

On peut également interroger très sérieusement sous cet angle les pratiques dites de « classe inversée » que l'on observe très sporadiquement dans les classes de philosophie et qui se veulent inspirées de la pratique universitaire des MOOC. Ce qui peut valoir à la rigueur pour des étudiants vaut-il forcément et de ce fait même pour de jeunes

élèves de Première et Terminale tout débutants en philosophie, à supposer que cela vaille tout court, au vu des nombreuses critiques et des doutes touchant la réalité de l'efficacité pédagogique et des effets de démocratisation attendus de ces nouveaux dispositifs numériques ? [80] De telle pratiques de classe inversée, au vu des observations que l'on peut en faire, sont presque systématiquement marquées par l'esprit le plus dogmatique : de façon tout à fait discutable, elles présupposent le cours de philosophie comme un corps de pensée qui pourrait être pré-constitué par le professeur et prêt à l'emploi, qu'il s'agirait ensuite simplement de « transmettre » aux élèves et dont il y aurait seulement lieu d'assurer « l'appropriation » sous la forme d'une reprise ou d'un commentaire suscités par leurs questions et au moyen d'exercices ou de travaux pratiques. De telles façons de faire cèdent en réalité aux mêmes présupposés que les distributions de cours polycopiés, partant aux mêmes illusions, et produisent le même type d'effets pervers sur les élèves : désordre, confusion, lenteur, ennui profond, désengagement d'élèves se croyant à tort prémunis, avantage objectif donné aux meilleurs élèves, déjà autonomes et capables d'apprendre par eux-mêmes. Ces constats n'interdisent pas cependant la conception et la mise en oeuvre de dispositifs inversés proposant par exemple aux élèves, en préparation de la leçon et pour en accompagner le déploiement, des dossiers de recherche disponibles en ligne, proposant des ressources numériques de toutes sortes (audiovisuelles ou textuelles) à consulter et à travailler, notamment sous la forme de lectures et d'exercices, mais qui s'abstienne scrupuleusement de distribuer le texte intégral ou un résumé de la leçon.

On peut mentionner enfin les heurs et malheurs de l'usage des diaporamas numériques en classe pour soutenir la prise de notes des élèves, qui constituent pour eux une aide véritable et efficace lorsqu'ils proposent un cadre souple et très allégé en contenu, laissant toute sa place au travail intellectuel de la classe et à l'activité de réflexion vivante, mais qui les pétrifient et les condamnent à une passivité forcée lorsqu'ils font défiler à toute vitesse le contenu du cours dans sa masse, produisant des effets de panique ou de découragement qui conduisent au résultat exactement inverse de celui qui était recherché en ce qu'ils dispensent et empêchent au bout du compte les élèves de prendre des notes, la mise en ligne de tels diaporamas sur l'ENT, pour couronner le tout, venant cumuler tous les travers.

Le problème auquel ces trois exemples renvoient est celui de la possibilité de « pédagogiser » ou non l'usage des outils numériques, et plus généralement technologiques (après tout la polycopie est une technologie) et des obstacles qu'elle rencontre de façon très variable en fonction de la nature propre et des formes spécifiques des enseignements, des disciplines ou des ensembles théoriques et pratiques dont qu'il s'agit précisément d'assurer l'apprentissage. Ce même problème se pose à propos des agents conversationnels en ligne : sont-ils ou non « pédagogisables » dans le cadre d'un enseignement de philosophie ? si oui comment ? avec quels obstacles et quelles limites ? Et les mêmes types riques que dans le cas des dépôts de cours et leçons intégralement rédigés sur les ENT, des dispositifs inversé ou de l'usage de diaporamas numériques menacent les usages possibles de ces nouveaux outils issus du développement des Intelligences Artificielles. L'enjeu est ici de savoir si certaines formes de pédagogisation peuvent déboucher sur des pratiques apportant une plus-value réelle pour l'apprentissage de la dissertation philosophique, ou bien si elles risquent de produire à des effets d'appauvrissement pédagogique et intellectuel, en conduisant à rabattre les pratiques d'apprentissage de la dissertation philosophique, sous l'apparence chatoyante et ludique du numérique, sur des façons de faire en réalité parfaitement archaïques. C'est là plus généralement le risque qui guette toute initiative pédagogique prétendant à l'innovation : s'imaginer que l'on fait oeuvre nouvelle alors que l'on réitère, sous une forme différente mais simplement dissimulatrice, des pratiques parfaitement convenues, traditionnelles, voire régressives.

2. Agents conversationnels et destins de la dialectique aristotélicienne

3. Le tchatbot comme simulateur doxique

4. Dialectiser les productions du tchatbot pour apprendre à disserter

5. Quelques recommandations et suggestions utiles

Les professeurs de philosophie n'ont pas à se sentir menacés par l'apparition de nouveaux moyens grâce auxquels certains élèves pourraient se dispenser de réaliser par eux-mêmes les exercices et devoirs qui leur sont prescrits. Des moyens de cette nature ont en réalité toujours existé. Le véritable enjeu pédagogique est de convaincre les élèves qu'il est préférable pour eux, pour leur bonne formation intellectuelle et philosophique, de ne pas recourir inconsidérément à ces moyens et de faire leur travail scolaire par eux-mêmes, partant de trouver des moyens ou des méthodes efficaces de les en convaincre.

Il convient par conséquent d'éviter toute posture obsidionale ou l'adoption d'un dispositif de lutte exclusivement constitué de mesures répressives, au profit de stratégies pédagogiques intégrant résolument l'existence des agents conversationnels en ligne dans l'accompagnement de l'apprentissage de la dissertation.

- **Le cadre institutionnel de réflexion et d'action touchant les Intelligences Artificielles et les agents conversationnels**

Le Ministère de l'Education Nationale et de la Jeunesse n'a pas attendu la récente mise en ligne en accès libre d'agents conversationnels pour s'engager dans une réflexion scientifique et pédagogique de fond sur les Intelligences Artificielles, leurs effets sur les apprentissages des élèves, les opportunités qu'elles présentent et les risques qu'elles comportent. Il est par conséquent recommandé aux professeurs de se reporter aux travaux institutionnels d'analyse et aux préconisations formulées par l'institution afin d'inscrire leur action pédagogique dans ce cadre.

> Voir Jean A. et alii, *Agents conversationnels en classe. Avancées et recommandations.* Conseil Supérieur de l'Education Nationale, 2021 : https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiSy9LSx479AhUBTKQEHSagCcUQFnoECBQQAQ&url=https%3A%2F%2Fwww.reseau-canope.fr%2Ffileadmin%2Fuser_upload%2FProjets%2Fconseil_scientifique_education_nationale%2FSynthese_IA_et_numerique__version_finale__.pdf&usg=AOvVaw26b0yYHwZT1aioMxxEdXjR

Il importe tout particulièrement de veiller, dans le choix et dans l'usage des outils numériques, à la question de la protection des droits d'auteurs relatifs aux données traitées par les agents conversationnels utilisés et des données personnelles et des professeurs, conformément au cadre de la RGPD.

> Sur ce point voir par exemple *Intelligence artificielle : les données à caractère personnel*, Réseau-Canopé, Bulletin de veille n°6, Direction de la Recherche et du Développement sur les Usages du Numérique Éducatif, Alexandra Coudray, chargée de la veille et de la prospective : https://www.reseau-canope.fr/fileadmin/user_upload/Projets/agence_des_usages/6855_BulletinVeille_6_IA.pdf.

- **Occuper le terrain : éduquer les élèves aux outils numériques et les prendre en compte dans l'enseignement de philosophie**

-
  - Montrer aux élèves des extraits bien choisis de la vidéo proposée en lien dans la première de cet article permettrait de leur expliquer précisément les mécanismes et les principes techniques sur la base desquels fonctionnent les agents conversationnels en ligne, en vue de les convaincre que le résultat, s'ils les utilisent, sera bien loin de ce qu'ils pouvaient en espérer.

- Les agents conversationnels constituent aussi un excellent exemple à exploiter et à travailler en classe avec les élèves, à la lumière de cette vidéo, dans une leçon qui entreprendrait de poser le problème de la réductibilité ou non de la pensée à des structures de langage.

- **Utiliser le chatbot comme simulateur endoxal**

(paragraphe en cours)

- **Exploiter un agent conversationnel dans le cadre de l'apprentissage de la dissertation : passer de la simulation à l'émulation**

- Premier exercice : comparer avec les élèves des exemples de "dissertations" écrites par un agent conversationnel et de bonnes dissertations écrites par des élèves ou construites en classe avec le professeur (l'exercice peut être réalisé « en aveugle », en mélangeant des copies d'élèves et des copies produites par l'IA sans dire aux élèves de quelle source elles proviennent) ; procéder à leur analyse et à leur évaluation, sous la forme de commissions d'entente comparables à celles auxquelles participent les correcteurs de l'épreuve du baccalauréat (demander aux élèves de les lire, de les évaluer et de les noter).

- Deuxième exercice : utiliser des matériaux produits par un agent conversationnel pour entraîner les élèves à la dissertation (par analogie avec l'entraînement des joueurs d'échecs au moyen d'Intelligences Artificielles) :

- mettre un sujet de dissertation (avec consignes rédactionnelles précises permettant d'obtenir un texte ressemblant fortement à une dissertation de philosophie) donné aux élèves dans l'agent conversationnel ;

- donner le résultat obtenu aux élèves et leur faire analyser les défauts de la copie ;

- leur demander de rédiger chacun une copie qui n'ait pas les défauts constatés : il s'agit de les mettre en concurrence avec la machine, en leur demandant de faire mieux qu'elle ; on pourrait voir là une façon originale de moderniser le vieux principe (d'origine jésuite) de l'émulation.

- Troisième exercice : utiliser des matériaux produits par un agent conversationnel pour amener les élèves à

s'approprier les méthodes d'élaboration d'une dissertation et à perfectionner leur mise en oeuvre :

- répartir les élèves par groupes de trois au quatre ;

- leur demander de commander une dissertation à un agent conversationnel sur un sujet donné (chaque groupe peut travailler sur un sujet différent, ou bien tous les groupes peuvent travailler sur un seul et même sujet) ;

- demander ensuite à chaque groupe de procéder à une amélioration du texte produit par l'IA à partir de son analyse critique.

Lien vers un exemple d'exercice proposé par Maryse Emel :

- Demandons au chatbot de réaliser une dissertation :
  - Exercice sur les limites du chatbot à propos du sujet : l'art imite-t-il la nature ?

6. Stratégies pédagogiques et éducation au numérique : comment lutter efficacement contre la prolétarisation des esprits ?

(Paragraphe à venir)

« Prosperity doth best discover vice, but adversity doth best discover virtue » : « la prospérité découvre nos vices, l'adversité nos vertus » écrivait Francis Bacon dans son *Essai* intitulé *Sur l'adversité*. Il est incontestable que l'irruption des agents conversationnels et plus généralement des ressources et outils numériques dans le champ pédagogique met l'enseignement de philosophie à l'épreuve. Elle lui offre cependant par cette adversité l'occasion de démontrer une fois de plus l'irréductibilité de ses vertus éducatrices. La responsabilité incombe à tous les professeurs de philosophie de réfléchir philosophiquement, de façon très précise et très informée, à la nature, à la puissance, aux limites et aux enjeux des outils que le développement de l'Intelligence Artificielle et de ses applications sous forme d'agents conversationnels en ligne mettent désormais à la disposition des élèves, parce qu'ils constituent l'un des visages de la réalité du monde contemporain, de saisir les difficultés d'évaluation que leur usage par les élèves impliquent, ainsi que les déplacements et les transformations profondes que le souci de leur trouver des solutions efficaces appelle dans leurs propres pratiques pédagogiques et dans leurs méthodes d'enseignement, d'inventer enfin, par l'intégration bien informée et résolue du facteur numérique dans leur pédagogie, des exercices et travaux à proposer aux élèves pour les accompagner dans leur apprentissage de la dissertation et de

l'explication de texte. Ce dernier point ouvre incontestablement à la discipline philosophie un horizon d'avenir, une voie de progrès et de renouvellement de ses méthodes d'enseignement. Les Inspectrices et Inspecteurs Pédagogiques Régionaux de Philosophie répondront toujours présents pour contribuer à cette oeuvre aux côtés de leurs collègues professeurs de philosophie : qu'elles et qu'ils sachent pouvoir compter sur leur aide et sur leur soutien.

Les articles présentés ci-dessus n'ont pas été écrits au moyen d'un agent conversationnel en ligne...

Eric Le Coquil, Inspecteur d'Académie - Inspecteur Pédagogique Régional de Philosophie

RESSOURCES PEDAGOGIQUES

Des exercices par Maryse Emel IAN de philosophie et webmestre du site académique de philosophie de l'académie de Créteil

Les ressources qui suivent sont en continuité avec l'article d'Eric Le Coquil

I. REPRESENTATIONS

- Exemples

On peut commencer par diffuser des émissions sur l'intelligence artificielle qui pour certaines peuvent prêter à sourire.

-
  - Montage d'archives projeté lors de la session « La science-fiction est-elle une science prospective ? », organisée par le Social Media Club le 23/11/2016 au NUMA Paris. Archives de 1947 à 1976. Avec (par ordre d'apparition) : Isaac Asimov, Jean d'Arcy, Forrest J Ackerman, Robert Silverberg, Emmanuel d'Astier de la Vigerie, John Brunner (en off 2min27)
    https://www.ina.fr/ina-eclaire-actu/video/man3152578291/science-fiction-une-science-prospective"
  - Les combats homme-ordinateur fascinent depuis des décennies. Déjà en 1959, une femme se confrontait à une machine dans un jeu des deux doigts. Discours qui en dit long sur la représentation de la femme plus que sur la machine !
    https://www.ina.fr/ina-eclaire-actu/video/man5676206958/intelligence-artificielle-en-1959
  - L'intelligence artificielle imaginée en 1968
    https://www.ina.fr/ina-eclaire-actu/video/s716326_001/l-intelligence-artificielle-imaginee-en-1968
  - Ce lien renvoie à plusieurs extraits d'émissions TV sur l'intelligence artificielle et questionne le débat sur le chatGPT
    https://www.ina.fr/recherche?q=intelligence+artificielle&espace=1&sort=pertinence&order=desc

II. Un exemple de dissertation

L'intelligence artificielle peut-elle prendre la place de l'homme

- L'exemple de Kasparov
- Des exemples cinématographiques : la question posée renvoie à ce qu'il y a de plus classique : la peur des machines animées d'une prétention à dépasser l'homme. Toutefois, avec le chatbot on a dépassé largement le stade de la machine entendue comme mécanique, animée d'un mouvement autonome.

a) Alex Garland consacre son premier film en tant que réalisateur, Ex Machina. Il leur donne l'apparence de femmes. Un choix qui fait naître chez les spectateurs et spectatrices un sentiment d'**inquiétante étrangeté** freudienne. Voir l'article : https://www.deuxiemepage.fr/2015/06/03/ex-machina-analyse/

b) On travaillera sur :

-
  - le film Blade Runner où est imaginé un test d'humanité de la machine ou encore ces références sur la machine comme créature. Dossier du film sur Transmettre le cinéma
  - A. I. Intelligence artificielle Steven Spielberg ; analyse audio https://www.forumdesimages.fr/les-programmes/toutes-les-rencontres/a.i.-intelligence-artificielle-de-steven-spielberg
  - Robert Wise, Jour où la Terre s'arrêta (Le) États-Unis (1951). Dossier
  - I, Robot Alex Proyas. Exercices à partir des Robots d'Asimov
  - Her, le film de Spike Jonze
    Spike Jonze aborde à travers son film, la question de la séparation (homme-machine ; âme-corps). Dossier Transmettre le cinéma

    : Her, le film de Spike JonzeDossier Transmettre le cinéma

    Pour l'essayiste Ariane Nicolas, le film Her montre le piège de l'idéal transhumaniste. L'humain peut croire échapper à son humanité par la technologie, mais c'est une impasse. « En tombant amoureux d'un logiciel, le héros pensait s'épargner la souffrance à laquelle une relation avec une personne humaine l'aurait exposé. Le réalisateur du film [...] suggère que la souffrance au contraire est l'expérience indispensable qui atteste de notre singularité en tant qu'êtres humains. Seul un être véritablement incarné est capable d'éprouver des émotions sincères et donc, in fine, de prendre conscience qu'il existe. »

Autres références :

Le Golem (Carl Boese et Paul Wegener, 1920)
Metropolis (Fritz Lang, 1927)
Blade runner (Ridley Scott, 1982)
Terminator (James Cameron, 1984)
A.I. Intelligence artificielle (Steven Spielberg, 2001)
Frankenstein Junior (Mel Brooks, 1974)
Frankenweenie (Tim Burton, 2012)
La Fiancée de Frankenstein (James Whale, 1935)
Dracula (Tod Browning, 1931)
Igor (Tony Leondis, 2008)
Docteur Frankenstein (Paul McGuigan, 2015)

III. SCIENCE et OPINION.

- Textes de Bachelard Introduction à la formation de l'esprit scientifique et Le nouvel esprit scientifique. Bachelard dresse les caractéristiques d'un discours, celui de la science, sans commune mesure avec l''opinion. Un texte qui permet de comprendre la rupture nécessaire des sciences et d'un ordinaire empirique. « *Or les instruments ne sont que des théories matérialisées. Il en sort des phénomènes qui portent de toutes parts la marque théorique*. » (Bachelard, *La formation de l'Esprit scientifique, introduction*). Voir avec ces textes le sens du concept d'obstacle épistémologique
- Travailler en classe sur des textes scientifiques afin de comprendre la spécificité de la démarche est essentiel pour détacher les élèves de la facilité.
  ex : Galilée, Newton...
  Il existe plusieurs vidéos avec Simondon (à rajouter)
- Une mise en ordre des mouvements de la science et des techniques. Travailler avec les élèves une approche nuancée de l'histoire des sciences. A ce sujet la collection Les Essentiels de la BNF présente l'histoire des sciences.
  Un autre dossier consacré à la vulgarisation des oeuvres. La fin du 19e siècle marque un âge d'or de l'ouverture des sciences vers un public toujours plus large. On assiste alors à un foisonnement des acteurs et des moyens mis en oeuvre pour mettre la science à la portée de tous. S'intéressant tour à tour à la pédagogie scientifique, l'astronomie, l'hygiène, les sciences naturelles, l'aéronautique ou encore l'électricité et les techniques, ce dossier permet de comprendre la place de la science dans la société de la fin du 19e siècle, ainsi que l'énergie qui fut mise dans **la vul-garisation scientifique.**
- l'illusion de la chronologie et de la ligne progressiste du temps
- Distinguer l'outil, la machine, l'instrument : à partir du film de Leroi-Gourhan qui interroge ses distinctions sur ce lien
- L'outil à partir de Stanley Kubrick 2001, l'Odyssée de l'espace
- Les automates dans la Grèce antique et dans les pays arabes
- Les robots d'Asimov
- Un florilège de textes philosophiques autour de la technique
- Les fins de la technique
- Progrès et technique

- Ecouter cette vidéo :

IV. Le chatGPT va-t-il prendre ma place ?

- Le chatbot se présente. Exercice à partir des propos du Chatbot.

L'expression "rend moi ma place" montre combien cette question est à la fois puérile et dans le même temps elle dégage une angoisse, une attente qui a tendance à construire un **discours de la fatalité** ôtant à l'homme sa liberté.
A partir de ces textes en ligneet de votre réflexion, tenter de trouver un véritable objet à la peur.
Sur la question de la peur, voir Spinoza *Ethique III*, 18 et Montaigne *De la peur*

- textes de Montaigne et Spinoza. A partir de là on construit des distinctions conceptuelles, ce qui doit affiner le jugement souvent dans la précipitation des élèves. La nuit du chasseur de Charles Laughton servira de point de départ sur le sens de la peur
- La différence entre les deux intelligences serait quantitative. Il y aurait gain ou perte, sur le modèle arithmétique de l'addition ou de la soustraction. L'intelligence artificielle serait affectée d'un "moins" en termes de sensibilité. L'intelligence peut-elle se développer sans émotions ?
- La fonction fabulatrice peut-elle être appliquée à la machine ?
  Bergson ou encore Bachelard peuvent apporter une contribution à une réflexion sur la fonction scientifique et philosophique de l'imagination.
- Les robots peuvent-ils ressentir des émotions,des sentiments et les communiquer ? ? Reconnaître celles des autres ?
  Le cinéma a souvent tenté de répondre à ces questions : en 1927, Fritz Lang présentait dans *Metropolis* Maria, un robot dont le personnage principal tombe amoureux. En 1982, l'adaptation au cinéma du roman éponyme de Philip K. Dick,

Blade Runner, les androïdes rêvent-ils de moutons électriques ? raconte la relation amoureuse entre Rachel, androïde cachée sur Terre, et Rick Deckard, chasseur de primes pour androïdes. Puis en 2014, *Ex Machina* met en scène la rencontre entre un jeune programmateur et la première intelligence artificielle au monde, celle-ci prenant la forme d'un robot féminin.

- **L'ingénium :**
  - Fichant Michel. L'Ingenium selon Descartes et le chiffre universel des règles pour la direction de l'esprit. In : Scepticisme et exégèse. Hommage à Camille Pernot ;
  - Denis Kambouchner, "Descartes et la force de l'imagination", Les Cahiers philosophiques de Strasbourg [Online], 48 | 2020, Online since 12 December 2020, connection on 26 March 2023. URL : http://journals.openedition.org/cps/4213 ; DOI : https://doi.org/10.4000/cps.4213

V. RENCONTRER LE CHATBOT

Mettre en oeuvre des exercices avec le chatbot est une façon de rencontrer cet instrument de travail :

-
  - L'enseignant pose un problème philosophique à l'élève qui doit le résoudre avec ChatGPT selon plusieurs scénarios. L'élève peut être le professeur du ChatGPT et le pousser à améliorer ses réponses en ajoutant des consignes.
-
  - Ou alors l'élève peut être dans une posture d'évaluation de la réponse du chat en recherchant des biais ou des omissions.
-
  - Les élèves d'un groupe comparent respectivement leurs réponses et mesurent les variations. Ceci peut se faire aussi sur la durée, les questions étant posées à intervalles réguliers.

â€" >Quel que soit le scénario, l'idée est de mettre l'élève dans un rôle actif et de considérer chatGPT comme un partenaire.

VI. La question du contrôle et de la domination. De la bêtise et de la censure

Bêtise

- un exercice du site consacré à la bêtise
- Un Recueil de textes
- Un exercice sur « le bon sens » de Descartes
- L'esprit critique

La fascination pour les machines : attrait pour le merveilleux

TRAVAIL DE RECHERCHE EN CLASSE
On distribuera des exposés sur :

- sur le merveilleux voir ce lien
  De la machine à la machination. Simuler et dissimuler
  On pourra travailler avec les élèves les concepts d'étonnement, de merveilleux et de stupéfaction, en passant par l'adverbe machinalement, qui renvoie à l'habitude et la routine
  On pourra prendre comme exemple des pièces de théâtre qui accordent une place importante aux effets de changements de décor et à la machinerie

  Blason de proue du pape Urbain VIII
  Par User : Bgabel sur wikivoyage shared, CC BY-SA 3.0, Lien

  Le genre est d'origine italienne : Le Bernin, par exemple, met en scène, dès 1638, L'Inondation du Tibre, dont la technique permet de produire des effets impressionnant le public
  Autre travail : le pouvoir politique et l'admiration.

Le langage du chatGPT

ChatGPT est un « modèle de langage étendu, un type d'intelligence artificielle qui utilise l'apprentissage profond (une forme d'apprentissage automatique) pour traiter et générer des textes en langage naturel (...) [Ce type de modèle est] formé sur des quantités massives de données textuelles, lui permettant d'apprendre les nuances et les complexités du langage humain" D'après Susnjak, 2022 [traduction]).

- **Le ChatGPT peut-il "doubler" l'homme ?**. Voir à ce propos Gorgias de Platon
- Si le chatGPT est en mesure de se servir d'une langue "naturelle", cela en fait-il un être humain, et de surcroît pensant ? Même si la réflexion philosophique s'enracine dans un travail sur et avec le langage, réfléchir philosophiquement de façon authentique ne consiste jamais seulement à produire du discours ou du texte, imitant le langage ou les langages des philosophes.
  Cela renvoie à un travail d'élucidation du formalisme à l'oeuvre dans les articles de vulgarisation à propos de l'IA, rapprochant ces discours d'une façon de dire propre à la rhétorique
- Exercice sur un texte de Kant sur l'art de la conversation
- Le dialogue selon Platon

Peut-on attribuer à la machine le statut de "sujet" ?
Elle procède par distinction, se donnant l'apparence d'exercer un jugement.
Cela permettra de mesurer les limites de cette démarche ;

Elle fonde son pouvoir sur la capacité à répondre, et répondre vite en l'occurrence.
Deux attitudes qui ne relèvent pas d'une attitude philosophique.
On examinera à partir d'exemples, la conception du "dialogue", nommé "conversation" que le chatbot défend.

- Un modèle cumulatif et continu du savoir qui instaure un temps de la continuité et du progrès.
  Dans la métaphore inouïe Ernesto Grassi écrit : " c'est la puissance métaphorique de la réalité qui dévoile sa duplicité en

opposition avec l'univocité du monde rationnel, qui aspire sans cesse à atteindre en vain, par un processus dialectique, la signification des étants abstraits, fossilisés"(p.176)

ARTICLES sur l'intelligence artificielle

- L'intelligence artificielle. Réflexion philosophique CYNTHIA FLEURY-PERKINS
  De l'humanisme à l'anthropotechnique, les définitions de l'humain et de l'humanisme ne sont pas closes © 2019 Publié par Elsevier Masson SAS
- ZARKA Yves Charles,« Éditorial. L'intelligence artificielle ou la maîtrise anonyme du monde », Cités, 2019/4 (N° 80), p. 3-8. DOI : 10.3917/cite.080.0003.

HUMANITE HUMANISME TRANSHUMANISME

- Posthumanisme, Humanité(s), humanisme(s)... Jean-Yves Goffi démêle le sac de noeuds des idées et des projets qui placent l'homme en leur centre.
  Entretien avec Jean-Yves Goffi
- Goffi, J.Y (2020), « Technique », version académique, dans M. Kristanek (dir.), l'Encyclopédie philosophique, URL : http://encyclo-philo.fr/technique-a/
- **Dossier sur le transhumanisme**
- Vers de nouvelles humanités ? 25 mars 2017
  L'homme remplacé
  - **Intelligence artificielle et robots**Introduction par le président de séance Ali Benmakhlouf, agrégé de philosophie et professeur à l'université de Paris-Est Val-de-Marne
  - Le robot : sans interaction culturelle et sociale, peut-on construire un vrai cerveau ? Raja Chatila, directeur de l'Institut des systèmes intelligents et de robotique, université Pierre-et-Marie-Curie.
  - Comment la robotisation transforme l'assurance ?. Pierre-Grégoire Marly, professeur de droit à l'Université du Maine.
  - Le droit des robots : quelle est l'autonomie de décision d'une machine ? Quelle protection mérite-t-elle ? Alain Bensoussan, avocat.
  - Comment protéger l'être humain face aux robots ? Nathalie Nevejans, maître de conférences à la Faculté de droit de Douai, membre du Comité d'éthique du CNRS (Comets).
  - Discussion, intelligence artificielle et robots

- **L'esprit dans une machine : le transfert de l'humain vers la machine**
  - Des humanités numériques à la singularité technologique : que reste-t-il de l'humanisme ? Jean-Gabriel Ganascia, professeur à l'université Pierre-et-Marie-Curie et président du comité d'éthique du CNRS (Comets)
  - Le droit des data Marie-Anne Frison-Roche, professeure à Sciences-Po Paris
  - L'être humain algorithmé, dépassement ou perte de soi
    Isabelle Falque-Pierrotin, présidente du CNIL
  - Penser le sujet de droit comme puissance : contingence et potentialité à l'ère du calcul intensif Antoinette Rouvroy, chercheure qualifiée du FNRS, Centre de recherche information, droit et société de l'université de Namur.
  - Le téléchargement de l'esprit : plus qu'une expérience de pensée ? Denis Forest, professeur de philosophie, université Paris X - Nanterre.
  - L'économie de l'éternité ?Pierre-Yves Geoffard, directeur de l'Ecole d'économie de Paris.
    Table-ronde - L'esprit dans une machine : le transfert de l'humain vers la machine
    Numérique et technique

- La culture numérique va-t-elle nous faire perdre le fil de l'histoire ? - Mardis des Bernardins

La révolution numérique restera-t-elle dans l'histoire comme un changement de paradigme radical pour l'humanité ?
« Avons-nous passé cet instant singulier où les machines sont devenues si puissantes qu'elles agissent profondément sur le cours de l'histoire humaine, sans retour possible ? »
Gilles Babinet, L'ère numérique, un nouvel âge de l'humanité, Cinq mutations qui vont bouleverser notre vie, Le Passeur, 2014

- Jean-Hugues Barthélémy, docteur en épistémologie
  Simondon aujourd'hui : genèse, histoire et normativité technique conférence en ligne sur la Forge Numérique de la MRSH de l'Université de Caen Normandie
  Date : 06/08/2013
  Lieu : CCIC Cerisy la Salle Durée : 47:23
  Cette conférence a été donnée dans le cadre du colloque intitulé « Gilbert Simondon et l'invention du futur » qui s'est tenu au Centre Culturel International de Cerisy du 5 au 15 août 2013, sous la direction de Jean-Hugues BARTH ÉLÉMY et Vincent BONTEMS.
- Robert Martine, « Le problème de la philosophie : une solution technique », Le Philosophoire 2/ 2003 (n° 20), p. 139-143
  DOI : 10.3917/phoir.020.0139

Conclusion provisoire

Rien de plus facile que de craindre. Texte Alain, Libres Propos, (1927).

> Il y a croire et croire, et cette différence paraît dans les mots croyance et foi. La différence va même jusqu'à l'opposition ; car selon le commun langage, et pour l'ordinaire de la vie, quand on dit qu'un homme est crédule, on exprime par là qu'il se laisse penser n'importe quoi, qu'il subit l'apparence, qu'il subit l'opinion, qu'il est sans ressort. Mais quand on dit d'un homme d'entreprise qu'il a la foi, on veut dire justement le contraire. [...] Dans le fait ceux qui refusent de croire sont des hommes de foi ; on dit encore mieux de bonne foi, car c'est la marque de la foi qu'elle est bonne. Croire à la paix, c'est foi ; il faut ici vouloir ; il faut se rassembler, tout comme un homme qui verrait un spectre, et qui se jurerait à luimême de vaincre cette apparence. Ici il faut croire d'abord, et contre l'apparence ; la foi va devant ; la foi est courage. Au contraire croire à la guerre, c'est croyance ; c'est pensée agenouillée et bientôt couchée. C'est avaler tout ce qui se dit ; c'est répéter ce qui a été dit et redit ; c'est penser mécaniquement. Remarquez qu'il n'y a aucun effort à faire pour être prophète de malheur ; toutes les raisons sont prêtes ; tous les lieux communs nous attendent. Il est presque inutile de lire un discours qui suit cette pente ; on sait d'avance ce qui sera dit, et c'est toujours la même chose. Quoi de plus facile que de craindre ?

BIBLIOGRAPHIE :

ARISTOTE, *Topiques*, 2 v., texte établi et traduit J. Brunschwig, Collection des Universités de France, Paris, Les Belles Lettres, 1967 (t.1), 2007 (t.2).

BIMBENET Etienne, *L'animal que je ne suis plus*, Folio-Essais, Paris, Gallimard, 2011.

CHOMSKY Noam, « La fausse promesse de ChatGPT », New York Times, 8 mars 2023 : https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

CUKIER Kenneth, MAYER-SCHÖNBERGER Viktor, *Big Data. La révolution des données est en marche*, Paris, Robert Laffont, 2014.

DE FONTENAY Elisabeth, *Le Silence des bêtes : la philosophie à l'épreuve de l'animalité*, Paris, Fayard, 1998.

DELEUZE Gilles, *La Philosophie critique de Kant*, Le Philosophie, Paris PUF, 1963 ; 5e édition, 1983.

- - - - - *Différence et répétition*, Bibliothèque de Philosophie Contemporaine, Histoire de la Philosophie et Philosophie Générale, Paris, PUF, 1968 ; 6e édition, 1989, p. 192 à 203.

- - - - - *Nietzsche et la Philosophie*, Paris PUF, 1962.

DERRIDA Jacques, *L'Animal que donc je suis*, Paris, Galilée, 2006.

DERRIDA Jacques, *Séminaire La bête et le Souverain*, Volume I (2001-2002), édition établie par Michel Lisse, Marie-Louise Mallet, Ginette Michaud, Paris, Galilée, 2008.

DESCARTES René, *Regulae ad directionem ingenii* in Ruvres, tome X : Discours de la méthode et Essais, texte établi par Charles Adam et Paul Tannery, Paris, Vrin (pp. 349-4878).

- - - - - *Règles pour la direction de l'esprit*, traduction et notes par J. Brunschwig, in *Ruvres philosophiques*, tome I (1618-1637), éd. par F. Alquié, Paris, Classiques Garnier, 1988 ? pp. 67 à 204.

- - - - - *Discours de la méthode* (1637) in Ruvres, tome VI : Discours de la méthode et Essais, texte établi par Charles Adam et Paul Tannery, Léopold Cerf, 1902 (p. 1-78) : https://fr.wikisource.org/wiki/Discours_de_la_m%C3%A9thode/%C3%89dition_Adam_et_Tannery,

- - - - - *Lettre au Marquis de Newcastle* du 23 novembre 1646

- - - - - *Lettre à Morus* du 5 février 1649.

DEVILLERS Laurence, *Des robots et des hommes : mythes, fantasmes et réalité : Mythes, fantasmes et réalité*, éditions Plon, 2017, 288 p.

- - - - - *Les Robots émotionnels : Santé, surveillance, sexualité... : et l'éthique dans tout ça ?*, éditions de l'Observatoire, coll. « Essais », 2020, 272 p.

- - - - - *La Souveraineté numérique dans l'après-crise*, éditions de l'Observatoire, coll. « Et après ? », 2020.

ESCAL Françoise, *Le compositeur et ses modèles*, Paris, PUF, 1984.

GANASCIA Jean-Gabriel, *L'intelligence artificielle.*- Flammarion (Collection Dominos), 1993.

- - - - - *L'Intelligence artificielle : vers une domination programmée ?*, éditions du Cavalier Bleu, 2017.

- - - - - *Le Mythe de la singularité : faut-il craindre l'intelligence artificielle ?*, éditions du Seuil, Collection Sciences Ouvertes, 2017.

- - - - - *Servitudes virtuelles*, éditions du Seuil, Collection Sciences Ouvertes, 2022.

GRAND Marie, "ChatGPT nous invite à un regain d'intelligence dans tous les domaines, dont l'enseignement", *Le Monde*, 27 mars 2023, 17:10 : https://www.lemonde.fr/idees/article/2023/03/27/chatgpt-nous-invite-a-un-regain-d-intelligence-dans-tous-les-domaines-dont-l-enseignement_6167101_3232.html.

HADOT Pierre, *Exercices spirituels et philosophie antique*, Paris, Albin Michel, 2002.

HEIDEGGER Martin, *Les Concepts fondamentaux de la métaphysique. Monde-finitude-solitude* (t. 29/30 de la Gesamstausgabe), trad. D. Panis, Paris, Gallimard, 1992

KANT Emmanuel, *Critique de la raison pure*, Traduction française avec notes par A. Tremesaygue et B. Pacaud, préface de Ch. Serrus, Quadrige, Paris, PUF, 1944 ; 10e édition, 1984, pp. 148-149.

- - - - - *Critique de la faculté de juger*, traduction par Alain Renaut, Paris GF, 1995.

- - - - - *Vers la paix perpétuelle, Que signifie s'orienter dans la pensée ? Qu'est-ce que le Lumières et autres textes*, introduction, notes, bibliographie et chronologie par Françoise Proust, traduction par Jean-Louis Poirier et Françoise Proust, Paris, GF, 1991.

LA METTRIE Julien Offray de, *L'Homme-machine*, Paris, Henry, 1865, sur Gallica : https://gallica.bnf.fr/ark:/12148/bpt6k6253039v/f9.image.

LECUN Yann, *L'apprentissage profond : une révolution en intelligence artificielle*, Leçon inaugurale, 4 février 2016, Chaire Informatique et sciences numériques, Collège de France : https://www.college-de-france.fr/agenda/lecon-inaugurale/apprentissage-profond-une-revolution-en-intelligence-artificielle/apprentissage-profond-une-revolution-en-intelligence-artificielle.

LECUN Yann, *L'apprentissage profond*, Cours au collège de France, 2016 : https://www.college-de-france.fr/agenda/cours/apprentissage-profond.

LECUN Yann, *L'apprentissage profond : théorie et pratique*, Séminaire au Collège de France, 2016 : https://www.college-de-france.fr/agenda/seminaire/apprentissage-profond-theorie-et-pratique.

MALHERBE Michel, *D'un pas de philosophe*, Matière étrangère, Paris, Vrin, 2013 ;

- - - - « Métaphysique du pied » : https://eduscol.education.fr/document/19777/download, Séminaire « Le pied sur terre : l'esprit nous vient-il d'en bas ? » ; Rencontres philosophiques de Langres 2013, *La Matière et l'Esprit*, https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=0CAMQw7AJahcKEwi41JeAzuX9AhUAAAAAHQAAAAQAw&url=https%3A%2F%2Feduscol.education.fr%2Fdocument%2F19768%2Fdownload&psig=AOvVaw0ZO0mPNrSgbegHviyuFl5j&ust=1679233388877832.

MARX Karl, *Manuscrits de 1844 (Economie, politique et philosophie)*, présentation, traduction et notes de Emile Bottigelli, Paris, Editions sociales, 1972.

MONTAIGNE Michel de, *Essais*, Livre III, chapitre VIII, De l'art de conférer, édition Pierre Villey revue par V. Saulnier, Paris, PUF, 2004.

PHILONENKO Alexis, *L'Ruvre de Kant*, 2 v., A la Recherche de la vérite, Paris, Vrin, 1988.

PLATON, *La République*, traduction Robert Baccou, Paris, GF, 1966.

ROGER Alain, *Bréviaire de la bêtise*, Bibliothèque des Idées, Paris, Gallimard, 2008.

STIEGLER Bernard, *États de choc, Bêtise et savoir au XXIe siècle*, Essais, Paris, Mille et une nuits, 2012.

VUILLEROD Jean-Baptiste, « Penser intelligemment la bêtise : de Deleuze à Stiegler », *Implications philosophiques*, 3 mars 2018, https://www.implications-philosophiques.org/penser-intelligemment-la-betise/

- **ARTICLES sur l'I A**

- Des collègues professeurs de philosophie corrigent des copies produites par ChatGPT : https://etudiant.lefigaro.fr/article/c-est-loin-de-faire-un-devoir-de-qualite-un-prof-corrige-le-bac-dephilo-de-chatgpt_a3cca6d4-8ddd-11ed-a86d-66d30ea5ec0a/

- Pour une réflexion philosophique sur l'intelligence artificielle, sa puissance réelle en son état actuel de développement technique, ses limites et ses véritables dangers (qui ne se situent certainement pas là où on le penserait) on peut écouter ce podcast récent, Avec Philosophie, sur France Culture : https://www.radiofrance.fr/franceculture/podcasts/avec-philosophie/l-intelligence-artificielle-estelle-vraiment-intelligente-2059509

- Les créateurs de l'I.A lancent la chasse aux tricheurs : https://www.liberation.fr/economie/economie-numerique/chatgpt-tu-peux-faire-mes-devoirs-les-createurs-de-lia-lancent-la-chasse-aux-tricheurs-20230203_RAQA55N54NFZZHEKSXKS6W5UVE/?at_campaign=NL_Lib%C3%A9_Matin_Samedi&at_email_type=acquisition&at_medium=email&at_creation=NL_Libe_Matin_samedi_2023-02-05&actId=ebwp0YMB8s1_OGEGSsDRkNUcvuQDVN7a57ET3fWtrS841mguA0zYUvG6YFUZoN3h&actCampaignType=CAMPAIGN_MAIL&actSource=522604

- **VIDEOS**

- Une vidéo dans laquelle un Data-Scientist explique de façon très claire les bases théoriques (mathématiques) du deeplearning, qui est au principe de l'Intelligence Artificielle, ainsi que leur développement historique : https://www.youtube.com/watch?v=XUFLq6dKQok

- https://www.bfmtv.com/tech/on-a-montre-le-travail-de-chat-gpt-a-une-prof-de-philo-voici-son-constat_AV-202301130473.html

- Un collègue explique et analyse les principes techniques, les modalités de fonctionnement, les possibilités et les limites de ChatGPT3 : https://www.youtube.com/watch?v=R2fjRbc9Sa0

NOTES

[1] Sur ce point : https://www.lemonde.fr/pixels/article/2023/02/06/google-annonce-le-lancement-du-robot-conversationnel-brard-riposte-du-geant-americain-a-chatgpt_6160781_4408996.html.

[2] Guillaume von der Weid, « Pourquoi ChatGPT est une mascarade », L'Obs, 19 février 2023 à 8h30 :

https://www.nouvelobs.com/opinions/20230219.OBS69747/pourquoi-chatgpt-est-une-mascarade.html.

[3] Gaspar Koenig, « La faillite épistémologique de ChatGPT », Les Echos, 22 févr. 2023 à 07:23 : https://www.lesechos.fr/idees-debats/editos-analyses/la-faillite-epistemologique-de-chatgpt-1908838

[4] Laurence Plazenet, « Pourquoi ChatGPT est complètement idiot », Le Figaro, mis à jour le 21/02/2023, https://amp.lefigaro.fr/vox/societe/pourquoi-chatgpt-est-completement-idiot-20230126

[5] Platon, *Phèdre* 266b

[6] Sur la philosophie comme exercice, voir par exemple : Pierre Hadot, *Exercices spirituels et philosophie antique*, Paris, Albin Michel, 2002

[7] 518b-d ; trad. V. Cousin.

[8] Jean-Gabriel Ganascia, *L'intelligence artificielle*, Idées, reçues, Le Cavalier Bleu éditions, 2007, p. 9.

[9] Turing, A.M. (1950). *Computing machinery and intelligence*. Mind, 59, 433-460 : https://academic.oup.com/mind/article/LIX/236/433/986238.

[10] J.-G. Ganascia, *opus cit.* p. 95.

[11] *Locus cit.*, p. 95.

[12] *Opus. cit.* p. 75.

[13] *Opus cit.* p. 25.

[14] Sur ce point voir : Françoise Escal, *Le compositeur et ses modèles*, Paris, PUF, 1984 ; Le plagiat, une longue tradition dans la musique classique, par Léopold Tobisch, publié le jeudi 8 octobre 2020 à 19h01 sur le site de *France Musique* : https://www.radiofrance.fr/francemusique/le-plagiat-une-longue-tradition-dans-la-musique-classique-3026632

[15] Sur le droit d'auteur et le droit d'exception pédagogique accordé aux professeurs dans le cadre de leurs missions d'enseignement : *Code de la propriété intellectuelle*, Légifrance ; *Protocole d'accord sur l'utilisation et la reproduction des livres, des oeuvres musicales éditées, des publications périodiques et des oeuvres des arts visuels à des fins d'illustration des activités d'enseignement et de recherche*, NOR : MENE1600684X - Protocole d'accord du 22-7-2016 - MENESR - DGESCO B1-1 - DGESCO B1-2, Bulletin officiel n°35 du 29 septembre 2016 ; *Mise en oeuvre du contrat du 2 juin 2014 concernant la reproduction par reprographie d'oeuvres protégées dans les établissements d'enseignement du premier degré public et privé sous contrat*, NOR : MENE1416581C - Circulaire n° 2014-094 du 18-7-2014 - MENESR - DGESCO B1-1,Bulletin officiel n° 31 du 28 août 2014.. Un tel acte relèverait vraisemblablement du périmètre juridique relatif au droit de copie privée et de la violation de ses limites

[16] Sur le principe d'exclusion des évaluations dites « formatives » des résultats des élèves cf. les travaux de Benjamin S. Bloom, fondateur du concept (behavioriste) d'évaluation formative : *Learning for mastery* (1968) UCLA - CSEIP - Evaluation Comment. Vol. 1. ; trad. franç. : *Apprendre pour maîtriser*, Payot,1972.) ; Benjamin S. Bloom, George F. Madaus, and J. Thomas Hastings, *Evaluation to Improve Learning*, New York : McGraw-Hill, Inc., 1981.

[17] Trad. J. Brunschwig, Collection des Universités de France, Paris, Les Belles Lettres, 1967, t.1, p. 2.

[18] Professeur à l'Université de New York et directeur de Facebook AI Research (FAIR), Professeur invité au Collège de France 2015-2016.

[19] Gustave Flaubert, *Dictionnaire des idées reçues*, L. Conard, 1910 (p. 415-452), Wikisource : https://fr.wikisource.org/wiki/Dictionnaire_des_id%C3%A9es_re%C3%A7ues/Texte_entier.

[20] In *Mélanges de logique*, traduction par Joseph Tissot, Librairie philosophique de Ladrange, 1862.

[21] trad. Alain Renaut, Paris GF, 1995 ; p. 218

[22] trad. Alain Renaut, Paris GF, 1995, pp. 278 à 281

[23] Alexis Philonenko, dans *L'Ruvre de Kant*, tome II, A la Recherche de la vérite, Paris, Vrin, 1988, p. 194, commente : « Qu'est-ce que la nature ? C'est l'oeuvre de nos jugements objectifs, qui découvrent la raison des choses, et celle-ci est la connaissance. Qu'est-ce que la connaissance ? C'est l'ensemble des concepts que nous forgeons et échangeons. Qu'est-ce que l'échange ? C'est la possibilité de la communication. Qu'est-ce enfin que communiquer ? C'est l'essence de notre savoir, et cette essence transparaît dans le jugement de goût, qui nous conduit à autrui. »

[24] CFJ, §16.

[25] Sur ces points : Gilles Deleuze, *La Philosophie critique de Kant*, Le Philosophie, Paris PUF, 1963 ; 5e édition, 1983, pp. 70 à 73.

[26] CFJ, §40, trad. Renaut, p.280

[27] René Descartes, *Discours de la méthode* (1637) in Ruvres, tome VI : Discours de la méthode et Essais, texte établi par Charles Adam et Paul Tannery, Léopold Cerf, 1902 (p. 1-78).

[28] Kant, CJF, §40, éd. cit. pp. 278-279.

[29] (Ibid. p. 279, note*)

[30] Voir par exemple la « copie » corrigée par un collègue d'histoire : https://www.youtube.com/watch?v=v5gatKRATj0.

[31] Kant, CFJ §40, éd. cit. p. 279.

[32] L'expression latine *secunda Petri* renvoie ici à la seconde partie de la *Dialectique* de Pierre de La Ramée et non à la seconde *Epître de Pierre* dans le *Nouveau Testament* ; cf. sur ce point : Rémi Brague, « Kant et la secunda Petri. Un contresens fréquent », *Revue de Métaphysique et de Morale*, n°3 (JUILLET-SEPTEMBRE 1999) pp. 419-422, PUF.

[33] Traduction française avec notes par A. Tremesaygue et B. Pacaud, préface de Ch. Serrus, Quadrige, Paris, PUF, 1944 ; 10e édition, 1984, pp. 148-149.

[34] Cette formule est une parodie du titre du film d'Arnaud Desplechin *Comment je me suis disputé... (ma vie sexuelle)* (1996).

[35] Maurine Briantais, Prometheus dans Bing : l'IA à la ChatGPT de Microsoft part en vrille, CCM, 19/02/23 09:08 https://www.commentcamarche.net/applis-sites/applications/27677-prometheus-dans-bing-l-ia-a-la-chatgpt-de-microsoft-part-en-vrille/ ; Microsoft Bing : ChatGPT dévoile son alter ego maléfique, Venom, 01net, 17 février 2023 à 10:35 : https://www.01net.com/actualites/microsoft-bing-chatgpt-alter-ego-malefique-venom.html ; Microsoft's Bing AI plotted its revenge and offered me furry porn : https://www.theverge.com/2023/2/16/23602965/microsoft-bing-ai-sydney-fury-furry-venom ; Bing Chat s'arrête enfin de délirer grâce à cette règle radicale, Pressecitron.net, 20 février 2023 à 13:45 : https://www.presse-citron.net/bing-chat-sarrete-enfin-de-delirer-grace-a-cette-regle-radicale// ; ChatGPT n'est qu'« un perroquet approximatif », selon le ministre délégué au numérique, Le Monde avec AFP, 20/02/2023 13h49 : https://www.lemonde.fr/pixels/article/2023/02/20/chatgpt-n-est-qu-un-perroquet-approximatif-selon-le-ministre-delegue-au-numerique_6162562_4408996.html.

[36] https://www.cnrtl.fr/definition/b%C3%AAtise

[37] Michel de Montaigne, *Essais*, Livre III, chapitre VIII, De l'art de conférer, édition Pierre Villey revue par V. Saulnier, Paris, PUF, 2004, p..

[38] Ibidem.

[39] Jean-Gabriel GANASCIA ,*Le Mythe de la singularité : faut-il craindre l'intelligence artificielle ?*, éditions du Seuil, Collection Sciences Ouvertes, 2017 ; Laurence DEVILLERS, *Des robots et des hommes : mythes, fantasmes et réalité : Mythes, fantasmes et réalité*, éditions Plon, 2017, 288 p.

[40] Comme c'est le cas par exemple chez Bernard Stiegler, *États de choc, Bêtise et savoir au XXIe siècle*, Essais, Paris, Mille et une nuits, 2012.

[41] Gilles Deleuze, *Différence et répétition*, Bibliothèque de Philosophie Contemporaine, Histoire de la Philosophie et Philosophie Générale, Paris, PUF , 1968 ; 6e édition, 1989, p. 192 à 203 ; Jacques Derrida, *Séminaire La bête et le Souverain*, Volume I (2001-2002), édition établie par Michel Lisse, Marie-Louise Mallet, Ginette Michaud, Paris, Galilée, 2008.

[42] *Discours de la méthode*, Cinquième partie : https://fr.wikisource.org/wiki/Discours_de_la_m%C3%A9thode/%C3%89dition_Adam_et_Tannery, *Lettre au Marquis de Newcastle* du 23 novembre 1646, *Lettre à Morus* du 5 février 1649

[43] *L'Homme-machine*, Paris, Henry, 1865, sur Gallica : https://gallica.bnf.fr/ark:/12148/bpt6k6253039v/f9.image

[44] Martin Heidegger, *Les Concepts fondamentaux de la métaphysique. Monde-finitude-solitude* (t. 29/30 de la Gesamstausgabe), trad. D. Panis, Paris, Gallimard, 1992 ; Elisabeth de Fontenay, *Le Silence des bêtes : la philosophie à l'épreuve de l'animalité*, Paris, Fayard, 1998 ; Jacques Derrida, *L'Animal que donc je suis*, Paris, Galilée, 2006 ; Etienne Bimbenet, *L'animal que je ne suis plus*, Folio-Essais, Paris, Gallimard, 2011.

[45] Gilles Deleuze, *Différence et répétition*, Bibliothèque de Philosophie Contemporaine, Histoire de la Philosophie et Philosophie Générale, Paris, PUF , 1968 ; 6e édition, 1989.

[46] Gilles Deleuze, *Différence et répétition*, p. 217.

[47] G. Deleuze, *Différence et répétition*, p. 217.

[48] Alain Roger, *Bréviaire de la bêtise*, Bibliothèque des Idées, Paris, Gallimard, 2008, p. 21.

[49] Descartes, *Discours de la méthode*, Cinquième partie, in *Ruvres philosophiques*, éd. par F. Alquié, tome 1, Paris Garnier, 1988, p. 628.

[50] Ibidem, p 628-629.

[51] Hervé Bazin, *L'Huile sur le feu*, Paris, Grasset, 1992.

[52] https://cnrtl.fr/definition/b%C3%AAte

[53] Gilles Deleuze, *Différence et répétition*, Bibliothèque de Philosophie Contemporaine, Histoire de la Philosophie et Philosophie Générale, Paris, PUF , 1968 ; 6e édition, 1989, p. 196.

[54] Philippe Kaenel, *Le Buffon de l'humanité. La zoologie politique de J.-J. Grandville (1803-1847)*, Revue de l'Art, Année 1986, 74, pp. 21-28, https://www.persee.fr/doc/rvart_0035-1326_1986_num_74_1_347591

[55] « (...) Äµ ´t ¿V½ ¿P À¬½Å Ä¹ Ã¿ÆxÂ b½ A À¼¸¸µzÂ (...) », Platon, *Protagoras* traduction par Émile Chambry, Paris, Garnier-Flammarion, 1967, p. 55.

[56] Cf. Bernard Stiegler, *Le Technique et le temps 1 La faute d'Epiméthée*, Paris, Galilée, 1994

[57] Anatole Bailly, *Dictionnaire grec-français*, Paris Hachette, 1895, p.

[58] Sur ce point, voir Michel Malherbe, *D'un pas de philosophe*, Matière étrangère, Paris, Vrin, 2013 ; « Métaphysique du pied » : https://eduscol.education.fr/document/19777/download, Séminaire « Le pied sur terre : l'esprit nous vient-il d'en bas ? » ; Rencontres philosophiques de

Langres 2013, *La Matière et l'Esprit*, https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=0CAMQw7AJahcKEwi41JeAzuX9AhUAAAAAHQAAAAQAw&url=https%3A%2F%2Feduscol.education.fr%2Fdocument%2F19768%2Fdownload&psig=AOvVaw0ZO0mPNrSgbegHviyuFl5j&ust=1679233388877832.

[59] Et c'est sans doute pourquoi Deleuze exprimer une profonde envers les animaux de compagnie, « familiers et familiaux » : cf. Gilles Deleuze, Claire Parnet, Pierre-André Boutang, *L'Abécédaire de Gilles Deleuze*, article A comme Animal, distribué par *Sub-til* depuis 2019 : https://www.youtube.com/watch?v=SlNYVnCUvVg&list=PLiR8NqajHNPbaX2rBoA2z6IPGpU0IPlS2&index=1

[60] https://www.cnrtl.fr/definition/p%C3%A9ristaltique

[61] G. Deleuze, Différence et répétion, loc.cit.

[62] G. Deleuze, *Différence et répétion*, p. 196.

[63] G. Deleuze, *Différence et Répétition*, p. 197.

[64] Alain ROGER, *Bréviaire de la bêtise*, Bibliothèque des Idées, Paris, Gallimard, 2008, p.

[65] Bernard Stiegler, *États de choc, Bêtise et savoir au XXIe siècle*, Essais, Paris, Mille et une nuits, 2012, p.

[66] G. Deleuze, *Nietzsche et la Philosophie*, Paris PUF, 1962, p.

[67] G. Deleuze, *Différence et répétition*, p. 193

[68] Ibid. p. 195.

[69] Sur ce point, voir Liming Hu, David Bell, Sameer Antani, Zhiyun Xue, Kai Yu, Matthew P Horning, Noni Gachuhi, Benjamin Wilson, Mayoore S Jaiswal, Brian Befano et alii, »An Observational Study of Deep Learning and Automated Evaluation of Cervical Images for Cancer Screening", JNCI : Journal of the National Cancer Institute, Volume 111, Issue 9, September 2019, Pages 923-932, https://doi.org/10.1093/jnci/djy225 ; Philipe Saltel, « L'intelligence artificielle bouleverse le diagnostic en cancérologie » in Pop'sciences Mag, Université de Lyon, 01/04/2019, https://popsciences.universite-lyon.fr/le_mag/lintelligence-artificielle-bouleverse-le-diagnostic-en-cancerologie/.

[70] G. Deleuze, *Différence et répétition*, p. 217.

[71] Gilles Deleuze, *Nietzsche et la philosophie*, Bibliothèque de Philosophie Contemporaine, Histoire de la Philosophie et Philosophie Générale, Paris, PUF, 1967, p. 120.

[72] G. Deleuze, *Différence et répétition*, p. 196.

[73] G. Deleuze, Locus cit.

[74] G. Deleuze, *Nietzsche et la philosophie*, p. 120.

[75] Nous renvoyons ici à Jean-Baptiste Vuillerod, « Penser intelligemment la bêtise : de Deleuze à Stiegler », *Implications philosophiques*, 3 mars 2018, https://www.implications-philosophiques.org/penser-intelligemment-la-betise/ : (...) lorsqu'un élève écrit purement et simplement que, pour Descartes, « je pense, donc je suis » ou bien « Dieu existe », en coupant ces propositions de leurs prémisses, alors certes il écrit quelque chose de vrai, mais cette vérité n'en est pas moins une bêtise et un non-sens. Ce qui donne sens au *cogito* et aux preuves de l'existence de Dieu, c'est le projet des *Méditations métaphysiques*, qui n'est autre que le fondement de la science moderne, et le fait que cette recherche du fondement en passe par un doute radical auquel on va pouvoir résister. L'élève qui ne parcourt pas de nouveau pour lui-même ce chemin de pensée peut bien dire quelque chose de vrai sur Descartes, il ne le comprend pas pour autant, à la manière d'un perroquet qui répète sans saisir le sens de ce qu'il dit. C'est cette

incompréhension du sens que désigne la bêtise."

[76] G. Deleuze, *opus, cit.* p. 198-199.

[77] G. Deleuze, opus cit. p. 197.

[78] Tome 1, p. XXXV.

[79] Microsoft annonce l'arrivée d'une version « plus puissante » de ChatGPT sur son moteur de recherche Bing, Thomas Leroy, article publié le 07/02/2023 à 19:40 | MAJ à 20:17 : https://www.bfmtv.com/tech/intelligence-artificielle/microsoft-confirme-l-arrivee-de-chat-gpt-sur-son-moteur-de-recherche-bing_AN-202302070690.html.

[80] Sur ce point voir : Nicolas Oliveri, Paul Rasse, « Les Mooc et leurs dérivés, ou l'imaginaire des technologies pédagogiques », *Hermès, La Revue* 2017/2 (n° 78), pages 110 à 117 ; CAIRN https://www.cairn.info/revue-hermes-la-revue-2017-2-page-110.htm?contenu=article.